Benjamin Fine
Gerhard Rosenberger

Number Theory

An Introduction via the Density of Primes

Second Edition

Birkhäuser

Birkhäuser

Benjamin Fine · Gerhard Rosenberger

Number Theory

An Introduction via the Density of Primes

Second Edition

Birkhäuser

Benjamin Fine
Department of Mathematics
Fairfield University
Fairfield, CT
USA

Gerhard Rosenberger
Universität Hamburg
Hamburg
Germany

ISBN 978-3-319-43873-3 ISBN 978-3-319-43875-7 (eBook)
DOI 10.1007/978-3-319-43875-7

Library of Congress Control Number: 2016947201

Mathematics Subject Classification (2010): 11A01, 11A03, 11M01, 11R04, 11Z05, 11T71, 11H01, 20A01, 20G01, 14G01, 08A01

Printed on acid-free paper

This book is published under the trade name Birkhäuser, www.birkhauser-science.com
The registered company is Springer International Publishing AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Preface to the Second Edition

We were very pleased with the response to the first edition of this book and we were very happy to do a second edition. In this second edition, we cleaned up various typos pointed out by readers and added some new material suggested by them. We have also included important new results that have appeared since the first edition came out. These results include results on the gaps between primes and the twin primes conjecture.

We have added a new chapter, Chapter 7, on p-adic numbers, p-adic arithmetic, and the use of Hensel's Lemma. This can be included in a year-long course.

We have extended the material on elliptic curves in Chapter 5 on primality testing.

We have added material in Chapter 4 on multiple-valued zeta functions.

As before, we would like to thank the many people who read or used the first edition and made suggestions. We would also especially like to thank Anja Moldenhauer and Anja Rosenberger who helped tremendously with editing and LATEX and made some invaluable suggestions about the contents.

Fairfield, USA
Hamburg, Germany

Benjamin Fine
Gerhard Rosenberger

Preface to the First Edition

Number theory is fascinating. Results about numbers often appear magical, both in their statements and in the elegance of their proofs. Nowhere is this more evident than in results about the set of prime numbers. The Prime Number Theorem, which gives the asymptotic density of the prime numbers, is often cited as the most surprising result in all of mathematics. It certainly is the result which is hardest to justify intuitively.

The prime numbers form the cornerstone of the theory of numbers. Many, if not most, results in number theory proceed by considering the case of primes and then pasting the result together for all integers by using the Fundamental Theorem of Arithmetic. The purpose of this book is to give an introduction and overview of number theory based on the central theme of the sequence of primes. The richness of this somewhat unique approach becomes clear once one realizes how much number theory and mathematics in general is needed to learn and truly understand the prime numbers. The approach provides a solid background in the standard material as well as presenting an overview of the whole discipline. All the essential topics are covered the fundamental theorem of arithmetic, theory of congruences, quadratic reciprocity, arithmetic functions, and the distribution of primes. In addition, there are firm introductions to analytic number theory, primality testing and cryptography, and algebraic number theory, as well as many interesting side topics. Full treatments and proofs are given to both Dirichlet's Theorem and the Prime Number Theorem. There is a complete explanation of the new AKS algorithm that shows that primality testing is of polynomial time. In algebraic number theory, there is a complete presentation of primes and prime factorizations in algebraic number fields.

The book grew out of notes from several courses given for advanced undergraduates in the United States and for teachers in Germany. The material on the Prime Number Theorem grew out of seminars also given both at the University of Dortmund and at Fairfield University. The intended audience is upper level undergraduates and beginning graduate students. The notes upon which the book was based were used effectively in such courses in both the United States and

Germany. The prerequisites are a knowledge of Calculus and Multivariable Calculus and some Linear Algebra. The necessary ideas from Abstract Algebra and Complex Analysis are introduced in the book. There are many interesting exercises ranging from simple to quite difficult. Solutions and hints are provided to selected exercises. We have written the book in what we feel is a user-friendly style with many discussions of the history of various topics. It is our opinion that it is also ideal for self-study.

There are two basic facts concerning the sequence of primes that are focused on in this book and from which much of the theory of numbers is introduced. The first fact is that there are infinitely many primes. This fact was of course known since at least the time of Euclid. However, there are a great many proofs of this result not related to Euclid's original proof. By considering and presenting many of these proofs, a wide area of modern number theory is covered. This includes the fact that the primes are numerous enough so that there are infinitely many in any arithmetic progression $an + b$ with a, b relatively prime (Dirichlet's Theorem). The proof of Dirichlet's Theorem allows us to first introduce analytic methods.

In distinction to there being infinitely many primes, the density of primes thins out. We first encounter this in the startling (but easily proved) result that there are arbitrarily large gaps in the sequence of primes. The exact nature of how the sequence of primes thins out is formalized in the Prime Number Theorem, which as already mentioned, many people consider the most surprising result in mathematics. Presenting the proof and the ideas surrounding the proof of the Prime Number Theorem allows us to introduce and discuss a large portion of analytic number theory.

Algebraic Number Theory arose originally as an attempt to extend unique factorization to algebraic number rings. We use the approach of looking at primes and prime factorizations to present a fairy comprehensive introduction to algebraic number theory.

Finally, modern cryptography is intimately tied to number theory. Especially crucial in this connection is primality testing. We discuss various primality testing methods, including the recently developed AKS algorithm and then provide a basic introduction to cryptography.

There are several ways that this book can be used for courses. Chapter 1 together with selections from the remaining chapters can be used for a one-semester course in number theory for undergraduates or beginning graduate students. The only prerequisites are a basic knowledge of mathematical proofs (induction, etc.) and some knowledge of Calculus. All the rest is self-contained, although we do use algebraic methods so that some knowledge of basic abstract algebra would be beneficial. A year-long course focusing on analytic methods can be done from Chapters 1, 2, 3, and 4 and selections from 5 and 6, while a year-long course focusing on algebraic number theory can be fashioned from Chapters 1, 2, 3, and 6 and selections from 4 and 5. There are also possibilities for using the book for one semester introductory courses in analytic number theory, centering on Chapter 4, or for a one semester introductory course in algebraic number theory, centering on Chapter 6. Some suggested courses:

Basic Introductory One Semester Number Theory Course: Chapter 1, Chapter 2, Sections 3.1, 4.1, 4.2, 5.1, 5.3, 5.4, 6.1

Year-Long Course Focusing on Analytic Number Theory: Chapter 1, Chapter 2, Chapter 3, Chapter 4, Sections 5.1, 5.3, 5.4, 6.1

Year-Long Course Focusing on Algebraic Number Theory: Chapter 1, Chapter 2, Chapter 3, Chapter 6, Sections 4.1, 4.2, 5.1, 5.3, 5.4

One-Semester Course Focusing on Analytic Number Theory: Chapter 1, Chapter 2 (as needed), Sections 3.1, 3.2, 3.3, 3.4, 3.5, Chapter 4

One-Semester Course Focusing on Algebraic Number Theory: Chapter 1, Chapter 2 (as needed), Chapter 6

We would like to thank the many people who have read through other preliminary versions of these notes and made suggestions. Included among these people are Kati Bencsath and Al Thaler, as well as the many students who have taken the courses. In particular, we would like to thank Peter Ackermann, who read through the whole manuscript both proofreading and making mathematical suggestions. Peter was also heavily involved in the seminars on the Prime Number Theorem from which much of the material in Chapter 4 comes.

Benjamin Fine
Gerhard Rosenberger

Contents

Chapter 1
Introduction and Historical Remarks

The theory of numbers is concerned with the properties of the **integers**, i.e., the class of whole numbers and zero, $0, \pm 1, \pm 2, \ldots$. The positive integers, $1, 2, 3 \ldots$ are called the **natural numbers**. The basic additive structure of the integers is relatively simple. Mathematically it is just an infinite cyclic group (see Chapter 2). Therefore the true interest lies in the multiplicative structure and the interplay between the additive and multiplicative structures. Given the simplicity of the additive structure, one of the enduring fascinations of the theory of numbers is that there are so many easily stated and easily understood problems and results whose proofs are either unknown or incredibly difficult. Perhaps the most famous of these was **Fermat's Big Theorem** which was stated about 1650 and only recently proved by A.Wiles. This result said that the equation $a^n + b^n = c^n$ has no nontrivial ($abc \neq 0$) integral solutions if $n > 2$. Wiles' proof ultimately involved the very deep theory of elliptic curves. Another result in this category is the **Goldbach conjecture** first given about 1740 and still open. This states that any even integer > 2 is the sum of two odd primes. We mention that since the first edition of this book appeared, the weak, or ternary Goldbach conjecture, has been proved by H.A. Helfgott [He]. This version states that any odd number greater than 7 is the sum of three odd primes. Another of the fascinations of number theory is that many results seem almost magical. The **prime number theorem** which describes the asymptotic distribution of the prime numbers has often been touted as the most surprising result in mathematics.

The cornerstone of the multiplicative theory of the integers is the series of primes and the **fundamental theorem of arithmetic** which states that any integer can be decomposed, essentially uniquely, as a product of primes. One of the basic modes of proof in the theory of numbers is to reduce to the case of a prime and then use the fundamental theorem to patch back together for all integers. This concept of a fundamental prime decomposition, which has its origin in the fundamental theorem of arithmetic, permeates much of mathematics. In many different disciplines one of the major techniques is to find the indecomposable building blocks (the "primes" in that discipline) and then use these as starting points in proving general results. The

B. Fine and G. Rosenberger, *Number Theory*,
DOI 10.1007/978-3-319-43875-7_1

idea of a **simple group** and the Jordan–Holder decomposition in group theory is one example (see [Ro]).

The purpose of this book is to give an introduction and overview of number theory based on the series of primes. It grew out of courses for advanced undergraduates in the United States and courses for teachers in Germany. There are many approaches to presenting this first material on number theory. We felt that this approach through the series of primes gave a solid background in standard material as well as presenting a wide overview of the whole discipline.

Modern number theory has essentially three branches, which overlap in many areas. The first is **elementary number theory**, which can be quite nonelementary, and which consists of those results concerning the integers themselves which do not use analytic methods. This branch has many subbranches: the theory of congruences, diophantine analysis, geometric number theory, quadratic residues to mention a few. The second major branch is **analytic number theory**. This is the branch of the theory of numbers that studies the integers by using methods of real and complex analysis. The final major branch is **algebraic number theory** which extends the study of the integers to other algebraic number fields. By examining the series of primes we will touch on all these areas.

In Chapter 2 we will consider the basic material in elementary number theory: the fundamental theorem of arithmetic, the theory of congruences, quadratic reciprocity and related results. One of the most important straightforward results is that there are an infinite collection of primes. In Chapter 3 we will look at a collection of proofs of this result. We will also look at Dirichlet's Theorem which says that there are infinitely many primes in any arithmetic progression and at the twin prime conjecture. Although there are an infinite number of primes their density tends to thin out. It was observed though that if $\pi(x)$ denotes the number of primes less than or equal to x then this function behaves asymptotically as the function $\frac{x}{\ln x}$. This result is known as the **prime number theorem**. Besides being a startling result, the proof of the prime number theorem, done independently by Hadamard and De la Valle Poussin, became the genesis for analytic number theory. We will discuss the prime number theorem and its proof as well as the Riemann hypothesis in Chapter 4. For larger integers determining if a number is a prime and determining its factorization becomes a nontrivial problem. The fact that factorization of large integers is so difficult has been used extensively in cryptography, especially public key cryptography, i.e., coding messages that cannot be hidden, such as privileged information sent over public access computer lines. In Chapter 5 we will discuss primality testing and hint at the uses in cryptography. The excellent book by Koblitz [Ko] is entirely devoted to the subject. Finally in Chapter 6 we discuss primes in algebraic number theory. We introduce the general idea of unique factorization and primes and prime ideals in number fields.

The history of number theory has been very well documented. The book by L.E. Dickson **The History of the Theory of Numbers** [D] gives a comprehensive history until the early part of the twentieth century. The book by O. Ore **Number Theory and its History** [O] gives a similar but not as comprehensive account and includes results up to the mid-twentieth century. Another excellent historical approach is the book by A.Weil **Number Theory: An Approach Through History.**

From Hammurapi to Legendre [W]. The Chapter Notes in Nathanson's book **Elementary Methods in Number Theory** [N] also provide good historical insights. In this book we will only touch on the history. For this introduction we give a very brief overview of some of the major developments.

Number theory arises from arithmetic and computations with whole numbers. Every culture and society has some method of counting and number representation. However it was not until the development of a place value system that symbolic computation became truly feasible. The numeration system that we use is called the Hindu-Arabic numeration system and was developed in India most likely during the period 600–800 A.D. This system was adopted by Arab cultures and transported to Europe via Spain. The adoption of this system in Europe and elsewhere was a long process and it was not until the Renaissance and after that symbolic computation widely superseded the use of abaci and other computing devices. We should remark that although mathematics is theoretical it often happens that abstract results are delayed without proper computation. Calculus and analysis could not have developed without the prior development of the concept of an irrational number.

Much of the beginnings of number theory came from straightforward observation and a great deal of number theoretic information was known to the Babylonians, Egyptians, Greeks, Hindus, and other ancient cultures. Greek mathematicians, especially the Pythagoreans (around 450 B.C.), began to think of numbers as abstractions and deal with purely theoretical questions. The foundation material of number theory—divisors, primes, gcd, lcm, the Euclidean algorithm, the fundamental theorem of arithmetic and the infinitude of primes—although not always stated in modern terms - are all present in **Euclid's Elements**. Three of Euclid's books, Book VII, Book VIII, and Book IX treat the theory of numbers. It is interesting that Euclid's treatment of number theory is still geometric in its motivation and most of its methods. It wasn't until the Alexandrian period, several hundred years later, that arithmetic was separated from geometry. The book **Introductio Arithmeticae** by Niomachus in the second century A.D. was the first major treatment of arithmetic and the properties of the whole numbers without geometric recourse. This work was continued by Diophantus of Alexandria about 250 A.D. His great work **Arithmetica** is a collection of problems and solutions in number theory and algebra. In this work he introduced a great deal of algebraic symbolism as well as the topic of equations with indeterminate quantities. The attempt to find integral solutions to algebraic equations is now called **Diophantine analysis** in his honor. Fermats' big theorem of solving $x^n + y^n = z^n$ for integers is an example of a Diophantine problem.

The improvements in computational techniques led mathematicians in the 1500s and 1600s to look more deeply at number theoretical questions. The giant of this period was Pierre Fermat who made enormous contributions to the theory of numbers. It was Fermat's work that could be considered the beginnings of number theory as a modern discipline. Fermat professionally was a lawyer and a judge and essentially only a mathematical amateur. He published almost nothing and his results and ideas are found in his own notes and journals as well as in correspondence with other mathematicians. Yet he had a profound effect on almost all branches of mathematics, not just number theory. He, as much as Descartes, developed analytic geometry. He did major work, prior to Newton and Leibniz, on the foundations of calculus. A series of letters between Fermat and Pascal established the beginnings of probability theory. In number theory, the work he did on factorization, congruences, and representations of integers by quadratic forms determined the direction of number theory until the nineteenth century. He did not supply proofs for most of his results but almost all of his work was subsequently proved (or shown to be false). The most difficult proved to be his big theorem which remained unproved until 1996. The attempts to prove this big theorem led to many advances in number theory including the development of algebraic number theory.

From the time of Fermat in the mid-seventeenth century through the eighteenth century a great deal of work was done in number theory but it was basically a series of somewhat disconnected, but often brilliant and startling, results. Important contributions were made by Euler, who proved and extended much of Fermat's results including Fermat's Two-Square Theorem (see Section 3.2). Euler also hinted at the law of quadratic reciprocity (see Section 2.6). This important result was eventually stated in its modern form by Legendre and the first complete proof was given by Gauss. During this period, certain problems were either stated or conjectured which became the basis for what is now known as **additive number theory**. The Goldbach conjecture and Waring's problem are two examples. We will not touch much on this topic in this book but refer an interested reader to [N].

In 1800 Gauss published a treatise on number theory called **Disquisitiones Arithmeticae**. This book not only standardized the notation used but also set the tone and direction for the theory of numbers up until the present. It is often joked that any new mathematical result is somehow inherent in the work of Gauss and in the case of number theory this is not really that far-fetched. Tremendous ideas and hints of things to come are present in Gauss' **Disquisitiones**. Gauss' work on number theory centered on three main concepts: the theory of congruences (see Chapter 2), the introduction of algebraic numbers (see Chapter 5) and the theory of forms, especially quadratic forms, and how these forms represent integers. Gauss, through his student Dirichlet, was also important in the infancy of analytic number theory. In 1837 Dirichlet proved, using analytic methods, that there are infinitely many primes in any arithmetic progression $\{a + nb; n \in \mathbb{N}\}$ with a, b relatively prime. We will discuss this result and its proof in Chapter 3. Euler and Legendre had both conjectured this theorem. Dirichlet's use of analysis really marks the beginning of analytic number theory. The main work in analytic number theory though, centered on the prime number theorem, was also conjectured by Gauss among others, including Euler and Legendre. This result deals with the asymptotic behavior of the function

$$\pi(x) = \text{ number of primes } \leq x.$$

The actual result says that

$$\lim_{x\to\infty} \frac{\pi(x)}{x/\ln x} = 1$$

and was proved in 1896 by Hadamard and independently by de la Valle Poussin. Both of their proofs used the behavior of the **Riemann zeta function**

$$\zeta(z) = \sum_{n=1}^{\infty} \frac{1}{n^z}$$

where $z = x + iy$ is a complex variable. Using this function, Riemann in 1859 attempted to prove the prime number theorem. In the attempted proof he hypothesized that all the zeros $z = x + iy$ of $\zeta(z)$ in the strip $0 \leq x \leq 1$ lie along the line $x = \frac{1}{2}$. This conjecture is known as the **Riemann hypothesis** and is still an open question.

Algebraic number theory also started basically with the work of Gauss. Gauss did an extensive study of the complex integers, that is the complex numbers of the form $a + bi$ with a, b integers. Today these are known as the **Gaussian integers**. Gauss proved that they satisfy most of the same properties as the ordinary integers including unique factorization into primes. In modern parlance he showed that they form a **unique factorization domain**. Gauss's algebraic integers were extended in many ways in attempt to prove Fermat's big theorem, and these extensions eventually developed into algebraic number theory. Kummer, a student of Gauss and Dirichlet, introduced in the 1840s a theory of algebraic integers and a set of ideal numbers from which unique factorization could be obtained. He used this to prove many cases of the Fermat theorem. Dedekind, in the 1870s, developed a further theory of algebraic numbers and unique factorization by ideals which extended both Gaussian integers and Kummer's algebraic and ideal numbers. Further work in the same area was done by Kronecker in the 1880s. We will discuss algebraic number theory and prime ideals in Chapter 6.

Modern number theory extends and uses all these classical ideas, although there have been many major new innovations. The close ties between number theory, especially diophantine analysis, and algebraic geometry led to Wiles' proof of the Fermat Theorem and to an earlier proof by Faltings of the Mordell conjecture, which is a related result. The vast area of mathematics used in both of these proofs is phenomenal. Probabilistic methods were incorporated into number theory by P. Erdos and studies in this area are known as **probabilistic number theory**. A great deal of recent work has gone into primality testing and factorization of large integers. These ideas have been incorporated extensively into cryptography (see [Ko]).

Chapter 2
Basic Number Theory

2.1 The Ring of Integers

The theory of numbers is concerned with the properties of the **integers**, that is, the class of whole numbers and zero, $0, \pm 1, \pm 2, \ldots$. We will denote the class of integers by $\mathbb{Z}$. The positive integers, $1, 2, 3, \ldots$ are called the **natural numbers**, which we will denote by $\mathbb{N}$. We will assume that the reader is familiar with the basic arithmetic properties of $\mathbb{Z}$ and in this section we will look at the abstract algebraic properties of the integers and what makes $\mathbb{Z}$ unique as an algebraic structure.

Recall that a **ring** R is a set with two binary operations, addition, denoted by $+$, and multiplication denoted by $\cdot$ or just by juxtaposition, defined on it satisfying the following six axioms:

1. Addition is commutative: $a + b = b + a$ for each pair a, b in R.
2. Addition is associative: $a + (b + c) = (a + b) + c$ for $a, b, c \in R$.
3. There exists an additive identity, denoted by 0, such that $a + 0 = a$ for each $a \in R$.
4. For each $a \in R$ there exists an additive inverse denoted $-a$, such that $a + (-a) = 0$.
5. Multiplication is associative: $a(bc) = (ab)c$ for $a, b, c \in R$.
6. Multiplication is distributive over addition: $a(b + c) = ab + ac$ and $(b + c)a = ba + ca$ for $a, b, c \in R$.

If in addition R satisfies

7. Multiplication is commutative: $ab = ba$ for each pair a, b in R

then R is a **commutative ring**, while if R satisfies

8. There exists a multiplicative identity denoted by 1 (not equal to 0) such that $a \cdot 1 = 1 \cdot a = a$ for each a in R

then R is a **ring with an identity**. A **commutative ring with identity** satisfies 1 through 8.

© Springer International Publishing AG 2016
B. Fine and G. Rosenberger, *Number Theory*,
DOI 10.1007/978-3-319-43875-7_2

A **field** K is a commutative ring with an identity in which every nonzero element has a multiplicative inverse, that is, for each $a \in K$ with $a \neq 0$ there exists an element $b \in K$ such that $ab = ba = 1$. In this case the set $K^\star = K \backslash \{0\}$ forms an abelian group with respect to the multiplication in K. $K^\star$ is called the **multiplicative group** of K.

A ring can be considered as the most basic algebraic structure in which addition, subtraction, and multiplication can be done. In any ring the equation $x + b = c$ can always be solved. Further a field can be considered as the most basic algebraic structure in which addition, subtraction, multiplication, and division can be done. Hence in any field, the equation $ax + b = c$ with $a \neq 0$ can always be solved.

Combining this definition with our knowledge of $\mathbb{Z}$ we get that

Lemma 2.1.1 *The integers $\mathbb{Z}$ form a commutative ring with identity.*

There are many examples of such rings (see Exercises), so to define $\mathbb{Z}$ uniquely we must introduce certain other properties. If two nonzero integers are multiplied together then the result is nonzero. This is not always true in a ring. For example, consider the set of functions defined on the interval [0, 1]. Under ordinary multiplication and addition, these form a ring (see Exercises) with the zero element being the function which is identically zero. Now let $f(x)$ be zero on $[0, \frac{1}{2}]$ and nonzero elsewhere and let $g(x)$ be zero on $[\frac{1}{2}, 0]$ and nonzero elsewhere. Then $f(x) \cdot g(x) = 0$ but neither is the zero function. We define an **integral domain** to be a commutative ring R with an identity and with the property that if $ab = 0$ with $a, b \in R$ then either $a = 0$ or $b = 0$. Two nonzero elements which multiply together to get zero are called **zero divisors** and hence an integral domain is a commutative ring with an identity and no zero divisors. Therefore, $\mathbb{Z}$ is an integral domain.

The integers are also ordered, that is, we can compare any two integers. We abstract this idea in the following manner. We say that an integral domain D is an **ordered integral domain** if there exists a distinguished set D^+, called the **set of positive elements**, with the properties that

(1) The set D^+ is closed under addition and multiplication.
(2) If $x \in D$ then exactly one of the following is true

   (a) $x = 0$
   (b) $x \in D^+$
   (c) $-x \in D^+$.

In any ordered integral domain D we can order the elements in the standard way. If $x, y \in D$ then $x < y$ means that $(y - x) \in D^+$. With this ordering D^+ can clearly be identified with those $x \in D$ such that $x > 0$. We then get

Lemma 2.1.2 *If D is an ordered integral domain then*
(1) $x < y$ and $y < z$ imply $x < z$.
(2) If $x, y \in D$ then exactly one of the following holds:

$$x = y \text{ or } x < y \text{ or } y < x.$$

We thus have that the integers are an ordered integral domain. Their uniqueness as such a structure depends on two additional properties of $\mathbb{Z}$ which are equivalent.

The Inductive Property *Let S be a subset of the natural numbers $\mathbb{N}$. Suppose $1 \in S$ and S has the property that if $n \in S$ then $(n+1) \in S$. Then $S = \mathbb{N}$.*

The Well-Ordering Property *Let S be a nonempty subset of the natural numbers $\mathbb{N}$. Then S has a least element.*

Lemma 2.1.3 *The inductive property is equivalent to the well-ordering property.*

Proof To prove this we must assume first the inductive property and show that the well-ordering property holds and then vice versa. Suppose the inductive property holds and let S be a nonempty subset of $\mathbb{N}$. We must show that S has a least element. Let T be the set

$$T = \{x \in \mathbb{N}; x \leq s, \forall s \in S\}.$$

Now $1 \in T$ since $S \subset \mathbb{N}$. If whenever $x \in T$ it would follow that $(x+1) \in T$ then by the inductive property $T = \mathbb{N}$ but then S would be empty contradicting that S is nonempty. Therefore, there exists an a with $a \in T$ and $(a+1) \notin T$. We claim that a is the least element of S. Now $a \leq s$ for all $s \in S$ since $a \in T$. If $a \notin S$ then every $s \in S$ would also satisfy $(a+1) \leq s$. This would imply that $(a+1) \in T$ a contradiction. Therefore, $a \in S$ and $a \leq s$ for all $s \in S$ and hence a is the least element. Therefore, the inductive property implies the well-ordering property.

Conversely, suppose that the well-ordering property holds and suppose $1 \in S$ and whenever $n \in S$ it follows that $(n+1) \in S$. We must show that $S = \mathbb{N}$. If $S \neq \mathbb{N}$ then $\mathbb{N} \backslash S$ is a nonempty subset of $\mathbb{N}$. Therefore, it must have a least element n. Hence $(n-1) \in S$. But then $(n-1)+1 = n \in S$, also which is a contradiction. Therefore, $\mathbb{N} \backslash S$ is empty and $S = \mathbb{N}$. □

The inductive property is of course the basis for **inductive proofs** which play a big role in the theory of numbers. To remind the reader, in an inductive proof we want to prove statements $\mathcal{P}(n)$ which depend on positive integers. In the induction we show that $\mathcal{P}(1)$ is true, then show that the truth of $\mathcal{P}(n+1)$ depends upon the truth of $\mathcal{P}(n)$. From the inductive property $\mathcal{P}(n)$ is then true for all positive integers n. We give an example which has an ancient history in number theory.

Example 2.1.1 Show that $1 + 2 + \cdots + n = \frac{(n)(n+1)}{2}$

Here for $n = 1$ we have $1 = \frac{(1)(2)}{2} = 1$. So its true for $n = 1$. Assume that the statement is true for $n = k$, that is

$$1 + 2 + \cdots + k = \frac{k(k+1)}{2}$$

and consider $n = k + 1$.

$$1+2+\cdots+k+(k+1) = (1+2+\cdots+k)+(k+1) = \frac{k(k+1)}{2}+(k+1) = \frac{(k+1)(k+2)}{2}$$

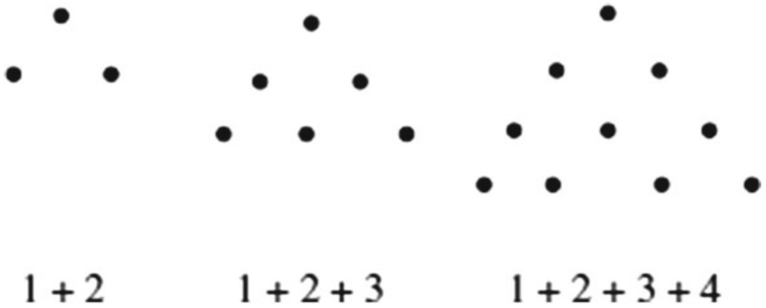


Fig. 2.1 Triangular Numbers

Hence the statement is true for $n = k + 1$ and hence true by induction for all $n \in \mathbb{N}$.

The series of integers

$$1, 1 + 2 = 3, 1 + 2 + 3 = 6, 1 + 2 + 3 + 4 = 10, \ldots$$

are called the **triangular numbers** since they are the sums of dots placed in triangular form as in Figure 2.1. These numbers were studied by the Pythagoreans in Greece in 500 B.C.

The inductive property is enough to characterize the integers among ordered integral domains up to **isomorphism.** Recall that if R and S are rings, a function $f : R \to S$ is a **homomorphism** if it satisfies:

1. $f(r_1 + r_2) = f(r_1) + f(r_2)$ for $r_1, r_2 \in R$.
2. $f(r_1 r_2) = f(r_1) f(r_2)$ for $r_1, r_2 \in R$.

If f is also a bijection, then f is an **isomorphism**, and R and S are **isomorphic**. Isomorphic algebraic structures are essentially algebraically the same. We have the following theorem.

Theorem 2.1.1 *Let R be an ordered integral domain which satisfies the inductive property (replacing $\mathbb{N}$ by the set of positive elements in R). Then R is isomorphic to $\mathbb{Z}$.*

We outline a proof in the exercises.

2.2 Divisibility, Primes, and Composites

The starting point for the theory of numbers is **divisibility**.

Definition 2.2.1 *If a, b are integers we say that a* **divides** *b, or that a is a* **factor** *or* **divisor** *of b, if there exists an integer q such that $b = aq$. We denote this by $a|b$. b*

is then a **multiple** *of* a*. If* $b > 1$ *is an integer whose only factors are* $\pm 1, \pm b$ *then* b *is a* **prime***, otherwise* $b > 1$ *is* **composite***.*

The following properties of divisibility are straightforward consequences of the definition:

Theorem 2.2.1 *(1)* $a|b \implies a|bc$ *for any integer* c*.*
(2) $a|b$ *and* $b|c$ *imply* $a|c$*.*
(3) $a|b$ *and* $a|c$ *imply that* $a|(bx + cy)$ *for any integers* x, y*.*
(4) $a|b$ *and* $b|a$ *imply that* $a = \pm b$*.*
(5) If $a|b$ *and* $a > 0, b > 0$ *then* $a \leq b$*.*
(6) $a|b$ *if and only if* $ca|cb$ *for any integer* $c \neq 0$*.*
(7) $a|0$ *for all* $a \in \mathbb{Z}$ *and* $0|a$ *only for* $a = 0$*.*
(8) $a| \pm 1$ *only for* $a = \pm 1$*.*
(9) $a_1|b_1$ *and* $a_2|b_2$ *imply that* $a_1a_2|b_1b_2$*.*

Proof We prove (2) and leave the remaining parts to the exercises.

Suppose $a|b$ and $b|c$. Then there exist x, y such that $b = ax$ and $c = by$. But then $c = axy = a(xy)$ and therefore $a|c$. □

If b, c, x, y are integers then an integer $bx + cy$ is called a **linear combination** of b, c. Thus part (3) of Theorem 2.2.1 says that if a is a **common divisor** of b, c then a divides any linear combination of b and c.

Further, note that if $b > 1$ is a composite then there exists $x > 0$ and $y > 0$ such that $b = xy$ and from part (5) we must have $1 < x < b, 1 < y < b$.

In ordinary arithmetic, given a, b we can always attempt to divide a into b. The next theorem, called the **division algorithm**, says that if $a > 0$ either a will divide b or the **remainder** of the division of b by a will be less than a.

Theorem 2.2.2 *(***Division Algorithm***) Given integers* a, b *with* $a > 0$ *then there exist unique integers* q *and* r *such that* $b = qa + r$ *where either* $r = 0$ *or* $0 < r < a$*.*

One may think of q and r as the **quotient** and **remainder**, respectively, when dividing b by a.

Proof Given a, b with $a > 0$ consider the set

$$S = \{b - qa \geq 0; q \in \mathbb{Z}\}.$$

If $b > 0$ then $b + a > 0$ and the sum is in S. If $b \leq 0$ then there exists a $q > 0$ with $-qa < b$. Then $b + qa > 0$ and is in S. Therefore, in either case S is nonempty. Hence S is a nonempty subset of $\mathbb{N} \cup \{0\}$ and therefore has a least element r. If $r \neq 0$ we must show that $0 < r < a$. Suppose $r \geq a$, then $r = a + x$ with $x \geq 0$ and $x < r$ since $a > 0$. Then $b - qa = r = a + x \implies b - (q + 1)a = x$. This means that $x \in S$. Since $x < r$ this contradicts the minimality of r which is a contradiction. Therefore, if $r \neq 0$ it follows that $0 < r < a$.

The only thing left is to show the uniqueness of q and r. Suppose $b = q_1 a + r_1$ also. By the construction above r_1 must also be the minimal element of S. Hence $r_1 \leq r$ and $r \leq r_1$ so $r = r_1$. Now

$$b - qa = b - q_1 a \implies (q_1 - q)a = 0$$

but since $a > 0$ it follows that $q_1 - q = 0$ so that $q = q_1$. □

The next ideas that are necessary are the concepts of **greatest common divisor** and **least common multiple**.

Definition 2.2.2 *Given nonzero integers a, b their* **greatest common divisor** *or* **GCD** $d > 0$ *is a positive integer which is a common divisor, that is, $d|a$ and $d|b$, and if d_1 is any other common divisor then $d_1|d$. We denote the greatest common divisor of a, b by either $gcd(a, b)$ or (a, b).*

The next result says that given any nonzero integers they do have a greatest common divisor and it is unique.

Theorem 2.2.3 *Given nonzero integers a, b their GCD exists, is unique, and can be characterized as the least positive linear combination of a and b.*

Proof Given nonzero a, b consider the set

$$S = \{ax + by > 0; x, y \in \mathbb{Z}\}$$

Now $a^2 + b^2 > 0$ so S is a nonempty subset of $\mathbb{N}$ and hence has a least element $d > 0$. We show that d is the GCD.

First, we must show that d is a common divisor. Now $d = ax + by$ and is the least such positive linear combination. By the division algorithm $a = qd + r$ with $0 \leq r < d$. Suppose $r \neq 0$. Then $r = a - qd = a - q(ax + by) = (1 - qx)a - qby > 0$. Hence r is a positive linear combination of a and b and therefore is in S. But then $r < d$ contradicting the minimality of d in S. It follows that $r = 0$ and so $a = qd$ and $d|a$. An identical argument shows that $d|b$ and so d is a common divisor of a and b. Let d_1 be any other common divisor of a and b. Then d_1 divides any linear combination of a and b and so $d_1|d$. Therefore, d is the GCD of a and b.

Finally, we must show that d is unique. Suppose d_1 is another GCD of a and b. Then $d_1 > 0$ and d_1 is a common divisor of a, b. Then $d_1|d$ since d is a GCD. Identically $d|d_1$ since d_1 is a GCD. Therefore, $d = \pm d_1$ and then $d = d_1$ since they are both positive. □

We note that as a consequence of Theorem 2.2.3 that if a, b, k are nonzero integers then the equation $ax + by = k$ has integer solutions x, y if and only if (a, b) divides k.

If $(a, b) = 1$ then we say that a, b are **relatively prime** or **coprime**. It follows that a and b are relatively prime if and only if 1 is expressible as a linear combination of a and b. We need the following three results:

Lemma 2.2.1 *If $d = (a, b)$ then $a = a_1 d$ and $b = b_1 d$ with $(a_1, b_1) = 1$.*

Proof If $d = (a, b)$ then $d|a$ and $d|b$. Hence $a = a_1 d$ and $b = b_1 d$. We have

$$d = ax + by = a_1 dx + b_1 dy.$$

Dividing both sides of the equation by d we obtain

$$1 = a_1 x + b_1 y.$$

Therefore, $(a_1, b_1) = 1$. □

Lemma 2.2.2 *For any integer c we have that $(a, b) = (a, b + ac)$.*

Proof Suppose $(a, b) = d$ and $(a, b + ac) = d_1$. Now d is the least positive linear combination of a and b. Suppose $d = ax + by$. d_1 is a linear combination of a, $b + ac$ so that

$$d_1 = ar + (b + ac)s = a(cs + r) + bs.$$

Hence d_1 is also a linear combination of a and b and therefore $d_1 \geq d$. On the other hand, $d_1|a$ and $d_1|(b + ac)$ and so $d_1|b$. Therefore, $d_1|d$ so $d_1 \leq d$. Combining these we must have $d_1 = d$. □

From this we easily see that $(a, b) = a$ if a, b are nonzero integers with $a|b$.

The next result, called the **Euclidean algorithm**, provides a technique for both finding the GCD of two integers and expressing the GCD as a linear combinations.

Theorem 2.2.4 *(***The Euclidean Algorithm***) Given integers b and $a > 0$ with $a \nmid b$ form the repeated divisions*

$$b = q_1 a + r_1, 0 < r_1 < a$$

$$a = q_2 r_1 + r_2, 0 < r_2 < r_1$$

$$\dots$$

$$r_{n-2} = q_n r_{n-1} + r_n, 0 < r_n < r_{n-1}$$

$$r_{n-1} = q_{n+1} r_n.$$

The last nonzero remainder r_n is the GCD of a, b. Further r_n can be expressed as a linear combination of a and b by successively eliminating the r_i's in the intermediate equations.

Proof In taking the successive divisions as outlined in the statement of the theorem each remainder r_i gets strictly smaller and still nonnegative. Hence it must finally end with a zero remainder. Therefore, there is a last nonzero remainder r_n. We must show that this is the GCD.

Now from Lemma 2.2.2, the GCD satisfies

$$(a, b) = (a, b - q_1 a) = (a, r_1) = (r_1, a - q_2 r_1) = (r_1, r_2).$$

Continuing in this manner we have then that $(a, b) = (r_{n-1}, r_n) = r_n$ since r_n divides r_{n-1}. This shows that r_n is the GCD.

To express r_n as a linear combination of a and b notice first that

$$r_n = r_{n-2} - q_n r_{n-1}.$$

Substituting this in the immediately preceding division we get

$$r_n = r_{n-2} - q_n(r_{n-3} - q_{n-1} r_{n-2}) = (1 + q_n q_{n-1}) r_{n-2} - q_n r_{n-3}.$$

Doing this successively, we ultimately express r_n as a linear combination of a and b. □

EXAMPLE 2.2.1 Find the GCD of 270 and 2412 and express it as a linear combination of 270 and 2412.

We apply the Euclidean algorithm

$$2412 = (8)(270) + 252$$

$$270 = (1)(252) + 18$$

$$252 = (14)(18)$$

Therefore, the last nonzero remainder is 18 which is the GCD. We now must express 18 as a linear combination of 270 and 2412.

From the first equation

$$252 = 2412 - (8)(270)$$

which gives in the second equation

$$270 = (2412 - (8)(270)) + 18 \implies 18 = (-1)(2412) + (9)(270)$$

which is the desired linear combination.

Now suppose that $d = (a, b)$ where $a, b \in \mathbb{Z}$ and $a \neq 0, b \neq 0$. Then we note that given one integer solution of the equation

$$ax + by = d$$

we can easily obtain all solutions.

Suppose without loss of generality that $d = 1$, that is, a, b are relatively prime. If not we can divide through by $d > 1$. Suppose that x_1, y_1 and x_2, y_2 are two integer solutions of the equation $ax + by = 1$, that is,

$$ax_1 + by_1 = 1$$

$$ax_2 + by_2 = 1.$$

Then

$$a(x_1 - x_2) = -b(y_1 - y_2).$$

Since $(a, b) = 1$ we get from Lemma 2.2.3 that $b|(x_1 - x_2)$ and hence $x_2 = x_1 + bt$ for some $t \in \mathbb{Z}$. Substituting back into the equations, we then get

$$ax_1 + by_1 = a(x_1 + bt) + by_2 \implies by_1 = abt + by_2.$$

Therefore, $y_2 = y_1 - at$. Hence all solutions are given by

$$x_2 = x_1 + bt$$

$$y_2 = y_1 - at$$

for some $t \in \mathbb{Z}$.

The final idea of this section is that of a **least common multiple**.

Definition 2.2.3 *Given nonzero integers a, b their* **least common multiple** *or* **LCM** *$m > 0$ is an positive integer which is a common multiple, that is, $a|m$ and $b|m$, and if m_1 is any other common multiple then $m|m_1$. We denote the least common multiple of a, b by either $lcm(a, b)$ or $[a, b]$.*

As for GCD's given any nonzero integers they do have a least common multiple and it is unique. First, we need the following result known as **Euclid's Lemma**. In the next section, we will use a special case of this applied to primes. We note that this special case is traditionally also called Euclid's lemma.

Lemma 2.2.3 *(**Euclid's Lemma**) Suppose $a|bc$ and $(a, b) = 1$, then $a|c$.*

Proof Suppose $(a, b) = 1$ then 1 is expressible as a linear combination of a and b. That is,

$$ax + by = 1.$$

Multiply by c, so that

$$acx + bcy = c.$$

Now $a|a$ and $a|bc$ so a divides the linear combination $acx + bcy$ and hence $a|c$. □

Theorem 2.2.5 *Given nonzero integers a, b their LCM exists and is unique. Further we have*

$$(a, b)[a, b] = ab.$$

Proof Let $d = (a, b)$ and let $m = |\frac{ab}{d}|$. We show that m is the LCM. Now $a = a_1 d, b = b_1 d$ with $(a_1, b_1) = 1$. Then $m = a_1 b_1 d$. Since $a = a_1 d$, $m = b_1 a$ so $a|m$. Identically, $b|m$ so m is a common multiple. Now let m_1 be another common multiple so that $m_1 = ax = by$. We then get

$$a_1 dx = b_1 dy \implies a_1 x = b_1 y \implies a_1 | b_1 y.$$

But $(a_1, b_1) = 1$ so from Lemma 2.2.3 $a_1 | y$. Hence $y = a_1 z$. It follows then that

$$m_1 = b_1 d(a_1 z) = a_1 b_1 dz = mz$$

and hence $m|m_1$. Therefore, m is an LCM.

The uniqueness follows in the same manner as the uniqueness of GCD's. Suppose m_1 is another LCM, then $m|m_1$ and $m_1|m$ so $m = \pm m_1$ and since they are both positive $m = m_1$. □

EXAMPLE 2.2.2 Find the LCM of 270 and 2412.
From Example 2.2.1, we found that (270, 2412) = 18. Therefore,

$$[270, 2412] = \frac{(270)(2412)}{(270, 2412)} = \frac{(270)(2412)}{18} = 36180.$$

2.3 The Fundamental Theorem of Arithmetic

In this section, we prove the fundamental theorem of arithmetic which is really the most basic number theoretic result. This result says that any integer $n > 1$ can be decomposed into prime factors in essentially a unique manner. First, we show that there always exists such a decomposition into prime factors.

Lemma 2.3.1 *Any integer $n > 1$ can be expressed as a product of primes, perhaps with only one factor.*

Proof The proof is by induction. $n = 2$ is prime so its true at the lowest level. Suppose that every integer $2 \leq k < n$ can be decomposed into prime factors, we must show that n then also has a prime factorization.

If n is prime then we are done. Suppose then that n is composite. Hence $n = m_1 m_2$ with $1 < m_1 < n$, $1 < m_2 < n$. By the inductive hypothesis both m_1 and m_2 can be expressed as products of primes. Therefore, n can also use the primes from m_1 and m_2, completing the proof. □

Before we continue to the fundamental theorem, we mention that this result can be used to prove that the set of primes is infinite. The proof we give goes back to Euclid and is quite straightforward. In the next chapter, we will present a whole collection of proofs, some quite complicated also show that the primes are an infinite set. Each of these other proofs will shed more light however on the nature of the integers.

Theorem 2.3.1 *There are infinitely many primes.*

Proof Suppose that there are only finitely many primes $p_1, \dots, p_n$. Each of these is positive so we can form the positive integer

$$N = p_1 p_2 \cdots p_n + 1.$$

From Lemma 2.3.1, N has a prime decomposition. In particular, there is a prime p which divides N. Then

$$p | (p_1 p_2 \cdots p_n + 1).$$

Since the only primes are assumed $p_1, p_2, \dots, p_n$ it follows that $p = p_i$ for some $i = 1, \dots, n$. But then $p | p_1 p_2 \cdots p_i \cdots p_n$ so p cannot divide $p_1 \cdots p_n + 1$ which is a contradiction. Therefore, p is not one of the given primes showing that the list of primes must be endless. □

A variation of Euclid's argument gives the following proof of Theorem 2.3.1. Suppose there are only finitely many primes $p_1, \dots, p_n$. Certainly $n \geq 2$. Let $P = \{p_1, \dots, p_n\}$. Divide P into two disjoint nonempty subsets P_1, P_2. Now consider the number $m = q_1 + q_2$ where q_i is a product of primes from P_1 and q_2 is a product of primes from P_2. Let p be a prime divisor of m. Since $p \in P$ it follows that p divides either q_1 or q_2 but not both. But then p does not divide m a contradiction. Therefore, p is not one of the given primes and the number of primes must be infinite.

Although there are infinitely many primes, a glance at the list of primes, shows that they appear to become scarcer as the integers get larger. If we let

$$\pi(x) = \text{ number of primes } \leq x$$

a basic question is what is the asymptotic behavior of this function. This question is the basis of the prime number theorem which will be discussed in Chapter 4. However, it is easy to show that there are arbitrarily large spaces or gaps within the set of primes.

Theorem 2.3.2 *Given any positive integer k there exists k consecutive composite integers.*

Proof Consider the sequence

$$(k+1)! + 2, (k+1)! + 3, \dots, (k+1)! + k + 1.$$

Suppose n is an integer with $2 \leq n \leq k+1$. Then $n | ((k+1)! + n)$. Hence each of the integers in the above sequence is composite. □

To show the uniqueness of the prime decomposition we need Euclid's Lemma, from the previous section, applied to primes.

Lemma 2.3.2 *(***Euclid's Lemma***) If p is a prime and $p|ab$ then $p|a$ or $p|b$.*

Proof Suppose $p|ab$. If p does not divide a then clearly a and p must be relatively prime, that is, $(a, p) = 1$. Then from Lemma 2.2.3, $p|b$. □

We now state and prove the **fundamental theorem of arithmetic**.

Theorem 2.3.3 *(***The Fundamental Theorem of Arithmetic***) Given any integer $n \neq 0$ there is a factorization*

$$n = cp_1 p_2 \cdots p_k$$

where $c = \pm 1$ and $p_1, \ldots, p_n$ are primes. Further this factorization is unique up to the ordering of the factors.

Proof We assume that $n \geq 1$. If $n \leq -1$ we use $c = -1$ and the proof is the same. We define the product of no primes, that is, when $k = 0$, to be 1. Then the statement certainly holds for $n = 1$ with $k = 0$. Now suppose $n > 1$. From Lemma 2.3.1, n has a prime decomposition

$$n = p_1 p_2 \cdots p_m.$$

We must show that this is unique up to the ordering of the factors. Suppose then that n has another such factorization $n = q_1 q_2 \cdots q_k$ with the q_i all prime. We must show that $m = k$ and that the primes are the same. Now we have

$$n = p_1 p_2 \cdots p_m = q_1 \cdots q_k$$

Assume that $k \geq m$. Then it follows that $p_1 | q_1 q_2 \cdots q_k$. From Lemma 2.3.2, then we must have that $p_1 | q_i$ for some i. But q_i is prime and $p_1 > 1$ so it follows that $p_1 = q_i$. Therefore, we can eliminate p_1 and q_i from both sides of the factorization to obtain

$$p_2 \cdots p_m = q_1 \cdots q_{i-1} q_{i+1} \cdots q_k.$$

Continuing in this manner, we can eliminate all the p_i from the left side of the factorization to obtain

$$1 = q_{i_1} \cdots q_{i_t}, \text{ with } t = k - m$$

If $q_{i_1}, \ldots, q_{i_t}$ were primes this would be impossible. Therefore, $m = k$ and each prime p_i was included in the primes $q_1, \ldots, q_m$ and vice versa. Therefore, the factorizations differ only in the order of the factors, proving the theorem. □

For any positive integer $n > 1$ we can combine all the same primes to write

$$n = p_1^{m_1} p_2^{m_2} \cdots p_k^{m_k} \text{ with } p_1 < p_2 < \cdots < p_k.$$

This is called the **standard prime decomposition**. Note that given any two positive integers a, b we can always write the prime decomposition with the **same** primes by allowing a zero exponent.

There are several easy consequences of the fundamental theorem.

Theorem 2.3.4 *Let a, b be positive integers > 1. Suppose*

$$a = p_1^{e_1} \cdots p_k^{e_k}$$

$$b = p_1^{f_1} \cdots p_k^{f_k}$$

where we include zero exponents for noncommon primes. Then

$$(a, b) = p_1^{min(e_1, f_1)} \cdot p_2^{min(e_2, f_2)} \cdots p_k^{min(e_k, f_k)}$$

$$[a, b] = p_1^{max(e_1, f_1)} \cdot p_2^{max(e_2, f_2)} \cdots p_k^{max(e_k, f_k)}$$

Corollary 2.3.1 *Let a, b be positive integers > 1, then $(a, b)[a, b] = ab$.*

We leave the proofs to the exercises but give an example.

EXAMPLE 2.3.1 Find the standard prime decompositions of 270 and 2412 and use them to find the GCD and LCM.

Recall that we found the GCD and LCM of these numbers in the previous section using the Euclidean algorithm. We note that in general it is very difficult as the size gets larger to determine the actual prime decomposition or even whether it is a prime or not. We will discuss primality testing in Chapter 5.

To find the prime decomposition we factor and then continue refactoring until there are only prime factors.

$$270 = (27)(10) = 3^3 \cdot 2 \cdot 5 = 2 \cdot 3^3 \cdot 5$$

which is the standard prime decomposition of 270.

$$2412 = 4 \cdot 603 = 4 \cdot 3 \cdot 201 = 4 \cdot 3 \cdot 3 \cdot 67 = 2^2 \cdot 3^2 \cdot 67$$

which is the standard prime decomposition of 2412. Hence we have

$$270 = 2 \cdot 3^3 \cdot 5 \cdot 67^0$$

$$2412 = 2^2 \cdot 3^2 \cdot 5^0 \cdot 67$$

$$\Longrightarrow \ (a, b) = 2 \cdot 3^2 \cdot 5^0 \cdot 67^0 = 2 \cdot 3^2 = 18$$

and

$$[a,b] = 2^2 \cdot 3^3 \cdot 5 \cdot 67 = 36180.$$

Note that the fundamental theorem of arithmetic can be extended to the rational numbers. Suppose $r = \frac{a}{b}$ with $a > 0, b \neq 0$ is a positive rational. Then

$$r = \frac{p_1^{e_1} \cdots p_k^{e_k}}{p_1^{f_1} \cdots p_k^{f_k}} = p_1^{e_1 - f_1} \cdots p_k^{e_k - f_k}.$$

Therefore, any positive rational has a standard prime decomposition

$$p_1^{t_1} \cdots p_k^{t_k} \text{ where } t_1, \ldots, t_k \text{ are integers.}$$

So, for example,

$$\frac{15}{49} = 3 \cdot 5 \cdot 7^{-2}.$$

This has the following interesting consequence.

Lemma 2.3.3 *If a is an integer which is not a perfect nth power then the nth root of a is irrational.*

Proof This result says, for example, that if an integer is not a perfect square then its square root is irrational. The fact that the square root of 2 is irrational was known to the Greeks.

Suppose b is an integer with standard prime decomposition

$$b = p_1^{e_1} \cdots p_k^{e_k}.$$

Then

$$b^n = p_1^{ne_1} \cdots p_k^{ne_k}$$

and this must be the standard prime decomposition for b^n. It follows that an integer a is an nth power if and only if it has a standard prime decomposition

$$a = q_1^{f_1} \cdots q_t^{f_t} \text{ with } n | f_i \text{ for every } i.$$

Suppose a is not an nth power then

$$a = q_1^{f_1} \cdots q_t^{f_t}$$

where n does not divide f_i for some i. Taking the nth root

$$a^{1/n} = q_1^{f_1/n} \cdots q_i^{f_i/n} \cdots q_t^{f_t/n}$$

But f_i/n is not an integer so $a^{1/n}$ cannot be rational by the extension of fundamental theorem to rationals. □

While induction and least well-ordering characterize the integers, unique factorization into primes does not. We close this section with a brief further discussion of unique factorization.

The concept of divisor and factor can be extended to any ring. $a|b$ is a ring R if there is a $c \in R$ with $b = ac$. We will restrict ourselves to integral domains. A **unit** in an integral domain is an element e with a multiplicative inverse. This means that there is an element e_1 in R with $ee_1 = 1$. Thus the only units in $\mathbb{Z}$ are ± 1. Two elements r, r_1 of an integral domain are **associates** if $r = er_1$ for some unit e. A **prime** in a general integral domain is an element whose only divisors are associates of itself or units. With these definitions, we can talk about factorization into primes.

We say that an integral domain D is a **unique factorization domain** or **UFD** if for each $d \in D$ then either $d = 0$, d is a unit or d has a factorization into primes which is unique up to ordering and unit factors. This means that if

$$r = p_1 \cdots p_m = q_1 \cdots q_k$$

then $m = k$ and each p_i is an associate of some q_j.

The fundamental theorem of arithmetic in more general algebraic language says that the integers $\mathbb{Z}$ are a unique factorization domain. However, they are far from being the only one. In the exercises, we outline a proof of the following.

Theorem 2.3.5 *Let F be a field and $F[x]$ the ring of polynomials in one variable over F. Then $F[x]$ is a UFD.*

This theorem is actually a special case of something even more general. An integral domain D is called a **Euclidean domain** if there exists a function $N : D\backslash\{0\} \to \mathbb{N} \cup \{0\}$ satisfying:

$$\text{For each } a, b \in D, a \neq 0 \text{ there exists } q, r \in D \text{ such that}$$

$$b = aq + r \text{ and either } r = 0 \text{ or } r \neq 0 \text{ and } N(r) < N(a).$$

Theorem 2.3.6 *Any Euclidean domain is a UFD.*

The proof of this essentially mimics the proof for the integers. See the exercises.

The **Gaussian integers** $\mathbb{Z}[i]$ are the complex numbers $a + bi$ where a, b are integers.

Lemma 2.3.4 *The integer $\mathbb{Z}$, the Gaussian integers $\mathbb{Z}[i]$, and the ring of polynomials $F[x]$ over a field F are all Euclidean domains.*

Corollary 2.3.2 *$\mathbb{Z}[i]$ and $F[x]$ with F, a field, are UFDs.*

Proofs of these results will be given in Chapter 6.

2.4 Congruences and Modular Arithmetic

Gauss based much of his number theoretical investigations around the theory of congruences. As we will see a **congruence** is just a statement about divisibility put into a more formal framework. In this section and the remainder of the chapter, we will consider congruences and in particular the solution of polynomial congruences. First, we give the basic definitions and properties.

2.4.1 Basic Theory of Congruences

Definition 2.4.1 *Suppose m is a positive integer. If x, y are integers such that $m|(x-y)$ we say that x is* **congruent to** *y* **modulo** *m and denote this by $x \equiv y \bmod m$. If m does not divide $x - y$ then x and y are* **incongruent** *modulo m.*

If $x \equiv y \bmod m$ then y is called a **residue** of x modulo m. Given $x \in \mathbb{Z}$ the set of integers $\{y \in Z; x \equiv y \bmod m\}$ is called the **residue class** for x modulo m. We denote this by $[x]$. Notice that $x \equiv 0 \bmod m$ is equivalent to $m|x$. We first show that the residue classes partition $\mathbb{Z}$, that is, each integer falls in one and only one residue class.

Theorem 2.4.1 *Given $m > 0$ then congruence modulo m is an equivalence relation on the integers. Therefore, the residue classes partition the integers.*

Proof Recall that a relation $\sim$ on a set S is an **equivalence relation** if it is **reflexive**, that is, $s \sim s$ for all $s \in S$; **symmetric**, that is, if $s_1 \sim s_2$ then $s_2 \sim s_1$; and **transitive**, that is, if $s_1 \sim s_2$ and $s_2 \sim s_3$ then $s_1 \sim s_3$. If $\sim$ is an equivalence relation then the equivalence classes $[s] = \{s_1 \in S; s_1 \sim s\}$ partition S.

Consider $\equiv \bmod m$ on $\mathbb{Z}$. Given $x \in \mathbb{Z}$, $x - x = 0 = 0 \cdot m$ so $m|(x - x)$ and $x \equiv x \bmod m$. Therefore, $\equiv \bmod m$ is reflexive.

Suppose $x \equiv y \bmod m$ then $m|(x - y) \implies x - y = am$ for some $a \in \mathbb{Z}$. Then $y - x = -am$ so $m|(y - x)$ and $y \equiv x \bmod m$. Therefore, $\equiv \bmod m$ is symmetric.

Finally suppose $x \equiv y \bmod m$ and $y \equiv z \bmod m$. Then $x - y = a_1m$ and $y - z = a_2m$. But then $x - z = (x - y) + (y - z) = a_1m + a_2m = (a_1 + a_2)m$. Therefore, $m|(x - z)$ and $x \equiv z \bmod m$. Therefore, $\equiv \bmod m$ is transitive and the theorem is proved. □

Hence given $m > 0$ every integer falls into one and only one residue class. We now show that there are exactly m residue classes modulo m.

Theorem 2.4.2 *Given $m > 0$ there exist exactly m residue classes. In particular,*

$$[0], [1], \ldots, [m-1]$$

gives a complete set of residue classes.

Proof We show that given $x \in \mathbb{Z}$, x must be congruent modulo m to one of $0, 1, 2, \ldots, m-1$. Further none of these are congruent modulo m. As a consequence

$$[0], [1], \ldots, [m-1]$$

give a complete set of residue classes modulo m and hence there are m of them.

To see these assertions suppose $x \in \mathbb{Z}$. By the division algorithm, we have

$$x = qm + r \text{ where } 0 \leq r < m$$

This implies that $r = x - qm$ or in terms of congruences that $x \equiv r \bmod m$. Therefore, x is congruent to one of the sets $\{0, 1, 2, \ldots, m-1\}$.

Suppose $0 \leq r_1 < r_2 < m$. Then $m \nmid r_2 - r_1$ so r_1 and r_2 are incongruent modulo m. Therefore, every integer is congruent to one and only one of $0, 1, \ldots, m-1$, and hence $[0], [1], \ldots, [m-1]$ give a complete set of residue classes modulo m. □

There are many sets of complete residue classes modulo m. In particular, a set of m integers $x_1, x_2, \ldots, x_m$ will comprise a **complete residue system** modulo m if $x_i \neq x_j \bmod m$ unless $i = j$. Given one complete residue system, it is easy to get another.

Lemma 2.4.1 *If* $\{x_1, \ldots, x_m\}$ *form a complete residue system modulo m and* $(a, m) = 1$ *then* $\{ax_1, \ldots, ax_m\}$ *also comprise a complete residue system.*

Proof Suppose $ax_i \equiv ax_j \bmod m$. Then $m|a(x_i - x_j)$. Since $(a, m) = 1$ then by Euclid's lemma $m|(x_i - x_j)$ and hence $x_i \equiv x_j \bmod m$. □

Finally, we will need the following:

Lemma 2.4.2 *If* $x \equiv y$ *mod m then* $(x, m) = (y, m)$.

Proof Suppose $x - y = am$ then any common divisor of x and m is also a common divisor of y. From this the result is immediate. □

2.4.2 The Ring of Integers Mod N

Perhaps the easiest way to handle results on congruences is to place them in the framework of abstract algebra. To do this we construct, for each $n > 0$ a ring, called the **ring of integers modulo n**. We will follow this approach. However we note, that although this approach simplifies and clarifies many of the proofs, historically purely number theoretical proofs were given. Often these purely number theoretical proofs inspired the algebraic proofs.

To construct this ring, we first need the following:

Lemma 2.4.3 *If $a \equiv b$ mod n and $c \equiv d$ mod n then*

1. $a + c \equiv b + d$ *mod* n
2. $ac \equiv bd$ *mod* n

Proof Suppose $a \equiv b \bmod n$ and $c \equiv d \bmod n$ then $a - b = q_1 n$ and $c - d = q_2 n$ for some integers q_1, q_2. This implies that $(a + c) - (b + d) = (q_1 + q_2)n$ or that $n|((a + c) - (b + d))$. Therefore, $a + c \equiv b + d \bmod n$.

We leave the proof of (2) to the exercises. □

We now define operations on the set of residue classes.

Definition 2.4.2 *Consider a complete residue system $x_1, \dots, x_n$ modulo n. On the set of residue classes $[x_1], \dots, [x_n]$ define*

1. $[x_i] + [x_j] = [x_i + x_j]$
2. $[x_i][x_j] = [x_i x_j]$

Theorem 2.4.3 *Given a positive integer $n > 0$, the set of residue classes forms a commutative ring with an identity under the operations defined in Definition 2.4.2. This is called the* **ring of integers modulo n** *and is denoted by $\mathbb{Z}_n$. The zero element is* $[0]$ *and the identity element is* $[1]$.

Proof Notice that from Lemma 2.4.3, it follows that these operations are well-defined on the set of residue classes, that is, if we take two different representatives for a residue class, the operations are still the same.

To show $\mathbb{Z}_n$ is a commutative ring with an identity we must show that it satisfies, relative to the defined operations, all the ring properties. Basically, $\mathbb{Z}_n$ inherits these properties from $\mathbb{Z}$. We show commutativity of addition and leave the other properties to the exercises.

Suppose $[a], [b] \in \mathbb{Z}_n$. Then

$$[a] + [b] = [a + b] = [b + a] = [b] + [a]$$

where $[a + b] = [b + a]$ since addition is commutative in $\mathbb{Z}$. □

This theorem is actually a special case of a general result in abstract algebra. In the ring of integers $\mathbb{Z}$ the set of multiples of an integer n forms an ideal (see [A] for terminology) which is usually denoted $n\mathbb{Z}$. The ring $\mathbb{Z}_n$ is the **quotient ring** of $\mathbb{Z}$ modulo the ideal $n\mathbb{Z}$, that is, $\mathbb{Z}/n\mathbb{Z} \cong \mathbb{Z}_n$.

We usually consider $\mathbb{Z}_n$ as consisting of $0, 1, \dots, n - 1$ with addition and multiplication **modulo n**. When there is no confusion we will denote the element $[a]$ in $\mathbb{Z}_n$ just as a. Below we give the addition and multiplication table modulo 5, that is, in $\mathbb{Z}_5$.

EXAMPLE 2.4.2.1 Addition and Multiplication Tables for $\mathbb{Z}_5$

+	0	1	2	3	4
0	0	1	2	3	4
1	1	2	3	4	0
2	2	3	4	0	1
3	3	4	0	1	2
4	4	0	1	2	3

.	0	1	2	3	4
0	0	0	0	0	0
1	0	1	2	3	4
2	0	2	4	1	3
3	0	3	1	4	2
4	0	4	3	2	1

Notice, for example, that modulo 5, $3 \cdot 4 = 12 \equiv 2 \bmod 5$ so that in $\mathbb{Z}_5$, $3 \cdot 4 = 2$. Similarly, $4 + 2 = 6 \equiv 1 \bmod 5$ so in $\mathbb{Z}_5$, $4 + 2 = 1$.

The question arises as to when the commutative ring $\mathbb{Z}_n$ is an integral domain and when is $\mathbb{Z}_n$ a field. The answer is when n is a prime and only when n is a prime.

Theorem 2.4.4 *(1) $\mathbb{Z}_n$ is an integral domain if and only if n is a prime.*
(2) $\mathbb{Z}_n$ is a field if and only if n is a prime.

Proof Since $\mathbb{Z}_n$ is a commutative ring with an identity for any n it will be an integral domain if and only if it has no zero divisors.

Suppose first that n is a prime and suppose that $ab = 0$ in $\mathbb{Z}_n$. Then in $\mathbb{Z}$ we have

$$ab \equiv 0 \bmod n \implies n|ab.$$

Since n is prime, by Euclid's lemma $n|a$ or $n|b$. In terms of congruences then

$$a \equiv 0 \bmod n \implies a = 0 \text{ in } \mathbb{Z}_n \text{ or } b \equiv 0 \bmod n \implies b = 0 \text{ in } \mathbb{Z}_n$$

Therefore, $\mathbb{Z}_n$ is an integral domain if n is prime.

Suppose n is not prime. Then $n = m_1 m_2$ with $1 < m_1 < n$, $1 < m_2 < n$. Then $n \nmid m_1$, $n \nmid m_2$ but $n|m_1 m_2$. Translating this into $\mathbb{Z}_n$, we have

$$m_1 m_2 = 0 \text{ but } m_1 \neq 0 \text{ and } m_2 \neq 0.$$

Therefore, $\mathbb{Z}_n$ is not an integral domain if n is not prime. These prove part (1).

Since a field is an integral domain, $\mathbb{Z}_n$ cannot be a field unless n is prime. To complete part (2), we must show that if n is prime then $\mathbb{Z}_n$ is a field. Suppose n is prime, since $\mathbb{Z}_n$ is a commutative ring with identity to show that its a field we must show that each nonzero element has a multiplicative inverse.

Suppose $a \in \mathbb{Z}_n, a \neq 0$. Then in $\mathbb{Z}$ we have $n \nmid a$ and hence since n is prime $(a, n) = 1$. Therefore, in $\mathbb{Z}$ there exists x, y such that $ax + ny = 1$. In terms of congruences this says that

$$ax \equiv 1 \bmod n$$

or in $\mathbb{Z}_n$,

$$ax = 1.$$

Therefore, a has an inverse in $\mathbb{Z}_n$ and hence $\mathbb{Z}_n$ is a field. □

The proof of the last theorem actually indicates a method to find the multiplicative inverse of an element modulo a prime. Suppose n is a prime and $a \neq 0$ in $\mathbb{Z}_n$. Use the Euclidean algorithm in $\mathbb{Z}$ to express 1 as a linear combination of a and n, that is,

$$ax + ny = 1.$$

The residue class for x will be the multiplicative inverse of a.

EXAMPLE 2.4.2.2 Find 6^{-1} in $\mathbb{Z}_{11}$.

Using the Euclidean algorithm

$$11 = 1 \cdot 6 + 5$$

$$6 = 1 \cdot 5 + 1$$

$$\implies 1 = 6 - (1 \cdot 5) = 6 - (1 \cdot (11 - 1 \cdot 6) \implies 1 = 2 \cdot 6 - 1 \cdot 11.$$

Therefore, the inverse of 6 modulo 11 is 2, that is, in $\mathbb{Z}_{11}$, $6^{-1} = 2$.

EXAMPLE 2.4.2.3 Solve the linear equation

$$6x + 3 = 1$$

in $\mathbb{Z}_{11}$.

Using purely formal field algebra, the solution is

$$x = 6^{-1}(1 - 3).$$

In $\mathbb{Z}_{11}$ we have

$$1 - 3 = -2 = 9 \text{ and } 6^{-1} = 2 \implies x = 2 \cdot 9 = 18 = 7.$$

Therefore, the solution in $\mathbb{Z}_{11}$ is $x = 7$. A quick check shows that

$$6 \cdot 7 + 3 = 42 + 3 = 45 = 1 \text{ in } \mathbb{Z}_{11}.$$

A linear equation in $\mathbb{Z}_{11}$ is called a **linear congruence** modulo 11. We will discuss solutions of such congruences in Section 2.5.

The fact that $\mathbb{Z}_p$ is a field for p a prime leads to the following nice result known as **Wilson's theorem**.

Theorem 2.4.5 *(Wilson's Theorem) If p is a prime then*

$$(p-1)! \equiv -1 \mod p.$$

Proof Now $(p-1)! = (p-1)(p-2)\cdots 1$. Since $\mathbb{Z}_p$ is a field each $x \in \{1, 2, \ldots, p-1\}$ has a multiplicative inverse modulo p. Further suppose $x = x^{-1}$ in $\mathbb{Z}_p$. Then $x^2 = 1$ which implies $(x-1)(x+1) = 0$ in $\mathbb{Z}_p$ and hence either $x = 1$ or $x = -1$ since $\mathbb{Z}_p$ is an integral domain. Therefore, in $\mathbb{Z}_p$ only $1, -1$ are their own multiplicative inverses. Further $-1 = p - 1$ since $p - 1 \equiv -1 \bmod p$.

Hence in the product $(p-1)(p-2)\cdots 1$ considered in the field $\mathbb{Z}_p$ each element is paired up with its distinct multiplicative inverse except 1 and $p-1$. Further the product of each with its inverse is 1. Therefore, in $\mathbb{Z}_p$ we have $(p-1)(p-2)\cdots 1 = p-1$. Written as a congruence then

$$(p-1)! \equiv p - 1 \equiv -1 \bmod p.$$

□

The converse of Wilson's theorem is also true, that is, if $(n-1)! \equiv -1 \bmod n$, then n must be a prime.

Theorem 2.4.6 *If $n > 1$ is a natural number and*

$$(n-1)! \equiv -1 \textit{ mod } n$$

then n is a prime.

Proof Suppose $(n-1)! \equiv -1 \bmod n$. If n were composite then $n = mk$ with $1 < m < n-1$ and $1 < k < n-1$. If $m \neq k$ then both m and k are included in $(n-1)!$. It follows that $(n-1)!$ is divisible by n so that $(n-1)! \equiv 0 \bmod n$ contradicting the assertion that $(n-1)! \equiv -1 \bmod n$. If $m = k \neq 2$ then $(n-1)! \equiv 0 \bmod m$ which is not congruent to $-1 \bmod m$. Therefore, n must be prime. If $m = k = 2$ then $n = 4$ and $(n-1)! = 6$ which is not congruent to $-1 \bmod 4$. □

2.4.3 Units and the Euler Phi Function

In a field F every nonzero element has a multiplicative inverse. If R is a commutative ring with an identity, not necessarily a field, then a **unit** is any element with a multiplicative inverse. In this case its inverse is also a unit. For example, in the integers $\mathbb{Z}$ the only units are ± 1. The set of units in a commutative ring with identity form an abelian group under ring multiplication called the **unit group** of R. Recall that a **group** G is a set with one operation which is associative, has an identity for that operation, and such that each element has an inverse with respect to this operation. If the operation is also commutative then G is an **abelian group**.

Lemma 2.4.4 *If R is a commutative ring with an identity then the set of units in R form an abelian group under ring multiplication. This is called the* **unit group** *of R denoted $U(R)$.*

Proof The commutativity and associativity of $U(R)$ follow from the ring properties. The identity of $U(R)$ is the multiplicative identity of R while the ring multiplicative inverse for each unit is the group inverse. We must show that $U(R)$ is closed under ring multiplication. If $a \in R$ is a unit we denote its multiplicative inverse by a^{-1}. Now suppose $a, b \in U(R)$. Then a^{-1}, b^{-1} exist. It follows that

$$(ab)(b^{-1}a^{-1}) = a(bb^{-1})a^{-1} = aa^{-1} = 1.$$

Hence ab has an inverse, namely $b^{-1}a^{-1}$ $(= a^{-1}b^{-1}$ in a commutative ring) and hence ab is also a unit. Therefore, $U(R)$ is closed under ring multiplication. □

The proof of Theorem 2.4.4 actually provides a method to classify the units in any $\mathbb{Z}_n$.

Lemma 2.4.5 *$a \in \mathbb{Z}_n$ is a unit if and only if $(a, n) = 1$.*

Proof Suppose $(a, n) = 1$. Then there exists $x, y \in \mathbb{Z}$ such that $ax + ny = 1$. This implies that $ax \equiv 1 \bmod n$ which in turn implies that $ax = 1$ in $\mathbb{Z}_n$ and therefore a is a unit.

Conversely, suppose a is a unit in $\mathbb{Z}_n$. Then there is an $x \in \mathbb{Z}_n$ with $ax = 1$. In terms of congruence then

$$ax \equiv 1 \bmod n \implies n|(ax - 1) \implies ax - 1 = ny \implies ax - ny = 1.$$

Therefore, 1 is a linear combination of a and n and so $(a, n) = 1$. □

If a is a unit in $\mathbb{Z}_n$ then a linear equation

$$ax + b = c$$

can always be solved with a unique solution given by $x = a^{-1}(c - b)$. Determining this solution is the same technique as in $\mathbb{Z}_p$ with p a prime. If a is not a unit the situation is more complicated. We will consider this case in Section 2.5.

EXAMPLE 2.4.3.1

Solve $5x + 4 = 2$ in $\mathbb{Z}_6$.

Since $(5, 6) = 1, 5$ is a unit in $\mathbb{Z}_6$. Therefore, $x = 5^{-1}(2-4)$. Now $2-4 = -2 = 4$ in $\mathbb{Z}_6$. Further $5 = -1$ so $5^{-1} = -1^{-1} = -1$. Then we have

$$x = 5^{-1}(2 - 4) = -1(4) = -4 = 2$$

Thus the unique solution in $\mathbb{Z}_6$ is $x = 2$.

Since an element a is a unit in $\mathbb{Z}_n$ if and only if $(a, n) = 1$ it follows that the number of units in $\mathbb{Z}_n$ is equal to the number of positive integers less than or equal to n and relatively prime to n. This number is given by the **Euler Phi Function**, our first look at a number theoretical function.

Definition 2.4.3 *For any $n > 0$,*

$$\phi(n) = \textit{ number of integers less than or equal to } n \textit{ and relatively prime to } n.$$

EXAMPLE 2.4.3.2
$\phi(6) = 2$ since among $1, 2, 3, 4, 5, 6$ only $1, 5$ are relatively prime to 6.

The following is immediate from our characterization of units:

Lemma 2.4.6 *The number of units in $\mathbb{Z}_n$, which is the order of the unit group $U(\mathbb{Z}_n)$, is $\phi(n)$.*

Definition 2.4.4 *Given $n > 0$ a* **reduced residue system** *modulo n is a set of integers $x_1, \dots, x_k$ such that each x_i is relatively prime to n, $x_i \not\equiv x_j \bmod n$ unless $i = j$ and if $(x, n) = 1$ for some integer x then $x \equiv x_i \bmod n$ for some i.*

Hence a reduced residue system is a complete collection of representatives of those residue classes of integers relatively prime to n. Hence it is a complete collection of units (up to congruence modulo n) in $\mathbb{Z}_n$. It follows that any reduced residue system modulo n has $\phi(n)$ elements.

EXAMPLE 2.4.3.3
A reduced residue system modulo 6 would be $\{1, 5\}$.

We now develop a formula for $\phi(n)$. As is the theme of this book, we first determine a formula for prime powers and then paste back together via the fundamental theorem of arithmetic.

Lemma 2.4.7 *For any prime p and $m > 0$,*

$$\phi(p^m) = p^m - p^{m-1} = p^m(1 - \frac{1}{p}).$$

Proof Recall that if $1 \le a \le p$ then either $a = p$ or $(a, p) = 1$. It follows that the positive integers less than p^m which are not relatively prime to p^m are precisely the multiples of p, that is, $p, 2p, 3p, \dots, p^{m-1}p$. All other positive $a < p^m$ are relatively prime to p^m. Hence the number of positive integers less than p^m and relatively prime to p^m is

$$p^m - p^{m-1}.$$

□

Lemma 2.4.8 *If $(a, b) = 1$ then $\phi(ab) = \phi(a)\phi(b)$.*

Proof Let $R_a = \{x_1, \dots, x_{\phi(a)}\}$ be a reduced residue system modulo a, $R_b = \{y_1, \dots, y_{\phi(b)}\}$ be a reduced residue system modulo b, and let

$$S = \{ay_i + bx_j; i = 1, \ldots, \phi(b), j = 1, \ldots, \phi(a)\}.$$

We claim that S is a reduced residue system modulo ab. Since S has $\phi(a)\phi(b)$ elements it will follow that $\phi(ab) = \phi(a)\phi(b)$.

To show that S is a reduced residue system modulo ab we must show three things: first, each $x \in S$ is relatively prime to ab; second, the elements of S are distinct; and finally, given any integer n with $(n, ab) = 1$ then $n \equiv s \bmod ab$ for some $s \in S$.

Let $x = ay_i + bx_j$. Then since $(x_j, a) = 1$ and $(a, b) = 1$ it follows that $(x, a) = 1$. Analogously, $(x, b) = 1$. Since x is relatively prime to both a and b we have $(x, ab) = 1$. This shows that each element of S is relatively prime to ab.

Next suppose that

$$ay_i + bx_j \equiv ay_k + bx_l \bmod ab.$$

Then

$$ab|((ay_i + bx_j) - (ay_k + bx_l)) \implies ay_i \equiv ay_k \bmod b.$$

Since $(a, b) = 1$ it follows that $y_i \equiv y_k \bmod b$. But then $y_i = y_k$ since R_b is a reduced residue system. Similarly, $x_j = x_l$. This shows that the elements of S are distinct modulo ab.

Finally, suppose $(n, ab) = 1$. Since $(a, b) = 1$ there exist x, y with $ax + by = 1$. Then

$$anx + bny = n.$$

Since $(x, b) = 1$ and $(n, b) = 1$ it follows that $(nx, b) = 1$. Therefore, there is an s_i with $nx = s_i + tb$. In the same manner $(ny, a) = 1$ and so there is an r_j with $ny = r_j + ua$. Then

$$a(s_i + tb) + b(r_j + ua) = n \implies n = as_i + br_j + (t + u)ab$$

$$\implies n \equiv as_i + br_j \bmod ab$$

and we are done. □

We now give the general formula for $\phi(n)$.

Theorem 2.4.7 *Suppose $n = p_1^{e_1} \cdots p_k^{e_k}$ then*

$$\phi(n) = (p_1^{e_1} - p_1^{e_1-1})(p_2^{e_2} - p_2^{e_2-1}) \cdots (p_k^{e_k} - p_k^{e_k-1}) = n \prod_i (1 - 1/p_i).$$

Proof From the previous lemma, we have

$$\phi(n) = \phi(p_1^{e_1})\phi(p_2^{e2}) \cdots \phi(p_k^{e_k})$$

$$= (p_1^{e_1} - p_1^{e_1-1})(p_2^{e_2} - p_2^{e_2-1}) \cdots (p_k^{e_k} - p_k^{e_k-1})$$

$$= p_1^{e_1}(1-1/p_1)\cdots p_k^{e_k}(1-1/p_k) = p_1^{e_1}\cdots p_k^{e_k}\cdot(1-1/p_1)\cdots(1-1/p_k)$$

$$= n\prod_i (1-1/p_i).$$

□

EXAMPLE 2.4.3.4
Determine $\phi(126)$. Now

$$126 = 2\cdot 3^2\cdot 7 \implies \phi(126) = \phi(2)\phi(3^2)\phi(7) = (1)(3^2-3)(6) = 36.$$

Hence there are 36 units in $\mathbb{Z}_{126}$.

An interesting result with many generalizations which we will look at later is the following.

Theorem 2.4.8 *For $n > 1$ and for $d \geq 1$*

$$\sum_{d|n}\phi(d) = n.$$

Proof As before we first prove the theorem for prime powers and then paste together via the fundamental theorem of arithmetic.

Suppose that $n = p^e$ for p a prime. Then the divisors of n are $1, p, p^2, \ldots, p^e$, so

$$\sum_{d|n}\phi(d) = \phi(1)+\phi(p)+\phi(p^2)+\cdots+\phi(p^e) = 1+(p-1)+(p^2-p)+\cdots+(p^e-p^{e-1}).$$

Notice that this sum telescopes, that is, $1+(p-1) = p$, $p+(p^2-p) = p^2$ and so on. Hence the sum is just p^e and the result is proved for n a prime power.

We now do an induction on the number of distinct prime factors of n. The above argument shows that the result is true if n has only one distinct prime factor. Assume that the result is true whenever an integer has less than k distinct prime factors and suppose $n = p_1^{e_1}\cdots p_k^{e_k}$ has k distinct prime factors. Then $n = p^e c$ where $p = p_1, e = e_1$ and c has fewer than k distinct prime factors. By the inductive hypothesis,

$$\sum_{d|c}\phi(d) = c.$$

Since $(c, p) = 1$ the divisors of n are all of the form $p^\alpha d_1$ where $d_1|c$ and $\alpha = 0, 1, \ldots, e$. It follows that

$$\sum_{d|n}\phi(d) = \sum_{d_1|c}\phi(c) + \sum_{d_1|c}\phi(pd_1) + \cdots + \sum_{d_1|c}\phi(p^e d_1)$$

Since $(d_1, p^\alpha) = 1$ for any divisor of c this sum equals

$$= \sum_{d_1|c} \phi(c) + \sum_{d_1|c} \phi(p)\phi(d_1) + \cdots + \sum_{d_1|c} \phi(p^e)\phi(d_1)$$

$$= \sum_{d_1|c} \phi(c) + (p-1)\sum_{d_1|c} \phi(d_1) + \cdots + (p^e - p^{e-1})\sum_{d_1|c} \phi(d_1)$$

$$= c + (p-1)c + (p^2 - p)c + \cdots + (p^e - p^{e-1})c$$

As in the case of prime powers this sum telescopes giving a final result

$$= p^e c = n.$$

□

EXAMPLE 2.4.3.5

Consider $n = 10$. The divisors are $1, 2, 5, 10$. Then $\phi(1) = 1, \phi(2) = 1, \phi(5) = 4, \phi(10) = 4$. Then

$$\phi(1) + \phi(2) + \phi(5) + \phi(10) = 1 + 1 + 4 + 4 = 10.$$

2.4.4 Fermat's Little Theorem and the Order of an Element

For any positive integer n the unit group $U(\mathbb{Z}_n)$ is a finite abelian group. Recall that in any group G each element $g \in G$ generates a **cyclic subgroup** consisting of all the distinct powers of g. If this cyclic subgroup is finite of order m then m is called the **order** of the element g. Equivalently, the order of an element $g \in G$ can be described as the least positive power m such that $g^m = 1$. If no such power exists then g has infinite order. We denote the order of the group G by $|G|$ and the order of $g \in G$ by $|g|$. If the whole group G is finite then each element clearly has finite order. We will apply these ideas to the unit group $U(\mathbb{Z}_n)$ but first we recall some further facts about finite groups.

Theorem 2.4.9 *(Lagrange's Theorem) Suppose G is a finite group of order n. Then the order of any subgroup divides n. In particular, the order of any element divides the order of the group.*

If $g \in G$ with $|G| = n$ then from Lagrange's theorem above there is an m with $g^m = 1$ and $m|n$. Hence $n = mk$ and so $g^n = g^{mk} = (g^m)^k = 1^k = 1$. Hence in any finite group, we have the following:

Corollary 2.4.1 *If G is a finite group of order n and $g \in G$ then $g^n = 1$.*

Theorem 2.4.10 *Let G be a finite abelian group with $|G| = n$ then*

1. *if $g_1, g_2 \in G$ with $|g_1| = a, |g_2| = b$ then $(g_1g_2)^{lcm(a,b)} = 1$,*

2. *if* $g_1, g_2 \in G$ *with* $|g_1| = a$, $|g_2| = b$ *and* $(a, b) = 1$ *then* $|g_1 g_2| = ab$,
3. *if* $n = p_1^{e_1} p_2^{e_2} \cdots p_k^{e_k}$ *is the prime factorization of* n *then*

$$G = H_1 \times H_2 \times \cdots \times H_k$$

where $|H_i| = p_i^{e_i}$.

The third part of the last theorem is part of the **Fundamental Theorem for Finitely Generated Abelian Groups** which plays the same role in abelian group theory as the fundamental theorem of arithmetic does in number theory.

With these facts in hand, consider a unit $a \in \mathbb{Z}_n$. Then $a \in U(\mathbb{Z}_n)$ and hence a has a **multiplicative order**, that is, there is an integer m with $a^m = 1$ in $\mathbb{Z}_n$. In terms of congruences this means that $a^m \equiv 1 \bmod n$. If $a \in \mathbb{Z}_n$ is not a unit then there cannot exist a power $m \geq 1$ such that $a^m \equiv 1 \bmod n$ for if such an m existed then a^{m-1} would be an inverse for a.

Lemma 2.4.9 *Given* $n > 0$ *then for an integer* a *there exists an integer* m *such that* $a^m \equiv 1$ *mod* n *if and only if* $(a, n) = 1$ *or equivalently* a *is a unit in* $\mathbb{Z}_n$.

Definition 2.4.5 *If* $(a, n) = 1$ *then the* **order** *of* a *modulo* n *is the least positive power* m *such that* $a^m \equiv 1$ *mod* n. *We will write* $order(a)$ *or alternatively* $| < a > |$ *or* $|a|$ *for the order of* a. *Equivalently, the order of* a *is the order of* a *considered as an element of the unit group* $U(\mathbb{Z}_n)$.

Since the order of $U(\mathbb{Z}_n) = \phi(n)$ we immediately get that the order of any element modulo n must divide $\phi(n)$.

Lemma 2.4.10 *If* $(a, n) = 1$ *then* $order(a)|\phi(n)$.

Applying Corollary 2.4.1 to the unit group $U(\mathbb{Z}_n)$ we get the following result, known as **Euler's theorem**.

Theorem 2.4.11 *(Euler's Theorem) If* $(a, n) = 1$ *then*

$$a^{\phi(n)} \equiv 1 \textit{ mod } n.$$

If $n = p$ a prime then any integer $a \neq 0 \bmod p$ is a unit in $\mathbb{Z}_p$. Further $\phi(p) = p - 1$, and hence we obtain the next corollary which is called **Fermat's theorem**. (This is often called **Fermat's Little theorem** to distinguish it from the result on $x^n + y^n = z^n$.)

Corollary 2.4.2 *If* p *is a prime and* $p \nmid a$ *then*

$$a^{p-1} \equiv 1 \textit{ mod } p.$$

If $(a, n) = 1$ and the order of a is exactly $\phi(n)$ then a is called a **primitive root** modulo n. In this case, the unit group is cyclic with a as a generator. For $n = p$ a prime there is always a primitive root.

Theorem 2.4.12 *For a prime p there is always an element a of order $\phi(p) = p-1$, that is, a primitive root. Equivalently, the unit group of $\mathbb{Z}_p$ is always cyclic.*

Proof Since every nonzero element in $\mathbb{Z}_p$ is a unit, the unit group $U(\mathbb{Z}_p)$ is precisely the multiplicative group of the field $\mathbb{Z}_p$. The fact that $U(\mathbb{Z}_p)$ is cyclic follows from the following more general result whose proof is also given. □

Theorem 2.4.13 *Let F be a field. Then any finite subgroup of the multiplicative group of F must be cyclic.*

Proof Suppose $G \subset F$ is a finite multiplicative subgroup of the multiplicative group of F. Suppose $|G| = n$. As has been our general mode of approaching results we will prove it for n a power of a prime and then paste the result together via the fundamental theorem of arithmetic.

Suppose $n = p^k$ for some k. Then the order of any element in G is p^α with $\alpha \le k$. Suppose the maximal order is p^t with $t < k$. Then the lcm of the orders is p^t. It follows that for every $g \in G$ we have $g^{p^t} = 1$. Therefore, every $g \in G$ is a root of the polynomial equation

$$x^{p^t} - 1 = 0.$$

However, over a field a polynomial cannot have more roots than its degree. Since G has $n = p^k$ elements and $p^t < p^k$, this is a contradiction. Therefore, the maximal order must be $p^k = n$. Therefore, G has an element of order $n = p^k$ and hence this element generates G and G must be cyclic.

We now do an induction on the number of distinct prime factors in $n = |G|$. The above argument handles the case where there is only one distinct prime factor. Assume the result is true if the order of G has less than k distinct prime factors. Suppose $n = p_1^{e_1} \cdots p_k^{e_k}$. Then $n = p^e c$ where c has less than k distinct prime factors. Since G is a finite abelian group with

$$|G| = n = p^e c \implies G = H \times K \text{ with } |H| = p^e,\ |K| = c.$$

By the inductive hypothesis, H and K are both cyclic so H has an element h of order p^e and K has an element k of order c. Since $(p^e, c) = 1$ the element hk has order $p^e c = n$ completing the proof. □

EXAMPLE 2.4.4.1 Determine a primitive root modulo 7.

This is equivalent to finding a generator for the multiplicative group of $\mathbb{Z}_7$. The nonzero elements are 1, 2, 3, 4, 5, 6 and we are looking for an element of order 6.

The table below list these elements and their orders

x	1	2	3	4	5	6
$\|x\|$	1	3	6	3	6	2

Therefore, there are two primitive roots 3 and 5 modulo 7. To see how these were determined powers were taken modulo 7 until a value of 1 was obtained. For example,

$$3^2 = 9 = 2, 3^3 = 2 \cdot 3 = 6, 3^4 = 3 \cdot 6 = 18 = 4, 3^5 = 3 \cdot 4 = 12 = 5,$$

$$3^6 = 3 \cdot 5 = 15 = 1$$

EXAMPLE 2.4.4.2 Show that there is no primitive root modulo 15.

The units in $\mathbb{Z}_{15}$ are $\{1, 2, 4, 7, 8, 11, 13, 14\}$. Since $\phi(15) = 8$ we must show that there is no element of order 8. The table below gives the units and their respective orders.

x	1	2	4	7	8	11	13	14
$\|x\|$	1	4	2	4	4	2	4	2

Therefore, there is no element of order 8.

Modulo a prime, there is always a primitive root but other integers can have primitive roots also. The fundamental result describing when an integer will have a primitive root is the following. We outline the proof in the exercises.

Theorem 2.4.14 *An integer n will have a primitive root modulo n if and only if*

$$n = 2, 4, p^k, 2p^k,$$

where p is a prime.

The order of an element, especially Fermat's theorem, provides a method for **primality testing**. Primality testing refers to determining for a given integer n whether it is prime or not. The simplest primality test is the following. If n were composite then $n = m_1 m_2$ with $1 < m_1 < n$, $1 < m_2 < n$. At least one of these factors must be $\leq \sqrt{n}$. Therefore, check all the integers less than or equal to the $\sqrt{n}$. If none of these divides n then n is prime. This can be improved using the fundamental theorem of arithmetic. If n has a divisor $\leq \sqrt{n}$ then it has a prime divisor $\leq \sqrt{n}$. It follows that in the above divisibility check, only the primes $\leq \sqrt{n}$ need be checked.

While this method always works it is often impractical for large n and other methods must be employed to see if a number is prime. By Fermat's theorem, if n were prime and $a < n$ then $a^{n-1} \equiv 1 \bmod n$. If a number a is found where this is not true then a cannot be prime. We give a trivial example.

EXAMPLE 2.4.4.3 Determine if 77 is prime.

If 77 were prime then $2^{76} \equiv 1 \bmod 77$. Now

$$2^{76} = 2^{38 \cdot 2} = 4^{38}.$$

Now we do computations mod 77

$$4^3 = 64 = -13 \implies 4^6 = 169 = 15 \implies 4^{12} = 225 = 71 = -6$$

$$\implies 4^{36} = (-6)^3 = -216 = -62 \implies 4^{38} = 4^2(-62) = -992 = -68 \neq 1.$$

Therefore, 77 is not prime.

This method can determine if a number n is **not** prime however it cannot determine if it is prime. There are numbers n for which $a^{n-1} \equiv 1 \bmod n$ is true for all $(a, n) = 1$ but n is not prime. These are called **pseudoprimes**. We will discuss primality testing further and in more detail in Chapter 5.

2.4.5 On Cyclic Groups

In the previous sections, we used some material from abstract algebra to prove results in number theory. Here we briefly reverse the procedure to use some number theory to develop and prove other ideas from algebra. After we do this we will turn the tables back again and use this algebra to give another proof of Theorem 2.4.8 on the Euler phi function.

Recall that a cyclic group G is a group with a a single generator say g. We denote a cyclic group G with generator g by $< g >$. The group G then consists of all the powers of g, that is, $G = \{1, g^{\pm 1}, g^{\pm 2}, \dots\}$. If G is finite of order n then $g^n = 1$ and n is the least positive integer x such that $g^x = 1$. It is then clear that if $g^m = 1$ for some power m it must follow that $m \equiv 0 \bmod n$, and if $g^k = g^l$ then $k \equiv l \bmod n$.

Let $H = (\mathbb{Z}_n, +)$ denote the additive subgroup of $\mathbb{Z}_n$. Then H is cyclic of order n with generator 1. If $G = < g >$ is also cyclic of order n then since multiplication of group elements is done via addition of exponents, it is fairly straightforward that the homomorphism $f : G \to (\mathbb{Z}_n, +)$ given by $g \to 1$ is actually an isomorphism (see the exercises). Further if $G = < g >$ is cyclic of infinite order then $g \to 1$ gives an isomorphism from G to the additive group of $\mathbb{Z}$.

Lemma 2.4.11 *(1) If G is a finite cyclic group of order n then G is isomorphic to* $(\mathbb{Z}_n, +)$. *In particular all finite cyclic groups of a given order are isomorphic.*
(2) If G is an infinite cyclic group then G is isomorphic to $(\mathbb{Z}, +)$.

Cyclic groups are abelian and hence their subgroups are also abelian. However as an almost direct consequence of the division algorithm, we get that any subgroup of a cyclic group must be cyclic.

Lemma 2.4.12 *Let G be a cyclic group. Then any subgroup of G is also cyclic.*

Proof Suppose $G = < g >$ and $H \subset G$ is a subgroup. Since G consists of powers of g, H also consists of certain powers of g. Let k be the least positive integer such that $g^k \in H$. We show that $H = < g^k >$, that is, H is the cyclic subgroup generated by g^k. This is clearly equivalent to showing that every $h \in H$ must be a power of g^k.

Suppose $g^t \in H$. We may assume that $t > 0$ and that $t > k$ since k is the least positive integer such that $g^k \in H$. If $t < 0$ work with $-t$. By the division algorithm, we then have

$$t = qk + r \text{ with } r = 0 \text{ or } 0 < r < k.$$

If $r \neq 0$ then $0 < r < k$ and $r = t - k$. Hence $g^r = g^{t-k} = g^t g^{-k}$. Now $g^t \in H$ and $g^k \in H$ and since H is a subgroup it follows that $g^{t-k} \in H$. But then $g^r \in H$ which is a contradiction since $0 < r < k$ and k is the least power of g in H. Therefore, $r = 0$ and $t = qk$. We then have

$$g^t = g^{qk} = (g^k)^q$$

completing the proof. □

Each element of a cyclic group G generates its own cyclic subgroup. The question is when does this cyclic subgroup coincide with all of G. In particular, which powers g^k are generators of G. The answer is purely number theoretic.

Lemma 2.4.13 *(1) Let $G = < g >$ be a finite cyclic group of order n. Then g^k with $k > 0$ is a generator of G if and only if $(k, n) = 1$, that is, k and n are relatively prime.*

(2) If $G = < g >$ is an infinite cyclic group then g, g^{-1} are the only generators.

Proof Suppose first that $G = < g >$ is finite cyclic of order n and suppose that $(k, n) = 1$. Then there exists integers x, y such that $kx + ny = 1$. It follows then that

$$g = g^1 = g^{kx+ny} = g^{kx}g^{ny} = (g^k)^x(g^n)^y.$$

But $g^n = 1$ so $(g^n)^y = 1$ and therefore

$$g = (g^k)^x.$$

Therefore, g is a power of g^k and hence every power of g is also a power of g^k. The whole group g then consists of powers of g^k and hence g^k is a generator for G.

Conversely, suppose that g^k is also a generator for G. Then there exists a power x such that $g = (g^k)^x = g^{kx}$. Hence $kx \equiv 1 \bmod n$ and so k is a unit mod n which implies from the last section that $(k, n) = 1$.

Suppose next that $G = < g >$ is infinite cyclic. Then there is no power of g which is the identity. Suppose g^k is also a generator with $k > 1$. Then there exists a power x such that $g = (g^k)^x = g^{kx}$. But this implies that $g^{kx-1} = 1$ contradicting that no power of g is the identity. Hence $k = 1$. □

Recall that $\phi(n)$ denotes the number of positive integers less than n which are relatively prime to n. This is then the number of generators of a cyclic group of order n.

Corollary 2.4.3 *Let G be a finite cyclic group of order n. Then there are $\phi(n)$ generators for G.*

By Lagrange's theorem (Theorem 2.4.9) for any finite group the order of a subgroup divides the order of a group, that is, if $|G| = n$ and $|H| = d$ with H a subgroup of G then $d|n$. However, the converse in general is not true, that is, if $|G| = n$ and $d|n$ there need not be a subgroup of order d. Further if there is a subgroup of order d there may or may not be other subgroups of order d. For a finite cyclic group G of order n however there is for each $d|n$ a **unique** subgroup of order d.

Theorem 2.4.15 *Let G be a finite cyclic group of order n. Then for each $d|n$ with $d \geq 1$ there exists a unique subgroup H of order d.*

Proof Let $G = < g >$ and $|G| = n$. Suppose $d|n$, then $n = kd$. Consider the element g^k. Then $(g^k)^d = g^{kd} = g^n = 1$. Further if $0 < t < d$ then $0 < kt < kd$ so $kt \not\equiv 0 \bmod n$ and hence $g^{kt} = (g^k)^t \neq 1$. Therefore, d is the least power of g^k which is the identity and hence g^k has order d and generates a cyclic subgroup of order d. We must show that this is unique.

Suppose $H = < g^t >$ is another cyclic subgroup of order d (recall that all subgroups of G are also cyclic). We may assume that $t > 0$ and we show that g^t is a power of g^k and hence the subgroups coincide. The proof is essentially the same as the proof of Lemma 2.4.12.

Since H has order d we have $g^{td} = 1$ which implies that $td \equiv 0 \bmod n$. Since $n = kd$ it follows that $t > k$. Apply the division algorithm

$$t = qk + r \text{ with } 0 \leq r < k.$$

If $r \neq 0$ then $0 < r < k$ and $r = t - qk$. Then

$$r = t - qk \implies rd = td - qkd \equiv 0 \bmod n.$$

Hence $n|rd$ which is impossible since $rd < kd = n$. Therefore, $r = 0$ and $t = qk$. From this

$$g^t = g^{qk} = (g^k)^q.$$

Therefore, g^t is a power of g^k and $H = < g^k >$. □

We now use this result to give an alternate proof of Theorem 2.4.8.

Theorem 2.4.16 *For $n > 1$ and for $d \geq 1$*

$$\sum_{d|n} \phi(d) = n.$$

Proof Consider a cyclic group G of order n. For each $d|n$, $d \geq 1$ there is a unique cyclic subgroup H of order d. H then has $\phi(d)$ generators. Each element in G

generates its own cyclic subgroup H_1, say of order d and hence must be included in the $\phi(d)$ generators of H_1. Therefore,

$$\sum_{d|n} \phi(d) = \text{ sum of the numbers of generators of the cyclic subgroups of } G.$$

But this must be the whole group and hence this sum is n. □

2.5 The Solution of Polynomial Congruences Modulo *m*

We are interested in solving **polynomial congruences** mod m. That is, solving polynomial equations

$$f(x) \equiv 0 \bmod m$$

where $f(x)$ is a nonzero polynomial with coefficients in $\mathbb{Z}_m$, the ring of integers modulo m. Typical examples might be

$$4x^2 + 3x - 2 \equiv 0 \bmod 12 \text{ or } 4x + 5 \equiv 0 \bmod 7.$$

Of course the solution of such congruences is given in terms of residue classes for if $x \equiv y \bmod m$ then $f(x) \equiv f(y) \bmod m$. Hence if x is a solution to a polynomial congruence then so is every integer congruent to its modulo m.

As has been our general procedure, we will reduce the solution of polynomial congruences to the solution modulo primes and then try to paste general solutions back together via the fundamental theorem of arithmetic. Suppose then that m has the prime factorization $m = p_1^{e_1} p_2^{e_2} \cdots p_k^{e_k}$ and that x_0 is a solution of $f(x) \equiv 0 \bmod m$. Then x_0 is also a solution of $f(x) \equiv 0 \bmod p_i^{e_i}$ for $i = 1, \ldots, k$. Then for each $i = 1, \ldots, k$ there is a y_i with $x_0 \equiv y_i \bmod p_i^{e_i}$. Conversely, suppose we are given y_i with $f(y_i) \equiv 0 \bmod p_i^{e_i}$ for $i = 1, \ldots, k$ then there is a technique based on what is called the **Chinese remainder theorem**, which we will discuss shortly, to piece these y_i together to get a solution x_0 of $f(x) \equiv 0 \bmod m$.

As a first step, we will describe the solution of linear congruences and the Chinese remainder theorem and then move on to higher degree congruences.

2.5.1 Linear Congruences and the Chinese Remainder Theorem

A **linear congruence** is of the form $ax + b = 0 \bmod m$ where $a \not\equiv 0 \bmod m$. In this section, we will consider solutions of linear congruences.

Before proceeding further, we note that solving a polynomial congruence

$$f(x) \equiv 0 \bmod m$$

is essentially equivalent to solving a polynomial equation

$$f(x) = 0$$

in the modular ring $\mathbb{Z}_m$. The solutions of the congruence are precisely the congruence classes modulo m.

For example, the congruence

$$2x \equiv 4 \bmod 5$$

is equivalent to the equation

$$2x = 4$$

in $\mathbb{Z}_5$. The unique solution in $\mathbb{Z}_5$ is $x = 2$, so that the solution of the congruence is $x \equiv 2 \bmod 5$. We will move freely between the two approaches to solving congruences, using $\equiv$ for congruence mod m and $=$ for equality in $\mathbb{Z}_m$.

Now we consider the linear congruence $ax + b \equiv 0 \bmod m$ where a is noncongruent to 0 mod m. For $m = p$, p a prime, the solution is immediate and it is unique. Since $\mathbb{Z}_p$ is a field and $a \neq 0$ the element a has an inverse. Therefore, the solution in $\mathbb{Z}_p$ is

$$x = a^{-1}(-b)$$

and any solution x_0 must be of the form $x_0 \equiv a^{-1}(-b) \bmod p$.

EXAMPLE 2.5.1.1 Solve $3x + 4 \equiv 0 \bmod 7$.

From the formal field properties, the solution is $x = 3^{-1} \cdot (-4)$. In $\mathbb{Z}_7$ we have $-4 = 3$ and since $3 \cdot 5 \equiv 1 \bmod 7$ it follows that $3^{-1} = 5$. Therefore, the solution is $x = 5 \cdot 3 = 15 \equiv 1 \bmod 7$.

Essentially the same method works if m is not prime but $(a, m) = 1$. In this case a is a unit in $\mathbb{Z}_m$ and the unique solution is $x = a^{-1}(-b)$. Consider the same equation as in Example 2.5.1.1 but modulo 8, that is

$$3x + 4 \equiv 0 \bmod 8 \implies x \equiv 3^{-1} \cdot (-4) \bmod 8.$$

However, modulo 8 we have $-4 = 4$ and $3^{-1} = 3$ so the solution is $x = 4{\cdot}3 = 12 = 4 \bmod 8$.

If $(a, m) \neq 1$ the situation becomes more complicated. We have the following theorem which describes the solutions and provides a technique for finding all solutions.

Theorem 2.5.1 *Consider $ax + b = 0$ mod m with $(a, m) = d > 1$. Then the congruence is solvable if and only if $d|b$. In this case there are exactly d solutions that are given by*

$$x = x_0 + \frac{tm}{d}, t = 0, 1, \dots, d-1$$

where x_0 is any solution of the reduced equation

$$\frac{a}{d}x + \frac{b}{d} \equiv 0 \, mod \, \frac{m}{d}.$$

Proof Let $d = (a, m)$. If x_0 is a solution then $b = -ax_0$ mod m or $b = -ax_0 + tm$ for some t. Therefore, $d|b$. Hence if d does not divide b there is no solution.

Suppose then that $d|b$. Then $(\frac{a}{d}, \frac{m}{d}) = 1$ and the reduced congruence

$$\frac{a}{d}x + \frac{b}{d} \equiv 0 \text{ mod } \frac{m}{d}$$

has a unique solution (mod $\frac{m}{d}$) say x_0. But then x_0 is also a solution mod m of the original congruence. Any integer x congruent to x_0 modulo $\frac{m}{d}$ and hence of the form $x = x_0 + \frac{tm}{d}$ is also a solution to the reduced congruence. However only d of these are incongruent modulo m. It is easy to check that each of $x = x_0 + \frac{tm}{d}$, $t = 0, 1, \dots, d-1$ are incongruent modulo m. □

The problem of solving a linear congruence is then reduced to finding a single solution of a congruence of the form $ax = b$ mod m with $(a, m) = 1$. The solution is then $x = a^{-1}b$ where a^{-1} is the inverse of a mod m. As explained in Section 2.4.3 this can be found using the Euclidean algorithm.

EXAMPLE 2.5.1.2 Solve $26x + 81 = 0$ mod 245

We apply the Euclidean algorithm to both determine if $(26, 245) = 1$ and if so to find the inverse of 26 mod 245

$$245 = (9)(26) + 11$$

$$26 = (2)(11) + 4$$

$$11 = (2)(4) + 3$$

$$4 = (1)(3) + 1.$$

Therefore, $(245, 26) = 1$. Working backward, we express 1 as a linear combination of 26 and 245

$$1 = 4 - (1)(3) = 4 - (11 - (2)(4)) = (3)(4) - (1)(11) = \cdots = (66)(26) - (7)(245)$$

Hence modulo 245 we have $66 \cdot 26 = 1$ and $26^{-1} = 66$. Therefore, the solution is

$$x = (26^{-1})(-81) \implies x = (66)(164) = 10824 \equiv 44 \bmod 245.$$

EXAMPLE 2.5.1.3 Solve $78x + 243 \equiv 0 \bmod 735$.

Using the Euclidean algorithm, we find that $(78, 735) = 3$ and $3|243$. The reduced congruence is

$$\frac{78}{3}x + \frac{243}{3} = 0 \bmod \frac{735}{3} \implies 26x + 81 \equiv 0 \bmod 245.$$

From the previous example, the solution to the reduced congruence is $x_0 = 44$ with $d = 3$. The solutions then mod 735 would be

$$x_0 + \frac{tm}{d}, t = 0, 1, \ldots, d-1 \implies x = 44 + \frac{735t}{3}, t = 0, 1, 2$$

$$\implies x \equiv 44, 289, 534 \bmod 735$$

The methods above provide techniques for solving linear congruences. Systems of linear congruences are handled by the next result which is called the **Chinese remainder theorem**.

Theorem 2.5.2 *(Chinese Remainder Theorem) Suppose that $m_1, m_2, \ldots, m_k$ are k positive integers that are relatively prime in pairs. If $a_1, \ldots, a_k$ are any integers then the simultaneous congruences*

$$x \equiv a_i \; mod \; m_i, i = 1, \ldots, k$$

have a common solution which is unique modulo $m_1 m_2 \cdots m_k$.

Proof The proof we give not only provides a verification but also provides a technique for finding the common solution.

Let $m = m_1 m_2 \cdots m_k$. Since the m_i are relatively prime in pairs we have $(\frac{m}{m_i}, m_i) = 1$. Therefore, there is a solution x_i to the reduced congruence

$$\frac{m}{m_i}x_i \equiv 1 \bmod m_i.$$

Further for x_i we clearly have

$$\frac{m}{m_j}x_i \equiv 0 \bmod m_i \text{ if } i \neq j.$$

Now let

$$x_0 = \sum_{i=1}^{k} \frac{m}{m_i} x_i a_i .$$

We claim that x_0 is a solution to the simultaneous congruences and that it is unique modulo m.

Now

$$x_0 = \sum_{i=1}^{k} \frac{m}{m_i} x_i a_i \equiv \frac{m}{m_j} x_j a_j \bmod m_j$$

sincex $\frac{m}{m_i} x_i \equiv 0 \bmod m_j$ if $i \neq j$. It follows then that

$$x_0 \equiv \frac{m}{m_j} x_j a_j \bmod m_j \equiv a_j \bmod m_j$$

since $\frac{m}{m_j} x_j \equiv 1 \bmod m_j$. Therefore, x_0 is a common solution. We must show the uniqueness part.

If x_1 is another common solution then $x_1 \equiv x_0 \bmod m_i$ for $i = 1, \ldots, k$. Therefore, $x_1 \equiv x_0 \bmod m$.

We note that if the integers m_i are not relatively prime in pairs there may be no solution to the simultaneous congruences. □

EXAMPLE 2.5.1.4 Solve the simultaneous congruences

$$x \equiv 6 \bmod 13$$

$$x \equiv 9 \bmod 45$$

$$x \equiv 12 \bmod 17.$$

Here $m_1 = 13, m_2 = 45, m_3 = 17$ so $m = 13 \cdot 45 \cdot 17$. We first solve

$$(17)(45)x \equiv 1 \bmod 13 \implies x \equiv 6$$

$$(13)(17)x \equiv 1 \bmod 45 \implies x \equiv 11$$

$$(13)(45)x \equiv 1 \bmod 17 \implies x \equiv 5.$$

To see how these solutions are found let us look at the second one:

$$(13)(17) \equiv 1 \bmod 45 \implies 221x \equiv 1 \bmod 45 \implies 41x \equiv 1 \bmod 45$$

since $221 \equiv 41 \bmod 45$. We now use the Euclidean algorithm;

$$45 = 1 \cdot 41 + 4, 41 = 10 \cdot 4 + 1 \implies 1 = (11)(41) - (10)(45) \implies 41^{-1} \equiv 11 \bmod 45.$$

Therefore using these solutions, the common solution is

$$x_0 = \frac{13 \cdot 45 \cdot 17}{13}(6)(6) + \frac{13 \cdot 45 \cdot 17}{45}(11)(9) + \frac{13 \cdot 45 \cdot 17}{17}(5)(12) =$$

$$\implies x_0 = 27540 + 21879 + 35100 = 84519 \equiv 4959 \bmod 9945$$

$$\implies x_0 = 4959.$$

The Chinese Remainder can also be used to piece together the solution of a single linear congruence.

EXAMPLE 2.5.1.5 Solve $5x + 7 \equiv 0 \bmod 468$.

Now $(468, 5) = 1$ so the solution is $x \equiv 5^{-1}(-7) \bmod 468$. The prime decomposition of $468 = 2^2 3^2 13$. Therefore, the solution can be considered as the simultaneous solution of

$$x \equiv 5^{-1}(-7) \bmod 2^2 \implies x \equiv 1 \bmod 4$$

$$x \equiv 5^{-1}(-7) \bmod 3^2 \implies x \equiv 4 \bmod 9$$

$$x \equiv 5^{-1}(-7) \bmod 13 \implies x \equiv 9 \bmod 13.$$

Letting $m_1 = 4, m_2 = 9, m_3 = 13$, and $m = 468$, then as before we first solve

$$(9)(13)x \equiv 1 \bmod 4 \implies x \equiv 1 \bmod 4$$

$$(4)(13)x \equiv 1 \bmod 9 \implies x \equiv 4 \bmod 9$$

$$(4)(9)x \equiv 1 \bmod 13 \implies x \equiv 4 \bmod 13$$

The common solution is

$$x_0 = (9)(13)(1)(1) + (4)(13)(4)(4) + (4)(9)(9)(4) \equiv 10201 \bmod 468$$

$$\implies x_0 = 373.$$

In the previous sections, we noted that for any natural number n, the additive group of $\mathbb{Z}_n$ and the group of units of $\mathbb{Z}_n$ are finite abelian groups. As an easy consequence of the Chinese remainder theorem, we have the following result.

Theorem 2.5.3 *For any natural number m let* $(\mathbb{Z}_m, +)$ *denote the additive group of* $\mathbb{Z}_m$ *and let* $U(\mathbb{Z}_m)$ *be the group of units of* $\mathbb{Z}_m$. *Let* $n = n_1 n_2 \cdots n_k$ *be a factorization* *of n with pairwise relatively prime factors. Then*

$$(\mathbb{Z}_n,+) \cong (\mathbb{Z}_{n_1},+) \times (\mathbb{Z}_{n_2},+) \times \cdots \times (\mathbb{Z}_{n_k},+)$$

$$U(Z_n) = U(Z_{n_1}) \times \cdots \times U(Z_{n_k}).$$

We leave the proof to the exercises.

2.5.2 *Higher Degree Congruences*

Now that we have handled linear congruences, we turn to the problem of solving higher degree polynomial congruences

$$f(x) \equiv 0 \bmod m \quad (2.5.3)$$

where $f(x)$ is a nonconstant integral polynomial of degree $k > 1$. Suppose that

$$f(x) = a_0 + a_1x + \cdots + a_kx^k \text{ and } g(x) = b_0 + b_1x + \cdots + b_kx^k$$

where $a_i \equiv b_i \bmod m$ for $i = 1, \ldots, k$. Then $f(c) \equiv g(c) \bmod m$ for any integer c and hence the roots of $f(x)$ modulo m are the same as those of $g(x)$ modulo m. Therefore, we may assume that in (2.5.2.1) the polynomial $f(x)$ is actually a polynomial with coefficients in $\mathbb{Z}_m$.

As remarked earlier if m has the prime factorization $m = p_1^{e_1}p_2^{e_2}\cdots p_k^{e_k}$ and x_0 is a solution of $f(x) \equiv 0 \bmod m$, then x_0 is also a solution of $f(x) \equiv 0 \bmod p_i^{e_i}$ for $i = 1, \ldots, k$. Then for each $i = 1, \ldots, k$ there is y_i with $x_0 \equiv y_i \bmod p_i^{e_i}$. Conversely, suppose we are given y_i with $f(y_i) \equiv 0 \bmod p_i^{e_i}$ for $i = 1, \ldots, k$ then the Chinese remainder theorem can be used to patch these y_i together to get a solution x_0 of $f(x) \equiv 0 \bmod m$. Specifically,

$$x_0 = \sum_{i=1}^{k} \frac{m}{p_i^{e_i}} z_i y_i$$

would give a solution where the z_i are determined so that $\frac{m}{p_i^{e_i}} z_i \equiv 1 \bmod p_i^{e_i}$.

EXAMPLE 2.5.2.1 Solve $x^2 + 7x + 4 \equiv 0 \bmod 33$.

Since $33 = 3 \cdot 11$ we consider $x^2 + 7x + 4 = 0 \bmod 3$ and $x^2 + 7x + 4 \bmod 11$. First,

$$x^2 + 7x + 4 \equiv 0 \bmod 3 \implies x^2 + x + 1 \equiv 0 \bmod 3 \implies x \equiv 1 \bmod 3.$$

and this is the only solution. Notice that in $\mathbb{Z}_3$ we have $(x+2)^2 = x^2 + x + 1$. Now modulo 11 we have

$$x^2+7x+4=0 \implies x^2-4x+4=0 \implies (x-2)^2=0 \implies x=2$$

is the only solution. Therefore, a solution modulo 33 would be given by the solution of the pair of congruences

$$x \equiv 1 \bmod 3$$

$$x \equiv 2 \bmod 11.$$

Now $11y \equiv 1 \bmod 3 \implies y=2$ and $3y \equiv 1 \bmod 11 \implies y=4$ so by the Chinese remainder theorem the solution modulo 33 is

$$x=(11)(2)(1)+(3)(4)(2)=46 \equiv 13 \bmod 33$$

Hence we have reduced the problem of solving polynomial congruences to the problem of solving modulo prime powers. From the algorithm using the Chinese remainder theorem, we can further give the total number of solutions. If $f(x)$ is a polynomial with coefficients in $\mathbb{Z}_m$ we let $N_f(m)$ denote the number of solutions of $f(x)=0 \bmod m$. Then

Theorem 2.5.4 *If $m = p_1^{e_1}p_2^{e_2}\cdots p_k^{e_k}$ is the prime decomposition of m then $N_f(m)=N_f(p_1^{e_1})N_f(p_2^{e_2})\cdots N_f(p_k^{e_k})$.*

The simplest case of solving modulo a prime power p^α is of course when $\alpha=1$. Then we are attempting to find solutions within $\mathbb{Z}_p$. Recalling that if p is a prime then $\mathbb{Z}_p$ is a field we can use certain basic properties of equations over fields to further simplify the problem. First recalling that in a field, a polynomial of degree n can have at most n distinct roots we get:

Theorem 2.5.5 *The polynomial congruence $f(x) \equiv 0$ mod p, p prime, has at most k solutions if the degree of $f(x)$ is k.*

Recall that from Fermat's theorem $x^p = x$ for any $x \in \mathbb{Z}_p$. This implies that every element of $\mathbb{Z}_p$ is a root of the polynomial $x^p - x$. Suppose that $f(x)$ is a polynomial of degree higher than p over $\mathbb{Z}_p$. Using the division algorithm for polynomials, we then have

$$f(x)=q(x)(x^p-x)+g(x) \text{ where } g(x)=0 \text{ or } deg(g(x))<p.$$

Since every element of $\mathbb{Z}_p$ is a solution of x^p-x it follows that the solutions of $f(x)=0$ are precisely the solutions of $g(x)=0$. Hence we can always reduce a polynomial congruence modulo p to a congruence of degree less than p.

Theorem 2.5.6 *If $f(x)$ has degree higher than p, p prime, then there exists a polynomial $h(x)$ of degree less than p such that the solutions of $f(x) \equiv 0$ mod p are exactly the solutions of $h(x) \equiv 0$ mod p.*

There is no general method to solve a polynomial congruence modulo a prime p. However for degree 2 and p an odd prime the quadratic formula holds. First, some more definitions.

Definition 2.5.1 *If $(a, m) = 1$ and and $x^2 \equiv a \bmod m$ has a solution then a is called a* **quadratic residue** *mod m. If $x^2 \equiv a \bmod m$ has no solution then a is a* **quadratic nonresidue**.

We will talk more about quadratic and nonquadratic residues in the next section. However, modulo a prime, we get something special. $x^2 - a$ is a quadratic polynomial and hence in a field it can have at most two solutions. Therefore,

Lemma 2.5.1 *Given $(a, p) = 1$ with p a prime. Suppose a is a quadratic residue mod p and $x_0^2 = a$ mod p. Then $-x_0$ is the only other solution and if p is odd, x_0 and $-x_0$ are distinct.*

If a is a quadratic residue mod p let $\sqrt{a}$ denote one of the two solutions to $x^2 \equiv a$ mod p. We then obtain the quadratic formula modulo any odd prime.

Theorem 2.5.7 *If p is an odd prime then the solutions to the quadratic congruence $ax^2 + bx + c \equiv 0$ mod p with $a \not\equiv 0$ mod p, are given by*

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

In particular, if $b^2 - 4ac$ is a quadratic nonresidue mod p then $ax^2 + bx + c = 0$ has no solutions mod p.

Proof The development of the quadratic formula is solely dependent on the field properties and so can be carried out purely symbolically in $\mathbb{Z}_p$. Suppose

$$ax^2 + bx + c = 0 \text{ then } x^2 + \frac{b}{a}x = \frac{-c}{a}.$$

Completing the square on the left side in the usual manner gives

$$x^2 + \frac{b}{a}x + \frac{b^2}{4a^2} = \frac{b^2}{4a^2} - \frac{c}{a}$$

where $\frac{b^2}{4a^2}$ is defined since $4 \neq 0$ and $a^2 \neq 0$ in $\mathbb{Z}_p$ (since p was odd). Then

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

where the squareroot has the meaning described above. Finally,

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

□

EXAMPLE 2.5.2.2 Solve $3x^2 + 5x + 1 \equiv 0 \bmod 7$.
First, we divide through by 3. Since $3 \cdot 5 = 1$ in $\mathbb{Z}_7$ then $3^{-1} = 5$ and so

$$3x^2 + 5x + 1 = 0 \implies x^2 + 25x + 5 = 0 \implies x^2 + 4x + 5 = 0.$$

Applying the quadratic formula

$$x = \frac{-4 \pm \sqrt{16 - 4(1)(5)}}{2} = \frac{3 \pm \sqrt{-4}}{2} = \frac{3 \pm \sqrt{3}}{2}.$$

Now 3 is a quadratic nonresidue mod 7 so the original congruence has no solutions modulo 7.

For prime power moduli p^α with $\alpha > 1$ the general idea is to first find solutions mod p, if possible, and then move, using the found solutions iteratively to solutions mod p^2, then solutions mod p^3, and so on. There is an algorithm, to handle this iterative procedure. We will not discuss this but refer the reader to [NZ] or [N] for more on this.

2.6 Quadratic Reciprocity

We close this chapter on basic number theory with a discussion of a famous result due originally to Gauss, called the **law of quadratic reciprocity**. There are now dozens of proofs of this result in print and the result has far ranging implications well beyond what might be expected. Further there are generalizations to algebraic number theory as well as applications to problems involving sums of squares.

Recall from the last section that if $x^2 \equiv a \bmod n$ has a solution then a is called a **quadratic residue** mod n. If $n = p$, an odd prime, then there are exactly two solutions mod p. Suppose that p, q are distinct odd primes. Then p might be, or might not be, a quadratic residue mod q. Similarly q might be, or might not be, a quadratic residue mod p. At first glance, there might seem to be no relationship between these two questions. Gauss discovered that there is a quite strong relationship and this is the quadratic reciprocity law. In particular, if either of p or q is congruent to 1 mod 4 then either both of $x^2 \equiv p \bmod q$ and $x^2 \equiv q \bmod p$ are solvable or both are nonsolvable. If both p and q are congruent to 3 mod 4 then one is solvable and the other is not. Before we state the theorem precisely, we introduce some terminology and machinery.

First, we give a criterion for an integer to be a quadratic residue modulo an odd prime.

Lemma 2.6.1 *If p is an odd prime and $(a, p) = 1$ then a is a quadratic residue mod p if and only if $a^{\frac{p-1}{2}} \equiv 1$ mod p. If a is a quadratic nonresidue then $a^{\frac{p-1}{2}} \equiv -1$ mod p.*

Proof Suppose $(a, p) = 1$. We do the computations in the field $\mathbb{Z}_p$. Since $a \neq 0$ then from Fermat's theorem $a^{p-1} = 1$ in $\mathbb{Z}_p$. This implies that $(a^{\frac{p-1}{2}} - 1)(a^{\frac{p-1}{2}} + 1) = 0$ in $\mathbb{Z}_p$. Since $\mathbb{Z}_p$ is a field it has no zero divisors and this implies that either $a^{\frac{p-1}{2}} = 1$ or $a^{\frac{p-1}{2}} = -1$. Hence either $a^{\frac{p-1}{2}} \equiv 1$ mod p or $a^{\frac{p-1}{2}} \equiv -1$ mod p. We show that in the former case and only in the former case is a a quadratic residue.

Suppose that $x^2 = a$ has a solution say x_0 in $\mathbb{Z}_p$. Then

$$a^{\frac{p-1}{2}} = (x_0^2)^{\frac{p-1}{2}} = x_0^{p-1} = 1.$$

It follows further that if $a^{\frac{p-1}{2}} = -1$ there can be no solution.

Conversely, suppose $a^{\frac{p-1}{2}} = 1$. Since the multiplicative group of $\mathbb{Z}_p$ is cyclic (see the last section) it follows that there is a $g \in \mathbb{Z}_p$ which generates this cyclic group and $a = g^t$ for some t. Hence $g^{\frac{t(p-1)}{2}} = 1$. However, the order of the multiplicative group of $\mathbb{Z}_p$ is $p - 1$ and therefore this implies that

$$\frac{t(p-1)}{2} \equiv 0 \bmod p - 1.$$

Therefore, t must be even $t = 2k$. Hence $a = g^{2k} = (g^k)^2$ and there is a solution to $x^2 = a$. □

To express the quadratic reciprocity law in a succinct manner, we introduce the **Legendre symbol**.

Definition 2.6.1 *If p is an odd prime and $(a, p) = 1$ then the* **Legendre symbol** *(a/p) is defined by*

1. *$(a/p) = 1$ if a is a quadratic residue mod p.*
2. *$(a/p) = -1$ if a is a quadratic nonresidue mod p.*

Thus the value of the Legendre symbol distinguishes quadratic residues from quadratic nonresidues. The next lemma establishes the basic properties of (a/p).

Lemma 2.6.2 *If p is an odd prime and $(a, p) = (b, p) = 1$ then*

1. *$(a^2/p) = 1$,*
2. *If $a \equiv b$ mod p then $(a/p) = (b/p)$,*
3. *$(a/p) \equiv a^{\frac{p-1}{2}}$ mod p,*
4. *$(ab/p) = (a/p)(b/p)$.*

Proof Parts (1) and (2) are immediate form the definition of the Legendre symbol. Part (3) is a direct consequence of Lemma 2.6.1.

To see part (4) notice that $(ab)^{\frac{p-1}{2}} = a^{\frac{p-1}{2}} b^{\frac{p-1}{2}}$ and use part (3). □

From part (4) of this last lemma, we see that to compute (a/p) we can use the prime factorization of a and then restrict to (q/p) where q is a prime distinct from p. The quadratic reciprocity law will allow us to compute this for odd primes and we will give a seperate result for $(2/p)$. After proving the quadratic reciprocity law, we will give examples on how to do this. We now give the theorem.

Theorem 2.6.1 *(Law of Quadratic Reciprocity) If p, q are distinct odd primes then*

$$(p/q)(q/p) = (-1)^{(\frac{p-1}{2})(\frac{q-1}{2})}.$$

Alternatively if p, q are distinct odd primes then
(1) If at least one of p,q is congruent to 1 mod 4 then

$$x^2 \equiv q \text{ mod } p \text{ and } x^2 \equiv p \text{ mod } q$$

are either both solvable or both unsolvable.
(2) If both p and q are congruent to 3 mod 4 then one of

$$x^2 \equiv q \text{ mod } p \text{ and } x^2 \equiv p \text{ mod } q$$

is solvable and the other is unsolvable.

Proof The proof we give is based on two lemmas due to Gauss and then a nice geometric argument due to Eisenstein.

Let p, q be distinct odd primes and set $h = \frac{p-1}{2}$. Consider the set

$$R = \{-h, \ldots, -2, -1, 1, 2, \ldots, h\}.$$

This is reduced residue system mod p and hence every integer a relatively prime to p, that is, with $(a, p) = 1$, is congruent to exactly one element of R. Let

$$S = \{q, 2q, \ldots, hq\}.$$

Since $(p, q) = 1$ any two elements of S are incongruent mod p and therefore each element of S is congruent to exactly one element of R. We first need the following lemma.

Lemma 2.6.3 *If n is the number of elements of S congruent mod p to negative elements of R then $(q/p) = (-1)^n$.*

Proof (Lemma 2.6.3) Suppose $a_1, \ldots, a_n$ are the negative elements of R congruent to elements of S and $b_1, \ldots, b_m$ with $m + n = h$ the positive elements congruent to the remaining elements of S. The product of the elements of S is $h!q^h$ so

$$h!q^h \equiv a_1 \cdots a_n b_1 \cdots b_m \bmod p.$$

Since any two elements of S are incongruent modulo p we cannot have $-a_i = b_j$ for some i, j, for if so then $a_i + b_j = 0 \equiv mq + nq \bmod p$ which would imply that $p|(m+n)q$ which is impossible since $m, n \leq \frac{p-1}{2}$. Therefore, $-a_1, \ldots, -a_n, b_1, \ldots, b_m$ give h distinct positive integers all less than or equal to h. Hence

$$\{-a_1, \ldots, -a_n, b_1, \ldots, b_m\} = \{1, \ldots, h\}.$$

It follows that

$$(-1)^n a_1 \cdots a_n b_1 \cdots b_m = h! \implies (-1)^n h! q^h \equiv h! \bmod p.$$

However $(h!, p) = 1$ then

$$(-1)^n q^h \equiv 1 \bmod p \implies q^h = q^{\frac{p-1}{2}} \equiv (-1)^n \bmod p.$$

From Lemma 2.6.2, we have

$$(q/p) \equiv q^{\frac{p-1}{2}} \bmod p \implies (q/p) \equiv (-1)^n \bmod p.$$

□

We are now going to count (q/p) in a different way. Let $[x]$ denote the greatest integer less than or equal to x. Notice that if $a, b \in \mathbb{Z}$ and $a = qb + r$ with $0 \leq r < b$ then $[\frac{a}{b}] = q$ and so $a = [\frac{a}{b}]b + r$. Consider now the sum

$$M = \sum_{i=1}^{h} [\frac{iq}{p}].$$

M is called a **Gauss sum**. The next lemma ties this Gauss sum to (q/p).

Lemma 2.6.4 *Let p, q be distinct odd primes and let M be defined as above. Then*

$$(q/p) = (-1)^M.$$

Proof As explained above for each i we have

$$iq = [\frac{iq}{p}]p + r_i, 0 < r_i < p.$$

Let R be as in Lemma 2.6.3. If iq is congruent to a negative element a_i of R then $r_i = p + a_i$ while if iq is congruent to a positive element b_i then $r_i = b_i$. Then

$$\sum_{i=1}^{h} iq = p\sum_{i=1}^{h}[\frac{iq}{p}] + \sum_{i=1}^{n}(a_i + p) + \sum_{i=1}^{m} b_i.$$

Further

$$\sum_{i=1}^{h} i = \frac{h(h+1)}{2} = \frac{p^2-1}{8}.$$

Let $P = \frac{p^2-1}{8}$ and plugging back into our sum over $\{iq\}$ we get

$$\sum_{i=1}^{h} iq = Pq = pM + np + \sum_{i=1}^{n} a_i + \sum_{i=1}^{m} b_i.$$

However as we saw in the proof of Lemma 2.6.3,

$$\{-a_1, \ldots, -a_n, b_1, \ldots, b_m\} = \{1, \ldots, h\} \implies -\sum_{i=1}^{n} a_i + \sum_{i=1}^{m} b_i = P.$$

Then

$$Pq = pM + np + P + 2\sum_{i=1}^{n} a_i \implies P(q-1) = (M+n)p + 2\sum_{i=1}^{n} a_i.$$

Since q is odd $q - 1 \equiv 0 \bmod 2$ and hence if we take the last sum mod 2 we get that

$$M + n \equiv 0 \bmod 2$$

which implies that M, n are both even or both odd. It follows that $(-1)^M = (-1)^n$. From Lemma 2.6.3 we have $(q/p) = (-1)^n$ and hence $(q/p) = (-1)^M$ proving the second lemma. □

We now interchange the roles of p and q. Let $k = \frac{q-1}{2}$ and let N be the Gauss sum for q,

$$N = \sum_{i=1}^{k}[\frac{ip}{q}].$$

Therefore from Lemma 2.6.4 applied to q, we have $(p/q) = (-1)^N$. Hence

$$(p/q)(q/p) = (-1)^M(-1)^N = (-1)^{M+N}.$$

We will show that

$$M + N = hk = (\frac{p-1}{2})(\frac{q-1}{2}) \quad 2.6.1$$

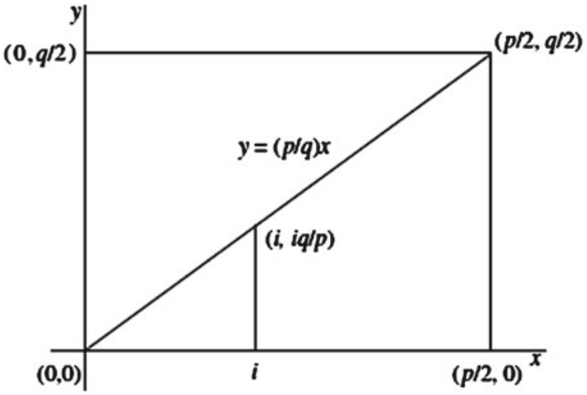


Fig. 2.2 Geometric argument for Quadratic Reciprocity

which will prove the quadratic reciprocity law.

To show (2.6.1) we will use a lovely geometric argument. Consider the lattice points, that is, points with integer coordinates, within the rectangle with corners at

$$(0,0),\ (\frac{p}{2},0),\ (\frac{p}{2},\frac{q}{2}),\ (0,\frac{q}{2})$$

as pictured in Figure 2.2.

Let T be the total number of lattice points within the rectangle. We will compute T in two different ways. First, notice that $T = hk$ since $[\frac{p}{2}] = h$ and $[\frac{q}{2}] = k$.

Now consider the number below the diagonal. Since the equation of the diagonal is $y = \frac{q}{p}x$ there are no lattice points on the diagonal. For an integer i, the vertical line $x = i$ hits the diagonal at the point $(i, \frac{q}{p}i)$ and hence the number of lattice points along the line $x = i$ and below the diagonal is $[\frac{iq}{p}]$. It follows that the total number of lattice points below the diagonal is

$$\sum_{i=1}^{h}[\frac{iq}{p}] = M.$$

An analogous argument shows that the total number of lattice points above the diagonal is N. Therefore, $T = M + N$. Hence

$$M + N = hk$$

and the quadratic reciprocity law is proved.

Before giving some examples we note that by modifying slightly the proof of Lemma 2.6.3 we get the following which allows us to compute $(2/p)$ for any odd prime p.

Theorem 2.6.2 *If p is an odd prime, then*

1. $(-1/p) = (-1)^{\frac{p-1}{2}}$ *and*
2. $(2/p) = (-1)^{\frac{p^2-1}{8}}$.

Proof The first part (1) follows directly from Lemmas 2.6.1 and 2.6.2 taking $a = -1$.

For (2), although we assumed that q was an odd prime in both Lemmas 2.6.3 and 2.6.4 the construction of the sets R and S and the Gauss sum M only required that $(q, p) = 1$. Now let $q = 2$. Then from the definition of the Gauss sum $M = 0$. Hence $\frac{p^2-1}{8} \equiv n \bmod p$. Then $(2/p) = (-1)^n = (-1)^{\frac{p^2-1}{8}}$. □

With the quadratic reciprocity law and Theorem 2.6.2 it is relatively easy to compute (a/p) for any a.

EXAMPLE 2.6.1 Determine (870/7).

The prime factorization of 870 is $870 = 2 \cdot 3 \cdot 5 \cdot 29$. Then

$$(870/7) = (2/7)(3/7)(5/7)(29/7).$$

First,

$$(2/7) = (-1)^{\frac{49-1}{8}} = (-1)^6 = 1$$

$$(3/7) = -(7/3) \text{ since both are congruent to } 3 \bmod 4$$

$$(7/3) = (1/3) = 1 \implies (3/7) = -1$$

$$(5/7) = (7/5) \text{ since } 5 \equiv 1 \bmod 4$$

$$(7/5) = (2/5) = (-1)^{\frac{24}{8}} = -1 \implies (5/7) = -1.$$

Finally,

$$(29/7) = (1/7) = 1.$$

Putting these all together

$$(870/7) = (2/7)(3/7)(5/7)(29/7) = (1)(-1)(-1)(1) = 1$$

and hence 870 is a quadratic residue mod 7.

This was just an illustration. For a small prime like 7 it would be easier to reduce mod 7 and do it directly.

$$870 \equiv 2 \bmod 7 \implies (870/7) = (2/7) = 1.$$

2.7 Exercises

2.1 Verify that the following are rings. Indicate which are commutative and which have identities. Which are integral domains?

(a) The set of rational numbers.

(b) The set of continuous functions on a closed interval $[a, b]$ under ordinary addition and multiplication of functions.

(c) The set of 2×2 matrices with integral entries.

(d) The set $n\mathbb{Z}$ consisting of all integers which are multiples of the fixed integer n.

2.2 (a) Show that in an ordered ring nonzero squares must be positive. Conclude that in an ordered ring with identity the multiplicative identity must be positive.

(b) Show that the complex numbers under the ordinary operations cannot be ordered.

2.3 Show that any ordered ring must be infinite. (Hint: Suppose $a > 0$ then $a + a > 0, a + a + a > 0$ and continue).

2.4 Prove by induction that there are 2^n subsets of a finite set with n elements.

2.5 Prove that $1^2 + 2^2 + \cdots + n^2 = \frac{n(n+1)(2n+1)}{6}$.

2.6 Let R be an ordered integral domain which satisfies the inductive property. Prove that R is isomorphic to $\mathbb{Z}$.

(Hint: Let 1 be the multiplicative identity in R. Define $2 \cdot 1 = 1 + 1$ and inductively $n \cdot 1 = (n - 1) \cdot 1 + 1$ in R. Define

$$\overline{R} = \{n \cdot 1 \in R; n \in \mathbb{Z}\}$$

and let $f : \mathbb{Z} \to R$ by $f(n) = n \cdot 1$. Show first that f is an isomorphism from $\mathbb{Z}$ to $\overline{R}$. Then use the inductive property in R to show that $\overline{R}$ is all of R.)

2.7 Prove the remaining parts of Theorem 2.2.1.

2.8 Find the GCD and LCM of the following pairs of integers and then express the GCD as a linear combination

(a) 78 and 30,

(b) 175 and 35,

(c) 380 and 127.

2.9 Prove that if $a = qb + r$ then $(a, b) = (b, r)$.

2.10 Prove that if $d = (a, b)$ then $\frac{a}{d}$ and $\frac{b}{d}$ are relatively prime.

2.11 Show that if $(a, b) = c$ then $(a^2, b^2) = c^2$. (Hint: The easiest method is to use the fundamental theorem of arithmetic.)

2.12 Redo Problem 2.8 using the prime decomposition of each integer.

2.13 Show that an integer is divisible by 3 if and only if the sum of its digits (in decimal expansion) is divisible by 3. (Hint: Write out the decimal expansion and take everything modulo 3.)

2.14 Let F be a field and let $F[x]$ denote the ring of polynomials over F. Prove that if $f(x), g(x) \in F[x]$ with $g(x) \neq 0$ then there exist unique polynomials $q(x), r(x) \in F[x]$ such that

$$f(x) = q(x)g(x) + r(x), r(x) = 0 \text{ or } deg(r(x)) < deg(g(x)).$$

This is the division algorithm for polynomials. (Hint: Model the proof on the proof for the integers.)

2.15 Suppose $p(x)$ is a polynomial over F and $p(r) = 0$. Show that $p(x) = (x - r)h(x)$ where $h(x)$ is another polynomial of degree one less. (Use the division algorithm.)

2.16 Let $g(x), f(x) \in F[x]$. Then their **greatest common divisor** or **GCD** is the monic polynomial $d(x)$ (leading coefficient 1) such that $d(x)$ divides both $f(x)$ and $g(x)$ and if $d_1(x)$ is any other common divisor of $g(x)$ and $f(x)$ the $d_1(x)$ divides $d(x)$. Show that the GCD of two polynomials exists and is the monic polynomial of least degree which can be expressed as a linear combination of $f(x)$ and $g(x)$. That is,

$$d(x) = h(x)f(x) + k(x)g(x)$$

and $d(x)$ has the least degree of any linear combination of this form. (Hint: Again model the proof on the proof for the integers.)

2.17 Prove Euclid's lemma for polynomials, that is, if $d(x)$ divides $f(x)g(x)$ and $(d(x), g(x)) = 1$ then $d(x)$ divides $f(x)$.

2.18 A polynomial $p(x)$ of positive degree over a field F is a **prime polynomial** or **irreducible polynomial** if it cannot be expressed as a product of two polynomials of positive degree over F. Prove that any nonconstant polynomial $f(x) \in F[x]$, where F is a field can be decomposed as a product of prime polynomials. Further this decomposition is unique except for ordering and unit factors. This is the **unique factorization theorem** for polynomial rings over fields. (Hint: Again model the proof on the proof of the fundamental theorem of arithmetic.)

2.19 Suppose $p(x)$ is a polynomial over F and the degree of $p(x)$ is n. Prove that $p(x)$ can have at most n distinct roots over F.

2.20 Mimic the results in Problems 2.14 through 2.18 for general Euclidean domains (see the definition on p. 21) and then use this to prove Theorem 2.3.6.

2.21 Show that the Gaussian integers $\mathbb{Z}[i]$ are Euclidean domain with $N(a+bi) = a^2 + b^2$. This shows that the Gaussian integers are a unique factorization domain.

2.22 Prove part (c) of Theorem 2.6.2: If $a \equiv b \bmod n$ and $c \equiv d \bmod n$ then $ac \equiv bd \bmod n$.

2.23 Verify the remaining ring properties to show that for any positive integer n, $\mathbb{Z}_n$ is a commutative ring with an identity.

2.24 Find the multiplicative inverse if it exists

(a) of 13 in $\mathbb{Z}_{47}$,
(b) of 17 in $\mathbb{Z}_{22}$,
(c) of 6 in $\mathbb{Z}_{30}$.

2.25 Solve the linear congruences

(a) $4x + 6 = 2$ in $\mathbb{Z}_7$,
(b) $5x + 9 = 12$ in $\mathbb{Z}_{47}$,
(c) $3x + 18 = 27$ in $\mathbb{Z}_{40}$.

2.26 Find $\phi(n)$ for

(a) $n = 17$,
(b) $n = 526$,
(c) $n = 138$.

2.27 Determine the units and write down the group table for the unit group $U(\mathbb{Z}_n)$ for

(a) $\mathbb{Z}_{12}$,
(b) $\mathbb{Z}_{26}$.

2.28 Verify Theorem 2.4.8 for

(a) $n = 26$,
(b) $n = 88$.

2.29 Prove Theorem 2.5.3, that is, for any natural number m let $(\mathbb{Z}_m, +)$ denote the additive group of $\mathbb{Z}_m$ and let $U(\mathbb{Z}_m)$ be the group of units of $\mathbb{Z}_m$. Let $n = n_1 n_2 \cdots n_k$ be a factorization of n with pairwise relatively prime factors. Then

$$(\mathbb{Z}_n, +) \cong (\mathbb{Z}_{n_1}, +) \times (\mathbb{Z}_{n_2}, +) \times \cdots \times (\mathbb{Z}_{n_k}, +)$$

$$U(\mathbb{Z}_n) = U(\mathbb{Z}_{n_1}) \times \cdots \times U(\mathbb{Z}_{n_k}).$$

2.30 Prove that if an integer is congruent to 2 modulo 3 then it must have a prime factor congruent to 2 modulo 3.

2.31 Prove that if p is an odd prime then there exist positive integers x, y such that $p = x^2 - y^2$.

2.32 Prove that if bc is a perfect square for integers b, c and $(b, c) = 1$ then both b and c are perfect squares.

2.33 Determine a primitive root modulo 11.

2.34 We outline a proof of Theorem 2.4.14: An integer n will have a primitive root modulo n if and only if

$$n = 2, 4, p^k, 2p^k$$

where p is a prime.

(a) Show that if $(m, n) = 1$ with $m > 2, n > 2$ then there is no primitive root modulo mn.

(b) Show that there is no primitive root modulo 2^k for $k > 2$.

(c) Prove that if p is an odd prime then there exists a primitive root a mod p such that a^{p-1} is not congruent to 1 modulo p^2. (Hint: Let a be a primitive root mod p. Then $a + p$ is also a primitive root. Show that either a or $(a + p)$ satisfies the result.)

(d) Prove that there exists a primitive root modulo p^k for any $k \geq 2$. (Hint: Let a be the primitive root mod p from part (c). Then this is a primitive root mod p^k for any $k \geq 2$.)

(e) Prove that if a is a primitive root mod p^k then, if a is odd, a is also a primitive root mod $2p^k$. If a is even then $a + p^k$ is a primitive root modulo $2p^k$.

2.35 Use the primality test based on Fermat's theorem to show that 1053 is not prime.

2.36 If $m > 2$ show that $\phi(m)$ is even.

2.37 Prove that $\phi(n^2) = n\phi(n)$ for any positive integer n.

2.38 Prove that if $n \geq 2$ then

$$\sum_{(m,n)=1, 0<m<n} m = \frac{n\phi(n)}{2}.$$

2.39 Prove that if n has k distinct odd factors then $2^k | \phi(n)$.

Chapter 3
The Infinitude of Primes

3.1 The Infinitude of Primes

The two most striking characteristics of the sequence of primes are that there are many of them but that their density is rather slim. From Euclid's theorem (Theorem 2.3.1) there are infinitely many primes, in fact there are infinitely many in any arithmetic sequence of integers. This latter fact was proved by Dirichlet and is known as **Dirichlet's Theorem**. However, despite the fact the primes are so numerous, their density among the natural numbers gets slim. As mentioned before if x is a natural number and $\pi(x)$ represents the number of primes less than or equal to x then asymptotically this function behaves as the function $\frac{x}{\ln x}$. This result is known as the **prime number theorem**. Besides being a startling result, the proof of the prime number theorem, done independently by Hadamard and De la Valle Poussin, became the genesis for analytic number theory. In this chapter we will discuss various aspects of the infinitude of primes. The prime number theorem will be introduced in the next chapter.

As a starting point we will give an array of proofs of the infinitude of primes: some are direct, some involve analysis and some come from quite different directions. Hopefully seeing these proofs will both shed some light on the nature of the sequence of primes and at the same time show the complexity of this rather straightforward result. Included among these will be several simple cases of Dirichlet's Theorem, which we will prove in its entirety in Section 3.3.

3.1.1 Some Direct Proofs and Variations

The purpose of this chapter is then to present a wide array of proofs that the set of primes is infinite. Each of these other proofs will hopefully shed further light on the nature of the primes and the nature of the integers. We first restate the basic theorem which was given in the last chapter as Theorem 2.3.1.

B. Fine and G. Rosenberger, *Number Theory*,
DOI 10.1007/978-3-319-43875-7_3

Theorem 3.1.1 *There are infinitely many primes.*

In the last chapter we gave two proofs of this result, the first of which goes back to Euclid. Recall that Euclid's argument went like this: suppose that there are only finitely many primes $p_1, \ldots, p_n$. Each of these is positive so we can form the positive integer

$$N = p_1 p_2 \cdots p_n + 1.$$

N has a prime decomposition so in particular there is a prime p which divides N. Then

$$p | (p_1 p_2 \cdots p_n + 1).$$

Since the only primes are assumed to be $p_1, p_2, \ldots, p_n$ it follows that $p = p_i$ for some $i = 1, \ldots, n$. But then $p | p_1 p_2 \cdots p_i \cdots p_n$ so p cannot divide $p_1 \cdots p_n + 1$ which is a contradiction. Therefore p is not one of the given primes showing that the list of primes must be endless. Notice that in this argument we could just as easily have worked with $N = p_1 \cdots p_n - 1$.

We also presented the following variation of Euclid's argument. Again suppose that there are only finitely many primes $p_1, \ldots, p_n$. Certainly $n \geq 2$. Let $P = \{p_1, \ldots, p_n\}$. Divide P into two disjoint nonempty subsets P_1, P_2. Now consider the number $m = q_1 + q_2$ where q_i is the product of all the primes from P_1 and q_2 is the product of all the primes from P_2. Let p be a prime divisor of m. Since $p \in P$ it follows that p divides either q_1 or q_2 but not both. But then p does not divide m giving a contradiction. Therefore p is not one of the given primes and the number of primes must be infinite.

We now give some further variations of Euclid's basic proof. All of these proofs do not use analysis. In the next section we prove Theorem 3.1.1 with some analytic ideas. These are precursors to both the proof of the prime number theorem and the proof of Dirichlet's theorem.

Proof (1a) (**Using Factorials**). Again suppose that $p_1, \ldots, p_n$ are the only primes and let $N = p_1 \cdots p_n$. Certainly $p_i < N$ for each i. Let q be the smallest prime divisor of $N! + 1$. If $q < N$ then q certainly divides $N!$ so q cannot divide $N! + 1$. Therefore $q > N$ and hence $q > p_i$ for $i = 1, \ldots, n$. Hence q is not one of the p_i and the sequence of primes is infinite.

Notice that the fact that the smallest prime divisor of $N! + 1$ is greater than N did not depend on N being a product of primes. Hence this proof can be varied as follows. □

Proof (1b) (**Again Using Factorials**) For each $n > 1$ let q_n be the smallest prime divisor of $n! + 1$. Exactly as in the previous proof we must have $q_n > n$ and hence there cannot be finitely many primes. □

We get another simple variation by using the sum $\sum_p \frac{1}{p}$ and assuming the set of primes is finite. In the next section we show that this sum actually diverges which also shows that the primes are infinite.

Proof (2) (**Using Sums**) As before suppose that $p_1, \ldots, p_n$ are the only primes and let $N = p_1 \cdots p_n$. Set

$$a = \sum_{i=1}^{n} \frac{1}{p_i} \text{ so that } aN = \sum_{i=1}^{n} \frac{N}{p_i}.$$

aN is an integer so it has a prime divisor which by assumption must be some p_j. Then $p_j | aN$ and $p_j | \frac{N}{p_i}$ for $i \neq j$. Since N is a product it follows that $p_j | \frac{N}{p_j}$ which is a contradiction. □

The next proof involves the use of the Euler phi function. Recall from Section 2.5 that for a positive integer n,

$$\phi(n) = \text{ number of positive integers } x \leq n \text{ with } (x, n) = 1.$$

For a prime p we have $\phi(p) = p - 1$ and if $(a, b) = 1$ then $\phi(ab) = \phi(a)\phi(b)$.

Proof (3) (**Using the Euler Phi Function**) Suppose that $p_1, \ldots, p_n$ are the only primes and let $N = p_1 \cdots p_n$. Notice that if $p_i > 2$ then $\phi(p_i) = p_i - 1 > 1$.

If $1 < n < N$ then n must have a prime divisor say p_j and hence p_j is a common divisor of n and N. It follows that $(n, N) \neq 1$, that is, they are not relatively prime. By definition then we must have $\phi(N) = 1$. On the other hand

$$\phi(N) = \phi(p_1 \cdots p_n) = \phi(p_1) \cdot \phi(p_2) \cdots \phi(p_n) = (p_1 - 1) \cdots (p_n - 1) > 1$$

a contradiction. □

The final proof of this first section is somewhat different than the others and involves integral polynomials. Let $\mathbb{Z}[x]$ denote the set of polynomials with integral coefficients and let $\mathbb{N}_0 = \mathbb{N} \cup \{0\}$.

Lemma 3.1.1 *For each nonconstant polynomial $f(x) \in \mathbb{Z}[x]$, the set of prime divisors of the integers $\{f(k); k \in \mathbb{N}_0\}$ is infinite. In particular the total number of primes is infinite.*

Proof Suppose that

$$f(x) = a_0 + a_1 x + \cdots + a_m x^m$$

and assume that for the set $\{f(k); k \in \mathbb{N}_0\}$, the number of prime divisors which occur for some $f(k)$, is finite. Let $U = \{p_1, \ldots, p_n\}$ be this set of prime divisors and let $D = p_1 \cdots p_n$. Without loss of generality suppose $a_0 \neq 0$. Choose an integer t so that p_i^t does not divide $f(0) = a_0$ for any i. Since the p_i are the only primes we must have $a_0 | D^t$, that is, $D^t = a_0 b$ for some $b \in \mathbb{Z}$. For $k \geq 1$ we have

$$f(kD^{2t}) = \sum_{j=1}^{m} a_j k^j D^{2tj} + a_0 = a_0(\sum_{j=1}^{m} a_j k^j b^{2j} a_0^{2j-1} + 1) = M.$$

For k large enough the integer M must have a prime divisor p which does not divide $a_0 b$ and hence $p \notin U$ a contradiction. □

3.1.2 Some Analytic Proofs and Variations

Both the proof of the prime number theorem and the proof of Dirichlet's theorem depend heavily on the use of analysis—both real and complex. The introduction of analytic methods into number theory can be traced back basically to the following two results of Euler which also imply that the sequence of primes is infinite.

Theorem 3.1.2 *The sum*

$$\sum_{p \text{ prime}} \frac{1}{p}$$

diverges. In particular the sequence of primes is infinite.

Proof Clearly, if the series

$$\sum_{p \text{ prime}} \frac{1}{p}$$

diverges, then there must be infinitely many primes, for otherwise this would be a finite sum.

We present two proofs that this sum diverges. The first is direct while the second introduces the Riemann zeta function which will be crucial in investigations of the density of primes.

Let $p_1, \dots, p_k, \dots$ be the sequence of primes in increasing order which at this point may or not be infinite. We first need the following fact:

Lemma 3.1.2 *If $p_1, \dots, p_k, \dots$ is the sequence of primes in increasing order then $p_n \le 2^{2^{n-1}}$ for all n and $p_n < 2^{2^{n-1}}$ for all $n > 1$.*

Proof (Lemma 3.1.2) By induction. $p_1 = 2 \le 2^1$ so its true for $n = 1$. Further no other prime is even so $p_k \ne 2^{2^k}$ if $k > 1$. Suppose then that $p_k < 2^{2^{k-1}}$ and consider p_{k+1}. Now, as in Euclid's proof of the infinitude of primes, $K = p_1 \cdots p_k + 1$ must have a prime divisor which is not one of $p_1, \dots, p_k$. Hence

$$p_{k+1} \le K = p_1 \cdots p_k + 1 < 2^2 2^{2^2} 2^{2^3} \cdots 2^{2^{k-1}} + 1 < 2^{2^k}.$$

Therefore the assumption is true for all n by induction. □

Proof Now we continue the proof of Theorem 3.1.2. Assume that

$$\sum_{p \text{ prime}} \frac{1}{p} = \sum_{i=1}^{\infty} \frac{1}{p_i}$$

converges. Note that we are not assuming here that there are infinitely many primes. If there are only finitely many then this is a finite sum. Since the series converges and the p_i are increasing there must be an N such that

$$\sum_{i=N+1}^{\infty} \frac{1}{p_i} < \frac{1}{2}.$$

Fix this value N, and let $Q_N(x)$, for any natural number x, be the number of positive integers less than or equal to x which are not divisible by any of the primes $p_{N+1}, p_{N+2}, \ldots$. For a given prime p the number of integers $n \leq x$ and divisible by p is smaller than $\frac{x}{p}$. Then it follows that for any integer x,

$$x - Q_N(x) < \frac{x}{p_{N+1}} + \frac{x}{p_{N+2}} + \cdots < \frac{x}{2}$$

since we assumed that

$$\sum_{i=N+1}^{\infty} \frac{1}{p_i} < \frac{1}{2}.$$

Therefore $\frac{x}{2} < Q_N(x)$. On the other hand if $n < x$ and n is not divisible by any of $p_{N+1}, p_{N+2}, \ldots$ then $n = n_1^2 m$ where m is squarefree. Hence $m = 2^{e_1} 3^{e_2} \cdots p_N^{e_N}$ where each $e_i = 0$ or 1. Hence there are at most 2^N choices for m. Further there are at most $\sqrt{x}$ choices for n_1. It follows then that

$$\frac{x}{2} < Q_N(x) < 2^N \sqrt{x}.$$

Since N is fixed this is a contradiction for x large enough and hence the sum

$$\sum_{p \text{ prime}} \frac{1}{p}$$

diverges. □

We now give a second proof of Theorem 3.1.2 which introduces the ideas of the Riemann Zeta Function and Euler products which are fundamental in some of our further discussions.

Proof (of Theorem 3.1.2) For a real variable $s > 1$ we define the **Riemann Zeta Function** by

$$\zeta(s) = \sum_{n=1}^{\infty} \frac{1}{n^s}.$$

From the classical p-series test this will converge if $s > 1$ and hence will define a function. When we discuss the prime number theorem in the next chapter we will extend this function to complex variables. Since $\sum_{n=1}^{\infty} \frac{1}{n}$ diverges it follows that as $s \to 1^+$ the sum $\zeta(s)$ will diverge. From the fundamental theorem of arithmetic each n can be expressed as a product of primes and hence the zeta function can be written as the following product

$$\zeta(s) = \prod_{p \text{ prime}} (1 + \frac{1}{p^s} + \frac{1}{p^{2s}} + \cdots).$$

However we have the geometric series converging so that

$$1 + \frac{1}{p^s} + \frac{1}{p^{2s}} + \frac{1}{p^{3s}} + \cdots = \frac{1}{1 - p^{-s}}.$$

Therefore

$$\zeta(s) = \prod_{p \text{ prime}} \left(\frac{1}{1 - p^{-s}}\right).$$

These last two products are called **Euler products** after Euler who first used them in his investigations.

Now if the sequence of primes was finite then the Euler product would be a finite number and hence $\zeta(s)$ would always converge. However as we pointed out $\zeta(s)$ diverges if $s \to 1^+$ and hence the sequence of primes is infinite.

For the second proof of Theorem 3.1.2 consider the inequality

$$\ln(\frac{1}{1-x}) = \sum_{n=1}^{\infty} \frac{x^n}{n} < \sum_{n=1}^{\infty} x^n = \frac{x}{1-x}$$

which holds if $0 < x < 1$ (see the exercises). It follows that for $0 < x < \frac{1}{2}$,

$$\ln(\frac{1}{1-x}) < 2x.$$

Then using the Euler product representation of $\zeta(s)$ and taking logarithms,

$$\ln(\zeta(s)) = \sum_{p \text{ prime}} \ln(1 - \frac{1}{p^s})^{-1} < 2 \sum_{p \text{ prime}} p^{-s}.$$

If $\sum_{p \text{ prime}} \frac{1}{p}$ were convergent then $2\sum_p p^{-s} < 2\sum_p p^{-1}$ for all $s > 1$ and it would follow that $\zeta(s)$ would not diverge if $s \to 1^+$ a contradiction. Therefore the sum diverges. □

The final results in this section give lower bounds on $\pi(x)$ the number of primes less than or equal to x. These lower bounds further imply the infinitude of primes.

Theorem 3.1.3 *For any natural number $x \geq 2$ we have*

$$\pi(x) > \ln \ln x.$$

Proof Let $p_1, \ldots, p_k, \ldots$ be the sequence of primes in increasing order. Recall that $p_n < 2^{2^{n-1}}$ for all $n > 1$. For a given x choose a k such that

$$2^{2^{k-1}} \leq x < 2^{2^k}.$$

Therefore since $p_k < 2^{2^{k-1}}$ we have

$$k \leq \pi(2^{2^{k-1}}) \leq \pi(x).$$

From $x < 2^{2^k} < e^{e^k}$ it follows that

$$\ln \ln x < k \leq \pi(x).$$

□

Using the Fundamental Theorem of Arithmetic we can arrive at a separate but similar lower bound.

Theorem 3.1.4 *For any natural number $x \geq 21$ we have*

$$\pi(x) > \frac{\ln x}{2 \ln \ln x}.$$

Proof For fixed x let p_i run over all the primes less than or equal to x. Then from the Fundamental Theorem of Arithmetic the number of integral solutions to the inequality

$$\prod_{p_i} p_i^{e_i} \leq x$$

for $e_i \geq 0$ is precisely x. On the other hand the number of solutions is the product of the number of choices for each e_i. Since for a solution $p_i^{e_i} \leq x$ we have

$$e_i \leq 1 + \frac{\ln x}{\ln p_i} \leq 1 + \frac{\ln x}{\ln 2} < (\ln x)^2$$

for $x > 20$. Therefore

$$x \leq \prod_{p_i} \left(1 + \frac{\ln x}{\ln p_i}\right) < ((\ln x)^2)^{\pi(x)}$$

$$\implies \pi(x) > \frac{\ln x}{2 \ln \ln x}.$$

□

Corollary 3.1.1 *$\pi(x) \to \infty$ as $x \to \infty$. In particular the sequence of primes is infinite.*

Proof From Theorem 3.1.3 we have $\pi(x) > \ln \ln x$ for $x \geq 2$. The latter sequence becomes infinite with x. Similarly from Theorem 3.1.4 we have $\pi(x) > \frac{\ln x}{2 \ln \ln x}$ for $x \geq 21$ and this latter sequence also becomes infinite with x. □

3.1.3 The Fermat and Mersenne Numbers

In the next several subsections we will examine primes in relation to certain special sequences of integers. Although not directly related, this path will lead ultimately to Dirichlet's Theorem.

The first such sequence we consider are called the **Fermat numbers**.

Definition 3.1.1 *The* **Fermat numbers** *are the sequence (F_n) of positive integers defined by*

$$F_n = 2^{2^n} + 1, n = 1, 2, 3, \ldots$$

If a particular F_m is prime it is called a **Fermat prime**.

Fermat believed that all the numbers in this sequence were primes. In fact F_1, F_2, F_3, F_4 are all prime but F_5 is composite and divisible by 641 (see exercises). It is still an open question whether or not there are infinitely many Fermat primes. It has been conjectured that there are only finitely many. On the other hand, if a number of the form $2^k + 1$ is a prime for some integer k then it must be a Fermat prime.

Theorem 3.1.5 *If $a \geq 2$ and $a^n + 1$ is a prime then a is even and $n = 2^m$ for some nonnegative integer m. In particular if $p = 2^k + 1$ is a prime then $k = 2^n$ for some n and p is a Fermat prime.*

Proof If a is odd then $a^n + 1$ is even and hence not a prime. Suppose then that a is even and $n = kl$ with k odd and $k \geq 3$. Then

$$\frac{a^{kl} + 1}{a^l + 1} = a^{(k-1)l} - a^{(k-2)l} \pm \cdots + 1$$

Therefore $a^l + 1$ divides $a^{kl} + 1$ if $k \geq 3$. Hence if $a^n + 1$ is a prime we must have $n = 2^m$. □

We now use the Fermat numbers to get another proof of the infinitude of primes. We first need the following.

Lemma 3.1.3 *Let (F_n) be the sequence of Fermat numbers. Then if $m \neq n$ we have $(F_n, F_m) = 1$.*

Proof Suppose that $n > m$ and suppose that $d|F_n, d|F_m$. Then

$$\frac{F_n - 2}{F_m} = \frac{2^{2^n} - 1}{2^{2^m} + 1} = (2^{2^m})^{2^{n-m}-1} - (2^{2^m})^{2^{n-m}-2} \pm \cdots - 1.$$

Therefore $F_m | F_n - 2$ and hence $d|F_n - 2$. Since $d|F_n$ it follows that $d|2$. But $d \neq 2$ since both F_n and F_m are odd. □

This now yields another proof of the infinitude of primes. Since the members of the infinite sequence (F_n) are pairwise coprime and each F_n must have at least one prime divisor it follows directly that the number of primes must be infinite.

We can also get the following variation of this method. Suppose $a \in \mathbb{N}$. Define the sequence $A_n = a^{2^n} + 1$. Then it can be proved that (see exercises)

(1) If $n > m \geq 1$ then $(a^{2^m} + 1)|(a^{2^n} - 1)$

(2) $(A_n, A_m) = 1$ with $n \neq m$ if a is even and $(A_n, A_m) = 2$ with $n \neq m$ if a is odd.

Then the same proof as used with the Fermat numbers goes through. In fact given any infinite integer sequence (a_k) with $(a_i, a_j) = 1$ for $i \neq j$ will yield a similar proof. As an example start with $(m, n) = 1$ and let $a_0 = m + n$. Then define inductively

$$a_{k+1} = a_k^2 - ma_k + m.$$

Then it can be proved that $(a_i, a_j) = 1$ if $i \neq j$ and this sequence can be used in the same proof.

The second sequence we consider is called the sequence of **Mersenne numbers**.

Definition 3.1.2 *The* **Mersenne numbers** *are the sequence (M_n) of positive integers defined by*

$$M_n = 2^n - 1, n = 1, 2, 3, \ldots \quad .$$

If a particular M_n is prime it is called a **Mersenne prime**.

The Mersenne numbers were introduced by the French clergyman and mathematician M. Mersenne who showed that if M_n is a prime than n must be a prime and claimed then that M_n is a prime for

$$n = 2, 3, 5, 7, 13, 17, 19, 31, 67, 127, 257$$

and composite for all others. It is now known that M_{67} and M_{257} are not primes while M_{61} and M_{89} are primes. Further M_p is prime for several large exponents and the search for larger and larger primes generally revolves around Mersenne primes. As in the case of the Fermat primes it is still an open question as to whether or not there are infinitely many Mersenne primes. However for the Mersenne primes it is conjectured that there are infinitely many. As of May 2013 there were 48 known Mersenne primes, the largest of which is $M_{6972593}$. Further information on the search for larger Mersenne primes can be found at the internet site www.mersenne.org.

Theorem 3.1.6 *Suppose a, n are positive integers. If $a^n - 1$ is prime then $a = 2$ and n is prime. In particular if a Mersenne number M_n is a Mersenne prime then n is prime.*

Proof Assume $a \geq 3$. Then $(a-1)|(a^n - 1)$. Therefore if $a^n - 1$ is prime we must have $a = 2$. If $n = kl$ with $2 \leq k, l < n$ then

$$(2^k - 1)|(2^n - 1).$$

Hence if $2^n - 1$ is prime n must be prime. □

As is the theme of this chapter we will now use the Mersenne numbers to derive the infinitude of primes.

Lemma 3.1.4 *For any pair of Mersenne numbers M_n, M_m we have*

$$(M_m, M_n) = (2^m - 1, 2^n - 1) = 2^{(m,n)} - 1.$$

Proof This is certainly correct if $m = n$ or $n = 1$ or $m = 1$. Assume that $m > n > 1$. From the Euclidean algorithm applied to m, n we have

$$\begin{aligned} m &= nq_0 + r_1 \\ n &= r_1q_1 + r_2 \\ &\ldots \\ r_{s-2} &= r_{s-1}q_{s-1} + r_s \\ r_{s-1} &= r_sq_s \end{aligned}$$

and $r_s = (m, n)$.

It follows then that

$$2^m - 1 = 2^{nq_0 + r_1} - 1 = 2^{r_1}(2^{q_0 n} - 1) + (2^{r_1} - 1)$$

$$2^n - 1 = 2^{r_2}(2^{q_1 r_1} - 1) + (2^{r_2} - 1)$$

$$\ldots$$

$$2^{r_{s-1}} - 1 = (2^{r_s} - 1)(2^{r_s(q_s-1)} + \cdots + 1).$$

This yields

$$(2^{r_s} - 1)|(2^{r_{s-1}} - 1) \text{ and } (2^{r_s} - 1)|(2^{r_{s-2}} - 1)$$

since also

$$2^{q_{s-1}r_{s-1}} - 1 = (2^{r_{s-1}} - 1)(2^{r_{s-1}(q_{s-1}-1)} + \cdots + 1).$$

Finally

$$(2^{r_s} - 1)|(2^n - 1) \text{ and } (2^{r_s} - 1)|(2^m - 1).$$

Suppose now that $d = (2^n - 1, 2^m - 1)$. It follows that $d|(2^{r_i} - 1)$ for $i = 1, \dots, s$. Therefore $d|(2^{r_s} - 1) = 2^{(m,n)} - 1$. □

Now let $P = \{p_1, \dots, p_n\}$ be a finite set of primes with

$$2 = p_1 < p_2 < \cdots < p_n.$$

Then

$$(2^{p_i} - 1, 2^{p_j} - 1) = (2^{(p_i,p_j)} - 1) = 1 \text{ if } i \neq j.$$

For $i = 1, \dots, n$ each $2^{p_i} - 1$ is odd and hence they have pairwise different odd prime divisors. Since there are only $n - 1$ odd primes in P it follows that there must be a prime number not in P.

The Mersenne numbers are closely tied to what are called the **perfect numbers**. A natural number n is a **perfect number** if it is equal to the sum of its proper divisors. That is

$$n = \sum_{d|n, d\geq 1, d\neq n} d.$$

For example the number 6 is perfect since its proper divisors are 1, 2, 3 which add up to 6.

If we denote by $\sigma(n)$ the sum of all positive divisors of n, that is

$$\sigma(n) = \sum_{d|n, d\geq 1} d$$

then $\sigma(n) = 2n$ if and only if n is perfect. The following result, the first part is from Euclid and the second part due to Euler, gives the relation between perfect numbers and Mersenne primes.

Theorem 3.1.7 *Let (M_n) be the sequence of Mersenne numbers. Then*

(1) (Euclid) If $M_p = 2^p - 1$ is a Mersenne prime then

$$n = 2^{p-1}(2^p - 1)$$

is a perfect number.

(2) (Euler) If $n \geq 2$ is a perfect number and even then

$$n = 2^{p-1}(2^p - 1)$$

and $M_p = 2^p - 1$ is a Mersenne prime.

Proof (1) Suppose $2^p - 1 = q$ is a prime and let $n = 2^{p-1}(2^p - 1)$. Then

$$\sigma(n) = 1 + 2 + \cdots + 2^{p-1} + q + 2q + \cdots + 2^{p-1}q$$

$$= (q + 1)(1 + 2 + \cdots + 2^{p-1}) = 2^p(2^p - 1) = 2(2^{p-1}(2^p - 1)) = 2n.$$

Therefore $\sigma(n) = 2n$ and hence n is a perfect number.

(2) Suppose n is a perfect number. Let $n = 2^t u$ with u odd. The divisors of n are of the form $2^s m$ with $0 \leq s \leq t$ and $m|u$. Consider s fixed and consider the divisors $2^s m$. Their contribution to the sum $\sigma(n)$ is equal to $2^s\sigma(u)$. It follows that

$$\sigma(n) = (1 + 2 + \cdots + 2^t)\sigma(u) = (2^{t+1} - 1)\sigma(u).$$

Since n is perfect we have $\sigma(n) = 2n$ and hence

$$2^{t+1}u = (2^{t+1} - 1)\sigma(u).$$

Since u is odd from Euclid's lemma we get

$$\sigma(u) = 2^{t+1}a \text{ and } u = (2^{t+1} - 1)a$$

for some natural number a. The number u has two different divisors a and $(2^{t+1} - 1)a > a$. Their sum is $2^{t+1}a = \sigma(u)$. This is possible only if $u = (2^{t+1} - 1)a$ has no other divisors, that is if $a = 1$ and $2^{t+1} - 1$ is prime. It follows that $t + 1$ must be a prime, $2^{t+1} - 1$ is a Mersenne prime and n has the required form. □

This completely characterizes in terms of Mersenne primes the even perfect numbers. It is still an open question whether there is an odd perfect number.

Finally we mention a result called the **Lucas–Lehmer Test** which is useful in testing for large Mersenne primes. We will give this result again, as well as its proof, in Chapter 5 on primality testing.

Theorem 3.1.8 *Let p be an odd prime and define the sequence (S_n) inductively by*

$$S_1 = 4 \text{ and } S_n = S_{n-1}^2 - 2.$$

Then the Mersenne number $M_p = 2^p - 1$ is a Mersenne prime if and only if M_p divides S_{p-1}.

3.1.4 The Fibonacci Numbers and the Golden Section

The next sequence of integers that we consider is called the **Fibonacci numbers**. This sequence has many remarkable properties, some of which we will explore in this section. The interest in this sequence, both by professional mathematicians and by amateurs, has been almost mystical and there is a whole journal **The Fibonacci Quarterly** devoted to results surrounding these numbers. In addition this sequence has an intricate tie to a number called the **golden section** or **golden ratio** which has tremendous and varied applications in geometry.

Definition 3.1.3 *The* **Fibonacci numbers** *are the sequence* (f_n) *defined recursively by* $f_1 = 1,\ f_2 = 1$ *and then*

$$f_n = f_{n-1} + f_{n-2}\, for\ n \geq 3.$$

Hence the first few terms of the sequence are

$$1, 1, 2, 3, 5, 8, 13, 21, \ldots\ .$$

This sequence was introduced by the Italian mathematician Leonardo Pisano or Leonardo of Pisa. He is better known as Fibonacci, son of Bonnaccio, via a problem in his book Liber Abaci published in 1202. In this problem he asked the following.

How many pairs of rabbits will be produced in a year, beginning with a single pair, if in every month each pair bears a new pair which becomes productive from the second month on.

This leads to the following scheme with A being a productive pair and B becomes a productive pair from the second month on (Figure 3.1).

Computing, we then get the following table

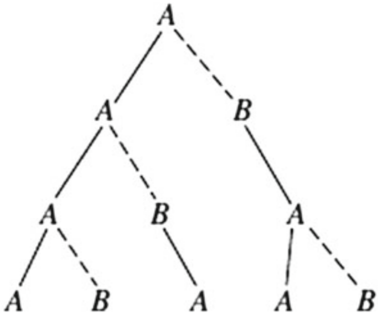


Fig. 3.1 Scheme for Fibonacci's Rabbit Problem

No. of A	No. of B	Total Number
1	0	1
1	1	2
2	1	3
3	2	5

and so on, which produces the recursive formula giving the Fibonacci numbers.

An alternative formulation of the Fibonacci numbers can be given by the next theorem.

Theorem 3.1.9 *Let $P_1 = P_2 = 1$ and for $n \geq 3$ let P_n be the number of 0-1 sequences of length $n - 2$ with no repeating 1's. Then $P_n = f_n$ for all n.*

Proof This is clear for $n = 3$. Then for $n > 3$ let q_n be the number of 0-1 sequences of length $n - 2$ with no repeating 1's and ending in 0 and let h_n be the number of 0-1 sequences of length $n - 2$ with no repeating 1's and ending in 1. For each such sequence of length $n - 2$ ending in 0 there are 2 new sequences of length $n - 1$ while there is only one new sequence for those ending in 1. Therefore

$$q_n = q_{n-1} + h_{n-1} \text{ and } h_n = q_{n-1}$$

and

$$P_n = q_n + h_n.$$

The result follows easily from this. □

The properties of the Fibonacci numbers are intricately tied to the number

$$\alpha = \frac{1 + \sqrt{5}}{2}.$$

This number is called the **golden section** or **golden ratio** and arises naturally in many geometric applications. Before continuing with the Fibonacci numbers we digress and discuss the golden section and its ties to geometry.

To define α, consider a line segment $\overline{AB}$, and let the point P be located so that it divides the line segment in **extreme to mean ratio**. By this we mean that

$$\frac{|AP|}{|PB|} = \frac{|AB|}{|AP|}.$$

If we let PB have length 1 as in Figure 3.2 then the length of AP is the golden section α.

To see that the value of α is $\frac{1+\sqrt{5}}{2}$ we have the ratio

$$\frac{\alpha}{1} = \frac{\alpha + 1}{\alpha}.$$

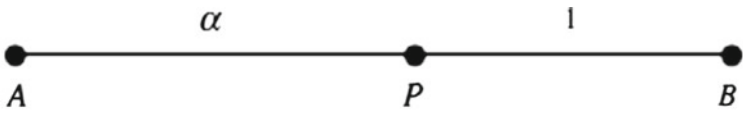
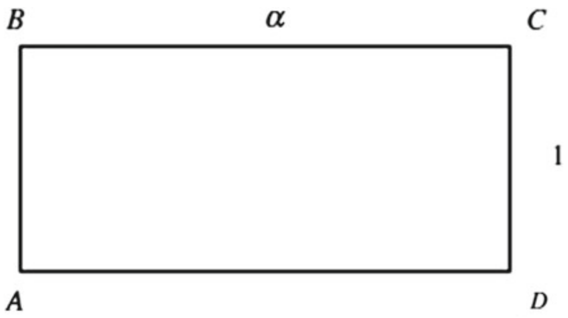


Fig. 3.2 Extreme to mean Ratio

Fig. 3.3 Golden Rectangle



This then gives the quadratic equation

$$\alpha^2 - \alpha - 1 = 0.$$

The two solutions are $\frac{1 \pm \sqrt{5}}{2}$ and since the golden ratio is positive we get that $\alpha = \frac{1+\sqrt{5}}{2}$ as desired.

If we have a rectangle $ABCD$ with $|BC| = \alpha$ and $|CD| = 1$ as in Figure 3.3 then this is a **golden rectangle**.

The classical Greeks regarded the golden rectangle as the most pleasing rectangular shape and built many of their temple fronts with this format.

If we begin with a golden rectangle $ABCD$ as in Figure 3.3 and remove the square $ABEF$, the remaining rectangle $ECDF$ is again a golden rectangle. To see this suppose that $|BC| = \alpha$ and $|CD| = 1$. Then

$$|EC| = \alpha - 1 \implies |CH| = \alpha - 1$$

and then

$$\frac{|DC|}{|EC|} = \frac{|DC|}{|CH|} = \frac{1}{\alpha - 1}$$

$$= \frac{1}{\frac{1+\sqrt{5}}{2} - 1} = \frac{1+\sqrt{5}}{2} = \alpha$$

This process of removing squares can be continued and each time we get a smaller golden rectangle as in Figure 3.4. Starting with A if opposite corners are connected

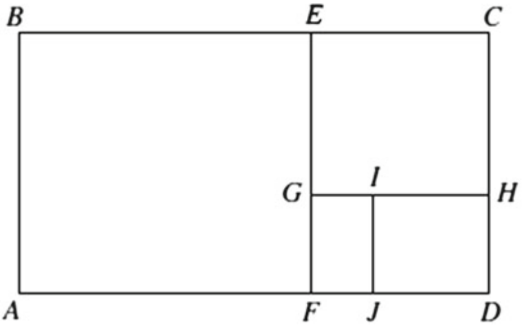


Fig. 3.4 Golden Spiral

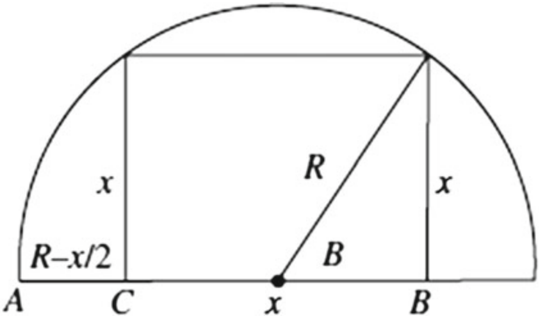


Fig. 3.5 Golden Section Relative to an Inscribed Square

by circular arcs with radius the side of the given square we get a logarithmic spiral called the **golden spiral**. Its equation in polar coordinates is $r = \alpha^{\frac{2\theta}{n}}$.

The golden section is of course an irrational number. However it can be constructed very easily with ruler and compass. To do this, start with a line segment AB of length 1, and a line segment AE of length $\frac{1}{2}$ and orthogonal to AB. Then the segment EB has length $\sqrt{1+\frac{1}{4}} = \frac{\sqrt{5}}{2}$. Adjoin to EB a line segment BC of length $\frac{1}{2}$ and EC has length α.

The golden section arises naturally in many geometric applications. We describe several of these. First, consider a square inscribed in a semicircle of radius R as pictured in Figure 3.5.

Suppose $|AB| = r$ and let x be the length of the side of the inscribed square. Then $r = R + \frac{x}{2}$. We then have from the Pythagorean theorem

$$x = \frac{2}{\sqrt{5}}R.$$

But then

$$|AB| = r = R(1 + \frac{1}{\sqrt{5}}) \text{ and } r - x = R(1 - \frac{1}{\sqrt{5}}).$$

Since

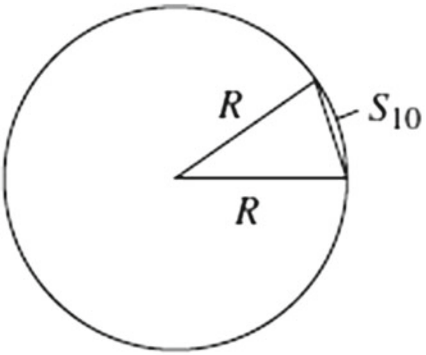


Fig. 3.6 Regular Decagon Inscribed in a Circle

$$\frac{(1+\frac{1}{\sqrt{5}})}{\frac{2}{\sqrt{5}}} = \frac{\frac{2}{\sqrt{5}}}{(1-\frac{1}{\sqrt{5}})}$$

we have

$$\frac{r}{x} = \frac{x}{r-x}$$

that is the point C divides the line segment AB by the golden ratio.

Next consider a regular decagon inscribed in a circle of radius R. A side S_{10} has length $2R\sin(\frac{\pi}{10})$ (Figure 3.6).

Using the trigonometric identities

$$\sin(\frac{2\pi}{10}) = 2\sin(\frac{\pi}{10})\cos(\frac{\pi}{10})$$

$$\cos(\frac{2\pi}{10}) = 1 - 2\sin^2(\frac{\pi}{10})$$

and

$$\sin(\frac{4\pi}{10}) = \cos(\frac{\pi}{10})$$

we get that

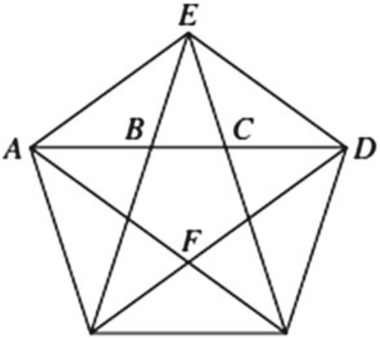
$$4\sin(\frac{\pi}{10})(1 - 2\sin^2(\frac{\pi}{10})) = 1.$$

Therefore the value of $\sin(\frac{\pi}{10})$ is a solution of the polynomial equation

$$4x(1-2x^2) = 1.$$

Since $\sin(\frac{\pi}{10}) > 0$ and $\sin(\frac{\pi}{10}) \neq \frac{1}{2}$ we obtain

Fig. 3.7 Regular Pentagon



$$\sin(\frac{\pi}{10}) = \frac{\sqrt{5}-1}{4} = \frac{1}{2(\alpha-1)}$$

where α is the golden section. Therefore

$$|S_{10}| = 2R\sin(\frac{\pi}{10}) = \frac{R}{\alpha-1} = R\alpha.$$

Hence the side of a regular decagon inscribed in a circle is the bigger section of the radius divided by the golden section.

Using this connection it is easy to construct regular decagons and regular pentagons with ruler and compass.

Next consider a regular pentagon. Its diagonals describe a regular star like the pentagon in Figure 3.7.

The angle $\angle AFD$ is $\frac{6\pi}{10}$ while the angle $\angle ADF$ is $\frac{2\pi}{10}$. From the law of sines we have

$$\frac{|AD|}{|AF|} = \frac{\sin(\frac{6\pi}{10})}{\sin(\frac{2\pi}{10})} = 2\cos(\frac{2\pi}{10}) = \alpha$$

since

$$2\cos(\frac{2\pi}{10}) = 2 - 4\sin^2(\frac{2\pi}{5}) = 2 - \frac{1}{\alpha^2} = \alpha,$$

Because $|AF| = |AC|$ we have $\frac{|AD|}{|AC|} = \alpha$ and hence the point C divides the line segment AD by the golden ratio.

Finally consider a rectangle as in Figure 3.8.

We wish to find the points P and Q so that the triangles $\triangle PAQ$ and $\triangle QBC$ and $\triangle CDP$ all have equal area.

If the triangles do have equal area we have the identities

$$xw = y(w+z) = z(x+y) \implies xw = yw + yz = xz + yz.$$

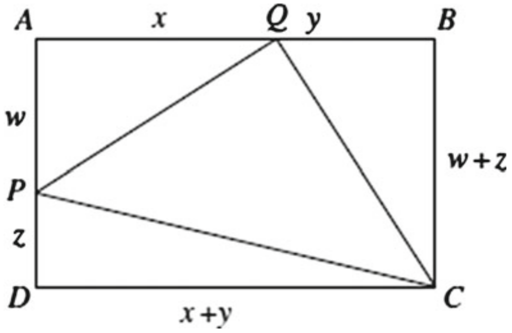


Fig. 3.8 Rectangle

This implies that

$$yw = xz \implies \frac{w}{z} = \frac{x}{y}.$$

Then from $xw = y(w+z)$ we get

$$\frac{x}{y} = \frac{w+z}{w} = 1 + \frac{z}{w} = 1 + \frac{1}{\frac{w}{z}} = 1 + \frac{1}{\frac{x}{y}}.$$

This means that

$$(\frac{x}{y})^2 - \frac{x}{y} - 1 = 0 \implies \frac{x}{y} = \frac{w}{z} = \alpha.$$

Hence the solution to the equal area problem is precisely the points P and Q which divide the sides AB and AD in the golden ratio.

We now return to the Fibonacci numbers and first show the tie to the golden section.

Theorem 3.1.10 *(Binet Formula) Let (f_n) be the Fibonacci sequence, let $\alpha = \frac{1+\sqrt{5}}{2}$ be the golden section and let $\beta = -\alpha^{-1} = \frac{1-\sqrt{5}}{2}$. Then for $n \geq 1$,*

$$f_n = \frac{\alpha^n - \beta^n}{\alpha - \beta}.$$

Proof The golden section α together with β as defined in the statement of the theorem is the zeros of the polynomial

$$x^2 - x - 1 = 0.$$

It follows that

$$\alpha^{n+2} = \alpha^{n+1} + \alpha^n$$

and

$$\beta^{n+2} = \beta^{n+1} + \beta^n \text{ for } n \geq 1$$

Further $\alpha - \beta = \sqrt{5} \neq 0$. We then have

$$f_1 = \frac{\alpha - \beta}{\alpha - \beta},$$

$$f_2 = \frac{\alpha^2 - \beta^2}{\alpha - \beta} = \alpha + \beta = 1,$$

and

$$f_{n+2} = \frac{\alpha^{n+1} - \beta^{n+1}}{\alpha - \beta} + \frac{\alpha^n - \beta^n}{\alpha - \beta} = f_{n+1} + f_n$$

for $n \geq 3$. □

Corollary 3.1.2 *If f_n and α are as above then*

$$\lim_{n \to \infty} \frac{f_{n+1}}{f_n} = \alpha = 1 + \cfrac{1}{1 + \cfrac{1}{1+ \ddots}}.$$

Proof From the Binet formula

$$\frac{f_{n+1}}{f_n} = \frac{\alpha^{n+1} - \beta^{n+1}}{\alpha^n - \beta^n} = \frac{1 - (\frac{\beta}{\alpha})^{n+1}}{\alpha^{-1}(1 - (\frac{\beta}{\alpha})^n)}.$$

Since $|\frac{\beta}{\alpha}| < 1$ this clearly goes to α as $n \to \infty$. Further by rearranging it is easily seen that

$$\frac{f_{n+1}}{f_n} = 1 + \frac{1}{\frac{f_n}{f_{n-1}}}.$$

□

We now list a collection of properties of the Fibonacci numbers. In addition to showing the rich theory of these numbers they will lead us to two more proofs of the infinitude of primes. Throughout all the remainder of this section the f_n are the Fibonacci numbers and α is the golden section.

Lemma 3.1.5 $f_1 + f_2 + \cdots + f_n = f_{n+2} - 1, n \geq 1$.

Proof This is correct for $n = 1$ and $n = 2$. For $n \geq 3$ we have

$$f_1 + \cdots + f_{n-1} + f_n = f_{n+1} - 1 + f_n = f_{n+2} - 1.$$

□

The next four results are again straightforward inductions, the first on n directly and the second fixing n and inducting on m. We leave the details to the exercises.

Lemma 3.1.6 $f_n f_{n+1} = f_1^2 + f_2^2 + \cdots + f_n^2$ *for* $n \geq 1$.

Lemma 3.1.7 $f_n^2 - f_{n-1} f_{n+1} = (-1)^n$ *for* $n \geq 1$.

Lemma 3.1.8 $f_{n+m} = f_{n-1} f_m + f_n f_{m+1}, n \geq 1$ *where* $f_0 = 0$.

Lemma 3.1.9 *(a) If r, s are positive integers then r dividing s implies that f_r divides f_s. Conversely if $m \geq 2$ then if $f_n | f_m$ it follows that $n | m$.*

(b) $(f_n, f_m) = f_{(m,n)}$. That is the gcd of f_n and f_m is the fibonacci number indexed by the gcd of the (m, n) term in the Fibonacci sequence. In particular f_n and f_m are relatively prime if m and n are relatively prime.

Proof (a) Recall that $\alpha\beta = -1$ and $\alpha + \beta = 1$. We then have

$$f_{rs} = \frac{\alpha^{rs} - \beta^{rs}}{\alpha - \beta}$$

$$= \frac{\alpha^s - \beta^s}{\alpha - \beta}(\alpha^{(r-1)s} + \alpha^{(r-2)s}\beta^s + \cdots + \alpha^s \beta^{(r-2)s} + \beta^{(r-1)s}).$$

Hence if $r | s$ then $f_r | f_s$.

We need part (b) in order to prove the converse. Suppose that $m > n$. Then by the Euclidean algorithm we have $r_t = (m, n)$ where

$$m = nq_0 + r_1 \text{ with } 0 \leq r_1 < n$$

$$n = r_1 q_1 + r_2 \text{ with } 0 \leq r_2 < r_1$$

$$\cdots$$

$$r_{t-2} = r_{t-1} q_{t-1} + r_t \text{ with } 0 \leq r_t < r_{t-1}$$

$$r_{t-1} = r_t q_t.$$

Then applying this to the corresponding Fibonacci numbers we have from Lemma 3.1.8:

$$(f_n, f_m) = (f_{nq_0 + r_1}, f_n) = (f_{nq_0 - 1} f_{r_1} + f_{nq_0} f_{r_1 + 1}, f_n)$$

$$= (f_{nq_0 - 1} f_{r_1}, f_n) = (f_{r_1}, f_n)$$

because $f_n | f_{nq_0}$ from the first part of part (a) and $(f_{nq_0}, f_{nq_0 - 1}) = 1$. (Clearly two neighboring Fibonacci numbers are relatively prime.)

Analogously

$$(f_{r_1}, f_n) = (f_{r_2}, f_{r_1}) = \cdots = (f_{r_t}, f_{r_{t-1}}) = f_{r_t}$$

since $f_{r_t} | f_{r_{t-1}}$. This completes the proof of part (b).

We now consider the second half of part (a). Suppose that $m \geq 2$ and that $f_n | f_m$. Then

$$f_n = (f_n, f_m) = f_{(m,n)}$$

from part (b). It follows that $n|m$ since $m \geq 2$ and $f_r < f_s$ if $2 \leq r < s$. □

Lemma 3.1.10 *(a)* $f_{2k} = f_k(f_{k+1} + f_{k-1}) = f_{k+1}^2 - f_{k-1}^2$.

(b) $f_{2k} = \sum_{i=0}^{k} \binom{k}{i} f_i$ *where* $\binom{k}{i}$ *is the binomial coefficient.*

(c) $f_{n+1} = \sum_{i=0}^{[\frac{n}{2}]} \binom{n-i}{i}$ *where* $[x]$ *is the greatest integer function.*

Proof These are all applications of the Binet formula. For part (a) we have

$$f_{2k} = f_k(\alpha^k + \beta^k) = f_k(\frac{\alpha^{k-1} - \beta^{k-1} + \alpha^{k+1} - \beta^{k+1}}{\alpha - \beta})$$

$$= f_k(f_{k-1} + f_{k+1}) = f_{k+1}^2 - f_{k-1}^2.$$

For part (b) apply the Binet formula to obtain

$$\sum_{i=0}^{k} \binom{k}{i} f_i = \frac{1}{\alpha - \beta}(\sum_{i=0}^{k} \binom{k}{i}(\alpha^i - \beta^i))$$

$$= \frac{1}{\alpha - \beta}((1+\alpha)^k - (1+\beta)^k) = \frac{1}{\alpha - \beta}(\alpha^{2k} - \beta^{2k}) = f_{2k}.$$

Finally for part (c), it clearly holds for $0 \leq n \leq 2$. Suppose now $n \geq 2$ and we proceed by induction. Then

$$f_{n+1} = f_n + f_{n-1} = \sum_{i=0}^{[\frac{n-1}{2}]} \binom{n-1-i}{i} + \sum_{i=0}^{[\frac{n-2}{2}]} \binom{n-2-i}{i}.$$

We first consider the case where $n = 2m$ with $m \geq 1$. Then $[\frac{n-1}{2}] = m - 1 = [\frac{2m-2}{2}]$ and hence from above

$$f_{n+1} = \sum_{i=0}^{m-1} \binom{2m-1-i}{i} + \sum_{i=0}^{m-1} \binom{2m-1-(i+1)}{(i+1)-1}$$

$$= \binom{2m-1}{0} + \binom{2m-1-m}{m-1} + \sum_{i=1}^{m-1}(\binom{2m-1-i}{i} + \binom{2m-1-i}{i-1})$$

$$= \sum_{i=0}^{m} \binom{2m-i}{i}$$

completing the even case.

Now suppose n is odd so $n = 2m + 1$ with $m \geq 1$. Then

$$[\frac{n-1}{2}] = m \quad [\frac{n-2}{2}] = m - 1 \quad [\frac{n}{2}] = m$$

and hence

$$f_{n+1} = \sum_{i=0}^{m} \binom{2m-i}{i} + \sum_{i=0}^{m-1} \binom{2m-1-i}{i+1-1}$$

$$= \binom{2m}{0} + \sum_{i=1}^{m} \binom{2m+1-i}{i}$$

$$= \sum_{i=0}^{m} \binom{2m+1-i}{i}$$

finishing the odd case and part (c). □

The next result and corollary deal with the relationship between the Fibonacci numbers and the primes. This will lead directly to another proof that there are infinitely many primes.

Theorem 3.1.11 *Let p be a prime. Then*

(1) $p|f_p$ if $p = 5$ and $p|f_{p-1}$ or $p|f_{p+1}$ if $p \neq 5$.

(2) $p|f_{p+1}$ if $p = 2$.

(3) $p|f_{p-1}$ if p is congruent to ± 1 modulo 10.

(4) $p|f_{p+1}$ if p is congruent to ± 3 modulo 10.

Proof If $p = 2$ then $f_3 = 2$ and hence $p|f_{p+1}$. If $p = 3$ then $f_4 = 3$ and $p|f_{p+1}$. If $p = 5$ then $f_5 = 5$ and $p|f_p$. Now let $p \geq 7$. By Binet's formula

$$f_n = \frac{1}{\sqrt{5}}(\frac{1+\sqrt{5}}{2})^n - \frac{1}{\sqrt{5}}(\frac{1-\sqrt{5}}{2})^n, n \geq 1$$

and by the binomial expansion

$$(1 \pm \sqrt{5})^n = 1 \pm \binom{n}{1}\sqrt{5} + \binom{n}{2}5 \pm \binom{n}{3}(\sqrt{5})^3 + \cdots + (-1)^n(\sqrt{5})^n.$$

If n is odd then

$$2^{n-1} f_n = \frac{1}{2\sqrt{5}}((1+\sqrt{5})^n - (1-\sqrt{5})^n) = n + \binom{n}{3}5 + \binom{n}{5}5^2 + \cdots + 5^{\frac{n-1}{2}}.$$

Now let $n = p$ be prime. Since $p|\binom{p}{i}$ if $1 \leq i < p$ we must have

$$f_p \equiv 5^{\frac{p-1}{2}} \bmod p \text{ and hence}$$

$$f_p^2 \equiv 1 \bmod p$$

by Fermat's theorem. Since

$$f_p^2 - f_{p-1} f_{p+1} = (-1)^{p-1} = 1$$

we get

$$0 \equiv f_p^2 - 1 \equiv f_{p-1} f_{p+1} \bmod p.$$

Therefore $p|f_{p+1}$ or $p|f_{p-1}$ since $(f_{p-1}, f_{p+1}) = f_{(p-1,p+1)} = f_2 = 1$. More concretely, we can use the above identities to show that

$p|f_{p-1}$ if p is congruent to ±1 modulo 10 and

$p|f_{p+1}$ if p is congruent to ±3 modulo 10 (see exercises). □

Corollary 3.1.3 *Let p be a prime greater than 7. Then each prime divisor of f_p is greater than p.*

Proof Let q be a prime divisor of f_p with $p \geq 7$ a prime. Assume $q \leq p$. If $q = p$ then $q = p = 5$ and hence we may assume that $q < p$. We then have

$$(f_p, f_q) = f_{(p,q)} = f_1 = 1,$$

$$(f_p, f_{q-1}) = f_{(p,q-1)} = f_1 = 1,$$

$$(f_p, f_{q+1}) = f_{(p,q+1)} = f_1 = 1.$$

Then from Lemma 3.1.10, either $q|f_q$ or $q|f_{q-1}$ or $q|f_{q+1}$. This gives a contradiction because $q|f_p$ and $q|f_q$ implies that $q|f_1 = 1$ and $q|f_p$ and $q|f_{q+1}$ or $q|f_{q-1}$ also implies that $q|1$. Therefore we must have that $q > p$. □

Based on the Fibonacci numbers, we can now give two more proofs of the fact that there are infinitely many primes.

Proof One: Let $M = \{p_1, \ldots, p_n\}$ be a finite set of distinct prime numbers and suppose that $p_1 < p_2 < \cdots < p_n$ with $p_n \geq 7$. Let p be a prime divisor of f_{p_n}. Then from Corollary 3.1.3 we must have $p > p_n$ and hence $p \notin M$. □

Proof Two: Suppose $\{p_1, \ldots, p_n\}$ with $p_1 = 2$ are all the prime numbers. We have $f_{p_i} > 1$ for $i = 2, \ldots, n$. Then at most one of the f_{p_i} for $i = 2, \ldots, n$ has two prime divisors for otherwise since $(f_{p_i}, f_{p_j}) = f_{(p_i,p_j)} = 1$ for $i \neq j$ we would already have $n + 1$ primes. This contradicts for example that

$$f_{19} = (37)(113) \text{ and } f_{31} = (557)(2417).$$

□

We note that many of the ideas concerning the Fibonacci numbers can be greatly generalized. For example suppose K is an arbitrary field and $x, y \in K$. Then we define

$$T_0(x, y) = 0, T_1(x, y) = 1 \text{ and then}$$

$$T_n(x, y) = xT_{n-1}(x, y) - yT_{n-2}(x, y)$$

This sequence in K will satisfy many of the same properties as the Fibonacci numbers. If A is a 2×2 invertible matrix over K with $tr(A) = x$ and $det(A) = y$ then

$$A^n = T_n(x, y)A + yT_{n-1}(x, y)I$$

where I is the identity matrix. In particular

$$T_n(x, y)^2 - T_{n+1}(x, y)T_{n-1}(x, y) = y^{n-1}, n \geq 1$$

If $x = 1$ and $y = -1$ then $T_n(x, y) = f_n$ for $n \geq 1$.

These generalized Fibonacci numbers are also related to the Chebyshev polynomials which play a role in the general approximation of functions. If $y = 1$ and $n \geq 1$ then

$$T_n(x, 1) = S_n(x)$$

where $S_n(x)$ is nth Chebyshev polynomial of the second kind. We have

$$S_{n+m}(x) = S_n(x)S_{m+1}(x) - S_m(x)S_{n-1}(x)$$

and

$$S_{nm}(x) = S_m(S_{n+1}(x) - S_{n-1}(x)) \cdot S_n(x)$$

for all natural numbers n, m. As polynomials in x these Chebyshev polynomials satisfy

$$S_{(m,n)}(x) = (S_n(x), S_m(x)).$$

For positive real values these Chebyshev polynomials have a particularly simple form. If $K = \mathbb{R}$ and $x \geq 0$ then let $x = 2\cos\theta < 2$. Then

$$S_n(x) = \frac{\sin(n\theta)}{\sin(\theta)}.$$

If $x = 2\cosh\theta > 2$ then

$$S_n(x) = \frac{2\sinh(n\theta)}{\sinh(\theta)}$$

while if $x = 2$ then

$$S_n(x) = n.$$

3.1.5 Some Simple Cases of Dirichlet's Theorem

Recall that Dirichlet's Theorem, which we will state and prove formally in Section 3.3, says that if a, b are positive integers with $(a, b) = 1$ then there are infinitely many primes of the form $an + b$. In this section we prove certain special cases of this result which can be handled by elementary methods. Most of these proofs depend on the following easy idea. Suppose $x \in \mathbb{N}$ has the prime factorization

$$x = p_1^{e_1} \cdots p_k^{e_k}.$$

Then if each $p_i \equiv 1 \bmod m$ then $x \equiv 1 \bmod m$. This fact is direct from the multiplicative property of congruences.

We first handle the case modulo 4.

Lemma 3.1.11 *There exist infinitely many primes of the form* $4n + 3$ *and infinitely many of the form* $4n + 1$.

Proof Suppose there are only finitely many primes of the form $4n + 3$, say $p_1, \ldots, p_k$, with p_k the largest. Let $q_1, \ldots, q_t$ be all the primes of the form $4n + 1$ less than p_k. Let

$$x = 4 \cdot 3 \cdot 7 \cdots p_k q_1 \cdots q_t - 1.$$

Then $x \equiv -1 \equiv 3 \bmod 4$ and hence x must be divisible by a prime $p \equiv 3 \bmod 4$. But then $p | (4 \cdot 3 \cdot 7 \cdots p_k q_1 \cdots q_t)$ so p cannot divide x and thus a contradiction. Therefore there are infinitely many primes of the form $4n + 3$.

To handle the case $4n + 1$ we must recall some facts about quadratic residues. From Section 2.6 it follows that if p is a prime greater than or equal to 3 then

$$(-1/p) = (-1)^{\frac{p-1}{2}}.$$

Hence -1 is a quadratic residue mod p only if $p \equiv 1 \bmod 4$. Equivalently if x is any positive integer then if $p | (x^2 + 1)$ it follows that $p \equiv 1 \bmod 4$. Now suppose that there are only finitely many primes of the form $4n + 1$ say $q_1, \ldots, q_k$. Let $x = q_1 \cdots q_k$ and let p be a prime divisor of $x^2 + 1$. Then $p \equiv 1 \bmod 4$. But $p|x$

so $p|x^2$ and hence p cannot divide x^2+1. Therefore a contradiction and there must exist infinitely many primes of the form $4n+1$. □

Essentially the same methods handle the situation modulo 8.

Lemma 3.1.12 *There exist infinitely many primes of each of the forms* $8n+1$, $8n+3$, $8n+5$ *and* $8n+7$.

Proof From the fact that $(2/p)=(-1)^{\frac{p^2-1}{8}}$ if $p\geq 3$ is prime (see Section 2.6) we can obtain the following results whose proofs we leave to the exercises. If x is any positive integer and $p\geq 3$ is a prime then

(1) If $p|(x^4+1)$ then $p\equiv 1 \bmod 8$.
(2) If $p|(x^2-2)$ then either $p\equiv 1 \bmod 8$ or $p\equiv 7 \bmod 8$.
(3) If $p|(x^2+2)$ then either $p\equiv 1 \bmod 8$ or $p\equiv 3 \bmod 8$.

Now suppose that there exists only finitely many primes of the form $8n+1$, say $p_1,\dots,p_k$, and let $x=p_1\cdots p_k$. Let p be a prime divisor of x^4+1. Then from above $p\equiv 1 \bmod 8$, but p is not one of $p_1,\dots,p_k$, and hence a contradiction. Therefore there exist infinitely many primes of the form $8n+1$.

Suppose next that there exists only finitely many primes of the form $8n+7$. As before call them $p_1,\dots,p_k$ and let $x=p_1\cdots p_k$. Now each $p_i\equiv -1 \bmod 8$ and so $x\equiv \pm 1 \bmod 8$ and so $x^2\equiv 1 \bmod 8$. Let p be a prime divisor of x^2-2. It must be congruent to either 1 or 7 modulo 8. If each prime divisor of x^2-2 is congruent to 1 mod 8 then x^2-2 is also congruent to 1 modulo 8. However x^2 is congruent to 1 modulo 8 and so x^2-2 is not congruent to 1 modulo 8. Therefore there must exist a prime divisor p of x^2-2 congruent to 7 modulo 8. This p cannot be one of $p_1,\dots,p_k$ and hence a contradiction.

The case of the form $8n+3$ is handled in an analogous manner (see exercises).

To handle the case $8n+5$ we first show the following. □

Lemma 3.1.13 *Let* a,b *be nonzero integers with* $(a,b)=1$. *Then each odd prime divisor of* a^2+b^2 *is of the form* $4n+1$.

Proof Let p be an odd prime divisor of a^2+b^2. Then there exists an n with

$$n^2=-1+kp$$

for some $k\in\mathbb{Z}$. Hence -1 is a quadratic residue mod p and therefore $p\equiv 1 \bmod 4$.

Now let p be the largest prime of the form $8n+5$ and let

$$x=3^2 5^2\cdots p^2+4$$

where $3,5,\dots,p$ are all the primes up to p and $p>7$. From Lemma 3.1.12 any prime divisor of x is congruent to 1 modulo 4 so then congruent to either 1 modulo 8 or 5 modulo 8. Since $(2m+1)^2+4=4m(m+1)+5$ it follows that x is congruent to 5 modulo 8. Therefore x must have a prime divisor of the form $8n+5$ which is larger then p. □

A slight modification and the use of quadratic reciprocity allow us to handle primes modulo 3.

Lemma 3.1.14 *There exist infinitely many primes of the form* $3n + 1$ *and infinitely many of the form* $3n + 2$.

Proof The case $3n + 2$ is handled directly. Suppose that $p_1, \dots, p_k$ are all the primes congruent to 2 modulo 3 and let $x = p_1 p_2 \cdots p_k$. If $x \equiv 1 \bmod 3$ then $x + 1 \equiv 2 \bmod 3$. Hence there must be a prime congruent to 2 mod 3 dividing $x + 1$. But as before $p|(p_1 \cdots p_k)$ so p cannot divide $x + 1$.

If $x \equiv 2 \bmod 3$, then $x + 3 \equiv 2 \bmod 3$. Then as before there must be a prime $p \equiv 2 \bmod 3$ dividing $x + 3$. But $p|x$ so p cannot divide $x + 3$. These two contradictions then imply that there are infinitely many primes of the form $3n + 2$.

To handle $3n + 1$ we must use quadratic reciprocity. Consider for an odd prime p

$$(-3/p) = (-1/p)(3/p).$$

Now $(-1/p) = (-1)^{\frac{p-1}{2}}$ and $(3/p) = (-1)^{\frac{p-1}{2}}(p/3)$ by quadratic reciprocity. Therefore

$$(-3/p) = (-1)^{\frac{p-1}{2}}(-1)^{\frac{p-1}{2}}(p/3) = (p/3).$$

Directly then

$$(p/3) = 1 \text{ if } p \equiv 1 \bmod 3$$

$$(p/3) = -1 \text{ if } p \equiv -1 \bmod 3.$$

Therefore -3 is a quadratic residue mod p only if $p \equiv 1 \bmod 3$. Equivalently for any integer x any odd prime divisor of $x^2 + 3$ must be congruent to 1 mod 3.

Now suppose that there are only finitely many primes of the form $3n + 1$ say $p_1, \dots, p_k$. Let $x = 2p_1 \cdots p_k$ and let p be a prime divisor of $x^2 + 3$. Then $p \equiv 1 \bmod 3$ but as before p cannot be one of the p_i. Hence there are infinitely many of the form $3n + 1$. □

The methods used in the preceding lemmas can handle many other special situations of Dirichlet's Theorem, for example $6n + 5$. However they cannot be extended to the whole result. We close this section with one general result which can be proved with the same kinds of elementary methods. The proof of this result is taken from [NZ] which was a modification of a result in [NP].

Theorem 3.1.12 *Let* m *be a positive integer. Then there exist infinitely many primes of the form* $mn + 1$.

Proof The theorem is actually a consequence of the next lemma which is interesting in its own right. □

Lemma 3.1.15 *Given a positive integer* m *then there exists a prime divisor of* $m^m - 1$ *which is congruent to* 1 *modulo* m.

Proof (Lemma 3.1.15) Suppose that given $m > 0$ there is no prime $p \equiv 1 \bmod m$ such that $p|(m^m - 1)$. For any prime factor q of $m^m - 1$ let h be the order of m modulo q that is h is the smallest positive integer such that $m^h \equiv 1 \bmod q$. Since the nonzero elements in $\mathbb{Z}_q$ form a multiplicative group it follows that $h|q-1$ and $h|m$ (see Chapter 2). If $h = m$ then $m|(q-1)$ and $q \equiv 1 \bmod m$ contrary to the assumption above. Therefore $h \neq m$ and $m = hc$ with $c > 1$. This holds, under the assumption, for possibly different h and c for any prime divisor of $m^m - 1$.

Suppose q^r is the highest power of q dividing $m^m - 1$. Then

$$m^m - 1 = (m^h - 1)(m^{ch-h} + m^{ch-2h} + \cdots + m^h + 1).$$

Since $m^h \equiv 1 \bmod q$ we have

$$m^{ch-h} + m^{ch-2h} + \cdots + m^h + 1 \equiv 1 + 1 + \cdots + 1 \equiv c \bmod q.$$

But q is a divisor of $m^m - 1$ so q is not a divisor of m or c and hence not of $m^{ch-h} + m^{ch-2h} + \cdots + m^h + 1$. Therefore q^r is also the highest power of q dividing $m^h - 1$. Further the same argument shows that if $s|m$ then q^r is also the highest power of q dividing $m^s - 1$.

Given a prime divisor q or m let h, c be defined as above and then let the distinct prime divisors of c and m be

$$p_1, \ldots, p_k \text{ and } p_1, \ldots, p_k, p_{k+1}, \ldots, p_n \text{ respectively}$$

with $1 \leq k \leq n$. Then h is not a divisor of any of the integers

$$\frac{m}{p_{k+1}}, \frac{m}{p_{k+2}}, \ldots, \frac{m}{p_n}.$$

Consider the integers of the form

$$\frac{m}{p_{i_1} p_{i_2} \cdots p_{i_t}}$$

where $1 \leq i_1 < i_2 < \ldots < i_t$. Let T be the set of integers of this form with t odd and U the set with t even. Define

$$Q = \frac{\prod_{s \in T}(m^s - 1)}{\prod_{s \in U}(m^s - 1)}.$$

We show that $Q = m^m - 1$ and then show that this is impossible leading to a contradiction and hence there must be a prime divisor congruent to 1 mod m.

To show first that $Q = m^m - 1$ we show that the prime power factors are the same. Each exponent s appearing in Q divides m and hence we need only consider prime factors of $m^m - 1$. If for a prime divisor q of $m^m - 1$ the corresponding $i_t > k$ then

h does not divide s. On the other hand if $i_t \leq k$ then the highest power of q dividing $m^s - 1$ is q^r also as shown above. Therefore q is a divisor of any term $m^s - 1$ in Q if and only if $h|s$ and this is true if and only if $i_t \leq k$. The number of factors of $m^s - 1$ in the numerator of Q having $i_t \leq k$ is

$$\binom{k}{1} + \binom{k}{3} + \binom{k}{5} + \cdots . \tag{3.1.5.1}$$

Similarly the number of factors of $m^s - 1$ in the denominator of Q having $i_t \leq k$ is

$$\binom{k}{2} + \binom{k}{4} + \binom{k}{6} + \cdots \tag{3.1.5.2}$$

If we subtract (3.1.5.1) from (3.1.5.2) we get the binomial expansion of $1 - (1 - 1)^k$ which clearly has value 1. It follows that Q must be an integer and the highest power of q dividing Q is q^r. Since this holds for every prime divisor q of $m^m - 1$ it must be the case that $Q = m^m - 1$.

We now show that this is impossible. Rewriting Q as $m^m - 1$ we get

$$(m^m - 1) \prod_{s \in U} (m^s - 1) = \prod_{s \in T} (m^s - 1).$$

Let b be the smallest integer of the form $\frac{m}{p_{i_1} p_{i_2} \cdots p_{i_t}}$ and consider the above equation modulo m^{b+1}. Every factor $m^s - 1$ is congruent to -1 modulo m^{b+1} except $m^b - 1$. Therefore the above equation reduces to

$$\pm(m^b - 1) \equiv \pm 1 \bmod m^{b+1}.$$

This then implies that

$$m^b \equiv 0 \bmod m^{b+1} \text{ or } m^b \equiv -2 \bmod m^{b+1}.$$

Both of these congruences are impossible since b is positive and $m \geq 2$. This contradiction establishes Lemma 3.1.15. □

We now prove Theorem 3.1.12.

Proof We want to show that given an m there are infinitely many primes of the form $mn + 1$. From Lemma 3.1.15 we know that in any progression of the form $1 + m, 1 + 2m, \ldots$ there is a prime that is a divisor of $m^m - 1$. Since this holds for any m it follows that in any arithmetic progression $1 + M, 1 + 2M, \ldots$ there must be a prime. Suppose then that for some m there are only finitely many primes of the form $mn + 1$ and let P be the product of these primes. From the observation above with $M = mP$ there is a prime q in the arithmetic progression $1 + mP, 1 + 2mP, \ldots, 1 + nmP, \ldots$. This prime is congruent to 1 modulo m but is not a divisor of the product P. Therefore a contradiction and hence there must be infinitely many primes of the form $nm + 1$. □

We note that the proof can be modified to also show that there infinitely many primes of the form $nm - 1$.

3.1.6 A Topological Proof and a Proof Using Codes

We close this section on elementary proofs of the infinitude of primes by presenting several more; one topological, one using codes and two more elementary analytic proofs.

We first look at the topological proof which is due to H. Furstenberg [Fu].

Proof **Using Topology**

We introduce a topology on the integers $\mathbb{Z}$. As a basis for the topology we take all arithmetic progressions from $-\infty$ to ∞. Each arithmetic progression is then open but also closed since its complement is a union of these arithmetic progressions. Hence each finite union of arithmetic progressions is closed.

Now let A_p be those arithmetic progressions consisting of multiples of a prime p, that is

$$A_p = \{..., -np, \dots, -p, 0, p, \dots, np, ...\} \text{ for } n \in \mathbb{N}.$$

Now let $A = \cup_p A_p$ where this union is taken over all primes p. The complement of A is $\{-1, 1\}$. Since $\{-1, 1\}$ is not open A is not closed. Hence A cannot be a finite union of closed sets. Therefore the number of primes must be infinite. □

A variation of this was given by S. Golomb [Go]. As a basis for the topology take all arithmetic progressions $an + b$ from $-\infty$ to ∞. Let A_{np} be those arithmetic progressions consisting of multiples of np where n is a positive integer and p is a prime. The progression $\{np\}$ with p a prime is closed and $X = \cup_p A_{np}$ is not closed. Then in the same manner as above the number of primes must be infinite.

We next give a proof using codes which is due to I. Stewart. We first need the following theorem.

Theorem 3.1.13 *If we have a finite set of* 2^N *elements and map it bijectively onto a set of binary strings then at least one string has length* $\geq N$.

Proof There are only $2^N - 1$ binary strings of length $< N$, the empty string, two of length 1, four of length 2, ..., 2^{N-1} of length $N - 1$. □

Now we can give our proof using codes.

Proof **Using Codes**

Assume that the set of primes is finite say $\{p_1, \dots, p_r\}$. We introduce a code via strings for each natural number together with zero. For 0 we choose the symbol 0. For each natural number n we write it as a product of primes and for each prime divisor we write down the multiplicity in the product. For the listing of these multiplicities we use brackets to start and end a listing. Suppose $r = 5$ then the primes are 2, 3, 5, 7, 11. Then we get the following codes for the first few natural numbers

$$\begin{aligned}
0 &\leftrightarrow 0\\
1 &\leftrightarrow [00000]\\
2 &\leftrightarrow [[00000]0000]\\
3 &\leftrightarrow [0[00000]000]\\
4 &\leftrightarrow [[[00000]0000]0000]\\
5 &\leftrightarrow [00[00000]00]\\
6 &\leftrightarrow [[00000][00000]000]
\end{aligned}$$

To analyze these codes we shorten each representation by canceling the closing brackets and take 1 for the starting bracket. Hence

$$\begin{aligned}
0 &\leftrightarrow 0\\
1 &\leftrightarrow 100000\\
2 &\leftrightarrow 11000000000\\
3 &\leftrightarrow 10100000000\\
4 &\leftrightarrow 1110000000000000\\
5 &\leftrightarrow 100100000100\\
6 &\leftrightarrow 1100000100000000
\end{aligned}$$

We next need the following lemma.

Lemma 3.1.16 *Assume that the first N nonnegative integers are coded all by strings of length less than t. Then the first 2^N nonnegative integers are coded by strings of length less than rt.*

Proof In their prime factorization, the first 2^N natural numbers have less than N times the factor 2. Analogously all r multiplicities in the decomposition are less than N. By assumption all the prime numbers $p_1, \ldots, p_r$ have codes of length less then t giving the result. □

We now show that r finite leads to a contradiction. If $N = 0$ then we can choose $t = 2$ since the length of the string 0 is 1 which is less than 2. Using the above Lemma we obtain by induction that the first $2^{2^{\cdot^{\cdot^{2}}}}$, the power being taken t times, natural numbers are coded all with strings less than $2(r^t)$. Choose $t = t_0$ large enough so that

$$\log_2(2^{2^{\cdot^{\cdot^{\cdot^{2}}}}}) = 2^{2^{\cdot^{\cdot^{\cdot^{2}}}}}_{\text{taken }(t_0-1)\text{times}} > 2r^{t_0}.$$

It follows that for

$$N_0 = 2^{2^{\cdot^{\cdot^{\cdot^{2}}}}}_{\text{taken }(t_0-1)\text{times}}$$

the first 2^{N_0} natural numbers can be coded by strings with length less than N_0. This contradicts Theorem 3.1.13 showing that there must be infinitely many primes.

The next proof is analytic and uses Stirling's approximation along with a formula due to Legendre. This proof appears in the book by Apostol [A].

Proof **Using Stirling's Approximation**

Stirling's approximation for $n!$ is given by (see [A])

$$n! \approxeq (\frac{n}{e})^n\sqrt{2\pi n} \text{ for large } n.$$

It follows then that

$$\lim_{n\to\infty}(n!)^{\frac{1}{n}} = \infty.$$

For $n \geq 1$ we have

$$n! = \prod_{p\leq n} p^{\alpha_p(n!)}$$

where p runs over all the primes less than n. From a formula of Legendre (see [A])

$$\alpha_p(n!) = \sum_{k>0}[\frac{n}{p^k}].$$

Now (see Cohen [C])

$$\alpha_p(n!) = \sum_{k>0,[\frac{n}{p^k}]\leq n}\sum_{k=1}^{\infty}\frac{1}{p^k} \leq \frac{n}{p-1}.$$

It follows that

$$(n!)^{\frac{1}{n}} = \prod_{p\leq n} p^{\frac{\alpha_p(n!)}{n}} \leq \prod_{p\leq n} p^{\frac{1}{p-1}}.$$

If the number of primes is finite it follows from the above that $(n!)^{\frac{1}{n}}$ stays finite in the limit as $n \to \infty$ contradicting the Stirling approximation. □

Proof **Another Analytic Proof**

This appears in the book of P. Ribenhoim [Ri]. Assume that there are only finitely many prime numbers

$$p_1 < p_2 < \cdots < p_r.$$

Suppose $t \in \mathbb{N}$ and let $N = p_r^t$. Each $m \leq N$ in $\mathbb{N}$ can be written as

$$m = p_1^{\alpha_1} p_2^{\alpha_2} \cdots p_r^{\alpha_r} \text{ with } \alpha_i \geq 0$$

and the sequence $(\alpha_1, \ldots, \alpha_r)$ is unique. We then have

$$p_i^{\alpha_i} \leq m \leq N = p_r^t.$$

Let $E = \frac{\ln p_r}{\ln p_1}$. Then $\alpha_i \leq tE$.

On the other hand N is at most equal to the number of sequences $(\alpha_1, \ldots, \alpha_r)$. Hence

$$p_r^t = N \leq (tE + 1)^r \leq t^r (E + 1)^r.$$

This gives a contradiction for t sufficiently large showing that there must be infinitely many primes. □

3.2 Sums of Squares

As we described in our historical overview much of the outline of the formal study of number theory was laid out in Gauss' work **Disquisitiones Arithmeticae**. He rested the study of number theory on three pillars—the theory of congruences, which we discussed in Chapter 2, the theory of algebraic integers which we will discuss in Chapter 6 and the theory of **forms**. In particular relative to this last topic Gauss considered the question of when an integer n can be represented by a **quadratic form** in other integers.

Here an **(integral) quadratic form** in n variables is a polynomial

$$f(x_1, \ldots, x_n) = \sum_{i,j=1}^{n} a_{ij} x_i x_j + \sum_i b_i x_i + c$$

where each a_{ij}, b_i and c are integers. A form is a **positive form** if the substitution of any integers other than $(0, 0, \ldots, 0)$ leads to a positive value. It is a **negative form** if the substitution of any integers other than $(0, 0, \ldots, 0)$ leads to a negative value. It is a **definite form** if it is either positive or negative. For example $f(x, y) = x^2 + y^2$ is a positive definite form.

In particular in two variables a quadratic from has the representation

$$f(x, y) = ax^2 + bxy + cy^2$$

where a, b, c are integers. The following lemma describes when such forms are positive definite.

Lemma 3.2.1 *The quadratic form $f(x, y) = ax^2 + bxy + cy^2$ is positive definite if and only if the discriminant $b^2 - 4ac$ is negative and $a > 0, c > 0$.*

Proof Suppose first that $f(x, y)$ is positive definite. Then $f(1, 0) = a > 0$ and $f(0, 1) = c > 0$. To show that the discriminant must be negative notice that $f(x, y)$ may be rewritten as

$$f(x, y) = \frac{1}{4a}((2ax + by)^2 + (4ac - b^2)y^2).$$

Using this rewritten from we see that $f(-b, 2a) = (4ac - b^2)a$. Since this must be positive and $a > 0$ it follows that $(4ac - b^2) > 0$ and hence the discriminant is negative.

Conversely suppose that the discriminant is negative and $a > 0$, $c > 0$. From the rewritten form for $f(x, y)$ above it is clear that $f(x, y) \geq 0$ for all integral pairs (x, y). If $f(x, y) = 0$ it follows that $2ax + by = 0$ and $(4ac - b^2)y^2 = 0$ from which one easily obtains that $x = y = 0$. Therefore $f(x, y)$ is positive. □

A quadratic form $f(x_1, \ldots, x_n)$ **represents** an integer m if there exists integers $(b_1, \ldots, b_n)$ such that $f(b_1, \ldots, b_n) = m$.

In this section we will look at the quadratic form question. Specifically we will consider the question of when an integer is represented as a sum of squares.

3.2.1 Pythagorean Triples

The oldest occurrence of sum of squares questions arises from integral solutions of the Pythagorean Theorem. Recall that a right triangle can have integral sides, for example $(3, 4, 5)$ or $(5, 12, 13)$. The question naturally arises as to finding, if possible, all such integer right triangles.

Definition 3.2.1 *A* **pythagorean triple** *is a triple (a, b, c) of integers with $a^2 + b^2 = c^2$. We consider c fixed and consider the triple (a, b, c) equivalent to the triple (b, a, c). A pythagorean triple (a, b, c) is called* **primitive** *if (a, b, c) are coprime.*

Now if $a^2 + b^2 = c^2$ then $(da)^2 + (db)^2 = (dc)^2$ for any integer d. Clearly then for the classification of pythagorean triples it is enough to consider primitive triples. The following theorem which in essence appeared in Diophantus' book **Arithmetica** written about 250 A.D. gives a complete classification of primitive pythagorean triples.

Theorem 3.2.1 *If n and m are two relatively prime integers with $n - m > 0$ and $n - m$ odd then $(2mn, n^2 - m^2, n^2 + m^2)$ is a primitive pythagorean triple. Further any primitive pythagorean triple can be obtained in this way.*

Proof Straightforward calculations show that if $a = 2nm$, $b = n^2 - m^2$ and $c = n^2 + m^2$ with $(n, m) = 1$ and $n - m = 2k + 1 > 0$ then (a, b, c) forms a primitive pythagorean triple (see the exercises).

Conversely we must show that any primitive pythagorean triple is obtained in this manner. Let (a, b, c) be a primitive pythagorean triple. Since (a, b, c) are coprime and $a^2 + b^2 = c^2$ it is easy to see that these integers must also be pairwise coprime. Hence no two can be even. Further suppose that both a and b are odd so that $a = 2m + 1, b = 2n + 1$. Then

$$c^2 = a^2 + b^2 = (2m + 1)^2 + (2n + 1)^2 = 2(2m^2 + 2n^2 + 2m + 2n + 1).$$

Then c^2 is even but c^2 is not divisible by 4, which is impossible. Hence a and b cannot both be odd. It follows that in (a, b, c) one of (a, b) must be even, the other odd and then c is odd.

Now suppose a is even and b and c are both odd. Then $c + b$ and $c - b$ are both even. Let

$$c + b = 2u \text{ and } c - b = 2v.$$

This implies directly that

$$b = u - v \text{ and } c = u + v.$$

Further $(u, v) = 1$ for otherwise $(b, c) \neq 1$. We now have

$$a^2 = c^2 - b^2 = (c + b)(c - b) = 4uv.$$

Since a is even $a = 2w$ which implies from the above that $w^2 = uv$ and hence uv is a perfect square. Since $(u, v) = 1$ it is then an easy consequence of the Fundamental Theorem of Arithmetic that both u and v must also be perfect squares (see Exercise 2.31). Hence $u = n^2$, $v = m^2$. Therefore we have

$$a = 2mn, b = n^2 - m^2, c = n^2 + m^2.$$

Thus (a, b, c) has the required from and we must show that n, m have the required properties.

Since $(u, v) = 1$ it follows that $(m, n) = 1$. Since $b > 0$ it follows that $u > v$ which implies that $n^2 > m^2$ which gives $n > m$ since both are positive. m and n cannot both be even and from the same argument as before they cannot both be odd. Therefore $n - m$ is odd completing the proof. □

There are many other questions concerning pythagorean triples that have been considered. For example we may ask when the (3, 4, 5) or (5, 12, 13) situation arises, that is, when does the hypotenuse differ from one of the legs by 1 or some fixed number d (see the exercises). Further as a corollary of the classification we get the following which is a special case of Fermat's Big Theorem and illustrates what has been called Fermat's **method of descent**. It is believed that Fermat's supposed proof of the big theorem was based on this technique.

Corollary 3.2.1 *The equation $x^4 + y^4 = z^2$ has no solutions in natural numbers. In particular the equation $x^4 + y^4 = z^4$ has no solutions in natural numbers.*

Proof Assume that there is a solution to $x^4 + y^4 = z^2$ for natural numbers (x_0, y_0, z_0). We then construct a further solution (x_1, y_1, z_1) with $z_1 < z_0$. As in the classification theorem we may assume that x_0, y_0, z_0 are coprime and then (x_0^2, y_0^2, z_0) is a primitive pythagorean triple. As in the proof of the classification one of (x_0, y_0) must be even the other odd and z_0 is then odd. Suppose then that y_0 is even. Then from the classification theorem there exist natural numbers a, b with $(a, b) = 1$ and

$$x_0^2 = a^2 - b^2, y_0^2 = 2ab, z_0 = a^2 + b^2.$$

a cannot be even because then b would be odd and it would follow that $x_0^2 \equiv 3 \bmod 4$. Hence a is odd and b is even and $x_0^2 + b^2 = a^2$. This implies that (x_0, b, a) is a primitive pythagorean triple with b even. It follows again from the classification theorem that

$$x_0 = c^2 - d^2, b = 2cd, a = c^2 + d^2$$

for coprime positive integers c, d with $c > d$ and $c + d$ odd.

Since $(a, b) = 1$ we obtain that c,d and $c^2 + d^2$ are pairwise coprime, that is

$$(c, d) = (c, c^2 + d^2) = (d, c^2 + d^2) = 1.$$

From

$$(\frac{1}{2}y_0)^2 = cd(c^2 + d^2)$$

we get a pairwise coprime triple (x_1, y_1, z_1) with

$$x_1^2 = c, y_1^2 = d, z_1^2 = c^2 + d^2.$$

This in turn implies that

$$c^2 + d^2 = x_1^4 + y_1^4 = z_1^2$$

and hence this triple gives another solution to the original equation. From

$$z_1 \leq z_1^2 = c^2 + d^2 = a < a^2 + b^2 = z_0$$

it follows that $z_1 < z_0$. Therefore if we assume that there is a solution $(x_0, y_0, z_0) \in \mathbb{N}^3$ of the equation $x^4 + y^4 = z^2$ then we can construct an infinite sequence (x_k, y_k, z_k), $k = 0, 1, 2, \ldots$ of solutions with $z_0 > z_1 > z_2 > \cdots > 0$. However by the well ordering of the natural numbers there must a minimal element and hence this is impossible and therefore a contradiction. □

3.2.2 Fermat's Two-Square Theorem

We have completely classified pythagorean triples (a, b, c) with $c^2 = a^2 + b^2$. We now consider the question of when an integer n, not necessarily a square, can be written as a sum of squares. That is, given n, when is $n = a^2 + b^2$ for integers a, b. In the language of forms we are asking when an integer n can be represented by the quadratic form $f(x, y) = x^2 + y^2$. The basic result is the following, generally called **Fermat's Two-Square Theorem**.

Theorem 3.2.2 *(Fermat's Two-Square Theorem) Let $n > 0$ be a natural number. Then $n = a^2 + b^2$ with $(a, b) = 1$ if and only if -1 is a quadratic residue modulo n.*

In this section we lay out a purely number theoretic proof of this theorem. In the course of developing this proof we will give several equivalent formulations of the theorem. In the next section we give a separate proof using the structure of the **Modular Group** $M = PSL_2(\mathbb{Z})$ (see the next section for an explanation). This second proof is interesting since it is in some sense independent of number theory.

We first consider the case of primes.

Lemma 3.2.2 *-1 is a quadratic residue modulo a prime p if and only if $p = 2$ or $p \equiv 1$ mod 4.*

Proof If $p = 2$ then $-1 \equiv 1 \equiv 1^2$ mod 2 and so -1 is a quadratic residue mod 2. Consider p now to be an odd prime. By Wilson's Theorem (Theorem 2.4.5) we have

$$(p-1)! \equiv -1 \bmod p \implies (1 \cdot 2 \cdots \frac{p-1}{2}) \cdot (\frac{p+1}{2} \cdots (p-1)) \equiv -1 \bmod p.$$

Now each number in the product $(\frac{p+1}{2} \cdots (p-1))$ is the negative modulo p of a number in the product $(1 \cdot 2 \cdots \frac{p-1}{2})$. For example modulo p, $-1 \equiv p-1$, $-2 \equiv p-2$ and so on. Therefore we can rewrite Wilson's Theorem as

$$(1 \cdot 2 \cdots \frac{p-1}{2}) \cdot (-(\frac{p-1}{2})(-\frac{p-3}{2}) \cdots (-1) \equiv -1 \bmod p.$$

But this implies

$$(-1)^{\frac{p-1}{2}}(1 \cdot 2 \cdots \frac{p-1}{2})^2 \equiv -1 \bmod p.$$

Let $x = 1 \cdot 2 \cdots \frac{p-1}{2}$ mod p. If $p \equiv 1$ mod 4 then $\frac{p-1}{2}$ is even and $(-1)^{\frac{p-1}{2}} = 1$. Hence

$$x^2 \equiv -1 \bmod p$$

and -1 is a quadratic residue mod p.

Conversely suppose $x^2 \equiv -1 \bmod p$ has a solution x_0. Then

$$x_0^2 \equiv -1 \bmod p \implies x_0^{2^{\frac{p-1}{2}}} \equiv (-1)^{\frac{p-1}{2}} \bmod p.$$

But $x_0^{2^{\frac{p-1}{2}}} = x_0^{p-1} \equiv 1 \bmod p$ by Fermat's theorem. It follows that $(-1)^{\frac{p-1}{2}} \equiv 1 \bmod p$. Since p is an odd prime, it follows that -1 is not congruent to 1 mod p so the above implies that $\frac{p-1}{2}$ is even and $p \equiv 1 \bmod 4$ completing the proof. □

We now tie this result to sums of squares.

Lemma 3.2.3 *If $p \equiv 1$ mod 4 then $p = a^2 + b^2$ with $(a, b) = 1$.*

Proof Note first that if $p = a^2 + b^2$ then a, b must be relatively prime for otherwise a common divisor of a and b would divide p.

Now suppose $p \equiv 1 \bmod 4$. Then from the previous Lemma -1 is a quadratic residue mod p. Let x_0 then be a solution to $x^2 \equiv -1 \bmod p$.

Let $K = [\sqrt{p}]$ be the greatest integer less than or equal to $\sqrt{p}$. Clearly then

$$K < \sqrt{p} < K + 1 \implies K^2 < p < (K+1)^2.$$

Consider the set of integers

$$S = \{u + x_0 v; 0 \leq u \leq K, 0 \leq v \leq K\}.$$

There are $K + 1$ choices for each of u and v and hence S has $(K + 1)^2$ elements. Since $p < (K + 1)^2$ and there are only p residue classes mod p we must have two distinct elements of S which are congruent modulo p. Hence there exists u_1, v_1, u_2, v_2 with

$$u_1 + x_0 v_1 \equiv u_2 + x_0 v_2 \bmod p.$$

Now if $u_1 = u_2$ we have $x_0 v_1 \equiv x_0 v_2 \bmod p$. But x_0 is a unit mod p so then $v_1 \equiv v_2 \bmod p$. Since both v_1, v_2 are less than p it follows that $v_1 = v_2$. Similarly if $v_1 = v_2$ it follows that $u_1 = u_2$. Since $u_1 + x_0 v_1$ is distinct from $u_2 + x_0 v_2$ it follows that $u_1 \neq u_2$ and $v_1 \neq v_2$.

From the congruence we may rewrite as

$$u_1 - u_2 \equiv x_0(v_2 - v_1) \bmod p.$$

Let $a = u_1 - u_2, b = v_2 - v_1$. Then $a \neq 0, b \neq 0$ and $a \equiv x_0 b \bmod p$. Therefore

$$a^2 \equiv x_0^2 b^2 \implies a^2 \equiv -b^2 \implies a^2 + b^2 \equiv 0 \bmod p.$$

Hence $p|(a^2 + b^2)$. We show that $p = a^2 + b^2$. Since $0 \leq u_1 \leq K$ and $0 \leq u_2 \leq K$ it follows that $-K \leq u_1 - u_2 \leq K$. Then $(u_1 - u_2)^2 = a^2 \leq K^2 < p$. Hence $a^2 < p$. Analagously $b^2 < p$. Therefore $0 < a^2 + b^2 < 2p$. However the only multiple of p within the range 0 to $2p$ is p itself. Therefore $p = a^2 + b^2$. □

Lemma 3.2.4 *Suppose $n = a^2 + b^2$ and q is a prime divisor of n. If $q \equiv 3$ mod 4 then $q^2|n$.*

Proof Suppose $q|(a^2 + b^2)$ with q a prime congruent to 3 mod 4. If $q \nmid a$ then a is a unit mod q. Then

$$a^2 + b^2 \equiv 0 \implies b^2 \equiv -a^2 \implies (ba^{-1})^2 \equiv -1 \bmod q.$$

Hence -1 is a quadratic residue mod q contradicting $q \equiv 3$ mod 4. Hence $q|a$. Similarly $q|b$. But then $q^2|(a^2 + b^2)$ and then $q^2|n$. □

Theorem 3.2.3 *Suppose $n \geq 2$ has the prime decomposition*

$$n = 2^{\alpha} p_1^{\beta_1} \cdots p_k^{\beta_k} q_1^{\gamma_1} \cdots q_t^{\gamma_t}$$

where $p_i \equiv 1$ mod 4 for $i = 1, \ldots, k$ and $q_j \equiv 3$ mod 4 for $j = 1, \ldots, t$. Then n can be expressed as the sum of two squares if and only if all the exponents γ_j of the primes congruent to 3 mod 4 are even.

We note that this theorem is also called Fermat's Two-Square Theorem.

Proof Notice first that for integers a, b, c, d we have

$$(a^2 + b^2)(c^2 + d^2) = (ac - bd)^2 + (bc + ad)^2.$$

Therefore if $m = uv$ and u is a sum of two squares and v is a sum of two squares then m is also a sum of two squares.

Now $2 = 1 + 1 = 1^2 + 1^2$ so any power of 2 is a sum of two squares. Similarly if $p \equiv 1$ mod 4 then from Lemma 3.2.3 p is the sum of two squares and hence any power of p is the sum of two squares. If $\gamma = 2k$ is even and $q \equiv 3$ mod 4 then $q^{\gamma} = q^{2k} = (q^k)^2 + 0^2$ and q^{γ} is a sum of two squares. Putting these all together we have that if each exponent of a prime congruent to 3 mod 4 is even in the prime decomposition of n then n is the sum of two squares.

Conversely if $n = a^2 + b^2$ and $q|n$ with $q \equiv 3$ mod 4 then from Lemma 3.2.4 $q^2|n$ and thus the exponent of q in n must be even. □

We now prove Theorem 3.2.2.

Proof (Theorem 3.2.2) Suppose $n = a^2 + b^2$ with $(a, b) = 1$. Then $(n, b) = 1$ for otherwise a common divisor of n and b would divide a. Hence b is a unit mod n and so b^{-1} exists mod n. Then

$$n = a^2 + b^2 \implies a^2 + b^2 \equiv 0 \implies (ab^{-1})^2 \equiv -1 \bmod n.$$

Therefore -1 is a quadratic residue mod n.

Conversely suppose -1 is a quadratic residue mod n. We show that $n = a^2 + b^2$ with $(a, b) = 1$ by using a modification of the proof of Lemma 3.2.3. Let x_0 be a solution of $x^2 \equiv -1 \bmod n$. Then there exist integers $(y, b) = 1$ with $0 < b \leq \sqrt{n}$ such that

$$|-\frac{x_0}{n} - \frac{y}{b}| < \frac{1}{b\sqrt{n}}$$

(see exercises). Now let

$$a = x_0 b + ny.$$

Then $a \equiv x_0 b \bmod n$ and hence $a^2 + b^2 \equiv 0 \bmod n$. Now $|a| < \sqrt{n}$ so

$$0 < a^2 + b^2 < 2n$$

and as in the proof of Lemma 3.2.3 the only multiple of n in this range is n itself and therefore $n = a^2 + b^2$. Further $(a, b) = 1$. To see this notice that we have

$$n = (x_0 b + ny)^2 + b^2 = (1 + x_0^2)b^2 + 2x_0 nby + n^2 y^2.$$

It follows that

$$1 = \frac{1 + x_0^2}{n} b^2 + x_0 by + x_0 by + ny^2 = ub + y(x_0 b + ny) = ub + ya.$$

□

Theorem 3.2.2 gives a criteria given n to determine if n is representable as a sum of two squares. A representation $n = a^2 + b^2$ with $(a, b) = 1$ is called a **primitive representation**. Combining the two forms for Fermat's Two Square Theorem we get the following corollary.

Corollary 3.2.2 *An integer n has a primitive representation as a sum of two squares if and only if $n = 2^\epsilon p_1^{\alpha_1} \cdots p_k^{\alpha_k}$ where $\epsilon = 0$ or $\epsilon = 1$, each $\alpha_i \in \mathbb{N}$ and each $p_i \equiv 1$ mod* 4.

Proof From Fermat's Two Square Theorem n has a primitive representation if and only if -1 is a quadratic residue mod n. Then -1 must be a quadratic residue mod p for any prime divisor of n. Therefore any odd prime divisor of n must be congruent to 1 mod 4. Further -1 is not a quadratic residue mod 2^α if $\alpha > 1$. Therefore the highest power of 2 which can divide n is 1. □

Theorems 3.2.2 and 3.2.3 characterize those integers n for which there is a representation as a sum of two squares. The question can then be asked how many different representations can there be? If we let

$$r(n) = \text{ the number of pairs } (a, b) \in \mathbb{Z}^2 \text{ with } n = a^2 + b^2$$

then the following can be proved (see [Za] or [NZ].) We leave the proof as an exercise (see Exercise 3.35).

Theorem 3.2.4 *Let $r(n)$ be defined as above. Then*
(1) $r(n) = 4\sum_{d|n}\chi(d)$ where

$$\begin{aligned}
\chi(d) &= 1 \textit{ if } n \equiv 1 \textit{ mod } 4,\\
\chi(d) &= -1 \textit{ if } n \equiv -1 \textit{ mod } 4,\\
\chi(d) &= 0 \textit{ if } n \equiv 0 \textit{ mod } 2.
\end{aligned}$$

(2) $\sum_{n=1}^{\infty}\frac{r(n)}{n} = 4\zeta(s)L(s)$ where

$$\zeta(s) = \sum_{n=1}^{\infty}\frac{1}{n^s}$$

and

$$L(s) = \sum_{n=1}^{\infty}\frac{\chi(n)}{n^s} \textit{ with } Re(s) > 1.$$

(3) $\frac{1}{4}r(mn) = \frac{1}{4}r(n)\frac{1}{4}r(m)$ if $(n, m) = 1$.

If $p \equiv 1 \bmod 4$ is a prime then

$$r(p) = 4\sum_{d|p}\chi(d) = 4(\chi(1) + \chi(p)) = 8.$$

For $p \equiv 3 \bmod 4$ then $r(p) = 0$. For example for $p = 5$ the 8 pairs are

$$(2, 1), (1, 2), (-1, 2), (2, -1), (1, -2), (-2, 1), (-1, -2), (-2, -1).$$

The function $\zeta(s)$ in the theorem is the Riemann zeta function which we introduced earlier and which will play a crucial role in the proof of the prime number theorem. The function $\chi(n)$ is called a **Dirichlet character** and the function $L(s)$ a **Dirichlet series**. These will play a role in the proof of Dirichlet's theorem.

3.2.3 The Modular Group

If R is any ring with an identity, then the set of invertible $n \times n$ matrices with entries from R forms a group under matrix multiplication called the **n-dimensional general linear group over R** (see [Ro]). This group is denoted by $GL_n(R)$. Since $\det(A)\det(B) = \det(AB)$ for square matrices A, B it follows that the subset of $GL_n(R)$ consisting of those matrices of determinant 1 forms a subgroup. This subgroup is called the **special linear group over R** and is denoted by $SL_n(R)$. In this section we concentrate on $SL_2(\mathbb{Z})$, or more specifically a quotient of it, $PSL_2(\mathbb{Z})$ and use properties of this group to give another, more direct, proof of Fermat's Two-Square Theorem.

The group $SL_2(\mathbb{Z})$ then consists of 2 × 2 integral matrices of determinant one:

$$SL_2(\mathbb{Z}) = \Big\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix}; a, b, c, d \in \mathbb{Z}, ad - bc = 1 \Big\}.$$

$SL_2(\mathbb{Z})$ is called the **homogeneous modular group** and an element of $SL_2(\mathbb{Z})$ is called a **unimodular matrix**.

If G is any group, its **center**, denoted by $Z(G)$, consists of those elements of G which commute with all elements of G;

$$Z(G) = \{g \in G; gh = hg, \forall h \in G\}.$$

It is easy to see that $Z(G)$ is a normal subgroup of G (see exercises) and hence we can form the factor group $G/Z(G)$. For $G = SL_2(\mathbb{Z})$ the only unimodular matrices that commute with all others are $\pm I = \pm \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$. Therefore $Z(SL_2(\mathbb{Z})) = \{I, -I\}$. The quotient

$$SL_2(\mathbb{Z})/Z(SL_2(\mathbb{Z})) = SL_2(\mathbb{Z})/\{I, -I\}$$

is denoted $PSL_2(\mathbb{Z})$ and is called the **projective special linear group** or **inhomogeneous modular group**. More commonly $PSL_2(\mathbb{Z})$ is just called the **Modular Group** and denoted by M.

M arises in many different areas of mathematics including number theory, complex analysis and Riemann surface theory and the theory of automorphic forms and functions. M is perhaps the most widely studied single finitely presented group. Complete discussions of M and its structure can be found in the books **Integral Matrices** by M. Newman [New 1] or **Algebraic Theory of the Bianchi Groups** by B. Fine [F].

Since $M = PSL_2(\mathbb{Z}) = SL_2(\mathbb{Z})/\{I, -I\}$ it follows that each element of M can be considered as $\pm A$ where A is a unimodular matrix. A **projective unimodular matrix** is then

$$\pm \begin{pmatrix} a & b \\ c & d \end{pmatrix}, a, b, c, d \in \mathbb{Z}, ad - bc = 1.$$

The elements of M can also be considered as linear fractional transformations over the complex numbers

$$z' = \frac{az + b}{cz + d}, a, b, c, d \in \mathbb{Z}, ad - bc = 1.$$

Thought of in this way, M forms a **Fuchsian group** which is a discrete group of isometries of the non-Euclidean hyperbolic plane. The book by Katok [K] gives a solid and clear introduction to such groups. This material can also be found in condensed form in [FR].

We will shortly describe the abstract structure of the group M. First though we use it to give a direct proof of Fermat's Two-Square Theorem. We need the following lemma. Recall that the **trace** of a matrix A is the sum of its diagonal elements. Trace is preserved under conjugation so that $tr(A) = tr(T^{-1}AT)$ for any square matrices A and invertible T. Recall also that in a group G two elements g, g_1 are **conjugate** if there exists an $h \in G$ such that $h^{-1}gh = g_1$. Conjugation is an equivalence relation on a group and the equivalence classes are called **conjugacy classes**.

Lemma 3.2.5 *Let A be a projective unimodular matrix with $tr(A) = 0$. Then A is conjugate within M to $X = \pm \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$. That is there exists $T \in M$ with $T^{-1}XT = A$.*

Proof Let $A = \pm \begin{pmatrix} \alpha & \beta \\ \gamma & -\alpha \end{pmatrix}$. Let S be the set of conjugates of A within M so that

$$S = \{T^{-1}AT; T \in M\}.$$

Since conjugation preserves trace S consists of matrices of trace zero. Let

$$Y = \pm \begin{pmatrix} a & b \\ c & -a \end{pmatrix}$$

be an element of S with $|a|$ minimal. This exists from the well ordering of $\mathbb{Z}$. We show that a must equal zero.

Suppose $a \neq 0$ then

$$-a^2 - bc = 1 \implies -bc = a^2 + 1 \implies |b||c| = a^2 + 1.$$

It follows then that $b \neq 0, c \neq 0$ and either $|b| < |a|$ or $|c| < |a|$. Assume first that $|c| < |a|$. We may assume that $a > 0$ and $c > 0$. Then

$$0 < a - c < a.$$

Now conjugate Y by $T = \pm \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$. Then $T^{-1} = \pm \begin{pmatrix} 1 & -1 \\ 0 & 1 \end{pmatrix}$ and

$$T^{-1}YT = \pm \begin{pmatrix} 1 & -1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} a & b \\ c & -a \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} = \pm \begin{pmatrix} a-c & 2a+b-c \\ c & c-a \end{pmatrix}.$$

But then $0 < a - c < a$ contradicting the minimality of $|a|$.

If $b < a$ assuming $a > 0, b > 0$ conjugate Y by $T = \pm\begin{pmatrix} 1 & 0 \\ -1 & 1 \end{pmatrix}$. Then $T^{-1} = \pm\begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}$ and

$$T^{-1}YT = \pm\begin{pmatrix} a-b & b \\ 2a+c-b & b-a \end{pmatrix}.$$

Again $0 < a - b < a$ contradicting the minimality of $|a|$.

Therefore in a minimal conjugate of A we must have $a = 0$ and hence $-bc = 1$. It follows that $b = \pm 1$ and c also and therefore

$$Y = \pm\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} = X$$

completing the proof. □

Now consider conjugates of X within M. Let $T = \pm\begin{pmatrix} a & b \\ c & d \end{pmatrix}$. Then

$$T^{-1} = \pm\begin{pmatrix} d & -b \\ -c & a \end{pmatrix}$$

and

$$TXT^{-1} = \pm\begin{pmatrix} a & b \\ c & d \end{pmatrix}\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}\begin{pmatrix} d & -b \\ -c & a \end{pmatrix} = \pm\begin{pmatrix} -(bd+ac) & a^2+b^2 \\ -(c^2+d^2) & bd+ac \end{pmatrix} \qquad (3.2.1)$$

Therefore any conjugate of X must have form (3.2.1).

We now reprove Fermat's Two-Square Theorem.

Theorem 3.2.5 *(Fermat's Two-Square Theorem) Let $n > 0$ be a natural number. Then $n = a^2 + b^2$ with $(a, b) = 1$ if and only if -1 is a quadratic residue modulo n.*

Proof Suppose -1 is a quadratic residue mod n. Then there exists an x with $x^2 \equiv -1 \mod n$ or $x^2 = -1 + mn$. This implies that $-x^2 - mn = 1$ so that there must exist a projective unimodular matrix

$$A = \pm\begin{pmatrix} x & n \\ m & -x \end{pmatrix}.$$

The trace of A is zero so by Lemma 3.2.5 A is conjugate within M to X and therefore A must have form (3.2.1). Therefore $n = a^2 + b^2$ since $n > 0$. Further $(a, b) = 1$ since in finding form (3.2.1) we had $ad - bc = 1$.

Conversely suppose $n = a^2 + b^2$ with $(a, b) = 1$. Then there exists $c, d \in \mathbb{Z}$ with $ad - bc = 1$ and hence there exists a projective unimodular matrix

$$T = \pm \begin{pmatrix} a & b \\ c & d \end{pmatrix}.$$

Then

$$TXT^{-1} = \pm \begin{pmatrix} \alpha & a^2 + b^2 \\ \gamma & -\alpha \end{pmatrix} = \pm \begin{pmatrix} \alpha & n \\ \gamma & -\alpha \end{pmatrix}.$$

This then has determinant one so

$$-\alpha^2 - n\gamma = 1 \implies \alpha^2 = -1 - n\gamma \implies \alpha^2 \equiv -1 \bmod n.$$

Therefore -1 is a quadratic residue mod n. □

This type of group theoretical proof can be extended in several directions. Kern-Isberner and Rosenberger [KR 1] considered groups of matrices of the form

$$U = \begin{pmatrix} a & b\sqrt{N} \\ c\sqrt{N} & d \end{pmatrix}, a, b, c, d, N \in \mathbb{Z}, ad - Nbc = 1$$

or

$$U = \begin{pmatrix} a\sqrt{N} & b \\ c & d\sqrt{N} \end{pmatrix}, a, b, c, d, N \in \mathbb{Z}, Nad - bc = 1.$$

They then proved that if

$$N \in \{1, 2, 3, 4, 5, 6, 8, 9, 10, 12, 13, 16, 18, 22, 25, 28, 37, 58\}$$

and $n \in \mathbb{N}$ with $(n, N) = 1$ then

(1) If $-N$ is a quadratic residue mod n and n is a quadratic residue mod N then n can be written as $n = x^2 + Ny^2$ with $x, y \in \mathbb{Z}$.

(2) Conversely if $n = x^2 + Ny^2$ with $x, y \in \mathbb{Z}$ and $(x, y) = 1$ then $-N$ is a quadratic residue mod n and n is a quadratic residue mod N.

The proof of the above results depends on the class number of $\mathbb{Q}(\sqrt{-N})$ (see [KR 1]).

In another direction Fine [F 1] and [F 2] showed that the Fermat Two-Square Property is actually a property satisfied by many rings R. These are called **sum of squares rings**. For example if $p \equiv 3 \bmod 4$ then $\mathbb{Z}_{p^n}$ for $n > 1$ is a sum of squares ring.

We close this subsection by describing the group theoretical structure of both $SL_2(\mathbb{Z})$ and $M = PSL_2(\mathbb{Z})$. This structure can be developed with only minimal number theory.

Theorem 3.2.6 *The group $SL_2(\mathbb{Z})$ is generated by the elements*

$$X = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \text{ and } Y = \begin{pmatrix} 0 & 1 \\ -1 & -1 \end{pmatrix}.$$

Further a complete set of defining relations for the group in terms of these generators is given by

$$X^4 = Y^3 = YX^2Y^{-1}X^{-2} = I.$$

In the language of combinatorial group theory we say that $SL_2(\mathbb{Z})$ *has the* **presentation**

$$< X, Y; X^4 = Y^3 = YX^2Y^{-1}X^{-2} = I > .$$

Proof We first show that $SL_2(\mathbb{Z})$ is generated by X and Y, that is every matrix A in the group can be written as a product of powers of X and Y.

Let

$$U = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}.$$

Then a direct multiplication shows that $U = XY$ and we show that $SL_2(\mathbb{Z})$ is generated by X and U which implies that it is also generated by X and Y. Further

$$U^n = \begin{pmatrix} 1 & n \\ 0 & 1 \end{pmatrix}$$

so that U has infinite order.

Let $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in SL_2(\mathbb{Z})$. Then we have

$$XA = \begin{pmatrix} -c & -d \\ a & b \end{pmatrix} \text{ and } U^k A = \begin{pmatrix} a + kc & b + kd \\ c & d \end{pmatrix}$$

for any $k \in \mathbb{Z}$. We may assume that $|c| \leq |a|$ otherwise start with XA rather than A. If $c = 0$ then $A = \pm U^q$ for some q. If $A = U^q$ then certainly A is in the group generated by X and U. If $A = -U^q$ then $A = X^2U^q$ since $X^2 = -I$. It follows that here also A is in the group generated by X and U.

Now suppose $c \neq 0$. Apply the Euclidean algorithm to a and c in the following modified way:

$$a = q_0c + r_1$$

$$-c = q_1r_1 + r_2$$

$$r_1 = q_2r_2 + r_3$$

$$\ldots$$

$$(-1)^n r_{n-1} = q_n r_n + 0$$

where $r_n = \pm 1$ since $(a, c) = 1$. Then

$$XU^{-q_n}\cdots XU^{-q_0}A = \pm U^{q_{n+1}} \text{ with } q_{n+1}\in\mathbb{Z}.$$

Then

$$A = X^m U^{q_0} XU^{q_1}\cdots XU^{q_n}XU^{q_{n+1}}$$

with $m = 0, 1, 2, 3; q_0, q_1, \ldots, q_{n+1} \in \mathbb{Z}$ and $q_0 \cdots q_n \neq 0$. Therefore X and U and hence X and Y generate $SL_2(\mathbb{Z})$.

We must now show that

$$X^4 = Y^3 = YX^2Y^{-1}X^{-2} = I \tag{3.2.2}$$

is a complete set of defining relations for $SL_2(\mathbb{Z})$ or that every relation on these generators is derivable from these (see [Ro] or [J] for a description of group presentations). It is straightforward to see that X and Y do satisfy these relations. Assume then that we have a relation

$$S = X^{\epsilon_1}Y^{\alpha_1}X^{\epsilon_2}Y^{\alpha_2}\cdots Y^{\alpha_n}X^{\epsilon_{n+1}} = I$$

with all $\epsilon_i, \alpha_j \in \mathbb{Z}$. Using the relations (3.2.2) we may transform S so that

$$S = X^{\epsilon_1}Y^{\alpha_1}\cdots Y^{\alpha_m}X^{\epsilon_{m+1}}$$

with $\epsilon_1, \epsilon_{m+1} = 0, 1, 2$ or 3 and $\alpha_i = 1$ or 2 for $i = 1, \ldots, m$ and $m \geq 0$. Multiplying by a suitable power of X we obtain

$$Y^{\alpha_1}X\cdots Y^{\alpha_m}X = X^{\alpha} = S_1$$

with $m \geq 0$ and $\alpha = 0, 1, 2$ or 3. Assume that $m \geq 1$ and let

$$S_1 = \begin{pmatrix} a & -b \\ -c & d \end{pmatrix}.$$

We show by induction that

$$a, b, c, d \geq 0, b + c > 0$$

or

$$a, b, c, d \leq 0, b + c < 0.$$

This claim for the entries of S_1 is true for

$$YX = \begin{pmatrix} 1 & 0 \\ -1 & 1 \end{pmatrix} \text{ and } Y^2X = \begin{pmatrix} -1 & 1 \\ 0 & -1 \end{pmatrix}.$$

Suppose it is correct for $S_2 = \begin{pmatrix} a_1 & -b_1 \\ -c_1 & d_1 \end{pmatrix}$. Then

$$YXS_2 = \begin{pmatrix} a_1 & -b_1 \\ -(a_1 + c_1) & b_1 + d_1 \end{pmatrix} \text{ and}$$

$$Y^2XS_2 = \begin{pmatrix} -a_1 - c_1 & b_1 + d_1 \\ c_1 & d_1 \end{pmatrix}.$$

Therefore the claim is correct for all S_1 with $m \geq 1$. This gives a contradiction, for the entries of X^α with $\alpha = 0, 1, 2$, or 3 do not satisfy the claim. Hence $m = 0$ and S can be reduced to a trivial relation by the given set of relations. Therefore they are a complete set of defining relations and the theorem is proved. □

Corollary 3.2.3 *The Modular Group $M = PSL_2(\mathbb{Z})$ has the presentation*

$$M =< x, y; x^2 = y^3 = 1 > .$$

Further x, y can be taken as the linear fractional transformations

$$x : z' = -\frac{1}{z} \textit{ and } y : z' = -\frac{1}{z+1}.$$

Proof The center of $SL_2(\mathbb{Z})$ is $\pm I$. Since $X^2 = -I$ setting $X^2 = I$ in the presentation for $SL_2(\mathbb{Z})$ gives the presentation for M. Writing the projective matrices as linear fractional transformations gives the second statement. □

In group theoretical language this corollary says that M is the **free product** of a cyclic group of order 2 and a cyclic group of order 3 (see [Ro]). From this structure it is easy to show that any element of M of order 2 must be conjugate within M to x. Further a straightforward calculation shows that a projective unimodular matrix has order 2 if and only if its trace is zero. Combining these two facts gives an easy proof of Lemma 3.2.5 which was the crux of the proof of Fermat's Two-Square Theorem.

3.2.4 Lagrange's Four Square Theorem

In the last section we considered when a natural number can be expressed as a sum of two squares. Here we prove the following theorem of Lagrange which shows that any natural number can be expressed as the sum of four squares. In the language of forms this says that any natural number is represented by the form $f(x, y, z, w) = x^2 + y^2 + z^2 + w^2$. The Lagrange Four-Square Theorem is actually a special case of **Waring's problem**. In 1770 Edward Waring stated, but did not prove, that every positive integer is a sum of nine cubes and also a sum of nineteen fourth powers. **Waring's problem** then became whether for each positive integer k there is an integer $s(k)$ such that every natural number is the sum of at most $s(k)$, kth powers. In this formulation, Lagrange's theorem says that $s(2) = 4$. Wieferich proved Waring's assertion about cubes that is every natural number can be written as a sum of nine cubes. D. Hilbert in 1909 proved Waring's problem for all exponents k. Subsequently there have been several other proofs given of this same result including ones by Hardy and Littlewood [HL], Vinogradov [V] and Linnik [Li]. Linnik's proof of the general result can be found in the book of Nathanson [N]. We give a proof of the four square result.

Theorem 3.2.7 *(Lagrange) Every natural number n can be represented as the sum of four squares*

$$n = a^2 + b^2 + c^2 + d^2$$

with $a, b, c, d \in \mathbb{Z}$.

Proof Now $1 = 1^2 + 0^2 + 0^2 + 0^2$ and $2 = 1^2 + 1^2 + 0^2 + 0^2$ so the theorem is clearly true for $n = 1, 2$. Further the product of two sums of four squares is again a sum of four squares. That is

$$(a^2 + b^2 + c^2 + d^2)(x^2 + y^2 + z^2 + w^2) = A^2 + B^2 + C^2 + D^2$$

where

$$A = ax + by + cz + dw, \qquad B = ay - bx - cw + dz,$$

$$C = az + bw - cx - dy, \qquad D = aw - bz + cy - dx.$$

This implies then that we need only prove the theorem for primes. Therefore let p be a prime $p \geq 3$.

We need the following lemma.

Lemma 3.2.6 *Let p be a prime. Then there exist $x, y \in \mathbb{Z}$ with $x^2 + y^2 \equiv -1$ mod p.*

Proof This is clear for $p = 2$ so assume $p \geq 3$. Consider the squares modulo p. That is consider the set

$$S = \{1^2, 2^2, \ldots, (p-1)^2\} \text{ modulo } p$$

Since $a^2 \equiv b^2$ mod p implies that $a \equiv \pm b$ mod p it follows that there are $\frac{p-1}{2}$ elements of S which are incongruent mod p. Therefore if we consider the integers

$$-x^2 - 1 \text{ for } x = 0, 1, \ldots, p-1$$

we must get some $x \in \{0, 1, 2, \ldots, p-1\}$ such that $-x^2 - 1 \equiv y^2$ mod p for some $y \in \{0, 1, 2, \ldots, p-1\}$. □

Proof (Theorem 3.2.7) From Lemma 3.2.6 there is a natural number m and integers x, y such that

$$mp = x^2 + y^2 + 1^2 + 0^2.$$

We may assume that $|x|, |y| \leq \frac{1}{2}p$ so that $m \leq \frac{1}{2}p$. If $m = 1$ then the theorem holds. Suppose then that $m > 1$.

From the above we have that for each prime $p \geq 3$ there is an m with $m \leq \frac{1}{2}p$ and

$$mp = x^2 + y^2 + z^2 + w^2, x, y, z, w \in \mathbb{Z}.$$

We will show that there is then a choice with $m = 1$.

Let a, b, c, d be the positive residues of x, y, z, w respectively mod m with the smallest absolute values. Then $|a|, |b|, |c|, |d|$ are all $\leq \frac{m}{2}$. Then

$$pm = x^2 + y^2 + z^2 + w^2 \equiv a^2 + b^2 + c^2 + d^2 \equiv 0 \bmod m.$$

Hence

$$a^2 + b^2 + c^2 + d^2 = mm'.$$

It follows then that

$$pm^2m' = (x^2 + y^2 + z^2 + w^2)(a^2 + b^2 + c^2 + d^2) = A^2 + B^2 + C^2 + D^2$$

where A, B, C, D are described as in the beginning of the proof. From these expressions since

$$a \equiv x, b \equiv y, c \equiv z, d \equiv w \bmod m$$

it follows that

$$A \equiv B \equiv C \equiv D \equiv 0 \bmod m.$$

Dividing through A^2, B^2, C^2, D^2 by m^2 we can then represent pm' as a sum of four squares.

Now from

$$m' = \frac{a^2 + b^2 + c^2 + d^2}{m} \text{ and } |a|, |b|, |c|, |d| \leq \frac{m}{2}$$

we get that $m' \leq m$. If $m' < m$ then we have a smaller multiple m' of p such that $m'p$ is a sum of four squares. Assume then that $m' = m$. We show that in this case p is a sum of four squares. $m = m'$ implies that

$$|a| = |b| = |c| = |d| = \frac{m}{2}.$$

Then

$$2a \equiv 2b \equiv 2c \equiv 2d \equiv 2x \equiv 2y \equiv 2z \equiv 2w \equiv 0 \bmod m.$$

It then follows that

$$4pm = 4x^2 + 4y^2 + 4z^2 + 4w^2 = vm^2$$

for some $v \in \mathbb{Z}$, $v \neq 0$. Hence $m|4p$. From $(m, p) = 1$ we get that $m|4$. Recall further that $1 < m \leq \frac{1}{2}p$.

If $m' = m = 4$ then x, y, z, w are all even so from above we get that

$$p = (\frac{x}{2})^2 + (\frac{y}{2})^2 + (\frac{z}{2})^2 + (\frac{w}{2})^2.$$

If $m = m' = 2$ then

$$4p = (1 + 1 + 0 + 0)2p = (1 + 1 + 0 + 0)(x^2 + y^2 + z^2 + w^2) = A^2 + B^2 + C^2 + D^2$$

with $A = x + y$,$B = y - x$, $C = z + w$, and $D = w - z$. Since A, B, C, D are all even we get a representation for p as a sum of four squares as above.

Therefore for each pm, $m > 1$ which is a sum of four squares we can find a pm' with $m' < m$ which is also a sum of four squares. Therefore the minimal m must be 1 and p itself is a sum of four squares proving the theorem. □

We note that we can further show that if a natural number n is not of the form $4^k(8n + 7)$ then n can be expressed as a sum of three squares. However if $n = 4^k(8n + 7)$ then four squares are necessary. This is related to the following extension of Waring's problem. Hilbert's solution showed that given k there exists an $s(k)$ such that every natural number can be represented as a sum of $s(k)$, kth powers. The extension asks to find the minimal value of $s(k)$. More details on this are in the book of Ribenhoim [Ri].

3.2.5 The Infinitude of Primes Through Continued Fractions

In this final part of Section 3.2 we give a proof of the infinitude of primes using **continued fractions**. A complete discussion of the theory of continued fractions can be found in [NZM]. We just touch on what we need for this proof.

Definition 3.2.2 *Let $a_0, a_1, \ldots, a_n$ be a finite sequence of integers all positive except possibly a_0. Then a* **finite simple continued fraction** *is the rational number defined by*

$$a_0 + \cfrac{1}{a_1 + \cfrac{1}{a_2 + \ddots}}.$$

If $a_0, a_1, \ldots, a_n, \ldots$ is an infinite sequence of integers all positive except possibly a_0, then a **infinite simple continued fraction** *is determined by the limit of the finite simple continued fractions formed up to a_n. Each of the finite simple continued fractions is called a* **convergent** *of the infinite simple continued fraction.*

The following can be proved (see [NZM]).

Theorem 3.2.8 *If $a_0, a_1, \ldots, a_n, \ldots$ is an infinite sequence of integers all positive except possibly a_0, then they determine a unique* **infinite simple continued fraction***, that is the limit of convergent exists. Further this value is always an irrational number.*

If the sequence defining a continued fraction becomes a periodic sequence after a certain point the resulting continued fraction is called a **periodic continued fraction**. Consider an infinite continued fraction with sequence $a_0, a_1, \ldots$ and let A_m,B_m be the numerator and denominator respectively for the mth convergent. We need the following results, the first being a theorem of Lagrange (see [P]).

Theorem 3.2.9 *A real irrational number which is a solution of the quadratic equation*

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

with $a, b, c, d \in \mathbb{Z}$ and not all zero, has a development as a periodic continued fraction.

As a special case of the above theorem we have that if

$$x = \frac{p + \sqrt{p^2 + 4}}{2}, \text{ with } p \neq 0, p \in \mathbb{Z}$$

then

$$x = p + \cfrac{1}{p + \cfrac{1}{p + \ddots}}$$

Lemma 3.2.7 *([P]) Suppose d is a positive squarefree integer. If the development of $\sqrt{d}$ as a periodic continued fraction has a period of length m then the equation $x^2 - dy^2 = -1$ has an integral solution and each positive solution x, y is of the form $x = A_i, y = B_i$ for $i = qm - 1$ with q odd.*

Using these we get the following proof of the infinitude of primes due to Barnes [B].

Proof (The sequence of Primes is infinite) As always assume there are only finitely many prime numbers

$$p_1 = 2 < p_2 = 3 < \cdots < p_r.$$

Let $p = p_1 \cdots p_r$ and $q = p_2 \cdots p_r = \frac{p}{2}$. Now let

$$x = \frac{p + \sqrt{p^2 + 4}}{2}.$$

Then

$$x = q + \sqrt{q^2 + 1}.$$

Since p_i does not divide $q^2 + 1$ for $i = 2, \ldots, r$ it follows that $q^2 + 1$ must be a power of 2. Further this power must be odd since x is irrational. Hence

$$q^2 + 1 = 2^{2t+1}, t \in \mathbb{N}.$$

This gives

$$q^2 - 2(2^t)^2 = -1$$

and hence the Diophantine equation

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

has a solution $x = q$, $y = 2^t$. From Lemma 3.2.7 then $\frac{q}{2^t}$ is an even convergent value of

$$\sqrt{2} = 1 + \cfrac{1}{2 + \cfrac{1}{2+ \ddots}}$$

It can be shown that

$$B_{m+1} = a_{m+1} B_m + B_{m-1}, m \geq 1$$

where as before B_k is the denominator of the kth convergent. From this it follows that for $m \geq 1$, B_{2m} is a positive odd integer > 1. Since 2^t is even we then must have $m = 0$ and hence

$$\frac{q}{2^t} = \frac{A_0}{B_0} = \frac{1}{1} = 1.$$

Then from $(q, 2^t) = 1$ we get $q = 1$ which is a contradiction since $q = p_2 \cdots p_2 > 1$. □

3.3 Dirichlet's Theorem

If $(a, b) = 1$ for natural numbers a and b then **Dirichlet's Theorem** states that there are infinitely many primes in the arithmetic progression $\{an + b; n \in \mathbb{N}\}$. On one hand, given the many proofs that we have exhibited of the infinitude of primes, this may not seem surprising. However when looked at in light of the prime number theorem which says that the density of primes gets scarcer and scarcer as x gets larger it is quite surprising. Since $an + b$ is linear in n the distribution of numbers in this sequence is uniform or regular on the integers. However since $\pi(x) \sim \frac{x}{\ln x}$ we have that $\frac{\pi(x)}{x} \sim \frac{1}{\ln x}$. We can interpret this as the probability of randomly choosing a prime $\leq x$ goes to zero as x goes to ∞.

Earlier in this chapter we presented several special cases of Dirichlet's Theorem. Specifically we showed that there were infinitely many primes of the form $3n + 1$, $3n + 2$, $4n + 1$, $4n + 3$, $8n + 1$, $8n + 3$, $8n + 5$, and $8n + 7$. Many other specific situations, such as $6n + 5$, can be proved by the same techniques. The most general case that we proved was Theorem 3.1.12 which showed that there were infinitely many primes of the form $mn + 1$ for any positive integer m. A complete proof of the full Dirichlet Theorem involves analysis and we present it in this section.

Theorem 3.3.1 *(Dirichlet's Theorem) Let a,b be natural numbers with $(a, b) = 1$. Then there are infinitely many primes of the form $an + b$ with $n > 0$.*

Dirichlet's proof rests on two concepts; **Dirichlet characters** and **Dirichlet series**. The basic idea is to build for each integer a, a series, which would converge if there were only finitely many primes congruent to b mod a and then show that this series actually diverges. We discuss characters first.

Definition 3.3.1 *For any positive integer k, a* **Dirichlet character** *modulo k, is a complex valued function on the integers $\chi : \mathbb{Z} \to \mathbb{C}$ satisfying*

1. $\chi(a) = 0$ *if* $(a, k) > 1$
2. $\chi(1) \neq 0$
3. $\chi(a_1a_2) = \chi(a_1)\chi(a_2)$ *for all* $a_1, a_2 \in \mathbb{Z}$
4. $\chi(a_1) = \chi(a_2)$ *whenever* $a_1 \equiv a_2$ *mod k.*

From (3) and (4) it is clear that a Dirichlet character can be considered as a multiplicative complex function on the set of residue classes modulo k. We will shorten the notation and use the word **character** to mean a Dirichlet character modulo k.

From a group theoretical point of view a Dirichlet character is just a character of a finite complex representation of the unit group $U(\mathbb{Z}_k)$. We will say more about this after our discussion of characters.

As an example consider the function

$$\chi_0(a) = 0 \text{ if } (a, k) > 1$$

$$\chi_0(a) = 1 \text{ if } (a, k) = 1.$$

It is easy to verify that this is a character. Thus, modulo k, there is always at least one character. The character above is called the **principal character** and exists as defined for each k. We will presently show that there are $\phi(k)$ characters, where ϕ is the Euler phi function, for each positive integer k.

We now describe some necessary properties of characters. In each of the following results when we say character we mean character modulo k, with $k > 0$ fixed.

Lemma 3.3.1 *(1) For every character* $\chi(1) = 1$.

(2) For every character if $(a, k) = 1$ *then* $|\chi(a)|^{\phi(k)} = 1$. *Hence* $|\chi(a)| = 1$ *and* $\chi(a)$ *is a* $\phi(k)$*-th root of unity.*

Proof (1) Since χ is multiplicative we have $\chi(1) = \chi(1)\chi(1)$. Since $\chi(1) \neq 0$ it follows that $\chi(1) = 1$.

(2) From Euler's Theorem (Theorem 2.4.11) we have that if $(a, k) = 1$ then

$$a^{\phi(k)} \equiv 1 \bmod k.$$

Since a character is multiplicative this implies

$$|\chi(a)|^{\phi(k)} = |\chi(a^{\phi(k)})| = |\chi(1)| = 1.$$

□

Lemma 3.3.2 *For every* $k > 0$ *there exist only finitely many characters mod* k.

Proof Given k there are only finitely many different residue classes mod k. If a is a positive residue mod k then from the previous lemma $\chi(a)$ is a $\phi(k)$-th root of unity. Hence there are only finitely many choices. □

For the time being we will let c denote the finite number of characters modulo k. After we prove certain orthogonality relations we will show that $c = \phi(k)$.

Lemma 3.3.3 *(1) If* χ_1 *and* χ_2 *are characters then so is* $\chi_1\chi_2$ *where*

$$(\chi_1\chi_2)(a) = \chi_1(a)\chi_2(a).$$

(2) If χ *is a character so is its complex conjugate* $\overline{\chi}$. *Further* $\chi(a)^{-1} = \overline{\chi(a)}$.

(3) If χ_1 *is a fixed character and* χ *runs over all characters then so does* $\chi_1\chi$.

Proof The proofs of (1) and (2) are straightforward verifications of the four properties in the definition of a character and we leave these to the exercises.

For part (3) suppose that $(a, k) = 1$ and $\chi_1(a)\chi_2(a) = \chi_1(a)\chi_3(a)$. Then since $\chi_1(a) \neq 0$ it follows that $\chi_2(a) = \chi_3(a)$. Hence if χ is a fixed character and we let χ_1 run over all c distinct characters then $\chi\chi_1$ are again c distinct characters and hence must be all of them. □

We need to prove certain orthogonality relations among the characters. The next Lemma is crucial to do this and contains much of the work in proving these results.

Lemma 3.3.4 *If* $d > 0$ *and* $(d, k) = 1$ *with* $d \not\equiv 1 \bmod k$ *then there exists a character for which* $\chi(d) \neq 1$.

Proof Since $\chi(a) = 0$ if $(a, k) > 1$ it follows that to determine a character for which $\chi(d) \neq 1$ we must only find a function satisfying properties (2), (3), (4) of the definition of a character for $(a, k) = 1$.

Let $k = p_1^{e_1} \cdots p_m^{e_m}$ be the prime decomposition of k. Since $d \not\equiv 1 \bmod k$ it follows that for one of the prime divisors p of k we have $d \not\equiv 1 \bmod p^t$ for some $t > 0$. Suppose first that p is an odd prime divisor of k satisfying this, that is $d \not\equiv 1 \bmod p^t$ where $p^t | k$. Then p does not divide d since $(d, k) = 1$.

Recall that the unit group modulo p^t is cyclic, that is, there is a primitive root g modulo p^t. There are $\phi(p^t)$ primitive roots so choose $g \neq d$. (See Theorem 2.4.3 and Section 2.4.4). If $(a, k) = 1$ then a is a unit modulo k and hence a power of g modulo k. That is

$$a \equiv g^b \bmod p^t \text{ with } b \geq 0.$$

Let σ be the root of unity given by

$$\sigma = e^{\frac{2\pi i}{\phi(p^t)}}$$

and define for each a with $(a, k) = 1$ with $a \equiv g^b$ as above

$$\chi(a) = \sigma^b.$$

Further if $(a, k) > 1$ define $\chi(a) = 0$. This defines a function on the residue classes mod k. We must show that χ is a character and that $\chi(d) \neq 1$.

Property (1) of the definition of a character is clear from the definition of χ. Now $\chi(1) = \sigma^0 = 1$ since $g^0 = 1$. Hence $\chi(1) \neq 0$. Further if $(a_1, k) = (a_2, k) = 1$ then $a_1 \equiv g^{b_1}$ and $a_2 \equiv g^{b_2} \bmod p^t$. This implies that $\chi(a_1) = \sigma^{b_1}$, $\chi(a_2) = \sigma^{b_2}$. But $a_1 a_2 = g^{b_1 + b_2} \bmod p^t$ and hence

$$\chi(a_1 a_2) = \sigma^{b_1 + b_2} = \sigma^{b_1} \sigma^{b_2} = \chi(a_1)\chi(a_2).$$

Therefore χ is multiplicative.

Finally if $a_1 \equiv a_2 \bmod p^t$ then $a \equiv g^b \equiv a_2$ and hence $\chi(a_1) = \chi(a_2)$. Therefore χ is a character. Since $d \not\equiv 1 \bmod p^t$ then $d \equiv g^r \bmod p^t$ for some r with $\phi(p^t)$ not dividing r. Therefore

$$\chi(d) = \sigma^r \neq 1.$$

The above proof works whenever we have an odd prime divisor p of k with $d \not\equiv 1 \bmod p^t$. This leaves only the prime 2. Now suppose that $d \not\equiv 1 \bmod 2^t$ where $2^t | k$. If $k = 2q$ with q odd then $d \equiv 1 \bmod 2$ since $(d, k) = 1$. Therefore if $d \not\equiv 1 \bmod k$ and $k = 2q$ with q odd there must exist an odd prime divisor of k with $d \not\equiv 1 \bmod p^s$ and we are back to the first case. Hence we have that $k = 2^t q$ with $t > 1$ and $d \not\equiv 1 \bmod 2^t$.

Now $d \equiv 1 \bmod 2$ and hence $d \equiv 1 \bmod 4$ or $d \equiv 3 \bmod 4$. We consider each of these cases separately.

If $d \equiv 1 \bmod 4$ then $t > 2$. If $(a, k) = 1$ then clearly $(a, 2) = 1$. Then it can be shown that (see exercises)

$$a \equiv (-1)^{\frac{a-1}{2}} 5^b \bmod 2^t \text{ for some } b \geq 0.$$

Now let

$$\sigma = e^{\frac{2\pi i}{2^{t-2}}}$$

and define $\chi(a) = \sigma^b$. Since b is determined mod 2^{t-2} it follows that χ is well defined on the residue classes mod k. As in the odd case if we define $\chi(a) = 0$ for $(a, k) > 1$ then it is straightforward to verify that χ is a character. Again as in the odd case since $d \not\equiv 1 \bmod 2^t$ and $d \equiv 1 \bmod 4$ then $d \equiv 5^r \bmod 2^t$ with r not divisible by 2^{t-2}. Hence $\chi(d) = \sigma^r \neq 1$.

If $d \equiv 3 \bmod 4$ then $d \equiv -1 \bmod 4$. For $(a, k) = 1$ define

$$\chi(a) = (-1)^{\frac{a-1}{2}}.$$

As in the other cases it is straightforward to verify that χ is a character. Here $\chi(d) = -1 \neq 1$. This completes the proof of Lemma 3.3.4. □

The next two theorems are called the **orthogonality relations** for Dirichlet characters. They are special cases of general results on characters of representations of finite groups.

Theorem 3.3.2 *(Orthogonality Relations I) (1) If χ is a fixed character and a runs over a complete set of residue classes mod k then*

$$\sum_a \chi(a) = \phi(k) \textit{ if } \chi = \chi_0$$

$$\sum_a \chi(a) = 0 \textit{ if } \chi \neq \chi_0.$$

(2) If $a > 0$ is an integer then if χ runs over the set of all c characters

$$\sum_\chi \chi(a) = c \textit{ if } a \equiv 1 \textit{ mod } k$$

$$\sum_\chi \chi(a) = 0 \textit{ if } a \not\equiv 1 \textit{ mod } k.$$

Proof (1) Let χ_0 be the principal character as defined immediately after Definition 3.3.1. That is

$$\chi_0(a) = 0 \text{ if } (a, k) > 1$$

$$\chi_0(a) = 1 \text{ if } (a,k) = 1.$$

If a runs over a complete set of k positive residue classes mod k then

$$\sum_a \chi_0(a)$$

has $\phi(k)$ terms each with value 1 and $(k - \phi(k))$ terms each with value 0. Hence

$$\sum_a \chi_0(a) = \phi(k).$$

If $\chi \neq \chi_0$ choose a d with $d > 0$, $(d,k) = 1$ and $\chi(d) \neq 1$. This exists by Lemma 3.3.4. Then as a runs over a complete residue system mod k so does da. Then

$$\sum_a \chi(a) = \sum_a \chi(da).$$

But χ is multiplicative so

$$\sum_a \chi(a) = \sum_a \chi(da) = \sum_a \chi(d)\chi(a) = \chi(d) \sum_a \chi(a).$$

Since $\chi(d) \neq 1$ it follows that $\sum_a \chi(a) = 0$.

(2) For $a \equiv 1 \bmod k$ the sum $\sum_\chi \chi(a)$ runs over c characters. From Lemma 3.3.1 each of these has value 1 and the sum has value c.

If $(a,k) > 1$ then each of the terms in the series is zero so the sum vanishes. If $(a,k) = 1$ but a is noncongruent to 1 modulo k then there exists a character (by Lemma 3.3.4) with $\chi_1(a) \neq 1$. Now as χ runs over all c characters then by Lemma 3.3.3 so does $\chi_1\chi$. Hence

$$\sum_\chi \chi(a) = \sum_\chi \chi_1(a)\chi(a) = \chi_1(a) \sum_\chi \chi(a).$$

Since $\chi_1(a) \neq 1$ it follows that $\sum_\chi \chi(a) = 0$. □

We can now prove that c, the number of distinct characters mod k, is exactly $\phi(k)$.

Corollary 3.3.1 *There exist exactly $\phi(k)$ characters modulo k.*

Proof There are exactly $\phi(k)$ positive residues a with $(a,k) = 1$. If we sum over all c characters and $\phi(k)$ residues we get using the orthogonality results above

$$\sum_{a,\chi} \chi(a) = \sum_a \sum_\chi \chi(a) = c + 0 + \cdots + 0 = c.$$

On the other hand

$$\sum_{a,\chi}\chi(a)=\sum_{\chi}\sum_{a}\chi(a)=\phi(k)+0+\cdots+0=\phi(k).$$

Therefore $c=\phi(k)$. □

Theorem 3.3.3 *(Orthogonality Relations II) (1) If χ_1 and χ_2 are characters mod k and a runs over a complete set of residue classes mod k then*

$$\sum_{a}\chi_1(a)\overline{\chi_2(a)}=\phi(k)\ \textit{if}\ \chi_1=\chi_2$$
$$\sum_{a}\chi_1(a)\overline{\chi_2(a)}=0\ \textit{if}\ \chi_1\neq\chi_2.$$

(2) If $a>0$ is an integer and $(a,k)=1$ then if χ runs over the set of all $\phi(k)$ characters

$$\sum_{\chi}\chi(t)\overline{\chi(a)}=\phi(k)\ \textit{if}\ a\equiv t\ \textit{mod}\ k$$
$$\sum_{\chi}\chi(t)\overline{\chi(a)}=\ \ 0\ \textit{if}\ a\not\equiv t\ \textit{mod}\ k.$$

Proof (1) From Lemma 3.3.3 we have that for any character $\chi^{-1}=\overline{\chi}$. Hence if $\chi_1=\chi_2$ then

$$\chi_1(a)\overline{\chi_2(a)}=\chi_1(a)\overline{\chi_1(a)}=\chi_0(a)$$

where χ_0 is the principal character. Therefore from Theorem 3.3.1

$$\sum_{a}\chi_1(a)\overline{\chi_2(a)}=\sum_{a}\chi_0(a)=\phi(k).$$

If $\chi_1\neq\chi_2$ then $\chi_1^{-1}\neq\overline{\chi_2}$ and hence $\chi_1\overline{\chi_2}\neq\chi_0$. Then again from Theorem 3.3.1

$$\sum_{a}\chi_1(a)\overline{\chi_2(a)}=0.$$

(2) The proof of the second part of the theorem follows in an analogous manner from Theorem 3.3.1. We leave the details to the exercises. □

Before moving on to Dirichlet series we mention that Theorems 3.3.1 and 3.3.2 are special cases of general results in group representation theory. If G is a finite group then a (matrix) **representation** of G is a homomorphism $\rho:G\to GL_n(R)$

(see Section 3.2) for some n and some ring R. Hence $\rho(g)$ is an invertible $n \times n$ matrix for $g \in G$. The **character** of the representation ρ is the function $\chi_\rho : G \to R$ given by $\chi_\rho(g) = tr(\rho(g))$. For any finite group G there are orthogonality relations on the set of characters which specialize in the case of finite abelian groups (for complex representations) to the theorems on Dirichlet characters. The book by Curtis and Reiner [CR] is a standard reference on representations of finite groups. A more elementary treatment can be found in the book by M. Newman [New 1].

The next ingredient in the proof of Dirichlet's Theorem is **Dirichlet series**.

Definition 3.3.2 *If χ is a character mod k then the* **Dirichlet L-series** *is defined for complex values s by*

$$L(s, \chi) = \sum_{n=1}^{\infty} \frac{\chi(n)}{n^s}.$$

A rough outline of the way these series lead to a proof of Dirichlet's Theorem is as follows. Consider $(a, b) = 1$ and consider Dirichlet characters mod a. It can be shown that for $s > 1$ the series $L(s, \chi)$ is an analytic function of s and further for $s > 1$ satisfies an analogue of the Euler product, (see Section 3.1.2 and [N]), that is,

$$L(s, \chi) = \prod_p (1 - \frac{\chi(p)}{p^s})^{-1}.$$

Then by logarithmic differentiation

$$-\frac{L'(s, \chi)}{L(s, \chi)} = \sum_p \frac{\chi(p) \ln p}{p^s - \chi(p)}.$$

If we introduce the function Λ on $\mathbb{N}$ by

$$\Lambda(n) = \begin{cases} \ln p & \text{for } n = p^c, c \geq 1 \\ 0 & \text{for all other } n > 0 \end{cases}$$

then the above can be rewritten as

$$-\frac{L'(s, \chi)}{L(s, \chi)} = \sum_{n=1}^{\infty} \frac{\chi(n)\Lambda(n)}{n^s}.$$

The function $\Lambda(n)$ is called the **von Mangoldt function** and will also play a role in the proof of the prime number theorem. Multiplying by $\overline{\chi(b)}$ and then summing over all other characters $\chi^\star$ we get by the orthogonality relations

$$\sum_{n \equiv b \bmod a} \frac{\Lambda(n)}{n^s} = \frac{1}{\phi(a)} \sum_{\chi^\star} \overline{\chi^\star(b)} \left(-\frac{L'(s, \chi^\star)}{L(s, \chi^\star)} \right).$$

As $s \to 1^+$ the left hand side becomes approximately

$$\sum_{p \equiv b \bmod a} \frac{\ln p}{p}.$$

What must be shown is that the right hand side becomes infinite. This would then imply that the number of primes congruent to b mod a must be infinite.

It can be shown that for the principal character we have $-\frac{L'(s,\chi_0)}{L(s,\chi_0)} \to \infty$ as $s \to 1^+$. It follows that to show that the right hand side above becomes infinite we must show that $\frac{L'(s,\chi)}{L(s,\chi)}$ remains bounded for any non-principal character. To show this we must show that $L(1, \chi) \neq 0$ for any non-principal character. We now outline a series of results which prove all these assertions.

Theorem 3.3.4 *For any character χ mod k the Dirichlet L-series is an analytic function for $s > 1$. Further it has an Euler product representation*

$$L(s, \chi) = \prod_p (1 - \frac{\chi(p)}{p^s})^{-1}.$$

The proof of this theorem follows from the following sequence of lemmas.

Lemma 3.3.5 *$L(s, \chi)$ is absolutely convergent for $s > 1$.*

Proof From Lemma 3.3.3 we know that $|\chi(n)| \leq 1$ and hence $\frac{|\chi(n)|}{n^s} \leq \frac{1}{n^s}$. Therefore

$$|L(s, \chi)| = |\sum_{n=1}^{\infty} \frac{\chi(n)}{n^s}| \leq \sum_{n=1}^{\infty} |\frac{\chi(n)}{n^s}| \leq \sum_{n=1}^{\infty} \frac{1}{n^s}$$

which converges for $s > 1$. Hence $L(s, \chi)$ is absolutely convergent for $s > 1$. □

Lemma 3.3.6 *The series*

$$\sum_{n=1}^{\infty} \frac{\chi(n) \ln n}{n^s}$$

converges absolutely for $s > 1$ and further in this range

$$L'(s, \chi) = -\sum_{n=1}^{\infty} \frac{\chi(n) \ln n}{n^s}.$$

Proof For $s > 1 + \epsilon$ we have

$$|\frac{\chi(n) \ln n}{n^s}| \leq \frac{\ln n}{n^{1+\epsilon}}.$$

However $\sum_{n=1}^{\infty} \frac{\ln n}{n^{1+\epsilon}}$ converges by the integral test. Thus the given series converges uniformly for $s > 1 + \epsilon$ and hence absolutely for $s > 1$. Now $L(s, \chi) = \sum_{n=1}^{\infty} \frac{\chi(n)}{n^s}$ so by uniform convergence we can differentiate termwise and therfore

$$L'(s, \chi) = -\sum_{n=1}^{\infty} \frac{\chi(n) \ln n}{n^s}.$$

(Recall that if $y = n^{-s}$ then $y' = -n^{-s} \ln n$.) □

Let μ be the **Möbius function** defined for natural numbers n by

$$\mu(n) = \begin{cases} 1 & \text{if } n = 1 \\ (-1)^r & \text{if } n = p_1 p_2 \cdots p_r \text{ with } p_1, \ldots, p_r \text{ distinct primes} \\ 0 & \text{otherwise.} \end{cases}$$

Then

Lemma 3.3.7 *The series*

$$\sum_{n=1}^{\infty} \frac{\chi(n)\mu(n)}{n^s}$$

converges absolutely for $s > 1$ and further in this range

$$L(s, \chi) \cdot \sum_{n=1}^{\infty} \frac{\chi(n)\mu(n)}{n^s} = 1.$$

It follows that $L(s, \chi) \neq 0$ for $s > 1$.

Proof As before $|\frac{\chi(n)\mu(n)}{n^s}| \leq \frac{1}{n^s}$ so the absolute convergence follows from the convergence of the series $\sum_{n=1}^{\infty} \frac{1}{n^s}$ for $s > 1$.

Now it can be shown that for the Möbius function $\mu(n)$ we have

$$\sum_{d|n} \mu(d) = \begin{cases} 1 & \text{if } n = 1 \\ 0 & \text{if } n > 1. \end{cases}$$

(See Theorem 2.4.8 for a similar type result and Section 3.6 for a proof.)

Using this above fact we then have

$$\sum_{m=1}^{\infty} \frac{\chi(m)}{m^s} \sum_{n=1}^{\infty} \frac{\chi(n)\mu(n)}{n^s} = \sum_{t=1}^{\infty} \sum_{mn=t} \frac{\chi(m)\chi(n)\mu(n)}{m^s n^s} = \sum_{t=1}^{\infty} \frac{\chi(t)}{t^s} \sum_{n|t} \mu(n) = 1.$$

Therefore

$$L(s,\chi)\cdot\sum_{n=1}^{\infty}\frac{\chi(n)\mu(n)}{n^s}=1.$$

□

We can now obtain the indicated Euler product representation for $L(s,\chi)$.

Lemma 3.3.8 *For $s>1$ we have the Euler product representation*

$$L(s,\chi)=\prod_p(1-\frac{\chi(p)}{p^s})^{-1}.$$

Proof For $m>1$ let S be the set of all positive integers n not divisible by any prime $p>m$. Then we have

$$\prod_{p\le m}(1-\frac{\chi(p)}{p^s})=\sum_{n\in S}\frac{\chi(n)\mu(n)}{n^s}.$$

All $n\le m$ are included in the set S and therefore

$$\prod_{p\le m}(1-\frac{\chi(p)}{p^s})=\sum_{1\le n\le m}\frac{\chi(n)\mu(n)}{n^s}+\sum_{n'>m}\frac{\chi(n')\mu(n')}{n'^s}$$

where the second sum runs over those $n'>m$ which are not divisible by any prime $p>m$. Now as $m\to\infty$ the first sum on the right goes to

$$\sum_{n=1}^{\infty}\frac{\chi(n)\mu(n)}{n^s}=\frac{1}{L(s,\chi)}$$

by Lemma 3.3.7. The second sum on the right approaches 0 since its absolute value is less than $\sum_{n>m}\frac{1}{n^s}$. Combining these

$$\prod_p(1-\frac{\chi(p)}{p^s})=\frac{1}{L(s,\chi)}\implies L(s,\chi)=\prod_p(1-\frac{\chi(p)}{p^s})^{-1}.$$

□

Recall that the von Mangoldt function $\Lambda(n)$ was defined for positive integers by

$$\Lambda(n)=\begin{cases}\ln p & \text{if } n=p^c, c\ge 1\\ 0 & \text{for all other } n>0.\end{cases}$$

We then get:

Theorem 3.3.5 *(1) For $s > 1$ we have*

$$-\frac{L'(s,\chi)}{L(s,\chi)} = \sum_{n=1}^{\infty} \frac{\chi(n)\Lambda(n)}{n^s}.$$

(2) As $s \to 1^+$ we have for the principal character χ_0,

$$-\frac{L'(s,\chi_0)}{L(s,\chi_0)} \to \infty.$$

Proof Since $|\chi(n)\Lambda(n)| \leq \ln n$ it follows that the series $\sum_{n=1}^{\infty} \frac{\chi(n)\Lambda(n)}{n^s}$ converges absolutely for $s > 1$.

Now it can be shown, in a similar manner as for the Möbius function, that

$$\sum_{d|n} \Lambda(d) = \ln n$$

(see exercises). Hence for $s > 1$.

$$L(s,\chi) \sum_{n=1}^{\infty} \frac{\chi(n)\Lambda(n)}{n^s} = \sum_{m=1}^{\infty} \frac{\chi(m)}{m^s} \sum_{n=1}^{\infty} \frac{\chi(n)\Lambda(n)}{n^s}$$

$$= \sum_{t=1}^{\infty} \frac{\chi(t)}{t^s} \sum_{n|t} \Lambda(n) = \sum_{t=1}^{\infty} \frac{\chi(t) \ln t}{t^s} = -L'(s,\chi).$$

For the principal character χ_0 we have $\chi_0(n) = 1$ if $(n,k) = 1$ and 0 otherwise. Therefore from the first part of the theorem it follows that

$$-\frac{L'(s,\chi_0)}{L(s,\chi_0)} = \sum_{n=1,(n,k)=1} \frac{\Lambda(n)}{n^s}$$

$$= \sum_{n=1}^{\infty} \frac{\Lambda(n)}{n^s} - \sum_{p|k} \ln p \sum_{m=1}^{\infty} \frac{1}{p^{ms}}$$

$$= \sum_{n=1}^{\infty} \frac{\Lambda(n)}{n^s} - \sum_{p|k} \frac{\ln p}{p^s - 1}.$$

As $s \to 1$ the second term on the right is finite. Hence to prove that $-\frac{L'(s,\chi_0)}{L(s,\chi_0)} \to \infty$ as $s \to 1^+$we must only show that the first term in the expression above diverges.

From Euler's proof of the infinitude of primes we know that $\sum_p \frac{1}{p}$ diverges. Since $\frac{\ln p}{p} > \frac{1}{p}$ it follows that $\sum_p \frac{\ln p}{p}$ diverges and hence so does $\sum_{n=1}^{\infty} \frac{\Lambda(n)}{n}$. Hence for every $t > 0$ there exists an $m = m(t)$ for which

$$\sum_{n=1}^{m} \frac{\Lambda(n)}{n} > t.$$

For $1 < s < 1 + \epsilon(t)$ we then have

$$\sum_{n=1}^{m} \frac{\Lambda(n)}{n^s} > t \implies \sum_{n=1}^{\infty} \frac{\Lambda(n)}{n^s} > t.$$

From this last inequality it follows clearly that the sum diverges. □

We now have one big brick of Dirichlet's proof in place that is that for the principal character

$$\frac{-L'(s,\chi_0)}{L(s,\chi_0)} \to \infty \text{ as } s \to 1^+.$$

As explained above we now need to know that $L(1,\chi)$ does not vanish for any non-principal character. This is the most difficult part of the proof.

First three more preliminary results are needed.

Lemma 3.3.9 *If* $t \geq m \geq 1$ *and* χ *is not the principal character then*

$$|\sum_{n=m}^{t} \chi(n)| \leq \frac{\phi(k)}{2}.$$

Proof By the orthogonality relations the sum $\sum \chi(a)$ over a complete set of residues is zero. Hence in the given sum we may assume that there are $\leq (k-1)$ terms. In a complete set of residues exactly $\phi(k)$ terms have $|\chi(a)| = 1$ and all the remaining terms have $|\chi(a)| = 0$. If between m and t there are at most $\frac{\phi(k)}{2}$ terms with $|\chi(a)| = 1$ then

$$|\sum_{n=m}^{t} \chi(n)| \leq \sum_{n=m}^{t} |\chi(n)| \leq \frac{\phi(k)}{2}.$$

If there are more than $\frac{\phi(k)}{2}$ such terms then

$$|\sum_{n=m}^{t} \chi(n)| = |\sum_{n=m}^{m+k-1} \chi(n) - \sum_{n=t+1}^{m+k-1} \chi(n)|$$

$$= |\sum_{n=t+1}^{m+k-1} \chi(n)| \leq \sum_{n=t+1}^{m+k-1} |\chi(n)| < \frac{\phi(k)}{2}.$$

□

Lemma 3.3.10 *For any character χ and $s > 1$ we have the inequality*

$$(L(s,\chi_0))^3|L(s,\chi)|^4|L(s,\chi^2)|^2 \geq 1.$$

Proof For real numbers x, y with $0 < x < 1$ we have the inequality

$$(1-x)^3|1-xe^{iy}|^4|1-xe^{2iy}|^2 < 1.$$

(See the exercises.)

If p is a prime which does not divide k let $\chi(p) = e^{iy}$ and let $x = \frac{1}{p^s}$. Applying the above inequality then gives

$$(1-\frac{\chi_0(p)}{p^s})^3|(1-\frac{\chi(p)}{p^s})|^4|(1-\frac{\chi^2(p)}{p^s})|^2 \leq 1.$$

Multiplying over all primes and using the Euler product representation of the L-series then gives the stated inequality. □

Lemma 3.3.11 *For any non-principal character χ we have $|L'(s,\chi)| < \phi(k)$ for $s \geq 1$.*

Proof From Lemma 3.3.6 we have

$$|L'(s,\chi)| = |\sum_{n=1}^{\infty}\frac{\chi(n)\ln n}{n^s}|$$

for $s > 1$ and so we work with the right hand sum.

It is straightforward to show that the function $f(t) = \frac{\ln t}{t^s}$ is a decreasing function for $t \geq 3$. Therefore from Lemma 3.3.9 we have for $t \geq m \geq 3$ the inequality

$$|\sum_{n=m}^{t}\frac{\chi(n)\ln n}{n^s}| \leq \frac{\phi(k)}{2}\frac{\ln m}{m^s} \leq \frac{\phi(k)}{2}\frac{\ln m}{m}.$$

Hence the series for $L'(s,\chi)$ converges uniformly for $s \geq 1$. In this range taking $m = 3$ and letting $t \to \infty$ it follows that

$$|\sum_{n=1}^{\infty}\frac{\chi(n)\ln n}{n^s}| \leq \frac{\ln 2}{2} + \frac{\phi(k)}{2}\frac{\ln 3}{3} < \frac{1}{2} + \frac{\phi(k)}{2} \leq \phi(k).$$

□

Theorem 3.3.6 *$L(1,\chi) \neq 0$ for any non-principal character and further for any non-principal character $\frac{L'(s,\chi)}{L(s,\chi_0)}$ is bounded for $s > 1$.*

Proof We break the proof into two pieces. The first for nonreal characters, that is characters which take complex values and the second for real, but not principal characters. This second part is the most difficult.

From Lemma 3.3.9 we have for any non-principal character

$$|\sum_{n=m}^{t}\chi(n)| \leq \frac{\phi(k)}{2}.$$

Therefore for any non-principal character with $s > 1$

$$|L(s,\chi)| < \phi(k)$$

letting $m = 1$ and $t \to \infty$ in the above inequality and using that $\frac{|\chi(n)|}{n^s} < |\chi(n)|$.

Assume first that χ is a nonreal character. Then χ^2 is not the principal character for if it were χ would have to be real. Then from the remark above we have for $s > 1$ that $|L(s,\chi^2)| < \phi(k)$. On the other hand if $1 < s < 2$ we have

$$L(s,\chi_0) = \sum_{n=1,(n,k)=1}^{\infty}\frac{1}{n^s} \leq \sum_{n=1}^{\infty}\frac{1}{n^s} < 1 + \int_1^{\infty}\frac{dz}{z^s} \tag{3.1}$$

$$= 1 + \frac{1}{s-1} = \frac{s}{s-1} < \frac{2}{s-1}. \tag{3.2}$$

Applying Lemma 3.3.10 we have

$$|L(s,\chi)| \geq \frac{1}{(L(s,\chi_0)^{\frac{3}{4}}}\frac{1}{|L(s,\chi^2)|^{\frac{2}{4}}} > \frac{(s-1)^{\frac{3}{4}}}{2^{\frac{3}{4}}}\frac{1}{\sqrt{\phi(k)}} > \frac{(s-1)^{\frac{3}{4}}}{2\sqrt{\phi(k)}}.$$

If $L(1,\chi) = 0$ then for $s > 1$

$$|L(s,\chi)| = |L(s,\chi) - L(1,\chi)| = |\int_1^s L'(t,\chi)dt| < \phi(k)(s-1).$$

Hence for $1 < s < 2$ we would have

$$(s-1)^{\frac{1}{4}} > \frac{1}{2\phi(k)^{\frac{3}{2}}}.$$

However this inequality is false for $s = 1 + \frac{1}{16\phi(k)^{\frac{3}{2}}}$. Therefore $L(1,\chi) \neq 0$ for χ any nonreal character.

Now assume that χ is a real character but not the principal character. As remarked earlier this is the most difficult part. To begin we define the function $f(n)$ on the positive integers n by

$$f(n) = \sum_{d|n} \chi(d).$$

Then we can prove that (see exercises) $f(n) \geq 0$ for all $n \geq 1$ and $f(n) \geq 1$ if $n = c^2$ a square.

Let $m = (4\phi(k))^6$ and $z = \sum_{n=1}^{m} 2(m-n)f(n)$. Applying the definition of $f(n)$ we have

$$z = \sum_{uv \leq m} 2(m - uv)\chi(v).$$

Since $f(n) \geq 0$ and $f(c^2) \geq 1$ we have

$$z \geq \sum_{v=1}^{\sqrt{m}} 2(m - v^2) \geq \sum_{v=1}^{\frac{\sqrt{m}}{2}} 2(m - v^2)$$

$$\geq \sum_{v=1}^{\frac{\sqrt{m}}{2}} 2(m - \frac{m}{4}) = \frac{3}{4}m^{\frac{3}{2}} = \frac{3}{4}(4\phi(k))^9.$$

Let

$$z_1 = \sum_{u=1}^{m^{\frac{1}{3}}} \sum_{m^{\frac{3}{2}} < v \leq \frac{m}{u}} 2(m - uv)\chi(v)$$

$$z_2 = \sum_{v=1}^{m^{\frac{2}{3}}} \sum_{0 < u \leq \frac{m}{v}} 2(m - uv)\chi(v).$$

Then it follows from $uv \leq m$ that either $u \leq m^{\frac{1}{3}}$, $v > m^{\frac{2}{3}}$ or $v \leq m^{\frac{2}{3}}$. This implies then that

$$z = z_1 + z_2.$$

Suppose that $z(n)$ is a complex valued function on the natural numbers. Let c be a natural number and for $t \geq c$ let $r(t) = \sum_{n=c}^{t} z(n)$. Let $r(u-1) = 0$. For $d \geq c$ let $\nu = \sum_{c \leq t \leq d} |r(t)|$ and let $\epsilon_c \geq \epsilon_{c+1} \geq \cdots \geq \epsilon_d \geq 0$. Then

$$\sum_{n=c}^{d} \epsilon_n z(n) = \sum_{n=c}^{d} \epsilon_n (r(n) - r(n-1)) = \sum_{n=c}^{d-1} r(n)(\epsilon_n - \epsilon_{n+1}) + r(d)\epsilon_d.$$

This then implies that

$$|\sum_{n=c}^{d}\epsilon_n z(n)| \leq \nu(\sum_{n=c}^{d-1}(\epsilon_n - \epsilon_{n+1}) + \epsilon_d) = \nu\epsilon_c. \qquad (3.3.1)$$

From Lemma 3.3.9

$$|\sum_{n=c}^{d}\chi(n)| \leq \frac{\phi(k)}{2}.$$

Applying the above remarks to this inequality with $\epsilon_n = \frac{1}{n^s}$ we get

$$|\sum_{n=c}^{d}\frac{\chi(n)}{n^s}| \leq \frac{\phi(k)}{2}.\frac{1}{c^s} \leq \frac{\phi(k)}{2c} \qquad (3.3.2)$$

Now applying the inequality (3.3.1) to the definition of z_1 gives us

$$z_1 \leq \sum_{u=1}^{m^{\frac{1}{3}}}|\sum_{m^{\frac{2}{3}}<v\leq\frac{m}{u}} 2(m-uv)\chi(v)| \leq \sum_{u=1}^{m^{\frac{1}{3}}} 2m\frac{\phi(k)}{2} = m^{\frac{4}{3}}\phi(k).$$

Now as defined

$$z_2 = \sum_{v=1}^{m^{\frac{2}{3}}}\sum_{0<u\leq\frac{m}{v}} 2(m-uv)\chi(v).$$

Let $\theta = \frac{m}{v} - [\frac{m}{v}]$ where [ ] is the greatest integer function. Then $0 \leq \theta < 1$ and

$$\sum_u(2m-2uv) = 2m\sum_u 1 - v\sum_u 2u = 2m[\frac{m}{v}] - v[\frac{m}{v}]([\frac{m}{v}]+1)$$

$$= 2m(\frac{m}{v}-\theta) - v((\frac{m}{v}-\theta)^2 + \frac{m}{v} - \theta)$$

$$= \frac{2m^2}{v} - 2m\theta - v(\frac{m^2}{v^2} - 2\theta\frac{m}{v} + \theta^2 + \frac{m}{v} - \theta)$$

$$= \frac{m^2}{v} - m + v(\theta - \theta^2).$$

Since $0 \leq \theta < 1$ we have $|\theta - \theta^2| \leq 1$ and hence

$$z_2 = m^2 \sum_{v=1}^{m^{\frac{2}{3}}} \frac{\chi(v)}{v} - m \sum_{v=1}^{m^{\frac{2}{3}}} \chi(v) + \sum_{v=1}^{m^{\frac{2}{3}}} \chi(v) v(\theta - \theta^2)$$

$$\leq m^2 (L(1,\chi) - \sum_{v=m^{\frac{2}{3}}+1}^{\infty} \frac{\chi(v)}{v^s}) + m\frac{\phi(k)}{2} + m^{\frac{2}{3}} \sum_{v=1}^{m^{\frac{2}{3}}} 1.$$

Applying the inequality

$$|\sum_{n=c}^{d} \frac{\chi(n)}{n^s}| \leq \frac{\phi(k)}{2} . \frac{1}{c^s} \leq \frac{\phi(k)}{2c}$$

and letting $c = m^{\frac{2}{3}} + 1$, $v \to \infty$ we obtain

$$z_2 < m^2 L(1,\chi) + m^2 \frac{\phi(k)}{2} \frac{1}{m^{\frac{2}{3}}} + m^{\frac{4}{3}} \frac{\phi(k)}{2} + m^{\frac{4}{3}} \phi(k)$$

$$= m^2 L(1,\chi) + m^{\frac{4}{3}} \phi(k)(\frac{1}{2} + \frac{1}{2} + 1)$$

$$= m^2 L(1,\chi) + 2m^{\frac{4}{3}} \phi(k).$$

It follows then, summarizing all these inequalities, that

$$\frac{3}{4}(4\phi(k))^9 \leq z < m^2 L(1,\chi) + 3m^{\frac{4}{3}} \phi(k) = m^2 L(1,\chi) + 3(4\phi(k))^8 \phi(k)$$

$$= m^2 L(1,\chi) + \frac{3}{4}(4\phi(k))^9.$$

This then clearly implies that $m^2 L(1,\chi) > 0$ and therefore $L(1,\chi) > 0$. Hence $L(1,\chi) \neq 0$ for χ a real non-principal character completing the proof that $L(1,\chi) \neq 0$ for any non-principal character.

We must now show that $\frac{L'(s,\chi)}{L(s,\chi)}$ remains bounded for $s > 1$. Since $L(1,\chi) \neq 0$ it follows that $\frac{1}{L(s,\chi)}$ is bounded for $s \geq 1$. From Lemma 3.3.11 $L'(s,\chi)$ is also bounded for $s \geq 1$ completing the proof. □

The final piece is

Theorem 3.3.7 *Suppose* $(t,k) = 1$, $t > 0$. *Then for* $s > 1$ *we have*

$$-\frac{1}{\phi(k)} \sum_{\chi} \overline{\chi(t)} \frac{L'(s,\chi)}{L(s,\chi)} = \sum_{n \equiv t \bmod k} \frac{\Lambda(n)}{n^s}.$$

Proof For $s > 1$ we have from Theorem 3.3.4 that

$$-\frac{L'(s,\chi)}{L(s,\chi)} = \sum_{n=1}^{\infty} \frac{\chi(n)\Lambda(n)}{n^s}.$$

Combining this with the orthogonality relations for characters we get

$$-\sum_{\chi} \frac{1}{\chi(t)} \frac{L'(s,\chi)}{L(s,\chi)} = \sum_{\chi} \frac{1}{\chi(t)} \sum_{n=1}^{\infty} \frac{\chi(n)\Lambda(n)}{n^s}$$

$$= \sum_{n=1}^{\infty} \frac{\Lambda(n)}{n^s} \sum_{\chi} \frac{1}{\chi(t)} \chi(n) = \sum_{n \equiv t \text{ mod} k} \frac{\Lambda(n)}{n^s} \phi(k).$$

□

We can now give the proof of Dirichlet's Theorem.

Proof We suppose that $(a, b) = 1$ and we want to show that there are infinitely many primes of the form $an + b$ or equivalently infinitely many primes congruent to b mod a. We consider the Dirichlet characters mod a. Apply Theorem 3.3.6 with $a = k$ and $b = t$ so that

$$-\frac{1}{\phi(a)} \sum_{\chi} \overline{\chi(b)} \frac{L'(s,\chi)}{L(s,\chi)} = \sum_{n \equiv b \text{ mod } a} \frac{\Lambda(n)}{n^s}.$$

As $s \to 1^+$ the left hand side approaches ∞ since the term for the principal character goes to $-\infty$ while the other $\phi(a) - 1$ terms remain bounded. Therefore we have as $s \to 1^+$ and with all congruences mod a

$$\sum_{p \equiv b} \frac{\ln p}{p^s} + \sum_{(p,m)_{p^m \equiv b, m > 1}} \frac{\ln p}{p^{ms}} \to \infty.$$

Now

$$\sum_{n=1}^{\infty} \frac{2 \ln n}{n^2} > \sum_{n=2}^{\infty} \frac{\ln n}{n(n-1)} \geq \sum_{p} \frac{\ln p}{p(p-1)}$$

$$\geq \sum_{p,m;m>1} \frac{\ln p}{p^m} > \sum_{p,m;m>1} \frac{\ln p}{p^{ms}}$$

$$\geq \sum_{(p,m)_{p^m \equiv b \text{ mod } a, m>1}} \frac{\ln p}{p^{ms}}, \; s > 1.$$

Therefore the second sum

$$\sum_{(p,m)_{p^m \equiv b \bmod a, m>1}} \frac{\ln p}{p^{ms}}$$

remains bounded as $s \to 1^+$. It follows that

$$\sum_{p \equiv b \bmod a} \frac{\ln p}{p^s} \to \infty.$$

Therefore the number of primes $\equiv b \bmod a$ must be infinite. □

Before leaving Dirichlet's Theorem we would like to mention a beautiful new result of Ben Green and Terence Tao [GT] also related to primes and arithmetic progressions. It is a classical conjecture that there are arbitrarily long arithmetic progressions of prime numbers. This conjecture was hinted at in the work of Lagrange and Waring in the late 1700's (see [D]). In 1939 Van der Corput [VC] established that there are infinitely many triples of primes in arithmetic progression. Green and Tao [GT] proved the following.

Theorem 3.3.8 *The prime numbers contain arithmetic progressions of length k for all k. That is, for all $k \in \mathbb{N}$ there exists $a, b \in \mathbb{N}$ with $(a, b) = 1$ such that*

$$a, a+b, a+2b, \ldots, a+(k-1)b$$

are all primes.

Their proof is probabilistic and nonconstructive and quite difficult.

3.4 Twin Prime Conjecture and Related Ideas

Twin primes are prime numbers p and q such that $|p-q| = 2$. For example $\{3, 5\}$, $\{5, 7\}$, $\{11, 13\}$ are all pairs of twin primes. Trivially 2, 3 is the only pair of primes that differ by one. It is not known whether or not there are infinitely many pairs of twin primes but an examination of the list of primes shows an abundance of such pairs and leads to the following conjecture.

Twin Primes Conjecture: *There are infinitely many pairs of twin primes.*

Despite the twin primes conjecture there is a remarkable theorem of Brun which says essentially that even if there are infinitely many twin primes the sum of their reciprocals converges. Recall that Euler proved that the sum $\sum_{p \text{ prime}} \frac{1}{p}$ diverges. This implied that the sequence of primes is infinite. Here let

$$S = \{p;\ p \text{ prime and } p+2 \text{ prime}\}.$$

That is S is the set of twin primes. Brun's Theorem is the following

Theorem 3.4.1 *(Brun) Let S be the set of twin primes then*

$$\sum_{p \in S} (\frac{1}{p} + \frac{1}{p+2})$$

converges.

Notice that if S is a finite set then certainly the sum converges. Brun's proof depends on a method known as Brun's sieve. We will look at this method as well as the proof of Theorem 3.4.1 in Chapter 5. We mention some elementary facts about twin primes—leaving the proofs to the exercises.

Lemma 3.4.1 *The integer* 5 *is the only prime appearing in two different twin prime pairs.*

Primes are those natural numbers which have only two possible positive divisors. The next Lemma gives a similar characterization of twin primes.

Lemma 3.4.2 *There is a one-to-one correspondence between twin prime pairs and those integers $n \geq 4$ for which $n^2 - 1$ has only four possible positive divisors.*

Lemma 3.4.3 *Suppose p, q are primes. Then $pq + 1$ is a square if and only if p and q are twin primes.*

Lemma 3.4.4 *If p, q are twin primes greater than* 3 *then $p + q$ is divisible by* 12.

Brun's Theorem has been extended to further pairs of primes separated by a constant $d > 2$. For example if $d = 4$ the pairs of primes of the form $(p, p + 4)$ are called **cousin primes**. Again it is open whether there are infinitely many of these (for each d or for any fixed d) but Segal [S] proved that for any given d the sums of the reciprocals of the pairs is also convergent.

In 2014 Y. Zhang [Zh] proved that there is a positive constant with the property that infinitely many pairs of primes differ by less than that constant. In 2015 J. Maynard [Ma] gave a numerical extension.

3.5 Primes Between x and $2x$

In Theorem 2.3.2 we saw that there are arbitrarily large gaps in the sequence of primes. Despite this fact, the next result, known as **Bertrand's Theorem**, says that for any integer x there must be a prime between x and $2x$. Bertrand verified this empirically for a large number of natural numbers and conjectured the result. The theorem was proved by Chebyshev.

Theorem 3.5.1 *(Bertrand's Theorem) For every natural number $n > 1$ there is a prime p such that $n < p < 2n$.*

Chebyshev's proof of Bertrand's conjecture used techniques which he also used in obtaining a simple asymptotic bound on $\pi(x)$. This bound was a step on the road to the prime number theorem. We will give a proof of Chebyshev's Theorem in the next chapter and defer a proof of Bertrand's Theorem until then.

3.6 Arithmetic Functions and the Möbius Inversion Formula

In the course of Chapters 2 and 3 we used several functions, such as the Euler phi function $\phi(n)$, the sum of the divisors function $\sigma(n)$, the Van Mangoldt function $\Lambda(n)$ and the Möbius function $\mu(n)$, whose domain was the natural numbers and whose range was contained in the complex numbers. Functions such as these are called **arithmetic functions** or **number theoretic functions** and play an extensive role in Number Theory. Several other functions of this type will be used in the proof of the prime number theorem. In this final section of this chapter we take a look at arithmetic functions in general and a very important result called the Möbius inversion formula.

Definition 3.6.1 *An* **arithmetic function** *or* **number theoretic function** *is a function $f : \mathbb{N} \to \mathbb{C}$, whose domain is the natural numbers and whose range is a subset of the complex numbers.*

Besides the arithmetic functions that we have mentioned already, very important examples are given by the **divisor functions**.

$$\begin{aligned}\tau(n) &= \text{ number of positive divisors of } n\\ \sigma(n) &= \text{ sum of the positive divisors of } n\\ \sigma_k(n) &= \text{ sum of the kth powers of the positive divisors of } n.\end{aligned}$$

These can also be written in the following form.

$$\begin{aligned}\tau(n) &= \sum_{d|n} 1\\ \sigma(n) &= \sum_{d|n} d\\ \sigma_k(n) &= \sum_{d|n} d^k.\end{aligned}$$

We saw in Section 2.4.3 that if ϕ is the Euler phi function and $(m, n) = 1$ then $\phi(mn) = \phi(m)\phi(n)$. This property is called **multiplicativity**.

Definition 3.6.2 *An arithmetic function f is* **multiplicative** *if*

$$f(mn) = f(m)f(n)$$

whenever $(m, n) = 1$.

If n has a prime decomposition $n = p_1^{e_1} \cdots p_k^{e_k}$ and f is a multiplicative arithmetic function then $f(n) = f(p_1^{e_1}) \cdots f(p_k^{e_k})$. Therefore multiplicative arithmetic functions are uniquely determined by their values on prime powers. Further notice that for any n we have $f(n) = f(n)f(1)$. Hence if there is any n with $f(n) \neq 0$ we must have $f(1) = 1$.

Multiplicativity is preserved under summing over divisors. More precisely we have the following theorem.

Theorem 3.6.1 *Suppose that $f(n)$ is a multiplicative arithmetic function and*

$$F(n) = \sum_{d|n} f(d).$$

Then $F(n)$ is also multiplicative.

Proof Suppose that $n = n_1 n_2$ with $(n_1, n_2) = 1$. If $d|n$ then since n_1 and n_2 are relatively prime it follows that $d = d_1 d_2$ with $d_1|n_1$, $d_2|n_2$ and $(d_1, d_2) = 1$. Conversely if $d = d_1 d_2$ with $d_1|n_1$ and $d_2|n_2$ then $d|n$. This establishes a one-to-one correspondence between the positive divisors of n and pairs of divisors d_1, d_2 of n_1, n_2 respectively. It follows that

$$f(n) = \sum_{d|n} f(d) = \sum_{d_1|n_1} \sum_{d_2|n_2} f(d_1 d_2).$$

The function f is assumed to be multiplicative and hence $f(d_1 d_2) = f(d_1)f(d_2)$. Therefore

$$F(n) = \sum_{d_1|n_1} f(d_1) \sum_{d_2|n_2} f(d_2) = F(n_1)F(n_2)$$

proving the theorem. □

This theorem can be used immediately to show that the divisor functions are multiplicative. It is clear from the Fundamental Theorem of Arithmetic and the definition that $\tau(n)$ is multiplicative. From the expressions

$$\sigma(n) = \sum_{d|n} d$$

$$\sigma_k(n) = \sum_{d|n} d^k.$$

it follows from the theorem that these are also multiplicative.

Lemma 3.6.1 *The divisor functions $\tau(n)$, $\sigma(n)$, $\sigma_k(n)$ are all multiplicative.*

The multiplicativity of $\phi(n)$ was used in Section 2.4.3 to derive a closed form formula for $\phi(n)$ in terms of the standard prime decompositions. The same can be done for $\tau(n)$ and $\sigma(n)$.

Theorem 3.6.2 *Suppose that $n = p_1^{e_1} \cdots p_k^{e_k}$. Then*

$$\tau(n) = (e_1 + 1) \cdots (e_k + 1)$$
$$\sigma(n) = (\frac{p_1^{e_1+1} - 1}{p_1 - 1})(\frac{p_2^{e_2+1} - 1}{p_2 - 1}) \dots (\frac{p_k^{e_k+1}}{p_k - 1}).$$

Proof We will exhibit the proof for $\tau(n)$ and leave the derivation of $\sigma(n)$ for the exercises.

As in the derivation of the formula for $\phi(n)$ we establish the formula first for prime powers. The general result then follows from the multiplicativity.

Suppose then that $n = p^e$ and consider

$$\tau(n) = \sum_{d|n} 1.$$

The divisors of p^e are $1, p, p^2, \dots, p^e$ and hence

$$\tau(n) = \tau(p^e) = \sum_{i=0}^{e} 1 = (e + 1).$$

This proves the first part of the theorem. □

EXAMPLE 3.6.1 Compute $\tau(250)$ and $\sigma(250)$.

Now

$$\tau(250) = \tau(2 \cdot 5^3) = \tau(2)\tau(5^3) = 2 \cdot 4 = 8.$$

Hence 250 has 8 positive divisors namely $1, 2, 5, 5^2, 5^3, 2 \cdot 5, 2 \cdot 5^2, 2 \cdot 5^3$. Next

$$\sigma(250) = \frac{2^2 - 1}{2 - 1} \frac{5^4 - 1}{5 - 1} = (3)(156) = 468.$$

An extremely important arithmetic function is the Möbius function that we introduced in Section 3.3 and used in the proof of Dirchlet's theorem. Recall that the **Möbius function** is defined for natural numbers n by

$$\mu(n) = \begin{cases} 1 & \text{if } n = 1 \\ (-1)^r & \text{if } n = p_1 p_2 \cdots p_r \text{ with } p_1, \dots, p_r \text{ distinct primes} \\ 0 & \text{otherwise.} \end{cases}$$

Lemma 3.6.2 *The Möbius function $\mu(n)$ is multiplicative.*

Proof Suppose that $(n, m) = 1$. If either n or m is not squarefree then mn is not squarefree. Hence in this case $\mu(mn) = 0$ and either $\mu(m) = 0$ or $\mu(n) = 0$ so that

$$\mu(mn) = \mu(n)\mu(m).$$

Hence we may assume that both n and m are squarefree. Assume

$$n = p_1 \cdots p_k \text{ and } m = q_1 \cdots q_t$$

with each having distinct sets of prime factors. Then $\mu(n) = (-1)^k$ and $\mu(n) = (-1)^t$. Since the sets of prime factors are disjoint the prime decomposition for nm is

$$nm = p_1 \cdots p_k q_1 \cdots q_t.$$

Therefore

$$\mu(nm) = (-1)^{k+t} = (-1)^k(-1)^t = \mu(n)\mu(m).$$

□

Using the multiplicativity we obtain the following theorem.

Theorem 3.6.3 *For the Möbius function $\mu(n)$,*

$$\sum_{d|n} \mu(d) = \begin{cases} 1 & \text{if } n = 1 \\ 0 & \text{if } n > 1. \end{cases}$$

Proof Clearly if $n = 1$

$$\sum_{d|n} \mu(d) = 1.$$

Since $\mu(n)$ is multiplicative from Theorem 3.6.1

$$F(n) = \sum_{d|n} \mu(d)$$

is also multiplicative. Therefore we need only prove the result for prime powers.

Let $n = p^e$ with $e > 0$. Then the positive divisors of n are $1, p, \ldots, p^e$ and hence

$$\sum_{d|n} \mu(d) = \sum_{i=1}^{e} \mu(p^i).$$

However $\mu(p^i) = 0$ if $i > 1$ and so

$$\sum_{d|n} \mu(d) = \mu(1) + \mu(p) = 1 + (-1) = 0$$

completing the proof. □

This result allows us to prove the following very important theorem which has far-ranging applications.

Theorem 3.6.4 *(Möbius Inversion Formula) Suppose that $f(n)$ is an arithmetic function and*

$$F(n) = \sum_{d|n} f(d).$$

Then

$$f(n) = \sum_{d|n} \mu(d) F(\frac{n}{d}).$$

Conversely if $F(n)$ is an arithmetic function and

$$f(n) = \sum_{d|n} \mu(d) F(\frac{n}{d})$$

then

$$F(n) = \sum_{d|n} f(d).$$

Proof Consider

$$\begin{aligned}\sum_{d|n} \mu(d) F(\frac{n}{d}) &= \sum_{d|n} \sum_{k|\frac{n}{d}} f(k) \\ &= \sum_{dk|n} \mu(d) f(k).\end{aligned}$$

This last sum is taken over all ordered pairs (d, k) with $dk|n$. This is symmetric in (d, k) so we can reverse the roles of d and k to obtain

$$\sum_{d|n} \mu(d) F(\frac{n}{d}) = \sum_{k|n} f(k) \sum_{d|\frac{n}{k}} \mu(d).$$

From Theorem 3.6.3

$$\sum_{d|\frac{n}{k}} \mu(d) = 0 \text{ unless } \frac{n}{k} = 1.$$

This would imply that $k = n$ and hence the sum on the right hand side reduces to $f(n)$, completing the first part.

Retracing the steps exactly in the opposite direction will prove the converse (see the exercises). □

The Möbius inversion formula is a special case of an **inversion formula** in mathematics. These arise in many different areas. An important continuous example is the **Fourier Inversion Theorem**. Suppose that $f(x)$ is an integrable function over the whole real line. Its **Fourier transform** is defined as the complex valued function given by

$$\hat{f}(w) = \int_{-\infty}^{\infty} f(u)e^{-iwu}du.$$

Then

Theorem 3.6.5 *(The Fourier Inversion Theorem) If $f(x)$ is an integrable function and $\hat{f}(w)$ is its Fourier transform then*

$$f(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \hat{f}(w)e^{iwx}du.$$

This inversion theorem is used in the solution of partial differential equations and also can be used in a proof of the famous central limit theorem from Mathematical Statistics (see [Gr]). The Fourier transform is an example of an integral transform. We will see and use another such transform, the Mellin transform, in the proof of the prime number theorem.

3.7 Exercises

3.1 Show that for any real number x with $0 < x < 1$ we have

$$\ln(\frac{1}{1-x}) = \sum_{n=1}^{\infty}\frac{x^n}{n} < \sum_{n=1}^{\infty} x^n = \frac{x}{1-x}.$$

(Hint: For the first part consider the Taylor series for $\ln(1-x)$. Start with the sum of a geometric series $\frac{1}{1-x} = 1 + x + x^2 + \cdots$ and integrate.)

3.2 Show that the Fermat numbers F_1, F_2, F_3 are all prime but that F_4 is composite (divisible by 641).

3.3 Prove that suppose (a_n) is any sequence of integers with $(a_n, a_m) = 1$ if $n \neq m$. Then there exist infinitely many primes.

3.4 If $A_n = a^{2^n} + 1$ then prove:
(a) If $n > m \geq 1$ then $(A_m|(A_n - 2)$,

(b) $(A_n, A_m) = 1$ if $n \neq m$ and a is even,
(b) $(A_n, A_m) = 2$ if $n \neq m$ and a is odd.

3.5 Determine using the same types of methods used to find the value of the golden section the value of

$$\sqrt{1+\sqrt{1+\sqrt{1+\cdots}}}.$$

3.6 Recall from Section 3.2.5 that a continued fraction is defined in the following way:

Let $a_0, a_1, \cdots, a_n$ be a finite sequence of integers all positive except possibly a_0. Then a **finite simple continued fraction** is the rational number defined by

$$a_0 + \cfrac{1}{a_1 + \frac{1}{a_2 + \ddots}}.$$

If $a_0, a_1, \ldots, a_n, \ldots$ is an infinite sequence of integers all positive except possibly a_0, then a **infinite simple continued fraction** is determined by the limit of the finite simple continued fractions formed up to a_n. Each of the finite simple continued fractions is called a **convergent** of the infinite simple continued fraction.

Find the values of the following infinite continued fractions.
(a) $a_n = 3$ for all n.
(b) $(a_n) = (1, 2, 1, 2, 1, 2, \ldots)$.

3.7 Prove Lemmas 3.1.6 and 3.1.7, that is prove:
(a) $f_n f_{n+1} = f_1^2 + f_2^2 + \cdots + f_n^2, n \geq 1$.
(b) $f_n^2 - f_{n-1} f_{n+1} = (-1)^n, n \geq 1$.
where f_n are the Fibonacci numbers.

3.8 Prove Lemma 3.1.8, that is, prove

$$f_{n+m} = f_{n-1} f_m + f_n f_{m+1}, n \geq 1,$$

where f_n are the Fibonacci numbers.

3.9 Prove
(a) $p | f_{p+1}$ if $p \equiv \pm 3 \bmod 10$ with p prime
(b) $p | f_{p-1}$ if $p \equiv \pm 1 \bmod 10$ with p prime
(Hint: Use the identities in the proof of Theorem 3.1.10.)

3.10 The real Chebyshev polynomials of the second kind can be defined by

$$S_0(x) = 0, S_1(x) = 1, S_{n+1}(x) = x S_n(x) - S_{n-1}(x)$$

Prove that

(a) If $x > 2, \quad x = 2\cos\theta < 2$ then

$$S_n(x) = \frac{\sin(n\theta)}{\sin\theta}.$$

(b) If $x \geq 2, \quad x = 2\cosh\theta > 2$ then

$$S_n(x) = \frac{\sinh(n)\theta)}{\sinh\theta}.$$

(c) If $x = 2$ then

$$S_n(x) = n.$$

(Hint: Use induction and trigonometric identities.)

3.11 Prove directly that there exists infinitely many primes of the form $8n + 3$.

3.12 Classify the pythagorean triples where the hypotenuse differs by one from one of the legs.

3.13 Show that given integers x_0, n with $x_0^2 \equiv -1 \bmod n$ then there exist integers y, b with $(y, b) = 1, 0 < b \leq \sqrt{n}$ and

$$|-\frac{x_0}{n} - \frac{y}{b}| < \frac{1}{b\sqrt{n}}.$$

3.14 Show that the number of representations of $m > 1$ as a sum $m = a^2 + b^2$ with $(a, b) = 1$ is equal to the number of solutions of

$$x^2 \equiv -1 \bmod m.$$

3.15 Determine the set of integers represented by the quadratic forms

(a) $f(x, y) = 2x^2 + 2y^2$

(b) $f(x, y) = 2x^2 - 2y^2$

3.16 Show that a projective matrix (see Section 3.2.3) $X \in PSL(2, \mathbb{Z})$ has order 2 if and only if its trace is zero.

3.17 If G is any group, its **center**, denoted by $Z(G)$ consists of those elements of G which commute with all elements of G;

$$Z(G) = \{g \in G; gh = hg, \forall h \in G\}.$$

Prove that $Z(G)$ is a normal subgroup of G.

3.18 Prove parts (1) and (2) of Lemma 3.3.5. That is prove that

(a) If χ_1 and χ_2 are characters then so is $\chi_1\chi_2$ where $(\chi_1\chi_2)(a) = \chi_1\underline{(a)\chi_2}(a)$.
(b) If χ is a character so is its complex conjugate $\overline{\chi}$. Further $\chi(a)^{-1} = \overline{\chi(a)}$.

3.19 Prove that if a is an odd integer and $t > 2$ then

$$a \equiv (-1)^{\frac{a-1}{2}} 5^b \bmod 2^t \text{ for some } b \geq 0$$

(Hint: Separate into two cases where $a \equiv 1 \bmod 4$ and $a \equiv 3 \bmod 4$. Then use the facts that 5^b represents exactly 2^{t-2} numbers incongruent mod 2^t and that 5^b is periodic mod 2^t with period 2^{t-2}.)

3.20 Fill in the details of the proof of the second part of Theorem 3.3.2. That is prove that if $a > 0$ is an integer and χ runs over the set of all $\phi(k)$ characters then

$$\sum_{\chi} \chi(t)\overline{\chi(a)} = \begin{cases} \phi(k) & \text{if } a \equiv t \bmod k \\ 0 & \text{if } a \not\equiv t \bmod k. \end{cases}$$

3.21 Consider the van Mangoldt function $\Lambda(n)$ defined for positive integers by

$$\Lambda(n) = \begin{cases} \ln p & \text{if } n = p^c, c \geq 1 \\ 0 & \text{for all other } n > 0 \end{cases}.$$

Prove that

$$\sum_{d|n} \Lambda(d) = \ln n.$$

3.22 Let χ be a real character mod k and define $f(n) = \sum_{d|n} \chi(d)$. Prove that $f(n) \geq 0$ for all $n \geq 1$ and $f(n) \geq 1$ if $n = c^2$ a square.

3.23 Prove Lemma 3.4.1, that is, prove: The integer 5 is the only prime appearing in two different twin prime pairs.

3.24 Prove Lemma 3.4.2, that is, prove: There is a one-to-one correspondence between twin prime pairs and those integers $n \geq 4$ for which $n^2 - 1$ has only four possible positive divisors.

3.25 Prove Lemma 3.4.3, that is prove: Suppose p, q are primes. Then $pq + 1$ is a square if and only if p and q are twin primes.

3.26 Prove Lemma 3.4.4, that is prove: If p, q are twin primes greater than 3 then $p+q$ is divisible by 12.

3.27 Prove that the divsior functions $\tau(n)$, $\sigma(n)$, $\sigma_k(n)$ are all multiplicative. (Fill in the details of the proof of Lemma 3.6.1.)

3.28 Prove that if $\sigma(n)$ is the sum of the positive divisors of n and $n = p_1^{e_1} \cdots p_k^{e_k}$ then

$$\sigma(n) = (\frac{p_1^{e_1+1} - 1}{p_1 - 1})(\frac{p_2^{e_2+1} - 1}{p_2 - 1}) \ldots (\frac{p_k^{e_k+1}}{p_k - 1})$$

(see Theorem 3.6.2.)

3.29 Compute $\tau(n)$ and $\sigma(n)$ for $n = 105, 72, 788$.

3.30 Prove that if $F(n)$ is an arithmetic function and

$$f(n) = \sum_{d|n} \mu(d) F(\frac{n}{d})$$

then

$$F(n) = \sum_{d|n} f(d).$$

3.31 Prove that for real numbers x, y with $0 < x < 1$ we have the inequality

$$(1-x)^3 |1 - xe^{iy}|^4 |1 - xe^{2iy}|^2 < 1.$$

3.32 Suppose that $f(n)$ and $g(n)$ are multiplicative arithmetic functions. Show that $F(n) = f(n)g(n)$ is also multiplicative.

3.33 Show that a natural number p is a prime if and only if $\sigma(p) = p + 1$.

3.34 Use the multiplicativity to derive a formula for $\sigma_k(n)$ the sum of the kth powers of the positive divisors of n.

3.35 Prove Theorem 3.2.4 by using the Möbius inversion formula. (Hint: First prove part (3) directly.) A group theoretic proof is in [KR 2].

Chapter 4
The Density of Primes

4.1 The Prime Number Theorem—Estimates and History

As we have seen, and proved in many different ways, there are infinitely many primes. In fact, as Dirichlet's Theorem shows, there are infinitely many primes in any arithmetic progression $an + b$ with $(a, b) = 1$. However, an examination of the list of positive integers shows that the primes become scarcer as the integers increase. This statement was quantified in Theorem 2.3.2, where we proved that there are arbitrarily large spaces or gaps within the sequence of primes. As a result of these observations the question arises concerning the distribution or density of the primes. The interest centers here on the **prime number function** $\pi(x)$ defined for positive integers x by

$$\pi(x) = \text{ number of primes } \leq x.$$

Clearly $\pi(x) \to \infty$ as $x \to \infty$ so the appropriate question on the distribution of primes is what is the rate of growth of this function. The **Prime Number Theorem** asserts that asymptotically $\pi(x)$ is given by $\frac{x}{\ln x}$. **Asymptotically** means as x goes to ∞. It has been touted as one of the most surprising results in mathematics given that it ties together the primes and the natural logarithm function in a simple way that is most unexpected. The proof of the prime number theorem, or more precisely the attempted proof by Riemann, is really considered the beginnings of modern **analytic number theory**. This refers to the use of analytic methods, especially complex analysis, in the study of number theory. However, as we saw relative to Dirichlet's theorem, the use of hard analysis actually precedes Riemann's work.

The prime number theorem was originally conjectured by both Gauss and Legendre although Euler also surmised the result. Gauss looked at the list of primes less than 3,000,000 and noticed that the prime number function is given very closely by the function $Li(x)$ which is defined by the integral

$$Li(x) = \int_2^x \frac{1}{\ln t} dt.$$

B. Fine and G. Rosenberger, *Number Theory*,
DOI 10.1007/978-3-319-43875-7_4

Gauss' observation was then that

$$\pi(x) \approxeq Li(x).$$

If integration by parts is used on the integral defining $Li(x)$ and we take the limit as $x \to \infty$ it is clear that this integral is asymptotically $\frac{x}{\ln x}$. Hence, Gauss's observation is then that (see Definition 4.2.2)

$$\lim_{x\to\infty} \frac{\pi(x)}{x/\ln x} = 1.$$

This is the prime number theorem which we now state formally.

Theorem 4.1.1 *(Prime Number Theorem) If $\pi(x)$ is the prime number function then*

$$\lim_{x\to\infty} \frac{\pi(x)}{x/\ln x} = 1.$$

Legendre, (actually published a bit earlier than Gauss), by looking at the list of primes up to 1,000,000 came up with a slightly different formula:

$$\pi(x) \approxeq \frac{x}{\ln x - 1.08366}.$$

Legendre's estimate is also asymptotically $\frac{x}{\ln x}$. Neither Gauss nor Legendre gave a proof of the prime number theorem nor an indication of how they arrived at their estimates. However in hindsight a possible explanation is as follows. Looking at tables of $\pi(10^n)$ it is observed that as n changes by 2 the ratio $\frac{x}{\pi(x)}$ changes by an almost constant amount 4.6 which is $2\ln(10)$. This would suggest that $\frac{10^n}{\pi(10^n)} \approxeq \ln(10^n)$. The figures are as below

x	10^2	10^4	10^6	10^8	10^{10}	10^{12}
$\pi(x)$	25	1229	78498	5761455	455052511	37607912018
$\frac{x}{\pi(x)}$	4.000	8.137	12.739	17.357	21.975	26.590
$\ln(x)$	4.605	9.210	13.816	18.421	23.026	27.361
$\frac{\ln(x)}{x/\pi(x)}$	1.151	1.132	1.085	1.061	1.048	1.039

The first real attempt to prove the prime number theorem was done by Chebyshev in 1848. He proved that there exist constants A_1 and A_2 with $.922 < A_1 < 1$ and $1 < A_2 < 1.105$ such that

$$A_1 < \frac{\pi(x)}{x/\ln(x)} < A_2.$$

Further he proved that if $\frac{\pi(x)}{x/\ln x}$ had a limit it would have to be 1. However, he could not prove that the function in the middle actually tends to a limit. In proving this result Chebyshev used the **Riemann zeta function**

$$\zeta(s) = \sum_{n=1}^{\infty} \frac{1}{n^s}$$

where $s > 1$ is a real variable. This function was introduced originally by Euler who used it to give a proof of the infinitude of primes (see Section 3.1.2). This was really the first use of analysis in number theory.

Chebyshev's inequality has been improved upon many times. Sylvester in 1882 improved it to $A_1 = .95695$ and $A_2 = 1.04423$ for sufficiently large x. It can now be shown that for all $x > 10$, $A_1 = 1$ can be used.

In 1859, Riemann attempted to give a complete proof of the prime number theorem using the zeta function for complex variables s. Although he was not successful in proving the prime number theorem he established many properties of the zeta function and showed that the prime number theorem depended on the zeros of the zeta function. He conjectured that all the zeros of $\zeta(s)$ in the strip $0 \leq Re(s) \leq 1$ are along the line $Re(s) = \frac{1}{2}$. This is known as the **Riemann hypothesis** and is still an open problem. We will discuss both the Riemann zeta function and the Riemann hypothesis is Section 4.4. In 1896, building on the work of Riemann, Hadamard, and independently C. de la Vallee Poussin proved the prime number theorem. Their proof relied heavily on complex analysis. It was felt for a long time that the prime number theorem was at least as complicated as the theory of complex variables. Most mathematicians doubted that a proof which did not heavily rely on the theory of analytic functions could be found. However in 1949, Selberg and later Erdos came up with an *elementary proof* of the prime number theorem. This proof is actually harder than the analytic proof but is elementary in that it does not use any complex analysis.

Although the proof of the prime number theorem is really considered the beginnings of analytic number theory we have seen that the use of analysis in proving results in number theory was done earlier. Euler introduced the zeta function in giving a proof that there are infinitely many primes. We presented this proof in Chapter 3. In his proof though the analysis was relatively easy. The first hard use of analysis was used by Dirichlet to prove **Dirichlet's theorem**. As we exhibited in Chapter 3, there are many special cases of this result that can be proved by very elementary methods. However no proof of the complete result is known without analysis.

Given that the prime number theorem has been established many other questions concerning it can be raised. First of all notice that if a is any constant then

$$\frac{x}{\ln x} \approxeq \frac{x}{\ln x - a} \text{ if } x \text{ were large.}$$

Hence the prime number theorem is equivalent to

$$\lim_{x \to \infty} \frac{\pi(x)}{x / \ln x - a} = 1$$

for any constant a. The question arises as to whether there is an optimal value for a. Empirical evidence is that $a = 1$ is an optimal choice and generally better for large x than Legendre's 1.08366 and better than Gauss' $Li(x)$. The table below compares the estimates.

x	$\pi(x)$	$\frac{x}{\ln x}$	$Li(x)$	$\frac{x}{\ln x - 1.08366}$	$\frac{x}{\ln x - 1}$
10^3	168	145	178	172	169
10^4	1229	1086	1246	1231	1218
10^5	9592	8686	9630	9588	9512
10^6	78498	72382	78628	78534	78030
10^7	664579	620420	664918	665138	661459
10^8	5761455	5428681	5762209	5769341	5740304

Observing the table above it is noticed that $Li(x) > \pi(x)$. The question arises as to whether this is always true. Littlewood in 1914 [Li] proved that $\pi(x) - Li(x)$ assumes both positive and negative values infinitely often. Te Riele in 1986 [Re] showed that there are greater than 10^{180} consecutive integers for which $\pi(x) > Li(x)$ in the range $6.62 \times 10^{370} < x < 6.69 \times 10^{370}$.

The prime number function $\pi(x)$ and the prime number theorem answer the basic questions concerning the density of primes. A related question concerns the function

$$p(n) = p_n$$

where p_n is the nth prime. That is the question of whether there is a closed form function which estimates the nth prime. The answer to this is yes and turns out to be equivalent to the prime number theorem. We state it below.

Theorem 4.1.2 *The nth prime p_n is given asymptotically by*

$$p_n \approxeq n \ln n.$$

Proof From the prime number theorem we have that $\pi(x) \approxeq \frac{x}{\ln x}$. Let

$$y = \frac{x}{\ln x} \implies \ln y = \ln x - \ln \ln x.$$

But $\ln \ln x$ is asymptotically small compared to $\ln x$ and hence

$$\ln y \approxeq \ln x.$$

Now

$$x = y \ln x \approxeq y \ln y.$$

This shows that the inverse function to $\frac{x}{\ln x}$ is asmptotically $x \ln x$. But by the prime number theorem this is asmptotically the inverse function of $\pi(x)$. □

Notice that if we had started with Theorem 4.1.2 we could have recovered the prime number theorem.

4.2 Chebyshev's Estimate and Some Consequences

The first significant progress in developing a proof of the prime number theorem was obtained by Chebyshev in 1848. He proved that the functions $\pi(x)$ and $\frac{x}{\ln x}$ are of the **same order of magnitude**, a concept we will explain in detail below and that if $\lim_{x\to\infty} \frac{\pi(x)}{x/\ln x}$ existed then the limit would have to be 1. At first glance it appeared that he was quite close to a proof of the prime number theorem. However, it would take another 50 years and the development of some completely new ideas from complex analysis to actually accomplish this. A proof, along the lines of Chebyshev's methods, without recourse to complex analysis, would not be done until the work of Selberg and Erdos in the late 1940's (see [N]).

Chebyshev proved the following result, now known as **Chebyshev's estimate**.

Theorem 4.2.1 *There exist positive constants A_1 and A_2 such that*

$$A_1 \frac{x}{\ln x} < \pi(x) < A_2 \frac{x}{\ln x}$$

for all $x \geq 2$.

The proof we will give is somewhat simpler than that of Chebyshev. The constants we arrive at in the proof given below are sufficient but nowhere near best possible. We will say more about this at the conclusion of the proof.

The proof depends on some properties and inequalities involving the **binomial coefficients** $\binom{n}{k}$. We have used these numbers in several instances in previous sections but here we begin by formally defining them and then reviewing some of their basic properties.

Definition 4.2.1 *Given nonnegative integers n, k with $n \geq 1$ and $n \geq k$ the* **binomial coefficient** $\binom{n}{k}$ *is defined as*

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}$$

Note that by convention $0! = 1$.

The first several results outline standard properties of the binomial coefficients and proofs can be found in any book on probability and statistics. We also outline proofs in the exercises.

Lemma 4.2.1 $\binom{n}{k}$ *represents the number of ways of choosing k objects out of n without replacement and without order.*

Clearly the number of ways of choosing k objects out of n objects also counts the number of possible subsets of size k in a finite set with n elements.

Corollary 4.2.1 $\binom{n}{k}$ = *the number of subsets of size k in a finite set with n elements.*

Lemma 4.2.2 *(The Binomial Theorem) For any real numbers a, b, and natural number n we have*

$$(a+b)^n = \sum_{k=0}^{n} \binom{n}{k} a^k b^{n-k}$$

Letting $a = b = 1$ in the Binomial Theorem we get

Corollary 4.2.2 $(1+1)^n = 2^n = \sum_{k=0}^{n} \binom{n}{k}$. *In particular* $\binom{n}{k} < 2^n$ *for all k with* $0 \le k \le n$.

Combining Corollaries 4.2.1 and 4.2.2 we obtain the well-known result that the number of subsets of a set with n element is 2^n. Consider a set with n elements. Then

$$\text{total number of subsets} =$$

$$\text{number of subsets of size } 0 + \cdots + \text{ number of subsets of size } n$$

$$= \binom{n}{0} + \binom{n}{1} + \cdots + \binom{n}{n} = \sum_{k=0}^{n} \binom{n}{k} = 2^n$$

Lemma 4.2.3 $\binom{n}{k} + \binom{n}{k-1} = \binom{n+1}{k}$.

This last lemma is the basis of **Pascal's Triangle** in which each row consists of the set of binomial coefficients for that numbered row.

$$\begin{array}{ccccccccccc}
 & & & & & 1 & & & & & \\
 & & & & 1 & & 1 & & & & \\
 & & & 1 & & 2 & & 1 & & & \\
 & & 1 & & 3 & & 3 & & 1 & & \\
 & 1 & & 4 & & 6 & & 4 & & 1 & \\
1 & & 5 & & & \ldots & & & 5 & & 1
\end{array}$$

Each subsequent row is formed by placing a one on the outside and each subsequent number is placed between 2 numbers in the previous row and is their sum. For example,

$$\begin{array}{ccccccccc}
 & 1 & & 3 & & 3 & & 1 & \\
1 & & 4 & & 6 & & 4 & & 1
\end{array}$$

Since

$$1+3=4, 3+3=6, 3+1=4.$$

The final standard idea we will need is that of **Stirling's approximation** (see Section 3.1.6)

$$n! \approxeq \sqrt{2\pi n}(\frac{n}{e})^n.$$

For Chebyshev's estimate we need the following results which are deeper and use number theory. $\pi(n)$ in the lemma is the prime number function.

Lemma 4.2.4 *(i)* $n^{\pi(2n)-\pi(n)} < \binom{2n}{n} \leq (2n)^{\pi(2n)}$,
(ii) $2^n \leq \binom{2n}{n} \leq 2^{2n}$.

Proof If p is a prime let e_p be the highest power such that $p^{e_p}|n!$. Then by an easy induction we have

$$e_p = \sum_{i=1}^{t_p}[\frac{n}{p^i}]$$

where [ ] is the greatest integer function and t_p is the first integer such that $p^{t_p+1} > n$. Clearly such a t_p exists for each prime p. Now consider

$$\binom{2n}{n} = \frac{(2n)!}{n!n!} = \frac{(2n)(2n-1)\cdots(n+1)}{n!} = \prod_{j=1}^{n}(\frac{n+j}{j}).$$

Given a prime p, let m_p be the highest power such that $p^{m_p}|\binom{2n}{n}$. From the observation above

$$m_p = \sum_{i=1}^{k_p}([\frac{2n}{p^i}] - 2[\frac{n}{p^i}])$$

where here k_p is the first integer such that $p^{k_p+1} > 2n$.

If $1 \leq i \leq k_p$ then

$$[\frac{2n}{p^i}] - 2[\frac{n}{p^i}] < \frac{2n}{p^i} - 2(\frac{n}{p^i} - 1) = 2.$$

Since $[\frac{2n}{p^i}]$ and $2[\frac{n}{p^i}]$ are integers it follows that

$$[\frac{2n}{p^i}] - 2[\frac{n}{p^i}] \leq 1$$

if $1 \leq i \leq k_p$. This then implies that

$$m_p = \sum_{i=1}^{k_p}([\frac{2n}{p^i}] - 2[\frac{n}{p^i}]) \leq \sum_{i=1}^{k_p} 1 = k_p.$$

Therefore

$$\binom{2n}{n} | \prod_{p \leq 2n} p^{k_p}$$

and hence

$$\binom{2n}{n} \leq \prod_{p \leq 2n} p^{k_p} \leq \prod_{p \leq 2n} (2n) = (2n)^{\pi(2n)}$$

giving one side of the first inequality.

On the other hand if $n < p \leq 2n$ then $p|(2n)!$ but p does not divide $n!$. It follows that

$$\prod_{n<p\leq 2n} p | \binom{2n}{n} \implies \prod_{n<p\leq 2n} p \leq \binom{2n}{n}.$$

Now

$$\prod_{n<p\leq 2n} p > \prod_{n<p\leq 2n} n = n^{\pi(2n)-\pi(n)}$$

since there are $\pi(2n) - \pi(n)$ primes in the range $p < n \leq 2n$. Therefore

$$n^{\pi(2n)-\pi(n)} < \binom{2n}{n}$$

establishing the other side of the first inequality.

For the second inequality we have

$$\binom{2n}{n} \leq (1+1)^{2n} = 2^{2n}$$

and from above

$$\binom{2n}{n} = \prod_{j=1}^{n}(\frac{n+j}{j}) \geq \prod_{j=1}^{n} 2 = 2^n.$$

Therefore

$$2^n \leq \binom{2n}{n} \leq 2^{2n}$$

establishing the second inequality.

□

We now give the proof of Chebyhsev's estimate.

Proof (Theorem 4.2.1) We have to show that there exist positive constants A_1 and A_2 such that

$$A_1 \frac{x}{\ln x} < \pi(x) < A_2 \frac{x}{\ln x}$$

for all $x \geq 2$.

From the previous lemma we have the inequalities

$$n^{\pi(2n)-\pi(n)} < \binom{2n}{n} \leq (2n)^{\pi(2n)}$$

$$2^n \leq \binom{2n}{n} \leq 2^{2n}.$$

Hence

$$n^{\pi(2n)-\pi(n)} < 2^{2n} \implies (\pi(2n) - \pi(n)) \ln n \leq 2n \ln 2$$

$$\implies \pi(2n) - \pi(n) \leq \frac{2n \ln 2}{\ln n}.$$

On the other hand

$$(2n)^{\pi(2n)} \geq 2^n \implies \pi(2n) \geq \frac{n \ln 2}{\ln(2n)}.$$

For a real variable $x \geq 2$ let $2n$ be the greatest even integer not exceeding x, so that $x \geq 2n$, $n \geq 1$ and $x < 2n + 2$. Then

$$\pi(x) \geq \pi(2n) \geq \frac{n \ln 2}{\ln(2n)} \geq \frac{n \ln 2}{\ln x}$$

$$\geq \frac{(2n+2) \ln 2}{4 \ln x} > \frac{\ln 2}{4} \frac{x}{\ln x}.$$

Therefore

$$\pi(x) > A_1 \frac{x}{\ln x}$$

for all $x \geq 2$ with $A_1 = \frac{\ln 2}{4}$.

To find the existence of A_2 let $2n = 2^t$ with $t \geq 3$. Then

$$\pi(2^t) - \pi(2^{t-1}) \leq \frac{2^t \ln 2}{(t-1) \ln 2} = \frac{2^t}{t-1}.$$

Consider the telescoping sum

$$\sum_{t=3}^{2j}(\pi(2^t) - \pi(2^{t-1})) = \pi(2^{2j}) - \pi(4).$$

Since $\pi(4) \leq 4 = \frac{2^2}{2-1}$ and $\pi(2^t) - \pi(2^{t-1}) \leq \frac{2^t}{t-1}$ we obtain using the telescoping sum that

$$\pi(2^{2j}) < \sum_{t=2}^{2j} \frac{2^t}{t-1} = \sum_{t=2}^{j} \frac{2^t}{t-1} + \sum_{t=j+1}^{2j} \frac{2^t}{t-1}.$$

Now

$$\sum_{t=2}^{j} \frac{2^t}{t-1} < \sum_{t=2}^{j} 2^t < 2^{j+1}$$

and

$$\sum_{t=j+1}^{2j} \frac{2^t}{t-1} \leq \sum_{t=j+1}^{2j} \frac{2t}{j} \leq \frac{1}{j}2^{2j+1}.$$

It follows that

$$\pi(2^{2j}) < 2^{j+1} + \frac{1}{j}2^{2j+1}.$$

Since $j < 2^j$ we have $2^{j+1} < \frac{2^{2j+1}}{j}$ and therefore for $j \geq 2$

$$\pi(2^{2j}) < 2(\frac{2^{2j+1}}{j}).$$

This implies that

$$\frac{\pi(2^{2j})}{2^{2j}} < \frac{4}{j} \text{ for all } j \geq 2.$$

Let $x \geq 2$ be a real variable. Then there exists an integer $j \geq 1$ such that $2^{2j-2} < x \leq 2^{2j}$. Hence

$$\frac{\pi(x)}{x} \leq \frac{\pi(2^{2j})}{2^{2j-2}} = \frac{4\pi(2^{2j})}{2^{2j}}.$$

Further

$$2j \geq \frac{\ln x}{\ln 2} \implies \frac{4}{j} \leq \frac{8 \ln 2}{\ln x}.$$

Implying the inequality for $\frac{\pi(2^{2j})}{2^{2j}}$ gives

$$\frac{\pi(2^{2j})}{2^{2j}} < \frac{4}{j} \implies \frac{\pi(x)}{x} < \frac{16}{j} \leq \frac{32 \ln 2}{\ln x}$$

$$\implies \pi(x) < (32\ln 2)\frac{x}{\ln x}$$

for all $x \geq 2$. Therefore

$$\pi(x) < A_2\frac{x}{\ln x}$$

for all $x \geq 2$ with $A_2 = 32\ln 2$ establishing Chebyshev's estimates.

□

We mention again that the proof is somewhat simpler than that originally given by Chebyshev and arrives at weaker constants. We obtained $A_1 = \frac{\ln 2}{4}$ and $A_2 = 32\ln 2$ which were sufficient for the theorem but nowhere near best possible. Chebyshev showed that $A_1 = .922$ and $A_2 = 1.105$ could be used. His proof actually involved a careful analysis of a form of Stirling's approximation. The values in the constants in Chebyshev's inequality have been improved upon many times. Sylvester in 1882 improved the values to $A_1 = .95695$ and $A_2 = 1.04423$ for sufficiently large x. It can now be shown that for all $x > 10$, $A_1 = 1$ can be used.

This following is an immediate corollary of the estimate, independent of the values of A_1 and A_2.

Corollary 4.2.3 $\frac{\pi(x)}{x} \to 0$ *as* $x \to \infty$.

Proof From Chebyshev's estimate we have

$$0 < \pi(x) < A_2\frac{x}{\ln x} \implies 0 < \frac{\pi(x)}{x} < \frac{A_2}{\ln x}.$$

Since A_2 is a constant $\frac{A_2}{\ln x} \to 0$ as $x \to \infty$ so clearly $\frac{\pi(x)}{x} \to 0$ also.

□

This corollary says that the primes become relatively scarcer as x gets larger. In probabilistic terms it says that the probability of randomly choosing a prime less than or equal to x goes to zero as x goes to infinity.

Before continuing and presenting some consequences of Chebyshev's result we introduce a convenient notation for describing the order of magnitude of a function.

Definition 4.2.2 *Suppose $f(x)$, $g(x)$ are positive real-valued functions. Then*

(1) $f(x) = O(g(x))$ (read $f(x)$ is big O of $g(x)$) if there exists a constant A independent of x and an x_0 such that

$$f(x) \leq Ag(x) \text{ for all } x \geq x_0.$$

(2) $f(x) = o(g(x))$ (read $f(x)$ is little o of $g(x)$) if

$$\frac{f(x)}{g(x)} \to 0 \text{ as } x \to \infty.$$

In other words $g(x)$ is of a **higher order of magnitude** *than $f(x)$.*

(3) If $f(x) = O(g(x))$ and $g(x) = O(f(x))$, that is, there exist constants A_1, A_2 independent of x and an x_0 such that

$$A_1 g(x) \le f(x) \le A_2 g(x) \textit{ for all } x \ge x_0,$$

then we say that $f(x)$ and $g(x)$ are of the **same order of magnitude** *and write*

$$f(x) \approxeq g(x).$$

(4) If

$$\frac{f(x)}{g(x)} \to 1 \textit{ as } x \to \infty$$

then we say that $f(x)$ and $g(x)$ are **asymptotically equal** *and we write*

$$f(x) \sim g(x).$$

In general, we write $O(g)$ or $o(g)$ to signify an unspecified function f such that $f = O(g)$ or $f = o(g)$. Hence, for example, writing $f = g + o(x)$ means that $\frac{f-g}{x} \to 0$ and saying that f is $o(1)$ means that $f(x) \to 0$ as $x \to \infty$.

It is clear that being $o(g)$ implies being $O(g)$ but not necessarily the other way around. Further it is easy to see that

$$f \sim g \text{ is equivalent to } f = g + o(g) = g(1 + o(1)).$$

In terms of the notation above Chebyshev's estimate can be expressed as

$$\pi(x) \approxeq \frac{x}{\ln x}.$$

Further the prime number theorem can be expressed by

$$\pi(x) \sim \frac{x}{\ln x}$$

or equivalently

$$\pi(x) = \frac{x}{\ln x}(1 + o(1)).$$

We will use this notation freely as we develop the proof of the prime number theorem.

We now present some consequences of Chebyshev's estimate. It was mentioned at the end of the previous section that the prime number theorem is equivalent to $p_n \sim n \ln n$, where p_n denotes the nth prime (Theorem 4.1.2). Chebyshev's estimate gives immediately that p_n and $n \ln n$ are of the same order of magnitude.

Theorem 4.2.2 *There exist positive constants B_1, B_2 such that*

$$B_1 n \ln n \leq p_n \leq B_2 n \ln n.$$

Equivalently

$$p_n \approxeq n \ln n.$$

Proof Let p_n be the nth prime. Then clearly $\pi(p_n) = n$. From Chebyshev's estimate

$$n = \pi(p_n) \leq A_2 \frac{p_n}{\ln p_n} \text{ for all } n \geq 2.$$

This implies

$$\frac{1}{A_2} n \ln p_n \leq p_n \text{ for all } n \geq 2.$$

However, $p_n > n$ so

$$\frac{1}{A_2} n \ln n < \frac{1}{A_2} n \ln p_n \leq p_n \text{ for all } n \geq 2.$$

Therefore In general we write

$$B_1 n \ln n \leq p_n$$

for all $n \geq 2$ with $B_1 = \frac{1}{A_2}$.

In the other direction, we have

$$n = \pi(p_n) \geq A_1 \frac{p_n}{\ln p_n}.$$

Since $p_n > n$ it follows that $\frac{\ln p_n}{\sqrt{p_n}} \to 0$ as $n \to \infty$. Therefore, there exists a constant k such that

$$\frac{\ln p_n}{\sqrt{p_n}} < A_1 \text{ if } n > k.$$

Hence

$$n \frac{\ln p_n}{p_n} \geq A_1 > \frac{\ln p_n}{\sqrt{p_n}} \text{ if } n > k.$$

It follows that $n > \sqrt{p_n}$ and so $\ln p_n < 2 \ln n$ if $n > k$. Let

$$B_2 = max\{\frac{2}{A_1}, \frac{p_2}{2 \ln 2}, \frac{p_3}{3 \ln 3}, \ldots, \frac{p_{k-1}}{(k-1) \ln(k-1)}\}.$$

Then

$$p_n \leq B_2 n \ln n \text{ for all } n \geq 2.$$

□

Note that we could have proved Theorem 4.2.2 and then deduced Chebyshev's estimate from it. This result also provides a very simple proof of Euler's Theorem given in Chapter 3 that the series $\sum_p \frac{1}{p}$ diverges.

Corollary 4.2.4 *The sum*

$$\sum_{p, prime} \frac{1}{p}$$

diverges.

Proof For $n \geq 2$ we have $\frac{1}{p_n} \leq \frac{1}{B_1 n \ln n}$ from the last theorem. However the series $\sum_{n=1}^{\infty} \frac{1}{n \ln n}$ diverges by the integral test. □

Although there are infinitely many primes and $\sum_p \frac{1}{p}$ diverges it still diverges very slowly. Using the methods applied in the proof of Chebyshev's estimate we can actually bound the growth of the series of reciprocals of the primes.

Theorem 4.2.3 *There exists a constant k such that*

$$\sum_{2<p\leq x} \frac{1}{p} < k \ln \ln x \textit{ if } x > 3.$$

Proof From Theorem 4.2.2 we have

$$p_n \geq B_1 n \ln n.$$

Therefore

$$\sum_{2<p\leq x} \frac{1}{p} = \sum_{n=2}^{\pi(x)} \frac{1}{p_n} < \sum_{n=2}^{\pi(x)} \frac{1}{B_1 n \ln n} < \frac{1}{B_1} \sum_{n=2}^{[x]} \frac{1}{n \ln n}.$$

However

$$\frac{1}{n \ln n} = \int_{n-1}^{n} \frac{dt}{n \ln n} \leq \int_{n-1}^{n} \frac{dt}{t \ln t}$$

since $\frac{1}{n \ln n} \leq \frac{1}{t \ln t}$ on $[n-1, n]$ if $n \geq 3$. Then

$$\sum_{2<p\leq x} \frac{1}{p} < \frac{1}{B_1} \sum_{n=2}^{[x]} \frac{1}{n \ln n} \leq \frac{1}{2B_1 \ln 2} + \frac{1}{B_1} \sum_{n=3}^{[x]} \int_{n-1}^{n} \frac{dt}{t \ln t}$$

$$\leq \frac{1}{2B_1 \ln 2} + \frac{1}{B_1} \int_{2}^{x} \frac{dt}{t \ln t} = \frac{1}{2B_1 \ln 2} + \frac{1}{B_1} \ln \ln x - \frac{1}{B_1} \ln \ln 2$$

$$= \frac{1}{B_1} \ln\ln x + C < k \ln\ln x$$

taking k large enough. □

In a similar vein we get the following result which bounds the product of all the primes p less than some given x.

Theorem 4.2.4 *If $x \geq 2$ then $\prod_{p \leq x} p < 4^x$.*

Proof The theorem is clear for $2 \leq x < 3$. Suppose the theorem is true for an odd integer n with $n \geq 3$. Then it is true for $n \leq x < n + 2$ since

$$\prod_{p \leq x} p = \prod_{p \leq n} p < 4^n < 4^x.$$

Therefore it is sufficient to prove the theorem for odd integers n. We do an induction on the odd integers. The theorem is true for $n = 3$ and so we assume that it is true for all odd integers less than or equal to $n \geq 5$. Let $k = \frac{n+1}{2}$ or $k = \frac{n-1}{2}$ chosen so that k is also odd. Then $k \geq 3$ and $n - k$ is even. Further $n - k = 2k \pm 1 - k \leq k + 1$. If p is a prime with $k < p \leq n$ then $p|n!$ but p does not divide either $k!$ or $(n - k)!$. Therefore $p|\binom{n}{k} = \frac{n!}{k!(n-k)!}$. It follows that the product of all such primes divides $\binom{n}{k}$ and hence

$$\prod_{k<p\leq n} p \leq \binom{n}{k}.$$

Since $\binom{n}{k} = \binom{n}{n-k}$ and both are in the binomial expansion of $(1 + 1)^n$ it follows that $\binom{n}{k} < 2^{n-1}$. Therefore using that $k < n$ and the inductive hypothesis

$$\prod_{p\leq n} p = \prod_{p\leq k} p \prod_{k<p\leq n} p < 4^k 2^{n-1} = 2^{n+2k-1} \leq 2^{2n} = 4^n.$$

□

Finally based on many of these estimates we can provide a proof of Bertrand's Theorem (actually proved by Chebyshev) which we introduced in the last chapter. Recall that this theorem says that given any natural number n there is always a prime between n and $2n$. The proof actually shows that given any real number $x > 1$ there exists a prime between x and $2x$.

Theorem 4.2.5 *(Bertrand's Theorem) For every natural number $n > 1$ there is a prime p such that $n < p < 2n$.*

Proof By direct computation the theorem is easily established for $n \leq 128$. Now suppose that for some $n > 128$ there is no prime between n and $2n$. For a prime p let m_p be the highest power of p dividing $\binom{2n}{n}$ and k_p the first power such that $p^{k_p+1} > 2n$

as in the proof of Chebyshev's Estimate. Then as in the proof of Chebyshev's estimate, since we assume no primes in the range n to $2n$ we have

$$\binom{2n}{n} = \prod_{p \le 2n} p^{m_p} = \prod_{p \le n} p^{m_p}, m_p \le k_p.$$

Here we use $[x]$ to indicate the greatest integer function, that is, the greatest integer less than or equal to x.

Now if $\frac{2n}{3} < p \le n$ we then have $p \ge 3$ and $2 \le \frac{2n}{p} < 3$ and therefore

$$m_p = [\frac{2n}{p}] - 2[\frac{n}{p}] = 2 - 2 = 0.$$

If $\sqrt{2n} < p \le \frac{2n}{3}$ then we have $p^2 > 2n$ and hence $k_p = 1$ and so $m_p \le 1$. Finally if $p \le \sqrt{2n}$ we have $p^{m_p} \le p^{k_p} \le 2n$. Therefore

$$\binom{2n}{n} = \prod_{p \le \sqrt{2n}} p^{m_p} \prod_{\sqrt{2n} < p \le \frac{2n}{3}} p^{m_p} \prod_{\frac{2n}{3} < p \le n} p^{m_p}$$

$$\le \prod_{p \le \sqrt{2n}} (2n) \prod_{\sqrt{2n} < p \le \frac{2n}{3}} p.$$

For a real number $x \ge 128$ we have $\pi(x) \le \frac{x+1}{2}$ since there are at most $\frac{x+1}{2}$ odd integers less than x so certainly no more than that primes. Further since $x \ge 128$ we have at least two odd non primes less than x so $\pi(x) \le \frac{x+1}{2} - 2 < \frac{x}{2} - 1$. It follows that $\pi(\sqrt{2n}) < \sqrt{\frac{n}{2}} - 1$ and hence

$$\prod_{p \le \sqrt{2n}} p < (2n)^{\sqrt{\frac{n}{2}} - 1}.$$

Further from Theorem 4.2.4 we have

$$\prod_{p \le \frac{2n}{3}} p < 4^{\frac{2n}{3}}.$$

Therefore

$$\binom{2n}{n} < (2n)^{\sqrt{\frac{n}{2}} - 1} 4^{\frac{2n}{3}}.$$

Now

$$2^{2n} = (1+1)^{2n} = 1 + \binom{2n}{1} + \cdots + \binom{2n}{n} + \cdots + \binom{2n}{2n-1} + 1.$$

There are $2n + 1$ terms in this expansion and $\binom{2n}{n}$ is the largest. Combining the two outside terms $(1 + 1 = 2)$ we have $2n$ terms each $\leq \binom{2n}{n}$ and therefore

$$2^{2n} < (2n)\binom{2n}{n} \Longrightarrow \binom{2n}{n} > (2n)^{-1}2^{2n}.$$

Combining these two inequalities gives

$$(2n)^{-1}2^{2n} < (2n)^{\sqrt{\frac{n}{2}}-1}4^{\frac{2n}{3}} \Longrightarrow 2^{\frac{2n}{3}} < (2n)^{\sqrt{\frac{n}{2}}}.$$

Taking logarithms then yields

$$n\frac{2}{3}\ln 2 < \sqrt{\frac{n}{2}}\ln(2n) \Longrightarrow \sqrt{8n}\ln 2 - 3\ln(2n) < 0.$$

We show that this is a contradiction for all $n \geq 128$.

Let $F(x) = \sqrt{8x}\ln 2 - 3\ln(2x)$. Then $F(128) = 8\ln 2 > 0$. Further

$$F'(x) = \ln 2\frac{\sqrt{8}}{2}\frac{1}{\sqrt{x}} - \frac{3}{x} = \ln 2\frac{\sqrt{2}\sqrt{x} - 3/\ln(2)}{x}.$$

This last expression is positive for $x \geq 128$ and hence $F(x)$ is an increasing function for $x \geq 128$. Since $F(128) > 0$ it follows that $F(x) > 0$ for all $x \geq 128$. Therefore

$$n(\frac{2}{3})\ln 2 > \sqrt{\frac{n}{2}}\ln(2n)$$

which implies that

$$\sqrt{8n}\ln 2 - 3\ln(2n) > 0$$

for $n \geq 128$. This is impossible and hence a contradiction. Therefore there must be a prime between n and $2n$ for any integer n. □

4.3 Equivalent Formulations of the Prime Number Theorem

The proof of the prime number theorem rests on the analysis of three additional functions besides the prime number function $\pi(x)$. The first and most important of these is the **Riemann zeta function** $\zeta(s)$. As was discussed in the previous chapter this function was introduced for real $s > 1$ by Euler in proving that there are infinitely many primes (see Section 3.3). The function was then modified by Dirichlet and used in proving that there are infinitely many primes of the form $an + b$ with $(a, b) = 1$.

Riemann extended the definition to allow the variable s to be complex and showed how knowledge of the location of the zeros of the now complex function $\zeta(s)$ in the complex plane would imply the prime number theorem. We will discuss the zeta function and describe its ties to the prime number theorem in the next section. The other two functions that must be analyzed are known as the **Chebyshev functions**. The first, denoted $\theta(x)$, is defined for a real variable x by

$$\theta(x) = \sum_{p \le x} \ln p \text{ with } p \text{ prime} \tag{4.3.1}$$

while the second, denoted $\psi(x)$, is defined, again for a real variable x, by

$$\psi(x) = \sum_{p^k \le x, k \ge 1} \ln p \text{ with } p \text{ prime} \tag{4.3.2}$$

These functions count respectively the number of primes $p \le x$ and the number of prime powers $p^k \le x$ weighted by $\ln p$. Recall that the van Mangoldt function $\Lambda(n)$ is defined for positive integers by

$$\Lambda(n) = \begin{cases} \ln p & \text{if } n = p^c, c \ge 1 \\ 0 & \text{for all other } n > 0. \end{cases}$$

Hence the Chebyshev function $\psi(x)$ is actually the summation function of $\Lambda(n)$. That is

$$\psi(x) = \sum_{n \le x} \Lambda(n).$$

Further for a given prime $p \le x$ the number of times $\ln p$ is counted in the sum for $\psi(x)$ is $[\frac{\ln x}{\ln p}]$. Hence $\psi(x)$ can also be expressed as

$$\psi(x) = \sum_{p \le x} [\frac{\ln x}{\ln p}] \ln p.$$

In the type of notation we have used in defining the Chebyshev functions the prime number function can be expressed as

$$\pi(x) = \sum_{p \le x, p \text{ prime}} 1 \tag{4.3.3}$$

There are certain immediate relationships between these three functions. First, if $p^k \le x$ then $p \le x$ so clearly

$$\theta(x) \le \psi(x).$$

Further since $1 \le \ln p$ for $p \ge 3$ we have

$$\pi(x) \leq \theta(x) \text{ for } x \geq 5.$$

Now if $p^k \leq x$ then $k \leq [\frac{\ln x}{\ln p}]$. It follows that

$$\psi(x) = \sum_{\substack{p^k \leq x, \\ k \geq 1}} \ln p = \sum_{p \leq x} (\sum_{\substack{p^k \leq x, \\ k \geq 1}} 1) \ln p$$

$$= \sum_{p \leq x} [\frac{\ln x}{\ln p}] \ln p \leq \sum_{p \leq x} \ln x = \pi(x) \ln x.$$

Therefore

$$\psi(x) \leq \pi(x) \ln x.$$

Now $\theta(x) = \sum_{p \leq x} \ln p = \ln(\prod_{p \leq x} p)$. However from Theorem 4.2.4 we have $\prod_{p \leq x} p < 4^x$. Therefore

$$\theta(x) < x(\ln 4)$$

and consequently

$$\theta(x) = O(x).$$

We will need the following lemma which says that relative to x, $\theta(x)$, and $\psi(x)$ have the same order of magnitude.

Lemma 4.3.1 $\psi(x) = \theta(x) + O(x^{\frac{1}{2}} (\ln x)^2)$

Proof By definition

$$\psi(x) = \sum_{\substack{p^k \leq x, \\ k \geq 1}} \ln p.$$

For a given prime $p \leq x$ let p^t be the highest power of p such that $p^t \leq x$. Then

$$p \leq x, p^2 \leq x, \ldots, p^t \leq x \implies p \leq x, p \leq x^{\frac{1}{2}}, \ldots, p \leq x^{\frac{1}{t}}.$$

It follows that

$$\psi(x) = \theta(x) + \theta(x^{\frac{1}{2}}) + \cdots + \theta(x^{\frac{1}{m}})$$

where m is the first integer such that $m + 1 > \frac{\ln x}{\ln 2}$. We have

$$\theta(x) = \sum_{p \leq x} \ln p \leq \sum_{p \leq x} \ln x \leq x \ln x \text{ if } x \geq 2.$$

It follows that

$$\theta(x^{\frac{1}{k}}) < x^{\frac{1}{k}} \ln x \leq x^{\frac{1}{2}} \ln x \text{ if } x \geq 2 \text{ and } k \geq 2.$$

In the sum

$$\sum_{k=2}^{m} \theta(x^{\frac{1}{k}})$$

there are $O(\ln x)$ terms since $m - 1 \leq \frac{\ln x}{\ln 2}$. This coupled with the fact that $\theta(x^{\frac{1}{k}}) \leq x^{\frac{1}{2}} \ln x$ gives that

$$\sum_{k=2}^{m} \theta(x^{\frac{1}{k}}) = O(x^{\frac{1}{2}}(\ln x)^2).$$

Therefore

$$\psi(x) = \theta(x) + O(x^{\frac{1}{2}}(\ln x)^2).$$

□

It follows immediately from this lemma and the fact that $x^{\frac{1}{2}}(\ln x)^2 = o(x)$ that if there exists a constant A with $\theta(x) < Ax$ then there exists a constant B such that $\psi(x) < Bx$ and if there exists a constant C with $Cx < \psi(x)$ then there exists a constant D with $Dx < \theta(x)$.

We extend these observations to show that $\theta(x)$ and $\psi(x)$ both have order of magnitude x.

Theorem 4.3.1 *There exist positive constants* A_1, A_2, B_1, B_2 *such that*

$$A_1 x \leq \theta(x) \leq A_2 x,$$

$$B_1 x \leq \psi(x) \leq B_2 x.$$

Thats is, $\theta(x) \approxeq x$ *and* $\psi(x) \approxeq x$.

Proof In light of the comments made preceding the theorem it suffices to bound $\theta(x)$ above and $\psi(x)$ below. From Theorem 4.2.4 we have that $\prod_{p \leq x} p < 4^x$. This implies that $\theta(x) = \sum_{p \leq x} \ln p < x \ln 4$ and hence $\theta(x) < Bx$ with $B = \ln 4$. This bounds $\theta(x)$ above.

We now show that we can bound $\psi(x)$ below. This is similar to the proof given for Chebyshev's estimate. As in that proof, if p is a prime, let m_p be the highest power of p such that $p^{m_p} | \binom{2n}{n}$ and let k_p be the first exponent such that $p^{k_p+1} > 2n$. Then as before

$$\binom{2n}{n} = \prod_{p \leq 2n} p^{m_p}$$

and

$$m_p \leq [\frac{\ln 2n}{\ln p}].$$

It follows that

$$\ln \binom{2n}{n} = \sum_{p \le 2n} m_p \ln p \le \sum_{p \le 2n} [\frac{\ln 2n}{\ln p}] \ln p = \psi(2n).$$

Further from Lemma 4.2.4(ii)

$$\binom{2n}{n} \ge 2^n \implies \psi(2n) \ge n \ln 2.$$

If $x \ge 2$ let $n = [\frac{x}{2}] \ge 1$ and then

$$\psi(x) \ge \psi(2n) \ge n \ln 2 > \frac{1}{4} x \ln 2.$$

Therefore $\psi(x) \ge Cx$ with $C = \frac{\ln 2}{4}$ completing the proof.

□

Considering again the result of Lemma 4.3.1 that

$$\psi(x) = \theta(x) + O(x^{\frac{1}{2}} (\ln x)^2)$$

coupled with the fact that $x^{\frac{1}{2}} (\ln x)^2 = o(x)$ we obtain that

$$\frac{\psi(x)}{x} = \frac{\theta(x)}{x} + o(1).$$

In particular this implies that

$$\lim_{x \to \infty} \frac{\psi(x)}{x} = 1 \text{ if and only if } \lim_{x \to \infty} \frac{\theta(x)}{x} = 1.$$

In the notation we introduced earlier this says that

$$\psi(x) \sim x \text{ if and only if } \theta(x) \sim x.$$

We show now that each of these statements is equivalent to the prime number theorem.

Theorem 4.3.2 *The following are all equivalent formulations of the prime number theorem*

(a) $\pi(x) \sim \frac{x}{\ln x}$,
(b) $\theta(x) \sim x$,
(c) $\psi(x) \sim x$.

Proof From the remarks immediately preceding the theorem we have that $\theta(x) \sim x$ if and only if $\psi(x) \sim x$. Therefore, it is sufficient to show that $\pi(x) \sim \frac{x}{\ln x}$ is equivalent to $\theta(x) \sim x$.

We have that $\theta(x) \leq \pi(x)\ln x$ and further $Ax \leq \theta(x)$ for some constant A. Therefore

$$\pi(x) \geq \frac{\theta(x)}{\ln x} \geq \frac{Ax}{\ln x}.$$

For any real ϵ with $0 < \epsilon < 1$ we have

$$\theta(x) \geq \sum_{x^{1-\epsilon}<p\leq x} \ln p > (1-\epsilon)\ln x \sum_{x^{1-\epsilon}<p\leq x} 1$$

$$= (1-\epsilon)\ln x(\pi(x) - \pi(x^{1-\epsilon})) \geq (1-\epsilon)\ln x(\pi(x) - x^{1-\epsilon})$$

since $x^{1-\epsilon} > \pi(x^{1-\epsilon})$.

It follows that

$$\pi(x) \leq x^{1-\epsilon} + \frac{\theta(x)}{(1-\epsilon)\ln x}$$

Combining these inequalities gives

$$\frac{Ax}{\ln x} \leq \frac{\theta(x)}{\ln x} \leq \pi(x) \leq x^{1-\epsilon} + \frac{\theta(x)}{(1-\epsilon)\ln x}$$

from which it follows that

$$1 \leq \frac{\pi(x)\ln x}{\theta(x)} \leq \frac{x^{1-\epsilon}\ln x}{\theta(x)} + \frac{1}{1-\epsilon}$$

Now $\theta(x) \geq Ax$ so

$$\frac{x^{1-\epsilon}\ln x}{\theta(x)} < \frac{\ln x}{Ax^{\epsilon}}.$$

Since ϵ is arbitrary in (0, 1) the value $\frac{1}{1-\epsilon}$ can be made arbitrarily close to 1. Further for a fixed ϵ the value $\frac{\ln x}{Ax^{\epsilon}}$ can be made arbitrarily small by choosing a large x. Therefore

$$\frac{x^{1-\epsilon}\ln x}{\theta(x)} + \frac{1}{1-\epsilon} < \epsilon_1 + 1$$

for x large enough and ϵ_1 arbitrarily small. Hence we have

$$1 \leq \frac{\pi(x)\ln x}{\theta(x)} < 1 + \epsilon_1$$

and thus

$$\lim_{x\to\infty} \frac{\pi(x)\ln x}{\theta(x)} = 1.$$

By definition then

$$\pi(x)\ln x \sim \theta(x) \implies \frac{\pi(x)\ln x}{x} \sim \frac{\theta(x)}{x}.$$

From this it is straightforward that as $x \to \infty$ we have

$$\frac{\theta(x)}{x} \to 1 \text{ if and only if } \frac{\pi(x)}{x/\ln x} \to 1$$

or

$$\theta(x) \sim x \text{ if and only if } \pi(x) \sim \frac{x}{\ln x}.$$

□

In this proof we will present the prime number theorem, we will actually show that $\psi(x) \sim x$ and then invoke the above result.

As we remarked in the last section Chebyhsev also proved that if

$$\lim_{x\to\infty} \frac{\pi(x)}{x/\ln x}$$

existed, then the limit would have to be one. Thus, he seemed very close to the prime number theorem. However, he could not actually prove this limit existed. We close this section by giving a proof of the result of Chebyshev. We need first the following result due to Mertens. This is one of several results in the area due to Mertens and known collectively as **Mertens' Theorems** (see [N]).

Theorem 4.3.3 *If* $\Lambda(n)$ *is the van Mangoldt function then*

$$\sum_{n\le x} \frac{\Lambda(n)}{n} = \ln x + O(1).$$

Proof Consider the sum

$$\sum_{n\le x} \ln(\frac{x}{n}).$$

Since $\ln x$ is an increasing function we have for $n \ge 2$

$$\ln(\frac{x}{n}) \le \int_{n-1}^{n} \ln(\frac{x}{t})dt.$$

From this it follows that

$$\sum_{n=2}^{[x]} \ln(\frac{x}{n}) \le \int_1^x \ln(\frac{x}{t})dt = x\int_1^x \frac{\ln u}{u^2}du < x\int_1^\infty \frac{\ln u}{u^2}du.$$

However the infinite integral $\int_1^\infty \frac{\ln u}{u^2} du$ is convergent so it has finite value A. Therefore

$$\sum_{n=2}^{[x]} \ln(\frac{x}{n}) < Ax \implies \sum_{n=2}^{[x]} \ln(\frac{x}{n}) = O(x).$$

Hence

$$\sum_{n\le x} \ln n = [x] \ln x + O(x) = x \ln x + O(x).$$

As in the proof of Chebyshev's estimate let

$$e_p = \sum_{m=1}^{t_p} [\frac{x}{p^m}]$$

so that

$$[x]! = \prod_p p^{e_p}.$$

Then taking logarithms we get

$$\ln([x]!) = \ln(\prod_p p^{e_p}) \implies \sum_{n\le x} \ln n = \sum_{p\le x} e_p \ln p$$

$$= \sum_{m\ge 1} \sum_{p^m \le x} [\frac{x}{p^m}] \ln p = \sum_{n\le x} [\frac{x}{n}] \Lambda(n)$$

where $\Lambda(n)$ is the van Mangoldt function. Further

$$\sum_{n\le x} (\frac{x}{n}) \Lambda(n) < \sum_{n\le x} [\frac{x}{n}] \Lambda(n) + \sum_{n\le x} \Lambda(n)$$

$$= \sum_{n\le x} [\frac{x}{n}] \Lambda(n) + \psi(x) = \sum_{n\le x} [\frac{x}{n}] \Lambda(n) + O(x)$$

since $\psi(x) = O(x)$. Combining these inequalities give us

$$\sum_{n\le x} (\frac{x}{n}) \Lambda(n) = \sum_{n\le x} \ln n + O(x) = x \ln x + O(x).$$

Removing the factor x yields finally

$$\sum_{n\le x}\frac{\Lambda(n)}{n}=\ln x+O(1).$$

□

As an immediate corollary we obtain.

Corollary 4.3.1 $\sum_{p\le x}\frac{\ln p}{p}=\ln x+O(1)$.

Proof By definition

$$\sum_{n\le x}\frac{\Lambda(n)}{n}=\sum_{m\ge 1}\sum_{p^m\le x}\frac{\ln p}{p^m}.$$

This implies that

$$\sum_{n\le x}\frac{\Lambda(n)}{n}-\sum_{p\le x}\frac{\ln p}{p}=\sum_{m\ge 2}\sum_{p^m\le x}\frac{\ln p}{p^m}<\sum_{p}(\frac{1}{p^2}+\frac{1}{p^3}+\cdots)\ln p$$

$$=\sum_{p}\frac{\ln p}{p(p-1)}\le\sum_{n=2}^{\infty}\frac{\ln n}{n(n-1)}.$$

This last infinite series converges to some value S. Hence

$$\sum_{n\le x}\frac{\Lambda(n)}{n}-\sum_{p\le x}\frac{\ln p}{p}<A$$

for some value A. Since from the previous theorem $\sum_{n\le x}\frac{\Lambda(n)}{n}=\ln x+O(1)$ it follows that

$$\sum_{p\le x}\frac{\ln p}{p}=\ln x+O(1).$$

□

Theorem 4.3.4 *If* $\lim_{x\to\infty}\frac{\pi(x)}{x/\ln x}$ *exists then* $lim_{x\to\infty}\frac{\pi(x)}{x/\ln x}=1$.

Proof Recall that $\psi(x)=\sum_{n\le x}\Lambda(n)$. Then

$$\sum_{n\le x}\frac{\Lambda(n)}{n}=\sum_{n\le x-1}\psi(n)(\frac{1}{n}-\frac{1}{n+1})+\frac{\psi(x)}{[x]}$$

which follows easily since $\Lambda(n)=\psi(n)-\psi(n-1)$. Since $\psi(x)=\psi(n)$ if $n\le x<n+1$ we have

$$\psi(n)(\frac{1}{n} - \frac{1}{n+1}) = \int_{n}^{n+1} \frac{\psi(t)}{t^2} dt.$$

Summing then yields

$$\sum_{n \le x-1} \psi(n)(\frac{1}{n} - \frac{1}{n+1}) = \int_{2}^{x} \frac{\psi(t)}{t^2} dt$$

since $\psi(1) = 0$. Hence

$$\sum_{n \le x} \frac{\Lambda(n)}{n} = \frac{\psi(x)}{x} + \int_{2}^{x} \frac{\psi(t)}{t^2} dt.$$

Since

$$\sum_{n \le x} \frac{\Lambda(n)}{n} = \ln x + O(1) \text{ and } \frac{\psi(x)}{x} = O(1)$$

it follows that

$$\int_{2}^{x} \frac{\psi(t)}{t^2} dt = \ln x + O(1).$$

Now suppose that $\liminf \frac{\psi(x)}{x} = 1 + \epsilon$ with $\epsilon > 0$. Then

$$\psi(x) > (1 + \frac{1}{2}\epsilon)x$$

for x sufficiently large say $x \ge x_0$. Then

$$\int_{2}^{x} \frac{\psi(t)}{t^2} dt = \int_{2}^{x_0} \frac{\psi(t)}{t^2} dt + \int_{x_0}^{x} \frac{\psi(t)}{t^2} dt > (1 + \frac{1}{2}\epsilon) \ln x - A$$

for some constant A. However, this contradicts that

$$\int_{2}^{x} \frac{\psi(t)}{t^2} dt = \ln x + O(1).$$

On the other hand, if $\limsup \frac{\psi(x)}{x} = 1 - \epsilon$ with $\epsilon > 0$ we obtain an analogous contradiction. Therefore

$$\limsup \frac{\psi(x)}{x} \ge 1 \text{ and } \liminf \frac{\psi(x)}{x} \le 1$$

and therefore, if the limit $\frac{\psi(x)}{x}$ existed as $x \to \infty$ the value would have to be one. Further since

$$\frac{\pi(x)}{x/\ln x} \sim 1 \text{ if and only if } \frac{\psi(x)}{x} \sim 1$$

this shows that if $\frac{\pi(x)}{x/\ln x}$ has a limit its value must be one also.

□

4.4 The Riemann Zeta Function and the Riemann Hypothesis

From Chebyshev's estimate and its consequences it seemed that a proof of the prime number theorem was close at hand. In 1860, B. Riemann attempted to prove this main result. Riemann eventually wrote only one paper in number theory, and although he failed in his primary goal of proving the prime number theorem, this paper had a profound effect on both number theory in particular and mathematics in general. Much as Gauss's **Disquisitiones Arithmeticae** set the direction for elementary and algebraic number theory, Riemann's work set the direction for analytic number theory. Riemann's basic new (and brilliant) idea was to extend the zeta function of Euler $\zeta(s)$ (see Section 3.1.2) to allow complex arguments that is to allow s to be a complex number. This idea of Riemann initiated the use of complex analysis, specifically, the theory of analytic functions and complex integration, into number theory and laid the ground work for modern analytic number theory. Recall that use of analysis begins with the Euler zeta function and continues through the work of Dirichlet. However, it is this paper of Riemann and the introduction of complex analytic methods that really is the beginning of analytic number theory.

Euler had introduced $\zeta(s)$ for real s in giving a proof that the primes were infinite and that the series $\sum \frac{1}{p}$ diverges. Dirichlet used a variation of this function, still for real s, in building the Dirichlet series used in the proof of his theorem on primes in arithmetic progressions (see Section 3.3). Riemann, in allowing complex s, showed that the resulting function $\zeta(s)$ is an analytic function for Re $(s) > 1$ and further can be **continued analytically** (see the next section) to a function, which we will also denote $\zeta(s)$, that is, analytic in all of $\mathbb{C}$ except $s = 1$. Further $s = 1$ is a simple pole with residue 1, that is,

$$\zeta(s) = \frac{1}{s-1} + H(s)$$

where $H(s)$ is an entire function. Riemann then showed that knowledge of the location of the complex zeroes of $\zeta(s)$ describes the density of primes. In particular, if there are no zeroes along the line Re $(s) = 1$, this would then imply the prime number theorem. This was precisely the main step in the proofs of Hadamard and de la Vallee Poussin (given independently) of the prime number theorem given 36 years after Riemann's paper.

4.4.1 The Real Zeta Function of Euler

Recall that the Euler zeta function was defined for real $s > 1$ by

$$\zeta(s) = \sum_{n=1}^{\infty} \frac{1}{n^s}.$$

From the classical p-series test this series converges absolutely for $s > 1$ and hence defines a real C^{∞} function in this range. Further, as $s \to 1, \zeta(s) \to \infty$ which implies through the Euler product representation that there are infinitely many primes (see Section 3.1.3).

As a direct consequence of the Fundamental Theorem of Arithmetic, Euler derived the following product decomposition (see Section 3.1.2)

$$\zeta(s) = \prod_{p \text{ prime}} (\frac{1}{1 - p^{-s}}).$$

This product decomposition will remain valid for complex s with $Re(s) > 1$ and hence it is clear that there are no real zeros of $\zeta(s)$ if $s > 1$.

There are ties between the zeta function and several of the other arithmetical functions which we have worked with in this chapter. First, from the Euler product decomposition we obtain by logarithmic differentiation

$$-\frac{\zeta'(s)}{\zeta(s)} = \sum_{p} \sum_{m=1}^{\infty} \frac{\ln p}{p^{ms}}.$$

Recall again that the van Mangoldt function $\Lambda(n)$ is defined for positive integers by

$$\Lambda(n) = \begin{cases} \ln p & \text{if } n = p^c, c \geq 1 \\ 0 & \text{for all other } n > 0. \end{cases}$$

Therefore

$$\sum_{p} \sum_{m=1}^{\infty} \frac{\ln p}{p^{ms}} = \sum_{n=1}^{\infty} \frac{\Lambda(n)}{n^s}$$

$$\implies -\frac{\zeta'(s)}{\zeta(s)} = \sum_{n=1}^{\infty} \frac{\Lambda(n)}{n^s}.$$

Next, again from the Euler product decomposition, we have for $s > 1$

$$\zeta(s)^{-1} = \prod_{p} (1 - p^{-s}).$$

Expanding the infinite product yields

$$\zeta(s)^{-1} = 1 - \sum_{p} p^{-s} + \sum_{p,q} (pq)^{-s} - \sum_{p,q,r} (pqr)^{-s} + \cdots$$

with $p, q, r, \ldots$ primes. In this summation only squarefree integers appear. Further for a squarefree integer n, the coefficient of n^{-s} in the above product is ± 1, depending on whether the number of prime factors of n is odd or even. This is precisely $\mu(n)$ where $\mu(n)$ is the Möbius function (see Sections 3.3 and 3.6). Therefore

$$\zeta(s)^{-1} = \sum_{n=1}^{\infty} \frac{\mu(n)}{n^s}.$$

Lemma 4.4.1 *For $s > 1$ we have the following relationships.*

1. $\zeta(s)^{-1} = \sum_{n=1}^{\infty} \frac{\mu(n)}{n^s}$*, where $\mu(n)$ is the Möbius function.*
2. $-\frac{\zeta'(s)}{\zeta(s)} = \sum_{n=1}^{\infty} \frac{\Lambda(n)}{n^s}$*, where $\Lambda(n)$ is the van Mangoldt function.*

Euler further determined the exact value of $\zeta(2)$ and showed that it was $\frac{\pi^2}{6}$. Originally this was done by a clever use of certain trigonometric identities (see [NZM]). Subsequently Euler developed a method to determine the values of $\zeta(s)$ at all even integers. We first give a proof of the basic result that $\zeta(2) = \frac{\pi^2}{6}$ using a different approach. Some basic ideas from the theory of Fourier series are needed.

Recall that a real or complex function $f(x)$ is periodic of period L if $f(x + L) = f(x)$ for all x. In the early 1800s, Fourier attempted to prove that any periodic function can be expressed as a trigonometric series that is a sum of sine functions and cosine functions. If $f(x)$ is periodic of period $2L$ then its **Fourier series** is

$$\overline{f} = a_0 + \sum_{n=1}^{\infty} (a_n \cos(\frac{n\pi x}{L}) + b_n \sin(\frac{n\pi x}{L})).$$

Using certain orthogonality relations between sines and cosines Fourier showed that if $f(x) = \overline{f}(x)$ then the coefficients a_0, a_n, b_n must be given by

$$a_0 = \frac{1}{2L} \int_{-L}^{L} f(x)dx$$

$$a_n = \frac{1}{L} \int_{-L}^{L} f(x) \cos(\frac{n\pi x}{L})dx, n = 1, 2, \ldots$$

$$b_n = \frac{1}{L} \int_{-L}^{L} f(x) \sin(\frac{n\pi x}{l})dx, n = 1, 2, \ldots$$

The a_n, b_n are called the **Fourier coefficients**. Fourier assumed that $\overline{f}(x) = f(x)$ but the situation was not definitively proved until the theory of Lebesgue integration was developed. What was then obtained is called the **Fourier Convergence Theorem**.

Theorem 4.4.1 *(Fourier Convergence Theorem) (see [Gr]) Let $f(x)$ be periodic of period $2L$. Then:*

1. *If both $f(x)$ and $f'(x)$ are piecewise continuous on $(-L, L)$ then the Fourier series converges pointwise to the mean value $\frac{f(x^+)+f(x^-)}{2}$.*
2. *If both $f(x)$ and $f'(x)$ are continuous on $(-L, L)$ then the Fourier series converges uniformly to $f(x)$.*

Therefore, a C^1 periodic function is everywhere represented by its Fourier series realizing Fourier's original idea. We now give Euler's result using Fourier series. In Section 4.7 we present a separate more general proof.

Theorem 4.4.2 $\zeta(2) = \frac{\pi^2}{6}$.

Proof Let $f(x) = x^2, -\pi < x < \pi$ and then continued periodically with period 2π. This is continuous everywhere and differentiable everywhere except at integer multiples of π. Therefore by the Fourier convergence theorem it is everywhere represented by its Fourier series.

We apply the formulas. First, $f(x)$ is an even function so there are only cosine terms and hence $b_n = 0$ for all n. Then

$$a_0 = \frac{1}{2\pi}\int_{-\pi}^{\pi} x^2 dx = \frac{\pi^2}{3}$$

and

$$a_n = \frac{1}{\pi}\int_{-\pi}^{\pi} x^2 \cos(nx) dx = (-1)^n \frac{4}{n^2}$$

using integration by parts and the fact that $\cos(n\pi) = (-1)^n$. Therefore the Fourier series for $f(x)$ is given by

$$x^2 = \frac{\pi^2}{3} + 4\sum_{n=1}^{\infty} \frac{(-1)^n}{n^2} \cos nx, \quad -\pi < x < \pi.$$

Now let $x = \pi$ and place this value into the Fourier expansion. Then

$$\pi^2 = \frac{\pi^2}{3} + 4\sum_{n=1}^{\infty} \frac{(-1)^n}{n^2} \cos(n\pi).$$

But $\cos(n\pi) = (-1)^n$ so

$$\pi^2 = \frac{\pi^2}{3} + 4\sum_{n=1}^{\infty} \frac{(-1)^n}{n^2}(-1)^n$$

$$\Longrightarrow \pi^2 = \frac{\pi^2}{3} + 4\sum_{n=1}^{\infty} \frac{1}{n^2} = \frac{\pi^2}{3} + 4\zeta(2)$$

$$\Longrightarrow \zeta(2) = \frac{\pi^2}{6}.$$

□

In Theorem 6.4.16, we will prove that π is a transcendental number. With this fact the above result leads directly to another proof that there are infinitely many primes.

Euler's method to find $\zeta(2)$ involved a detailed look at certain trigonometric identities (see [NZM] or [Na]). Subsequently, he developed a technique to determine the value of $\zeta(s)$ for s an even positive integer. In particular, he tied the values of $\zeta(2n)$ to the **Bernoulli numbers** B_n. These numbers are defined in terms of the coefficients of the Taylor series expansion about $x = 0$ of the function $f(x) = \frac{x}{e^x-1}$ with $f(0) = 1$. Specifically

$$\frac{x}{e^x - 1} = \sum_{n=0}^{\infty} \frac{B_n}{n!} x^n.$$

Euler proved the following. For a proof of this result see Section 4.7.

Theorem 4.4.3 $\zeta(2n) = \frac{(-1)^{n-1}B_{2n}}{2(2n)!}(2\pi)^{2n}$.

Substitution in this formula and using that $B_2 = \frac{1}{6}$, $B_4 = -\frac{1}{30}$ yields $\zeta(2) = \frac{\pi^2}{6}$ and $\zeta(4) = \frac{\pi^4}{90}$. Euler himself determined up to $\zeta(26)$ for even n. From Euler's formula and the fact that π is transcendental it follows that $\zeta(2n)$ is transcendental for any even positive integer $2n$. On the other hand very little is known about the arithmetic nature of $\zeta(s)$ for $s = 2n + 1$ an odd positive integer. It was shown by R. Apery (also by DeBranges) that $\zeta(3)$ is irrational and Apery also gave the following formula

$$\zeta(3) = \frac{5}{2}\sum_{k=1}^{\infty} \frac{(-1)^{k-1}}{k^3\binom{2k}{k}}.$$

The number $\zeta(3)$ is called **Apery's constant** and has an approximate value of 1.202057. Euler's result has also been recovered using Fourier series methods along the lines of the proof we gave for $\zeta(2) = \frac{\pi^2}{6}$.

There are several equivalent analytic expressions for $\zeta(s)$ for real $s > 1$. We mention one such expression here because of the ties to the analytic continuation of the complex Riemann zeta function. This will be discussed shortly. In order to introduce this expression we must first describe the Gamma function.

Definition 4.4.1 *If $s > 0$ the* **Gamma function** *is given by*

$$\Gamma(s) = \int_0^\infty x^{s-1} e^{-x} dx.$$

By a straightforward integration by parts (see exercises) we obtain the following.

Lemma 4.4.2 $\Gamma(s+1) = s\Gamma(s)$.

It is easy to determine that $\Gamma(1) = 1$. Hence

$$\Gamma(2) = 1\Gamma(1) = 1, \Gamma(3) = 2\Gamma(2) = 2!, \Gamma(4) = 3\Gamma(3) = 3!, \ldots$$

An easy induction then gives that:

Corollary 4.4.1 $\Gamma(n) = (n-1)!$ *for any* $n \geq 1, n \in \mathbb{N}$.

The Gamma function is then the extended factorial function.

The functional equation $\Gamma(s+1) = s\Gamma(s)$ allows us to extend the definition of $\Gamma(s)$ to all nonpositive real numbers s except for 0 and the negative integers. Further $\lim_{s \to -n} \Gamma(s) = \infty, n \in \mathbb{N}$.

Another important result whose proof we will outline in the exercises is the following.

Lemma 4.4.3 $\Gamma(\frac{1}{2}) = \sqrt{\pi}$.

The relation we wish to show for $\zeta(s)$ is given in the next theorem.

Theorem 4.4.4 *For real $s > 1$*

$$\zeta(s) = \frac{1}{\Gamma(s)} \int_0^\infty \frac{t^{s-1}}{e^t - 1} dt.$$

Proof For $s > 1$ let

$$G(s) = \frac{1}{\Gamma(s)} \int_0^\infty \frac{t^{s-1}}{e^t - 1} dt.$$

We show that $G(s) = \zeta(s)$. Recall that the sum of a geometric series with ratio r is given by

$$\sum_{k=0}^\infty r^k = \frac{1}{1-r} \text{ if } |r| < 1.$$

It follows then that

$$\frac{1}{1-e^{-t}} = \sum_{k=0}^\infty e^{-kt}.$$

Now

$$\frac{t^{s-1}}{e^t - 1} = e^{-t}t^{s-1}\frac{1}{1-e^{-t}} = e^{-t}t^{s-1}\sum_{k=0}^{\infty} e^{-kt} = t^{s-1}\sum_{k=1}^{\infty} e^{-kt}.$$

It follows that

$$\int_0^{\infty} \frac{t^{s-1}}{e^t - 1}dt = \sum_{k=1}^{\infty}(\int_0^{\infty} e^{-kt}t^{s-1}dt).$$

Now let $y = kt$ so that $dt = \frac{1}{k}dy$ and substitute

$$G(s) = \frac{1}{\Gamma(s)}\sum_{k=1}^{\infty}(\int_0^{\infty} e^{-kt}t^{s-1}dt) = \frac{1}{\Gamma(s)}\sum_{k=1}^{\infty}(\int_0^{\infty} e^{-y}(\frac{y}{k})^{s-1}\frac{1}{k}dy)$$

$$= \frac{1}{\Gamma(s)}(\sum_{k=1}^{\infty}\frac{1}{k^s})\int_0^{\infty} y^{s-1}e^{-y}dy.$$

However $\int_0^{\infty} y^{s-1}e^{-y}dy = \Gamma(s)$ and therefore

$$G(s) = \sum_{k=1}^{\infty}\frac{1}{k^s} = \zeta(s).$$

□

4.4.2 Analytic Functions and Analytic Continuation

Riemann introduced complex analysis, specifically the theory of analytic functions and the theory of complex integration, into the study of number theory. In this section, we briefly go over the basic necessary ideas.

If $w = f(z)$ is a complex function then the **complex derivative** is defined in exactly the same formal manner as the real derivative.

Definition 4.4.2 *If $f(z)$ is any complex function, then its* **derivative** *$f'(z_0)$ at $z_0 \in \mathbb{C}$ is*

$$f'(z_0) = \lim_{\Delta z \to 0}\frac{f(z_0 + \Delta z) - f(z_0)}{\Delta z}$$

whenever this limit exists. If $f'(z_0)$ exists, then $f(z)$ is **differentiable** *there. $f(z)$ is differentiable on a whole region if it is differentiable at each point of the region.*

The complex function $w = f(z)$ is **analytic** or **holomorphic** at z_0 if $f(z)$ is differentiable in a circular neighborhood of z_0. $f(z)$ is analytic in a domain U if it is analytic at each point of U. If $f(z)$ is analytic throughout $\mathbb{C}$, then it is called an **entire function**. Many of the standard functions from analysis: polynomials, e^z, $\sin z$, $\cos z$, appropriately defined for complex arguments, are entire.

If $f(z)$ is a complex function defined on a region U containing the curve

$$\gamma(t) = x(t) + iy(t), \qquad t_0 \le t \le t_1$$

then the **complex contour integral** $\int_\gamma f(z)dz$ is defined by

$$\int_\gamma f(z)dz = \int_{t_0}^{t_1} f(\gamma(t))\gamma'(t)dt.$$

Most of complex analysis deals with the properties and implications of complex integration of analytic functions. One of the cornerstones of this theory is **Cauchy's Theorem**.

Theorem 4.4.5 *(Cauchy's Theorem) Let $f(z)$ be analytic throughout a simply connected domain U and suppose γ is a simple closed curve entirely contained in U. Then*

$$\int_\gamma f(z)dz = 0.$$

As a consequence of Cauchy's Theorem one obtains (via the Cauchy integral formulae) that analytic functions have the property that they have derivatives of all possible orders. That is, if $f(z)$ is analytic at z_0 then $f'(z_0), f''(z_0), \dots, f^{(n)}(z_0), \dots$ all exist. Further in a neighborhood of z_0 the function $f(z)$ is then given by a convergent Taylor series centered on z_0:

$$f(z) = \sum_{n=0}^{\infty} \frac{f^{(n)}(z_0)}{n!}(z - z_0)^n \text{ for } |z - z_0| < R.$$

The derivatives $f^{(n)}(z_0)$ are given by the **Cauchy integral formula** as

$$f^{(n)}(z_0) = \frac{n!}{2\pi i} \int_\gamma \frac{f(z)}{(z - z_0)^{n+1}} dz$$

where γ is any simple closed curve around z_0 within a simply connected domain U where $f(z)$ is analytic. Recall that a **simply connected domain** in $\mathbb{C}$ is a region where every simple closed curve can be continuously shrunk to a point, that is, a region which has no holes in it (see [Ah]). Hence, the values of a complex analytic function and its derivatives within U are determined by its values on the boundary. Hence the interior values are a type of **average** of the boundary values. Although we will not pursue this further, the idea has been exploited extensively in number theory and analysis. The next theorem summarizes all these comments.

Theorem 4.4.6 *Suppose $f(z)$ is analytic in a simply connected domain U containing z_0 and γ is a simple closed curve within U. Then:*

1. *$f(z)$ has derivatives of all possible orders at z_0.*
2. *There exists an $R > 0$ such that $f(z)$ is given by a convergent Taylor series centered on z_0:*

$$f(z) = \sum_{n=0}^{\infty} \frac{f^{(n)}(z_0)}{n!}(z - z_0)^n \text{ for } |z - z_0| < R.$$

3. *The derivatives are given by the Cauchy integral formulae as*

$$f^{(n)}(z_0) = \frac{n!}{2\pi i} \int_{\gamma} \frac{f(z)}{(z - z_0)^{n+1}} dz.$$

We note that Theorem 4.4.6 is in distinction to the situation for real differentiable functions. A function $y = f(x)$ with $x, y \in \mathbb{R}$ can have one derivative but not two, two derivatives but not three and so on. Further there are real function which are C^{∞}, that is, they have infinitely many derivatives, but which are not given by convergent Taylor series. A real function which has a convergent Taylor series centered on x_0 is said to be **real analytic** at x_0.

An extremely important concept in studying the zeta function is that of **analytic continuation**. The basic idea is the following: suppose a complex analytic function $f(z)$ is given by an analytic expression which holds in a domain S in $\mathbb{C}$. Suppose that this is equivalent within S or within a subset of S to another analytic expression which holds in a larger domain S_1. Then, the second expression can be used to analytically extend or continue $f(z)$ to the larger domain S_1. We make this precise.

Suppose that $f_1(z)$ is analytic on a domain S_1 and $f_2(z)$ is analytic on a domain S_2. Suppose that $S_1 \cap S_2 \neq \emptyset$ and $f_1(z) = f_2(z)$ on $S_1 \cap S_2$. Then $(f_2(z), S_2)$ is said to be a **direct analytic continuation** of $(f_1(z), S_1)$. The individual pairs (f_1, S_1) and (f_2, S_2) are called **function elements**. A function element (f, S) is an **analytic continuation** of (f_1, S_1) if there is a chain (f_i, S_i) of function elements connecting (f_1, S_1) to (f, S) and with each neighboring pair a direct analytic continuation. A **global analytic function** is a nonempty collection of function elements $F = \{(f_\alpha, S_\alpha)\}$ such that any two in this collection are analytic continuations of each other. A global analytic function is **complete** if it contains all analytic continuations of any of its function elements.

Finally, analytic continuation is essentially unique in the sense that two analytic functions which agree on a sufficiently large domain, for example, an open neighborhood of a curve, are identical.

As an example of a type of analytic continuation, consider the Gamma function

$$\Gamma(s) = \int_0^{\infty} t^{s-1} e^{-t} dt.$$

This has meaning only for real $s > 0$. However, Euler proved that for real $s > 0$

$$\Gamma(s) = \frac{e^{-\gamma s}}{s} \prod_{n=1}^{\infty} (1 + \frac{s}{n})^{-1} e^{\frac{s}{n}} \tag{4.4.2.1}$$

where γ is **Euler's constant** and has an approximate value of .57722. The expression in (4.4.2.1) is valid now for complex s with $Re(s) > 0$ and can be used for the definition of the complex Gamma function $\Gamma(z)$. Using the relation

$$\Gamma(z+1) = z\Gamma(z)$$

the complex function can be continued to a function which is analytic except at $z = 0, z = -1, z = -2, \ldots$.

If $f(z)$ is not analytic at z_0 but is analytic in a neighborhood of z_0 then z_0 is called an **isolated singularity**. Isolated singularities are classified as either **removable** in which case $\lim_{z \to z_0} f(z)$ exists and is not infinite; a **pole**, in which case $\lim_{z \to z_0} f(z) = \infty$; or an **essential singularity** in which case $\lim_{z \to z_0} f(z)$ does not exist. For a pole z_0 there exists an integer $m \geq 1$ such that $f(z) = \frac{h(z)}{(z-z_0)^m}$ with $h(z)$ analytic at z_0. The minimal integer m with that property is called the **order of the pole**. If $m = 1$ then z_0 is a **simple pole**. The value

$$\frac{1}{(m-1)!} \lim_{z \to z_0} \frac{d^{n-1}(z-z_0)^n f(z)}{dz^{n-1}}$$

is the **residue** of $f(z)$ at the pole z_0. The residue is equal to

$$\frac{1}{2\pi i} \int_{\gamma} f(z) dz$$

where γ is any simple closed curve around z_0 within a domain around z_0 where $f(z)$ is analytic.

If $f(z)$ has a simple pole at z_0 with residue w_0 then the function $h(z)$ given by

$$h(z) = f(z) - \frac{w_0}{z - z_0}$$

is analytic at z_0.

A function $f(z)$ is **meromorphic** in a domain S if it is analytic except for poles which by definition are isolated. We will see in the next section that via analytic continuation the zeta funciton $\zeta(s)$ can be considered as a meromorphic function in the whole complex plane with a simple pole at $z = 1$ and with residue 1. Hence

$$\zeta(z) - \frac{1}{z-1} = H(z)$$

where $H(z)$ is an entire function.

4.4.3 The Riemann Zeta Function

The **Riemann zeta function** starts with the Euler zeta function $\zeta(s)$ and extends it by allowing complex arguments s. That is

$$\zeta(s) = \sum_{n=1}^{\infty} \frac{1}{n^s} \text{ where } s = \sigma + it \text{ and } \sigma, t \in \mathbb{R}. \tag{4.4.3.1}$$

Recall that for real numbers x and t we have

$$x^{it} = e^{ix \ln t} = \cos(x \ln t) + i \sin(x \ln t).$$

It follows that $|x^{it}| = 1$. Therefore for each natural number n and $s = \sigma + it$ we have

$$|\frac{1}{n^s}| = |\frac{1}{n^{\sigma+it}}| = |\frac{1}{n^\sigma}||\frac{1}{n^{it}}| = |\frac{1}{n^\sigma}| = |\frac{1}{n^{\text{Re }(s)}}|.$$

Consequently by the p-series test the series in (4.4.3.1) converges absolutely for Re $(s) > 1$ and hence defines $\zeta(s)$ as an analytic function in this domain.

Since the basic formulas concerning the Euler product decomposition and those tying $\zeta(s)$ to the Van Mangoldt function hold on a connected arc (the part of the real line $s > 1$), by analytic continuation they are still valid for complex arguments within the domain of analyticity Re $s > 1$. Thus, we have

$$\zeta(s) = \prod_{p \text{ prime}} (\frac{1}{1 - p^{-s}}), s \in \mathbb{C}, \text{Re } s > 1;$$

$$-\frac{\zeta'(s)}{\zeta(s)} = \sum_{n=1}^{\infty} \frac{\Lambda(n)}{n^s}, s \in \mathbb{C}, \text{Re } s > 1;$$

and

$$\zeta(s)^{-1} = \sum_{n=1}^{\infty} \frac{\mu(n)}{n^s}, s \in \mathbb{C}, \text{Re } s > 1.$$

From the Euler product decomposition it is clear that $\zeta(s)$ has no zeros for Re $s > 1$.

The initial step in studying the zeta function and applying it to the proof of the Prime Number Theorem is to show that it can be continued analytically to a function, also denoted $\zeta(s)$, which is meromorphic in all of $\mathbb{C}$. This is accomplished in several steps but we next state the whole result.

Theorem 4.4.7 *The Riemann zeta function $\zeta(s)$ can be analytically continued to a function, also denoted $\zeta(s)$, which is meromorphic in the whole plane. The only singularity of $\zeta(s)$ is a simple pole at $s = 1$ with residue 1, that is,*

$$\zeta(s) = \frac{1}{s-1} + H(s)$$

where $H(s)$ is an entire function.

As remarked above, for $\operatorname{Re} s > 1$, it follows from the basic definition that $\zeta(s)$ is analytic. The first step is to analytically continue to a function that is analytic for $\operatorname{Re} s > 0$ except $s = 1$. To do this suppose first that $\operatorname{Re} s > 2$. Then

$$\zeta(s) = \sum_{n=1}^{\infty} \frac{1}{n^s} = \sum_{n=1}^{\infty} \frac{n}{n^s} - \sum_{n=1}^{\infty} \frac{n-1}{n^s}$$

$$= \sum_{n=1}^{\infty} \frac{n}{n^s} - \sum_{n=1}^{\infty} \frac{n}{(n+1)^s}$$

$$= \sum_{n=1}^{\infty} n(\frac{1}{n^s} - \frac{1}{(n+1)^s}) = \sum_{n=1}^{\infty} ns \int_n^{n+1} x^{-s-1} dx = s \sum_{n=1}^{\infty} \int_n^{n+1} [x] x^{-s-1} dx$$

$$= s \int_1^{\infty} [x] x^{-s-1} dx.$$

This final integral defines an analytic function of s for $\operatorname{Re} s > 1$ and therefore by the uniqueness of analytic continuation this integral formulation of $\zeta(s)$ holds for $\operatorname{Re} s > 1$.

Now consider the integral

$$s \int_1^{\infty} (x) x^{-s-1} dx = \frac{s}{s-1} = 1 + \frac{1}{s-1}.$$

Combining this with the integral representation of $\zeta(s)$ gives

$$\zeta(s) = \frac{1}{s-1} + 1 + s \int_1^{\infty} ([x] - x) x^{-s-1} dx. \tag{4.4.3.2}$$

The integral on the right-hand side converges for $\operatorname{Re} s > 0$ and hence for $\operatorname{Re} s > 0$ the right-hand side provides a meromorphic function with a simple pole at $s = 1$ with residue 1. Therefore, this provides an analytic continuation of $\zeta(s)$ to such a meromorphic function in the whole half-plane $\operatorname{Re} s > 0$.

To proceed further we need the following functional relation involving $\zeta(s)$ and $\zeta(1-s)$ and which ties the Riemann zeta function to the complex Gamma function (see Theorem 4.4.4).

Theorem 4.4.8 *The Riemann zeta function satisfies the functional relation*

$$\pi^{-s/2}\Gamma(\frac{s}{2})\zeta(s) = \pi^{-(1-s)/2}\Gamma(\frac{1-s}{2})\zeta(1-s)$$

or equivalently

$$\zeta(s) = 2^s\pi^{s-1}\sin(\frac{\pi s}{2})\Gamma(1-s)\zeta(1-s),\ s \neq 0, 1.$$

Proof The proof uses certain facts about the complex Gamma function and another function known as the **Jacobi theta function**. This latter function is defined as

$$\theta(u) = \sum_{n=-\infty}^{\infty} e^{-\pi n^2 u}.$$

Using the theory of Fourier transforms applied to the function $f(x) = e^{-\pi u x^2}$ it can be shown that the Jacobi theta function satisfies the functional relation

$$\theta(\frac{1}{u}) = \sqrt{u}\theta(u).$$

Now recall that

$$\Gamma(s) = \int_0^\infty x^{s-1}e^{-x}dx$$

so that

$$\Gamma(\frac{s}{2}) = \int_0^\infty x^{(s/2)-1}e^{-x}dx.$$

Applying the change of variables $y = \frac{x}{\pi n^2}$ this becomes

$$\pi^{-s/2}\Gamma(\frac{s}{2})n^{-s} = \int_0^\infty y^{(s/2)-1}e^{-\pi n^2 y}dy.$$

This will hold for each positive integer $n > 1$. Summing over all the positive integers we get

$$\pi^{-s/2}\Gamma(\frac{s}{2})\zeta(s) = \int_0^\infty \frac{1}{2}(\theta(y)-1)y^{(s/2)-1}dy = \int_0^\infty \theta_1(y)y^{(s/2)-1}dy \quad (4.4.3.3)$$

where $\theta_1(y) = \frac{1}{2}(\theta(y)-1)$.

If we make the new change of variable $z = \frac{1}{y}$ then we have from the functional relation on θ that

$$\theta(\frac{1}{y}) = \sqrt{y}\theta(y) \implies \theta(z) = \frac{\theta(\frac{1}{z})}{\sqrt{z}}.$$

Splitting the integral at $y = 1$ and using this change of variable gives us

$$\int_0^1 \theta_1(y)y^{(s/2)-1}dy = \frac{1}{s(s-1)} + \int_1^\infty \theta_1(z)z^{-(s+1)/2}dz.$$

Substituting this back into (4.4.3.3) we have

$$\pi^{-s/2}\Gamma(\frac{s}{2})\zeta(s) = \frac{1}{s(s-1)} + \int_1^\infty \theta_1(x)(x^{-(s+1)/2} + x^{(s/2)-1})dx. \qquad (4.4.3.4)$$

The integral on the right-hand side of (4.4.3.4) converges and hence defines an analytic function of s. Hence, the whole right-hand side defines a meromorphic function which is invariant under the transformation $s \to 1-s$. Therefore, the left-hand side must also be invariant under this transformation implying that

$$\pi^{-s/2}\Gamma(\frac{s}{2})\zeta(s) = \pi^{-(1-s)/2}\Gamma(\frac{1-s}{2})\zeta(1-s) \qquad (4.4.3.5)$$

which is the desired functional relation.

To obtain the equivalent formulation given in the statement of the theorem we use two properties of the Gamma function. The first is called the **formula of complements** and is given by

$$\Gamma(s)\Gamma(1-s) = \frac{\pi}{\sin(\pi s)}.$$

The second is called the **duplication formula** and is given by

$$\Gamma(s)\Gamma(s+\frac{1}{2}) = \sqrt{\pi}2^{1-2s}\Gamma(2s).$$

The duplication formula was originally given by Legendre. Using these formulas in (4.4.3.5) the relation becomes

$$\zeta(s) = 2^s\pi^{s-1}\sin(\frac{\pi s}{2})\Gamma(1-s)\zeta(1-s), s \neq 0, 1.$$

We leave the details to the exercises. □

Note that the functional relation has the form

$$\zeta(s) = K(s)\zeta(1-s)$$

where

$$K(s) = 2^s \pi^{s-1} \sin(\frac{\pi s}{2})\Gamma(1-s).$$

The transformation $s \to 1-s$ has $s = \frac{1}{2}$ as its center of symmetry. Therefore, since $\zeta(s)$ is defined for Re $s \geq \frac{1}{2}$ the functional equation can be used to continue $\zeta(s)$ to a function defined for Re $s \leq \frac{1}{2}$ and hence defined over the whole complex plane.

From the analytic continuation of the Gamma function it follows that the function $K(s)$ has singularities, that is, becomes infinite at the positive odd integers $2n+1$, $n \geq 1$. However, $\zeta(2n+1)$ is finite for all $n \geq 1$. Hence from the functional relation this is possible only is $\zeta(1-s) = 0$ if $s = 2n+1$. Therefore $\zeta(s) = 0$ at all the negative even integers $-2, -4, \ldots$. These are called the **trivial zeros** of $\zeta(s)$.

The functional equation also establishes that $s = 1$ is the only singularity of $\zeta(s)$ in the whole complex plane. This follows from the fact that $\zeta(s)$ has only a simple pole at $s = 1$ for Re $s \geq \frac{1}{2}$ and the only singularities of $K(s)$ are at the positive odd integers. Hence by analytic continuation this is true over the whole plane. Further the fact that $s = 1$ is a simple pole and that the residue is 1 follows from the integral representation of $\zeta(s)$ (4.4.3.2). These last comments complete the proof of Theorem 4.4.7.

What becomes crucial in applying the zeta function to the proof of the prime number theorem is the location of its zeros. In particular, we will see in the next section that the fact that $\zeta(s)$ has no zeros on the line Re $s = 1$ is equivalent to the prime number theorem. We have already seen that $\zeta(s)$ has zeros at $s = -2, -4, \ldots$. These are called the **trivial zeros**. Riemann in his original paper showed that any nontrivial zeros must fall in the **critical strip** $0 \leq$ Re $s \leq 1$. Further, he conjectured that all the nontrivial zeros lie along the line Re $s = \frac{1}{2}$ which is called the **critical line**. This is called the **Riemann hypothesis** and is still an open question. It has resisted solution for almost a hundred and fifty years and has had tremendous impact on both Number Theory and other branches of mathematics. Now that Fermat's last theorem has been settled, the Riemann hypothesis can be considered the outstanding open problem in mathematics. We will say more about the Riemann hypothesis after we show that there are no zeros on the line Re $s = 1$. This fact was the fundamental step in the proofs of both Hadamard and de la Valle Poussin of the prime number theorem. Their proofs were independent and appear different but are essentially the same (see [Na]).

Theorem 4.4.9 *The Riemann zeta function $\zeta(s)$ has no zeros on the line Re $s = 1$.*

Proof The proof we give is a simplification of the proofs of Hadamard and De La Valle Poussin and was given by Mertens in 1898 (see [Na]). The starting off point is the inequality

$$3 + 4\cos\theta + \cos(2\theta) = 2(1+\cos(\theta))^2 \geq 0 \text{ for all real } \theta.$$

Now suppose that $\zeta(1+it)=0$ for t real and $t\neq 0$. Then let

$$\phi(s)=\zeta^3(s)\zeta^4(s+it)\zeta(s+2it).$$

Since the pole at $s=1$ of $\zeta^3(s)$ cannot cancel the zero of $\zeta^4(s+it)$ it would follow that $\phi(s)$ is analytic and that

$$\ln|\phi(s)|\to-\infty \text{ as } s\to 1.$$

Now take s to be real and with $s>1$. By the Euler product decomposition

$$\ln|\zeta(s+it)|=\text{Re }(|\ln(\zeta(s+it))|)=-\text{Re }(\sum_p \ln(1-p^{-s-it}))=$$

$$\text{Re }(\sum_p(p^{-s-it}+\frac{1}{2}(p^2)^{-s-it}+\frac{1}{3}(p^3)^{-s-it}\cdots))$$

$$=\text{Re }(\sum_1^\infty a_n n^{-s-it}) \text{ with } a_n\geq 0.$$

Then

$$\ln|\phi(s)|=\text{Re }(\sum_1^\infty a_n n^{-s}(3+4n^{-it}+n^{-2it}))$$

$$=\sum_1^\infty a_n n^{-s}(3+4\cos(t\ln n)+\cos(2t\ln n)).$$

However, this last sum is ≥ 0 by the trigonometric inequality given at the beginning of the proof, contradicting the fact that the limit must go to $-\infty$. This contradiction then implies that $\zeta(1+it)\neq 0$.

□

Theorem 4.4.9 will imply the prime number theorem in roughly the following manner. This will be made precise in the next section. Recall that the prime number theorem is equivalent to $\psi(x)\sim x$ where $\psi(x)$ is the Chebyshev function. Therefore, we want to show that $\psi(x)\sim x$. Now

$$\psi(x)=\sum_{n\leq x}\Lambda(n) \text{ and } [x]=\sum_{n\leq x}1.$$

Therefore, we want to show that roughly as $x\to\infty$ the van Mangoldt function $\Lambda(n)$ looks like 1. We have further

$$-\frac{\zeta'(s)}{\zeta(s)} = \sum_{n=1}^{\infty} \frac{\Lambda(n)}{n^s}.$$

If Re $s > 1$ we can obtain an integral representation of this

$$-\frac{\zeta'(s)}{\zeta(s)} = s \int_1^{\infty} \psi(x) x^{-s-1} dx.$$

If there are no zeros of $\zeta(s)$ on the line Re $s = 1$ then by complex integration this integral can be handled and in turn used to show that $\psi(x) \sim x$.

Before closing this section we make some further comments on the zeros and on the Riemann hypothesis. Hardy in 1914 proved that $\zeta(s)$ has infinitely many zeros along the line Re $s = \frac{1}{2}$. As of 2002, it is known that at least the first billion and a half nontrivial zeros of $\zeta(s)$ lie along the critical line.

Selberg in 1942 showed that a positive proportion of the nontrivial zeros lie along the critical line. Levinson in 1974 improved this to show that at least $\frac{1}{3}$ of the nontrivial zeros are on the critical line. This has subsequently been improved to at least 40 % of the nontrivial zeros are on the critical line.

There are several quantitative statements that are equivalent to the Riemann hypothesis. Koch in 1901 showed that the Riemann hypothesis was equivalent to

$$\pi(x) = Li(x) + O(\sqrt{x} \ln x)$$

where $Li(x)$ is the logarithmic integral function of Gauss

$$Li(x) = \int_2^x \frac{1}{\ln t} dt.$$

In a similar manner, the Riemann hypothesis can be shown to be equivalent to

$$\pi(x) = Li(x) + O(x^{\frac{1}{2}+\epsilon}) \ \ \forall \epsilon > 0.$$

An entirely elementary formulation of the Riemann hypothesis is the following (see [P]). Define a positive squarefree integer n to be **red** if it is the product of an even number of distinct primes and **blue** if it is the product of an odd number of distinct primes. Let $R(n)$ be the number of red integers not exceeding n and $B(n)$ the number of blue integers not exceeding n. The Riemann hypothesis is equivalent to the statement that for any $\epsilon > 0$ there exists an N such that for all $n > N$

$$|R(n) - B(n)| < n^{\frac{1}{2}+\epsilon}.$$

We mention one major extension of the Riemann hypothesis. Recall that for an integer k a Dirichlet L-series is defined by

$$L(s,\chi)=\sum_{n=1}^{\infty}\frac{\chi(n)}{n^s}$$

where χ is a character mod k and s is a complex variable (see Chapter 3). Recall further that Dirichlet L-series also have Euler product representations. The **generalized Riemann hypothesis** is that the nontrivial zeros of any Dirichlet L-series also lie along the critical line Re $s=\frac{1}{2}$.

4.5 The Prime Number Theorem

We are now ready to prove the prime number theorem.

Theorem 4.5.1 $\pi(x)\sim\frac{x}{\ln x}$.

As we have already mentioned the proof is dependent on the fact that $\zeta(s)$ has no zeros on the line Re $s=1$. The original proofs were given by Hadamard and De La Valle Poussin and were quite complicated. An exposition and commentary on the original proofs can be found in the book of Narkuwieiz [Na]. The proof was somewhat simplified by Wiener and others but still remained quite complicated. In 1980, D.J. Newman found a way to give a proof using only fairly straightforward facts about complex integration and which allowed a relatively short proof to be presented. The proof we give is based on Newman's method.

In another direction in 1949 Selberg and then Erdos came up with an "elementary proof" of the prime number theorem along the lines that Chebyshev had begun a century earlier. This proof is elementary only in the sense that it does not use complex analysis and is in fact more complex, meaning complicated that the complex analytic proofs. We will say more about the elementary proof in the next section.

Newman's method is based on the following theorem and the subsequent corollary. We will state them and then show how they imply the proof of the prime number theorem. After this we will go back and prove them.

Theorem 4.5.2 *Let $F(t)$ be bounded on $(0,\infty)$ and integrable over every finite subinterval and suppose that the Laplace transform*

$$G(s)=\int_0^\infty F(t)e^{-st}dt$$

is well-defined and analytic throughout the open half-plane Re $s>0$. Suppose further that $G(s)$ can be continued analytically to a neighborhood of every point of the imaginary axis. Then

$$\int_0^\infty F(t)dt$$

exists and equals $G(0)$.

Corollary 4.5.1 *Let $f(x)$ be nonnegative, nondecreasing and $O(x)$ on $[1, \infty)$ so that the function*

$$g(s) = s \int_1^\infty f(x) x^{-s-1} dx$$

is well-defined and analytic throughout the half-plane Re $s > 1$. ($g(s)$ is called the **Mellin transform** *of $f(x)$). Suppose further that for some constant c the function*

$$G(s) = g(s) - \frac{c}{s-1}$$

can be continued analytically to a neighborhood of every point on the line Re $s = 1$. Then

$$\frac{f(x)}{x} \to c \textit{ as } x \to \infty.$$

The proof of the prime number theorem now follows easily from the corollary.

Proof (Theorem 4.5.1). Recall that the prime number theorem is equivalent to $\psi(x) \sim x$, that is,

$$\frac{\psi(x)}{x} \to 1 \text{ as } x \to \infty.$$

Take $f(x)$ in the corollary to be $\psi(x)$. $\psi(x)$ is nonnegative, nondecreasing and $O(x)$ on $[1, \infty)$ so we must show that the other conditions of the corollary apply. We have already seen (see Section 4.4) that

$$g(s) = s \int_1^\infty \psi(x) x^{-s-1} dx = -\frac{\zeta'(s)}{\zeta(s)}.$$

Since $\zeta(s)$ has a simple pole with residue 1 at $s = 1$ the same is then true of $g(s)$. The analyticity of $\zeta(s)$ at the points of Re $s = 1$ with $s \neq 1$ and its nonvanishing on this line then imply that $g(s)$ can be continued analytically to a neighborhood of each point on this line. Hence

$$G(s) = g(s) - \frac{1}{s-1}$$

has an analytic continuation to the closed half-plane Re $s \geq 1$. Therefore, the conditions of the corollary are met (with $c = 1$) and hence

$$\frac{\psi(x)}{x} \to 1 \text{ as } x \to \infty.$$

□

We now give the proofs of Theorem 4.5.2 and Corollary 4.5.1.

Proof (Theorem 4.5.2) We suppose that $F(t)$ is bounded on $(0, \infty)$ and that its Laplace transform

$$G(s) = \int_0^\infty F(t)e^{-st}dt$$

is well-defined and analytic throughout $\operatorname{Re} s > 0$. We suppose further that $G(s)$ can be continued analytically to a neighborhood of every point of the imaginary axis. Therefore, we have an analytic function, which we will also call $G(s)$ which is analytic on a neighborhood of $\operatorname{Re} s \geq 0$. Hence there is a $\delta > 0$, chosen small enough, such that $G(s)$ is analytic for $\operatorname{Re} s \geq -\delta$.

Since $f(t)$ is bounded, without loss of generality, we may assume that $|F(t)| \leq 1$ for $t > 0$. For $\lambda > 0$ let

$$G_\lambda(s) = \int_0^\lambda F(t)e^{-st}dt.$$

Since this is a finite integral and $F(t)$ is bounded, $G_\lambda(s)$ is analytic for all s and for all finite λ. We must show that

$$G_\lambda(0) = \int_0^\lambda F(t)dt \to G(0) \text{ as } \lambda \to \infty.$$

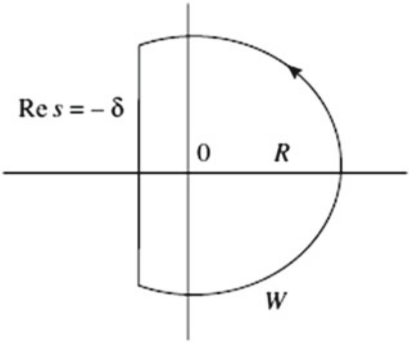
For $R > 0$ choose a $\delta = \delta(R)$ so that $G(s)$ is analytic on and within the closed curve W where W is given by the arc of the circle $|z| = R$ for $\operatorname{Re} s \geq -\delta$ together with the line segment $\operatorname{Re} s = -\delta$. We picture this in Figure 4.1.

We orient W to go counterclockwise and let W_+ be the part of W for $\operatorname{Re} s > 0$ and W_- the part of W for $\operatorname{Re} s < 0$.

Now for each λ the function $G(s) - G_\lambda(s)$ is analytic at $s = 0$. Therefore by the Cauchy integral formula (Theorem 4.4.6 part (3)) we have

$$G(0) - G_\lambda(0) = \frac{1}{2\pi i}\int_W \frac{G(z) - G_\lambda(z)}{z}dz. \qquad (4.5.1)$$

Fig. 4.1 Curve W



We have the following inequalities which will be needed to evaluate the final limit. First for $x = \text{Re } s > 0$,

$$|G(s) - G_\lambda(s)| = |\int_\lambda^\infty F(t)e^{-st}dt| \leq \int_\lambda^\infty e^{-xt}dt = \frac{1}{|x|}e^{-\lambda x}.$$

Next for $x = \text{Re } s < 0$

$$|G_\lambda(s)| = |\int_0^\lambda F(t)e^{-st}dt| \leq \int_0^\lambda e^{-xt}dt \leq \frac{1}{|x|}e^{-\lambda x}.$$

Next, if we let $H(z) = e^{\lambda z}G(z)$ and $H_\lambda(z) = e^{\lambda z}G_\lambda(z)$ then clearly $H(0) = G(0)$ and $H_\lambda(0) = G_\lambda(0)$ so

$$H(0) - H_\lambda(0) = G(0) - G_\lambda(0).$$

Further, within and on W, the function $\frac{(G(s)-G_\lambda(s))e^{\lambda s}s}{R^2}$ is analytic so that

$$\int_W \frac{(G(z) - G_\lambda(z))e^{\lambda z}z}{R^2}dz = 0$$

by Cauchy's Theorem. Therefore combining these observations with (4.5.1) we get

$$G(0) - G_\lambda(0) = H(0) - H_\lambda(0) = \frac{1}{2\pi i}\int_W (G(z) - G_\lambda(z))e^{\lambda z}(\frac{1}{z} + \frac{z}{R^2})dz.$$

On the circle $|z| = R$ we have

$$\frac{1}{z} + \frac{z}{R^2} = \frac{2x}{R^2}$$

and hence on W_+

$$|(G(z) - G_\lambda(z))e^{\lambda z}(\frac{1}{z} + \frac{z}{R^2})| \leq \frac{1}{x}e^{-\lambda x}e^{\lambda x}(\frac{2x}{R^2}) = \frac{2}{R^2}.$$

It follows that

$$|\frac{1}{2\pi i}\int_{W_+}(G(z) - G_\lambda(z))e^{\lambda z}(\frac{1}{z} + \frac{z}{R^2})dz| \leq \frac{1}{2\pi}\frac{2}{R^2}\pi R = \frac{1}{R}.$$

Now we consider the integral over W_-. Since $G_\lambda(s)$ is analytic for all s we may replace, using Cauchy's Theorem, the W_- path by the corresponding integral over the semicircle $W_-^* = |z| = R$, Re $z < 0$. Then by Cauchy's Theorem and our previous inequalities

$$|\frac{1}{2\pi i}\int_{W_-} G_\lambda(z)e^{\lambda z}(\frac{1}{z}+\frac{z}{R^2})dz| = |\frac{1}{2\pi i}\int_{W_-^*} G_\lambda(z)e^{\lambda z}(\frac{1}{z}+\frac{z}{R^2})dz| < \frac{1}{R}.$$

Now consider

$$|\frac{1}{2\pi i}\int_{W_-} G(z)e^{\lambda z}(\frac{1}{z}+\frac{z}{R^2})dz|. \tag{4.5.2}$$

Since $G(s)$ is analytic on W_- there exists a constant B depending on δ and on R such that

$$|G(s)(\frac{1}{s}+\frac{s}{R^2})| \leq B \text{ on } W_-.$$

It follows that

$$|G(s)e^{\lambda s}(\frac{1}{s}+\frac{s}{R^2})| \leq Be^{\lambda x} \text{ on } W_-.$$

Therefore on W_- where $x \leq -\delta < 0$ the integrand in (4.5.2) tends to zero uniformly as $\lambda \to \infty$. On the remaining small part of W_- (take $\delta_1 < \delta$ small) the integrand is bounded by B. Hence given a fixed W chosen as above the integral in (4.5.2) tends to zero as $\lambda \to \infty$.

Now we put all of this together. Given $\epsilon > 0$ choose $R = \frac{1}{\epsilon}$. Choose δ as above so that $G(s)$ is analytic within and on W. Finally determine a value λ_1 so that (4.5.1) is bounded by ϵ for all $\lambda > \lambda_1$. Combining then all the inequalities we get

$$|G(0) - G_\lambda(0)| < 3\epsilon \text{ for } \lambda > \lambda_1.$$

Therefore

$$G_\lambda(0) \to G(0) \text{ as } \lambda \to \infty.$$

□

The corollary follows in a relatively straightforward manner from this theorem.

Proof (Corollary 4.5.1) We suppose that $f(x)$ and $G(x)$ satisfy the conditions given in Corollary 4.5.1. That is $f(x)$ is nonnegative, nondecreasing, and $O(x)$ on $[1, \infty)$ and

$$g(s) = s\int_1^\infty f(x)x^{-s-1}dx$$

is well-defined and analytic throughout the half-plane Re $s > 1$. Further, there is constant c so that the function

$$G(s) = g(s) - \frac{c}{s-1}$$

can be continued analytically to a neighborhood of every point on the line Re $s = 1$.

Now let $x = e^t$ and define

$$F(t) = e^{-t}f(e^t) - c.$$

From the conditions on $f(x)$ it follows that $F(t)$ is bounded on $(0, \infty)$. The Laplace transform of $F(t)$ is given by

$$G(s) = \int_0^\infty (e^{-t}f(e^t) - c)e^{-st}dt$$

$$= \int_1^\infty f(x)x^{-s-2}dx - \frac{c}{s} - c = \frac{1}{s+1}(g(s+1) - \frac{c}{s} - c).$$

From the conditions on $g(s)$ it follows that $G(s)$ can be continued analytically to a neighborhood of every point of the imaginary axis.

Now let $t = -\ln x$ and apply Theorem 4.5.2 to $G(s)$. From this it follows that the improper integrals

$$\int_0^\infty (e^{-t}f(e^t) - c)dt = \int_1^\infty \frac{f(x) - cx}{x^2}dx \tag{4.5.3}$$

exist. Since $f(x)$ is an increasing function this implies that $\frac{f(x)}{x} \to c$ as $x \to \infty$.

To see this last assertion suppose that $\limsup \frac{f(x)}{x} > c$. Then there would exist a $\delta \geq 0$ such that for certain arbitrarily large y

$$f(y) > (c + 2\delta)y.$$

Since $f(x)$ is increasing it would then follow that

$$f(x) > (c + 2\delta)y > (c + \delta)x \text{ for } y < x < \sigma y$$

where $\sigma = \frac{c+2\delta}{c+\delta}$. Then

$$\int_y^{\sigma y} \frac{f(x) - cx}{x^2}dx > \int_y^{\sigma y} \frac{\delta}{x}dx = \delta \ln \sigma.$$

But this is bounded away from zero for arbitrarily large y contradicting that the improper integral in (4.5.3) converges. Therefore $\limsup \frac{f(x)}{x} \leq c$.

Next suppose that $\liminf \frac{f(x)}{x} < c$. Then in a similar manner there exists an interval $\sigma y < x < y$ with $\sigma < 1$ and $f(x) < (c - \delta)x$ on this interval. Applying this to the integral we obtain

$$\int_{\sigma y}^{y} \frac{f(x) - cx}{x^2}dx < \int_{\sigma y}^{y} (-\frac{\delta}{x})dx = \delta \ln \sigma.$$

This is negative and again bounded away from zero contradicting the convergence of the improper integrals. It follows that $\liminf \frac{f(x)}{x} \geq c$.

Since $\liminf \frac{f(x)}{x} \leq \limsup \frac{f(x)}{x}$ it follows that

$$\liminf \frac{f(x)}{x} = \limsup \frac{f(x)}{x} = c$$

and therefore, the limit exists and equals c completing the proof of the corollary. □

We have seen that the absence of zeros of $\zeta(s)$ on the line $\text{Re}\, s = 1$ implied the prime number theorem. It was pointed out by Wiener that the converse is also true and hence the prime number theorem is equivalent to the fact that there are no zeros of $\zeta(s)$ on $\text{Re}\, s = 1$.

Theorem 4.5.3 *The prime number theorem is equivalent to the fact that there are no zeros of $\zeta(s)$ on the line Re $s = 1$.*

Proof We have already seen that the absence of zeros implies the prime number theorem. Suppose now that $\psi(x) \sim x$ and $\zeta(1+it) = 0$ with t real and $t \neq 0$. Then, if the order of the zero is m we have the expansion

$$\zeta(s) = c(s-(1+it))^m + \cdots$$

which is valid on a neighborhood of $1+it$. Let

$$g(s) = -\frac{\zeta'(s)}{\zeta(s)} = \sum_{n=1}^{\infty} \frac{\Lambda(n)}{n^s}.$$

The expansion above would imply that

$$\lim_{\text{Re}\, s \to 1^+} (s-1)g(s+it) = -m.$$

Further

$$g(s) = \frac{s}{s-1} + s\int_1^{\infty} (\psi(y)-y)\frac{1}{y^{s+1}}dy \text{ with Re } s > 1.$$

Then since $\psi(y) \sim y$

$$|(s-1)g(s)| \leq |(s-1)s|(\frac{1}{|t|} + \int_0^{\infty} o(y^{-\text{Re}\, s})dy) = o(1)$$

as $\text{Re}\, s \to 1^+$. This would imply that $m = 0$ contradicting the existence of a zero on the line $\text{Re}\, s = 1$. □

4.6 The Elementary Proof

As we have noted Chebyshev's theorem (Theorem 4.2.1) appeared to be quite close to the Prime Number Theorem. It provided the right bounds and further Chebyshev showed that if $\lim_{x\to\infty} \frac{\pi(x)\ln x}{x}$ existed then the value of the limit must be one. Chebyshev's methods were elementary in the sense that they involved no analysis more complicated than simple real integration and the properties of the logarithmic function (although the proofs themselves were complicated). This would seem appropriate for a proof of a theorem about primes since primes are in the realm of arithmetic and should not require deep analytic notions. However Chebyshev could not establish that the limit exists and then Riemann, ten years or so later, tried a different approach using the theory of complex analytic functions. As discussed in the last section, the proof of the prime number theorem was reduced to knowing the location of the zeros of the complex analytic Riemann zeta function. Still, even with Riemann's ideas, the proof resisted solution for another 36 years and during this time many mathematicians began to doubt that the limit $\lim_{x\to\infty} \frac{\pi(x)\ln x}{x}$ exists. These doubts were put to rest with the proofs of Hadamard and de La Valle Poussin. As we have proved (Theorem 4.5.3) the prime number theorem, a result seemingly arising in arithmetic, is equivalent to the result that there are no zeros of the Riemann zeta function $\zeta(s)$ along the line $Re(s) = 1$, a result really in complex analysis. This raised the question of the actual relationship between the distribution of primes and complex function theory. This led to the further question of whether there could exist an elementary proof of the prime number theorem along the lines of Chebyshev's methods.

The opinion that came to prevail was that it was doubtful that such a proof existed. The feeling was that complex analysis was somehow *deeper* than real analysis and in view of the equivalence mentioned above it would be unlikely to prove the prime number theorem using just the methods of real analysis. On the other hand it was felt that if such a proof existed it would open up all sorts of new avenues in Number Theory.

The English mathematician G.H. Hardy, who made major contributions to the study of the relationship between the prime number function $\pi(x)$ and Gauss's logarithmic integral function $Li(x)$, described the situation this way in a lecture in 1921 (see [N]).

G.H. Hardy: *No elementary proof of the prime number theorem is known and one may ask whether it is reasonable to expect one. Now we know that the theorem is roughly equivalent to a theorem about an analytic function, the theorem that Riemann's zeta function has no roots on a certain line. A proof of such a theorem, not fundamentally dependent upon the ideas of the theory of functions, seems to me to be extraordinarily unlikely. It is rash to assert that a mathematical theorem cannot be proved in a particular way; but one thing seems quite clear. We have certain views about the logic of the theory; we think that some theorems, as we say "lie deep" and others nearer to the surface. If anyone produces an elementary proof of the prime number theorem, he will show that these views are wrong that the subject does not* *hang together in the way we have supposed, and that it is time for the books to be cast aside and for the theory to be rewritten.*

However what actually occurred was even more surprising. Selberg and then Erdos and then Erdos and Selberg together in 1948 developed elementary proofs of the prime number theorem along the lines of Chebyshev's methods. All of these proofs depended on asymptotic estimates for an extension of the von Mangoldt function. These asymptotic estimates are now called **Selberg formulae**. The discovery of this elementary proof put to rest the discussion of the relative profoundness of complex analysis versus real analysis. However, despite the brilliance of the Selberg-Erdos approach, it did not produce the startling consequences in understanding both the distribution of primes and the zeros of the Riemann zeta function that were predicted. There are now many so-called elementary proofs and the techniques involved have become standard in analytic number theory. It may be that in time these methods will lead to a deeper understanding of the basic questions.

In this section, we will state the Selberg formulae (without proof) and then outline (also without proof) how this formula leads to a proof of the prime number theorem. A complete exposition of Selberg's original proof can be found in the book of Nathanson [N] while a self-contained exposition of another elementary proof is in the book of Tenenbaum and Mendes-France [TMF]. A slightly different approach based on Selberg's methods can also be found in Hardy and Wright [HW].

The Selberg formula from which the elementary proof can be derived is the following.

Theorem 4.6.1 (Selberg Formula) *For $x \geq 1$,*

$$\sum_{p \leq x} (\ln p)^2 + \sum_{p,q \leq x} \ln p \ln q = 2x \ln x + O(x)$$

where p, q run over all the primes $\leq x$.

Several alternative formulations of this result are used in the elementary proof. First, the formula can be expressed in terms of the von Mangoldt function which we used in our other (nonelementary) proof. In particular:

Theorem 4.6.2 *(Selberg Formula) For $x \geq 1$,*

$$\sum_{n \leq x} \Lambda(n) \ln n + \sum_{n,m \leq x} \Lambda(n)\Lambda(n) = 2x \ln x + O(x)$$

where $\Lambda(n)$ is the von Mangoldt function.

To show that these are equivalent the two sums are considered separately. We give a partial demonstration. Consider, the first sum $\sum_{n \leq x} \Lambda(n) \ln n$. Since $\Lambda(n) = 0$ if $n \neq p^k$ for a prime p and $\Lambda(p^k) = \ln p$ we have

$$\sum_{n\le x}\Lambda(n)\ln n=\sum_{p\le x}(\ln p)^2+\sum_{p^k\le x,k\ge 2}k(\ln p)^2.$$

If $p^k\le x$ with $k\ge 2$ then $p\le\sqrt{x}$. Hence

$$\sum_{p^k\le x,k\ge 2}k(\ln p)^2=\sum_{p\le\sqrt{x}}(\ln p)^2\sum_{k=2}^{[\frac{\ln x}{\ln p}]}k$$

$$\le\sum_{p\le\sqrt{x}}(\ln p)^2(\frac{\ln x}{\ln p})^2\le\sqrt{x}(\ln x)^2.$$

However, clearly

$$\sqrt{x}(\ln x)^2=O(x)$$

and therefore, it follows that

$$\sum_{n\le x}\Lambda(n)\ln n=\sum_{p\le x}(\ln p)^2+O(x).$$

In a similar manner (see the outline in the exercises)

$$\sum_{n,m\le x}\Lambda(n)\Lambda(m)=\sum_{p,q\le x}\ln p\ln q+O(x).$$

Hence for $x\ge 1$,

$$\sum_{n\le x}\Lambda(n)\ln n+\sum_{n,m\le x}\Lambda(n)\Lambda(m)=2x\ln x+O(x)$$

if and only if

$$\sum_{p\le x}(\ln p)^2+\sum_{p,q\le x}\ln p\ln q=2x\ln x+O(x).$$

Therefore, the two versions given of Selberg's formula are equivalent.

If we introduce a generalization of the von Mangoldt function, Selberg's formula can be expressed in a very succinct manner. To do this, we must introduce some operations on the set of arithmetic functions.

Recall that a **number theoretic function** is any complex-valued function whose domain is the set of natural numbers $\mathbb{N}$ (see Section 3.6). We have introduced numerous examples of such functions: the von Mangoldt function, the Möbius function and the Euler phi function to name just a few. On the set of number theoretic functions we define addition in the standard way pointwise. That is, if $f(n)$, $g(n)$ are number theoretic functions then

$$(f+g)(n) = f(n) + g(n).$$

The function given by $0(n) = 0$ for all $n \in \mathbb{N}$ is then an additive identity for this addition.

We define a multiplication in the following manner.

Definition 4.6.1 *If $f(n), g(n)$ are number theoretic functions then their* **Dirichlet convolution** *is the number theoretic function given by*

$$f \star g(n) = \sum_{d|n} f(d)g(\frac{n}{d}).$$

If we define

$$\delta(n) = \begin{cases} 1 \text{ if } n = 1 \\ 0 \text{ if } n \geq 2 \end{cases}$$

then $\delta(n)$ is a multiplicative identity for Dirichlet convolution. With these operations the set of number theoretic functions becomes a ring.

Theorem 4.6.3 *The set of number theoretic functions with addition defined pointwise and multiplication given by Dirichlet convolution forms a commutative ring with an identity.*

The proof is a straightforward calculation (see the exercises).

We need the idea of **Möbius inversion** (see Section 3.6). Recall that the **Möbius function** μ is defined for natural numbers n by

$$\mu(n) = \begin{cases} 1 & \text{if } n = 1 \\ (-1)^r & \text{if } n = p_1p_2\cdots p_r \text{ with } p_1, \ldots, p_r \text{ distinct primes} \\ 0 & \text{otherwise.} \end{cases}$$

For number theoretic functions we then have the following formula known as the **Möbius inversion formula** which was stated and proved in Section 3.6.

Theorem 4.6.4 *(Theorem 3.6.4) (Möbius Inversion Formula) Let $f(n)$ be a number theoretic function. Define*

$$g(n) = \sum_{d|n} f(d).$$

Then

$$f(n) = \sum_{d|n} \mu(d)g(\frac{n}{d}).$$

Based on Dirichlet convolution and using Möbius inversion we define a generalization of the von Mangoldt function. First, define

$$L(n) = \ln n \text{ for all } n \in \mathbb{N}.$$

We then have:

Lemma 4.6.1 *$\Lambda(n) = \mu \star L(n)$ where μ is the Möbius function.*

Proof Let $1(n) = n$ for all $n \in \mathbb{N}$. Then, if $n = p_1^{e_1} \cdots p_k^{e_k}$ we have

$$1 \star \Lambda(n) = \sum_{d|n} d\Lambda(\frac{n}{d}) = \sum_{d_1 d_2 = n} d_1 \Lambda(d_2)$$

$$= e_1 \ln p_1 + \cdots + e_k \ln p_k = \ln n = L(n).$$

Therefore $1 \star \Lambda = L$ and so from the Möbius inversion formula

$$\mu \star L = \Lambda.$$

□

Definition 4.6.2 *For each $r \geq 1$ define the* **generalized von Mangoldt function** *$\Lambda_r = \mu \star L^r$.*

The tie to the Selberg formula is the following.

Lemma 4.6.2 *For each natural number n,*

$$\Lambda_2(n) = \Lambda(n) \ln n + \Lambda \star \Lambda(n).$$

Selberg's formula can now be expressed concisely as

Theorem 4.6.5 *(Selberg formula) For all $x \geq 1$*

$$\sum_{n \leq x} \Lambda_2(n) = 2x \ln x + O(x).$$

The elementary proof requires two more equivalent formulations which tie the Selberg formula to the Chebyshev functions $\theta(x)$ and $\psi(x)$.

Theorem 4.6.6 *(Selberg Formula) For $x \geq 1$*

$$(1)\ \ \theta(x) \ln x + \sum_{p \leq x} (\ln p) \theta(\frac{x}{p}) = 2x \ln x + O(x),$$

$$(2)\ \ \psi(x) \ln x + \sum_{n \leq x} \Lambda(n) \psi(\frac{x}{n}) = 2x \ln x + O(x).$$

In Theorem 4.3.2, we showed that the prime number theorem is equivalent to $\theta(x) \sim x$ and to $\psi(x) \sim x$. In our earlier (nonelementary) proof we actually showed that $\psi(x) \sim x$ to establish the prime number theorem. In Selberg's elementary proof he showed that $\theta(x) \sim x$. In particular, if we let $R(x) = \theta(x) - x$ then the Selberg proof shows that $R(x) = o(x)$ which clearly implies that $\theta(x) \sim x$. More precisely, in the proof it is shown that there exists sequences (a_n), (b_n) of positive real numbers such that

$$|R(x)| \leq a_n x \text{ for all } x \geq b_n$$

and $\lim_{n \to \infty} a_n = 0$.

This is proved via a series of estimates whose proofs all work with, or start with, the Selberg formula (in one of its formulations), and then use tricky and difficult manipulation of series. The lengthy details of a completely elementary (again not simple but no complex analysis) proof due to Selberg can be found in the book of Nathanson [N]. A separate proof along the same lines but using some analysis is in the book of Hardy and Wright [HW]. Finally, a separate elementary proof (again using some analysis) is in the notes of Tenenbaum and Mendes-France [TMF].

It is an easy consequence of the prime number theorem that if p_n is the nth prime then

$$\lim_{n \to \infty} \frac{p_{n+1}}{p_n} = 1. \tag{4.6.1}$$

This fact however plays a role in the history of the elementary proof. When Selberg first gave his formula Erdos used it to give an elementary proof of (4.6.1). Selberg then used his formula along with the methods of Erdos' proof to develop the first elementary proof of the prime number theorem. Erdos then gave a second elementary proof. There now exist several elementary proofs of the prime number theorem that do not depend on Selberg's formula. A nice survey on the use of elementary methods in the study of primes was written by Diamond [Di].

4.7 Multiple Zeta Values

Throughout this chapter we have seen the importance of the zeta function

$$\zeta(s) = \sum_{n=1}^{\infty} \frac{1}{n^s}.$$

For real values of s this is the Euler zeta function and we presented Euler's proof of the infinitude of primes based on this. For complex values of s this is the Riemann zeta funciton and was essential in the proof of the prime number theorem. Recall that Riemann reduced the proof of the prime number theorem to showing that the zeros of the Riemann zeta function are off the line Re $s = 0$. The Riemann hypothesis developed from Riemann's work and states that all the nontrivial zeros of the Riemann zeta function are on the line Re $s = \frac{1}{2}$ (see Section 4.4). The Riemann hypothesis is among the most important open problems in mathematics. However, there are many other important problems concerning $\zeta(s)$ and there is an entire line of research devoted to these as well as to generalizations of $\zeta(s)$. In this section, we introduce and briefly discuss a generalization of $\zeta(s)$ called **multiple zeta values** or **MZV**. This generalization, besides being of independent interest, also sheds light on the zeta function itself. For further information, as well as for the proofs of the results in this section, we refer to the paper by Burgos Gil, Fresan and Kühn [BFK], or the survey articles [Wa] and [Zud]. Our short discussion follows the description in [BFK].

Before introducing the MZV we look back at certain results on the zeta function and repeat some of the material that we looked at in Section 4.4. Euler considered the problem of determining the value of $\zeta(m)$ for an integer m. The *Basel problem*, solved by Euler in 1735, asked for the value of $\zeta(2)$. In Section 4.4, we showed that $\zeta(2) = \frac{\pi^2}{6}$ by using Fourier series.

However, as mentioned, Euler proved a great deal more. In particular, he determined the values of $\zeta(m)$ for all even integers $m = 2k$. The result depends on the **Bernoulli numbers** which are rational numbers defined by the power series identity

$$\frac{t}{e^t - 1} = 1 + \sum_{k=1}^{\infty} B_k \frac{t^k}{k!}. \tag{4.7.1}$$

Note that the function

$$f(t) = \frac{t}{e^t - 1} + \frac{1}{2}t = \frac{t(1+e^t)}{2(e^t - 1)}$$

is even, that is, $f(t) = f(-t)$. It follows that $B_1 = -\frac{1}{2}$ and $B_k = 0$ for all odd integers $k \geq 3$. The first Bernoulli numbers are easily computed:

k	2	4	6	8	10	12
B_k	$\frac{1}{6}$	$-\frac{1}{30}$	$\frac{1}{42}$	$-\frac{1}{30}$	$\frac{5}{66}$	$-\frac{691}{2730}$

We now give Euler's theorem and a general proof which, given a certain trigonometric identity, is straightforward.

Theorem 4.7.1 *(Euler, 1735) The values of the zeta function at even positive integers are given by*

$$\zeta(2k) = (-1)^{k-1} \frac{(2\pi)^{2k}}{2(2k)!} B_{2k},$$

where B_{2k} are the Bernoulli numbers.

Proof The key ingredient is the following identity for the cotangent function, also due to Euler (see exercises): for $x \in \mathbb{C} \setminus \mathbb{Z}$,

$$\pi \cot(\pi x) = \frac{1}{x} + \sum_{n=1}^{\infty} \frac{2x}{x^2 - n^2}. \tag{4.7.2}$$

Expanding the quotient inside the sum sign as a geometric series and interchanging the order of summation, we obtain

$$\pi \cot(\pi x) = \frac{1}{x} - 2\sum_{k=1}^{\infty} \zeta(2k) x^{2k-1}. \tag{4.7.3}$$

On the other hand, we have

$$\frac{1}{e^t - 1} = \frac{e^{-\frac{t}{2}}}{e^{\frac{t}{2}} - e^{-\frac{t}{2}}} \quad \text{and} \quad -\frac{1}{e^{-t} - 1} = \frac{e^{\frac{t}{2}}}{e^{\frac{t}{2}} - e^{-\frac{t}{2}}},$$

from which the identity

$$\frac{e^{\frac{t}{2}} + e^{-\frac{t}{2}}}{e^{\frac{t}{2}} - e^{-\frac{t}{2}}} = \frac{2}{t} + 2\sum_{k=1}^{\infty} \frac{B_{2k} t^{2k-1}}{(2k)!}$$

follows, using this and the vanishing of B_k for odd k. Hence:

$$\pi \cot(\pi x) = \pi i \frac{e^{\frac{2\pi i x}{2}} + e^{-\frac{2\pi i x}{2}}}{e^{\frac{2\pi i x}{2}} - e^{-\frac{2\pi i x}{2}}} = \frac{1}{x} + \sum_{k=1}^{\infty} \frac{(2\pi i)^{2k} B_{2k}}{(2k)!} x^{2k-1} \tag{4.7.4}$$

and we conclude by identifying the coefficients in (4.7.3) and (4.7.4). □

Euler's formula implies the following equality of subrings of $\mathbb{R}$:

$$\mathbb{Q}[\zeta(2), \zeta(4), \ldots] = \mathbb{Q}[\pi^2].$$

Thanks to the functional equation (see Section 4.4)

$$\pi^{-\frac{s}{2}} \Gamma\left(\frac{s}{2}\right) \zeta(s) = \pi^{-\frac{1-s}{2}} \Gamma\left(\frac{1-s}{2}\right) \zeta(1-s),$$

it follows that $\zeta(-k) = -\frac{B_{k+1}}{k+1}$ for all $k \geq 1$.

By contrast, no one has been able to determine closed analytic formulas for the values of the zeta function at $s = 3, 5, 7, \ldots$, in terms of previously known real numbers like π. This leads to the following conjecture:

Conjecture 4.7.1 *The numbers* $1,\ \pi,\ \zeta(3),\ \zeta(5), \ldots$ *are algebraically independent.*

Real numbers $k_1, k_2, ..., k_n$ are **algebraically independent** if

$$P(k_1, ..., k_n) \neq 0$$

for each $n \geq 1$ and each nonzero polynomial $P \in \mathbb{Z}[x_1, \ldots, x_n]$.

This conjecture seems to be completely out of reach. The transcendence of π was proved by Lindemann in 1882. By Euler's result it follows that the numbers $\zeta(2k)$ are transcendental. However, we do not know whether $\zeta(3)$ is transcendental, not to speak of the algebraic independence with π. The are a few known results about the nature of the numbers $\zeta(2k+1)$. We summarize them (see [BFK]):

- *Apéry* proved that $\zeta(3)$ is irrational. Different proofs are known, but no one has been able to generalize them to show that for example $\zeta(5)$ is irrational.
- *Rivoal* and *Ball* and *Rivoal* proved that, if n is an odd integer ≥ 3, then

  $$\dim_{\mathbb{Q}}\langle 1, \zeta(3), \zeta(5), \ldots, \zeta(n)\rangle \geq \frac{1}{3}\log(n).$$

  In particular, infinitely many $\zeta(2k+1)$ are irrational.
- *Zudilin* proved that at least one out of the four numbers $\zeta(5)$, $\zeta(7)$, $\zeta(9)$ and $\zeta(11)$ is irrational.

Besides the algebraic independence conjecture, the values of the Riemann zeta function at the integers are linked to many other interesting problems in mathematics (see [BFK]).

In order to investigate possible relations among zeta values, Euler examined the algebraic structure of these numbers. If we multiply two Riemann zeta values we obtain a new kind of interesting sum;

$$\begin{aligned}\zeta(s_1)\cdot\zeta(s_2) &= \left(\sum_{n_1\geq 1}\frac{1}{n_1^{s_1}}\right)\cdot\left(\sum_{n_2\geq 1}\frac{1}{n_2^{s_2}}\right)\\ &= \sum_{n_1,n_2\geq 1}\frac{1}{n_1^{s_1}n_2^{s_2}}\\ &= \sum_{n_1>n_2\geq 1}\frac{1}{n_1^{s_1}n_2^{s_2}} + \sum_{n_2>n_1\geq 1}\frac{1}{n_2^{s_2}n_1^{s_1}} + \sum_{n=n_1=n_2\geq 1}\frac{1}{n^{s_1+s_2}}. \qquad (4.7.5)\end{aligned}$$

The first two terms in the last line are called *double zeta values* and have the various representations

$$\begin{aligned}\zeta(s_1, s_2) &= \sum_{n_1>n_2\geq 1} \frac{1}{n_1^{s_1} n_2^{s_2}}\\ &= \sum_{n=2}^{\infty} \frac{1}{n^{s_1}}\left(1 + \frac{1}{2^{s_2}} + \cdots + \frac{1}{(n-1)^{s_2}}\right)\\ &= \sum_{m,n\geq 1} \frac{1}{(n+m)^{s_1} n^{s_2}}.\end{aligned}$$

With this notation, equation (4.7.5) can be rewritten as

$$\underbrace{\zeta(s_1)\cdot\zeta(s_2)}_{\text{product of zeta values}} = \underbrace{\zeta(s_1, s_2) + \zeta(s_2, s_1) + \zeta(s_1 + s_2)}_{\substack{\text{linear combination of zeta}\\ \text{and double zeta values}}}. \tag{4.7.6}$$

As we have seen, products of two zeta values are linear combinations of zeta and double zeta values. To handle products of more factors we shall need to generalize double zeta values to multiple zeta values (MZV). These new numbers satisfy many linear relations with rational coefficients and the main goal of the theory of MZV is to fully understand them.

We now define multiple zeta values and begin to study its basic properties. Of great importance is that multiple zeta values can be written both as series and as integrals. There are two important operations uses in the study of MZV, the **stuffle product** and the **shuffle product**. From the series representation of MZV one obtains the stuffle product, whereas the integral representation gives the shuffle product. By comparing both products one obtains many relations among multiple zeta values. We will not delve into these products in this short introduction. The definitions of the stuffle and shuffle product can be found in [BFK].

In order to define MZV we introduce the following terminology: a tuple

$$\mathbf{s} = (s_1, \ldots, s_l) \in \mathbb{Z}^l$$

is said to be **positive** if $s_i \geq 1$ for all $i = 1, \ldots, l$ and **admissible** if, in addition, $s_1 \geq 2$. By convention, the empty tuple will also be considered to be admissible.

Definition 4.7.1 *Let* $\mathbf{s} = (s_1, s_2, \ldots, s_l) \in \mathbb{Z}^l$ *be an admissible tuple. The* **multiple zeta value** *associated to* $\mathbf{s}$ *is the real number*

$$\zeta(\mathbf{s}) = \zeta(s_1, s_2, \ldots, s_l) = \sum_{n_1>n_2>\cdots>n_l\geq 1} \frac{1}{n_1^{s_1}\cdot n_2^{s_2}\cdots n_l^{s_l}},$$

with the convention $\zeta(\emptyset) = 1$.

Note that if $\mathbf{s}$ is an admissible tuple, then the series $\zeta(\mathbf{s})$ converges absolutely.

We define the **weight** of $\zeta(\mathbf{s})$ as $s_1 + s_2 + \cdots + s_l$ and the **length** of $\zeta(\mathbf{s})$ (also called the **depth** in the literature) as l, and we write

$$\mathrm{wt}(\zeta(\mathbf{s})) = \mathrm{wt}(\mathbf{s}) = s_1 + \cdots + s_l$$

$$l(\zeta(\mathbf{s})) = l(\mathbf{s}) = l.$$

In particular, $\mathrm{wt}(1) = l(1) = 0$.

Definition 4.7.2 *We will denote by $\mathcal{Z}$ the $\mathbb{Q}$-vector space generated by all multiple zeta values*

$$\mathcal{Z} = \mathbb{Q}[MZV] = \langle 1, \zeta(2), \zeta(3), \zeta(2,1), \zeta(4), \ldots \rangle_{\mathbb{Q}}.$$

We also define the following subvector spaces of $\mathcal{Z}$:

$$\mathcal{Z}_k = \langle \zeta(\mathbf{s}) \mid wt(\mathbf{s}) = k \rangle_{\mathbb{Q}},$$

$$\mathcal{F}_l\mathcal{Z} = \langle \zeta(\mathbf{s}) \mid l(\mathbf{s}) \leq l \rangle_{\mathbb{Q}},$$

$$\mathcal{F}_l\mathcal{Z}_k = \langle \zeta(\mathbf{s}) \mid wt(\mathbf{s}) = k,\ l(\mathbf{s}) \leq l \rangle_{\mathbb{Q}}.$$

Observe that there is an obvious inclusion $\mathcal{F}_l\mathcal{Z}_k \subset \mathcal{F}_l\mathcal{Z} \cap \mathcal{Z}_k$. This is actually expected to be an equality, but not known so far.

This is the first indication that the $\mathbb{Q}$-vector space $\mathcal{Z}$ has the structure of an algebra.

Theorem 4.7.2 *Multiplication of real numbers induces an algebra structure on $\mathcal{Z}$ which is compatible with the weight and the length, that is:*

$$\mathcal{F}_{l_1}\mathcal{Z}_{k_1} \cdot \mathcal{F}_{l_2}\mathcal{Z}_{k_2} \subset \mathcal{F}_{l_1+l_2}\mathcal{Z}_{k_1+k_2}$$

for any integers l_1, l_2, k_1 and k_2.

The theorem shows in particular, that every product of zeta or multiple zeta values can be written as a linear combination of MZV.

Corollary 4.7.1 *Every polynomial relation between Riemann zeta values $\zeta(k)$ gives rise to a linear relation between multiple zeta values.*

Thus, the problem of finding algebraic relations among zeta values is reduced to the problem of finding linear relations among MZV: we have *linearized* the algebraic independence conjecture.

The task of finding linear relations among multiple zeta values by elementary methods was first done by reordering multiple sums by means of a partial fraction decomposition. Nielsen proved the following reduction formula in 1906.

Theorem 4.7.3 (Reduction formula, Nielsen 1906). *Let $p \geq 2$ and $q \geq 1$ be integers. Then*

$$\begin{aligned}\zeta(p,q) = \sum_{k=0}^{q-2}(-1)^k\tbinom{p+k-1}{p-1}\zeta(q-k)\zeta(p+k)\\ + (-1)^q\sum_{k=0}^{p-2}\tbinom{q+k-1}{q-1}\zeta(p-k,q+k)\\ + (-1)^{q-1}\tbinom{p+q-2}{p-1}[\zeta(p+q)+\zeta(p+q-1,1)].\end{aligned}$$

Making $q = 1$ we immediately get:

Corollary 4.7.2 (Euler's sum formula). *If $s \geq 3$, then*

$$\zeta(s) = \sum_{j=1}^{s-2}\zeta(s-j,j). \tag{4.7.7}$$

In particular $\zeta(3) = \zeta(2,1)$.

This ties the individual values of the zeta function to the values of the MZV. Nielsen extended this.

Corollary 4.7.3 *(see [BFK]) If $n \geq 2$, the following equalities hold:*

$$\sum_{r=1}^{n-1}\zeta(2r,2n-2r) = \frac{3}{4}\zeta(2n),$$

$$\sum_{r=1}^{n-1}\zeta(2r+1,2n-2r-1) = \frac{1}{4}\zeta(2n).$$

Corollary 4.7.4 *In $\mathcal{Z}$ we have the following linear relations:*

1. *in weight 3:*
$$\zeta(3) = \zeta(2,1).$$
2. *in weight 4:*
$$\zeta(4) = 4\zeta(3,1), \quad \zeta(2,2) = 3\zeta(3,1).$$
3. *in weight 5:*
$$\zeta(5) = -4\zeta(4,1) + 2\zeta(2,3), \quad \zeta(3,2) = \zeta(2,3) - 5\zeta(4,1).$$
4. *in weight 6:*
$$\zeta(6) = 4\zeta(5,1) + 4\zeta(3,3), \quad \zeta(2,4) = \frac{13}{3}\zeta(5,1) + \frac{7}{3}\zeta(3,3),$$

$$\zeta(4,2) = -\frac{4}{3}\zeta(5,1) + \frac{2}{3}\zeta(3,3).$$

Numerical experiments show the following further relations:

- $\zeta(2,2) = \frac{3}{4}\zeta(4)$;
- $\zeta(2,1,1) = \zeta(4)$;
- $\zeta(5) = 4\zeta(3,2) + 6\zeta(2,3)$;
- $\zeta(5) = \zeta(2,1,1,1)$;
- $\zeta(4,1) = \zeta(3,1,1)$;
- $\zeta(2,1,2) = \zeta(2,3)$;
- $\zeta(2,1,1) = \zeta(3,2)$.

From these identities we can obtain upper bounds for the dimension of the $\mathbb{Q}$-vector space $\mathcal{F}_2\mathcal{Z}_k$ generated by zeta and double zeta values of weight k:

Theorem 4.7.4 *Let $k > 3$, then $\mathcal{F}_2\mathcal{Z}_k$ is spanned by $\zeta(k)$ and $\zeta(r, k-r)$ for $r \geq (k-1)/2$. In particular, we have*

$$\dim \mathcal{F}_2\mathcal{Z}_k \leq \lceil \frac{k-2}{2} \rceil$$

the smallest integer greater than or equal to $\frac{k-2}{2}$.

As we have seen there are many linear relations between MZV. A major line of research is to determine the structure of the algebra of MZV and find all possible linear relations among them. After extensive experimentation by many mathematicians, no nontrivial linear relation between MZV of different weight has been found. That is, all known relations among multiple zeta values are generated by homogeneous relations (see [BFK]).

An important conjecture concerning the structure of the algebra of MZV was given by Zagier. The conjecture concerns the dimension of the space of multiple zeta values and there is large numerical evidence for it. In order to state the Zagier conjecture we introduce some Fibonacci type numbers. Set $d_0 = 1$, $d_1 = 0$, $d_2 = 1$ and, for $k \geq 3$,

$$d_k = d_{k-2} + d_{k-3}$$

These numbers are determined by the power series identification

$$\sum_{k=0}^{\infty} d_k t^k = \frac{1}{1 - t^2 - t^3}.$$

Conjecture 4.7.2 *(Zagier) (see [BFK])*

1. *The weight defines a graduation on $\mathcal{Z}$. That is,*

$$\mathcal{Z} = \bigoplus_{k \geq 0} \mathcal{Z}_k$$

and, in particular, $\mathcal{Z}_k \cap \mathcal{Z}_{k'} = 0$ if $k \neq k'$.

2. $\dim_{\mathbb{Q}} \mathcal{Z}_k = d_k$.

There have been several extensions and refinements of Zagier's conjecture (see [BFK]). The following two results, the first by Terasoma and by Delgne and Goncharov and the second by Brown show what is known about the Zagier conjecture and its extensions (see [BFK] for references).

Theorem 4.7.5 *We have* $\dim \mathcal{Z}_k \leq d_k$.

Theorem 4.7.6 *The MZV with only* 2 *and* 3 *in their index do generate $\mathcal{Z}$, and in particular we obtain* $\dim \mathcal{Z}_k \leq d_k$.

4.8 Some Extensions and Comments

In Chapter 3, we looked at a large number of ways to prove that there are infinitely many primes and our look led us to a large array of number theoretical ideas. Basic congruences and the fundamental theorem of arithmetic handled many of the proofs but we used some elementary analysis to show that $\sum \frac{1}{p}$ diverges. We then used some more difficult analysis to prove that there are infinitely many primes in any arithmetic progression $\{an + b; n \in \mathbb{N}\}$ with $(a, b) = 1$. However, despite the fact the set of primes is infinite it is clear that the density of primes among the natural numbers thins out as the natural numbers get larger. In fact we showed (Theorem 2.3.2) that there are arbitrarily large gaps in the sequence of primes. Hence in this chapter we looked at the density of the sequence of primes. The major result was the prime number theorem which said that $\pi(x) \sim \frac{x}{\ln x}$ as $x \to \infty$ where $\pi(x)$ is the number of primes less than or equal to x. However, we have just touched the tip of the iceberg relative to the study of the distribution of primes. In this section, we mention some further results and conjectures on primes and their distribution which are in the same spirit as the results and proofs of the last two chapters.

By far the most important open problem surrounding the distribution of primes and the Prime Number Theorem is the Riemann hypothesis. We introduced this at the end of Section 4.4 but here we repeat what we said at that point and extend somewhat our comments and observations. Recall that the Riemann zeta function was defined for all $s > 1$ by

$$\zeta(s) = \sum_{n=1}^{\infty} \frac{1}{n^s}.$$

This could be continued analytically to a meromorphic function also denoted $\zeta(s)$ which is analytic for all complex $s \neq 1$ and which has a simple pole at $s = 1$. This fact follows from the fact that $\zeta(s)$ satisfies a functional relation

$$\zeta(s) = K(s)\zeta(1-s)$$

where

$$K(s) = 2^s \pi^{s-1} \sin(\frac{\pi s}{2})\Gamma(1-s).$$

This functional relation also establishes that $\zeta(s) = 0$ at all the negative even integers $-2, -4, \ldots$. These are called the **trivial zeros** of $\zeta(s)$. Riemann in his original paper showed that any nontrivial zeros must fall in the **critical strip** $0 \leq \text{Re } s \leq 1$. He further showed that if $\zeta(s)$ has no zeros on the line $\text{Re } s = 1$ this was sufficient to prove the prime number theorem. This final fact was proven by Hadamard and de la Valle Poussin. In the course of this investigation Riemann conjectured that all the nontrivial zeros lie along the line $\text{Re } s = \frac{1}{2}$ which is called the **critical line**. This is the common form of the Riemann hypothesis.

Riemann Hypothesis: *All the nontrivial zeros of the Riemann zeta function lie along the line* $\text{Re } (s) = \frac{1}{2}$.

The Riemann hypothesis has resisted solution for almost a hundred and fifty years and has had tremendous impact on both Number Theory and other branches of mathematics. Now that Fermat's last theorem has been settled the Riemann hypothesis can be considered the outstanding open problem in mathematics. There are various further results concerning the Riemann hypothesis and the zeros of the zeta function. Hardy in 1914 proved that $\zeta(s)$ has infinitely many zeros along the critical line $\text{Re } s = \frac{1}{2}$. As of 2002 it is known that at least the first billion and a half nontrivial zeros of $\zeta(s)$ lie along the critical line.

Selberg in 1942 showed that a positive proportion of the nontrivial zeros lie along the critical line. Levinson in 1974 improved this to show that at least $\frac{1}{3}$ of the nontrivial zeros are on the critical line. This has subsequently been improved to at least 40% of the nontrivial zeros are on the critical line.

There are several quantitative statements that are equivalent to the Riemann hypothesis. Koch in 1901 showed that the Riemann hypothesis was equivalent to

$$\pi(x) = Li(x) + O(\sqrt{x} \ln x) \tag{4.8.1}$$

where $Li(x)$ is the logarithmic integral function of Gauss

$$Li(x) = \int_2^x \frac{1}{\ln t} dt.$$

In a similar manner the Riemann hypothesis can be shown to be equivalent to

$$\pi(x) = Li(x) + O(x^{\frac{1}{2}+\epsilon}) \ \ \forall \epsilon > 0.$$

The equality (4.8.1) was also conjectured by Riemann in his original paper and is often called the prime number theorem form of the Riemann hypothesis.

There are many other computational variations of both the prime number theorem and the Riemann hypothesis. Many of these are discussed in the excellent book by Crandall and Pomerance [CP]. Several of these involve the Möbius function $\mu(n)$ and **Merten's function** defined by

$$M(x) = \sum_{n \leq x} \mu(x).$$

Merten's function is related to the Riemann zeta function by (see Section 4.4.3)

$$\frac{1}{\zeta(s)} = \sum_{n=1}^{\infty} \frac{\mu(n)}{n^s} = s \int_1^{\infty} \frac{M(x)}{x^{s+1}} dx.$$

Van Mangoldt proved the following.

Theorem 4.8.1 *The prime number theorem is equivalent to the statement*

$$\sum_{n=1}^{\infty} \frac{\mu(n)}{n} = 0.$$

Further the following is also known.

Theorem 4.8.2 *If $M(x)$ is Merten's function then:*
(1) the prime number theorem is equivalent to

$$M(x) = o(x).$$

(2) the Riemann hypothesis is equivalent to

$$M(x) = O(x^{\frac{1}{2}+\epsilon}) \textit{ for any fixed } \epsilon > 0.$$

One of the questions that arises from the prime number theorem is which function exactly is the "best approximation" to $\pi(x)$. Note that for any positive real numbers A, B we have that $\frac{x}{A \ln x + B}$ is asymptotically equal to $Li(x)$. Hence

(1) $\pi(x) \sim \frac{x}{\ln x}$,
(2) $\pi(x) \sim \frac{x}{\ln x - a}$ for $a > 0$,
(3) $\pi(x) \sim \frac{x}{\ln x - 1.08366}$ (Legendre's estimate),
(4) $\pi(x) \sim Li(x)$ (Gauss),

are all equivalent to the prime number theorem. The question arises as to whether there is an optimal value for a in (2) above. Empirical evidence is that $a = 1$ is an optimal choice and generally better for large x than Legendre's 1.08366 and better than Gauss' $Li(x)$. The table below compares the estimates.

x	$\pi(x)$	$\frac{x}{\ln x}$	$Li(x)$	$\frac{x}{\ln x - 1.08366}$	$\frac{x}{\ln x - 1}$
10^3	168	145	178	172	169
10^4	1229	1086	1246	1231	1218
10^5	9592	8686	9630	9588	9512
10^6	78498	72382	78628	78534	78030
10^7	664579	620420	664918	665138	661459
10^8	5761455	5428681	5762209	5769341	5740304

Observing the table above it is noticed that $Li(x) > \pi(x)$. Riemann proposed that this is true for all sufficiently large x. This turned out to be incorrect. In 1914 Littlewood [Li] proved the following.

Theorem 4.8.3 *The difference $\pi(x) - Li(x)$ assumes both positive and negative values infinitely often.*

Littelwood's proof was interesting in that it used the following technique which has become extremely useful in analytic number theory. First, he assumed that the Riemann hypothesis is true and proved that $\pi(x) - Li(x)$ changes sign infinitely often. He then showed that the same is true if the Riemann hypothesis is assumed to be false. A complete but somewhat simplified proof of Littelwood's result can be found in [P]. More recently, Te Riele in 1986 [Re] showed that there are greater than 10^{180} consecutive integers for which $\pi(x) > Li(x)$ in the range $6.62 \times 10^{370} < x < 6.69 \times 10^{370}$.

In light of trying to improve the approximation to $\pi(x)$ afforded by $Li(x)$ Riemann's work suggested (see Zagier [Za]) that $\frac{\pi(x)}{x}$ would be closer to $\frac{1}{\ln x}$, that is, the probability of choosing a prime randomly less than x would be closer to $\frac{1}{\ln x}$ if one counted not only the primes but also the "weighted powers" of the primes. That is counting a p^2 as half a prime, p^3 as a third of a prime and so on. This would lead to an approximation for $Li(x)$ given by

$$Li(x) \approxeq \pi(x) + \frac{1}{2}\pi(x^{\frac{1}{2}}) + \frac{1}{3}\pi(x^{\frac{1}{3}}) + \cdots .$$

Upon inverting this

$$\pi(x) \approxeq Li(x) - \frac{1}{2}Li(x^{\frac{1}{2}}) - \frac{1}{3}Li(x^{\frac{1}{3}}) \cdots .$$

Based on these ideas Riemann proposed the following **explicit formula** for $\pi(x)$,

$$\pi(x) = \sum_{n=1}^{\infty} \frac{\mu(n)}{n} Li(x^{\frac{1}{n}}). \tag{4.8.2}$$

The series on the right side of (4.8.2) can be shown to converge for $x \geq 2$ and is called the **Riemann function** $R(x)$, that is,

$$R(x) = \sum_{n=1}^{\infty} \frac{\mu(n)}{n} Li(x^{\frac{1}{n}}), x \geq 2.$$

Riemann's conjecture was then that $\pi(x) = R(x)$. The equality given in (4.8.2) is not true, however it is asymptotically correct. That is

Theorem 4.8.4 *We have* $\pi(x) \sim R(x)$ *where* $R(x)$ *is the Riemann function.*

In fact this approximation is remarkably close for large x. For $x = 400, 000, 000$ we have

$$\pi(400, 000, 000) = 21, 336, 326 \text{ and } R(400, 000, 000) = 21, 355, 517$$

while for $x = 1, 000, 000, 000$ we have

$$\pi(1, 000, 000, 000) = 50, 847, 534 \text{ and } R(1, 000, 000, 000) = 50, 847, 455.$$

Related to Riemann's explicit formula it can be shown that the distribution of the number of zeros of the Riemann zeta function along the critical line can be given asymptotically by

$$N(t) = \frac{t}{2\pi} \ln(\frac{t}{2\pi}) - \frac{t}{2\pi}$$

where $N(t)$ is the number of zeros z with $z = \frac{1}{2} + is$ along the critical line with $0 < s < t$.

There are also some surprising relationships between some physical phenomena and the location of the zeros of the Riemann zeta function. The article [BK] discusses some of these which are far afield from our present presentation.

An entirely elementary formulation of the Riemann hypothesis is the following (see [P]). Define a positive squarefree integer n to be **red** if it is the product of an even number of distinct primes and **blue** if it is the product of an odd number of distinct primes. Let $R(n)$ be the number of red integers not exceeding n and $B(n)$ the number of blue integers not exceeding n. The Riemann hypothesis is equivalent to the statement that for any $\epsilon > 0$ there exists an N such that for all $n > N$

$$|R(n) - B(n)| < n^{\frac{1}{2}+\epsilon}.$$

As we mentioned in Section 4.1 if p_n denotes the nth prime then it is a straightforward consequence of the prime number theorem that

$$p_n \sim n \ln n$$

and hence

$$\lim \frac{p_{n+1}}{p_n} = 1$$

even though there are arbitrarily large gaps in the primes. It was noted in the last section that when Selberg first gave his formula Erdos then used it to give an elementary proof of the second fact above. Subsequently Selberg then used his formula along with the methods of Erdos' proof to develop the first elementary proof of the prime number theorem.

There are two well-known conjectures concerning the difference $p_{n+1} - p_n$. The first is called **Cramer's conjecture**.

Cramer's Conjecture: $p_{n+1} - p_n \leq (1 + o(1))(\ln n)^2$.

It follows from Koch's equivalence to the Riemann hypothesis that if the Riemann hypothesis is true then

$$p_{n+1} - p_n = O(p_n^{\frac{1}{2}+\epsilon}) \text{ for any fixed } \epsilon > 0.$$

The second conjecture is called **Lindelof's hypothesis**.

Lindelof's Hypothesis: $\sum_{p_n \leq x} (p_{n+1} - p_n)^2 \leq x^{1+o(1)}$.

It can be shown that the Riemann hypothesis implies the Lindelof hypothesis.

Dirichlet's theorem, giving that there are infinitely many primes in any arithmetic progression $an + b$ with $(a, b) = 1$, extended the result that there are infinitely many primes. Dirchlet's proof (see Chapter 3) used L-series and then an Euler product formula. Recall that for an integer k, a Dirichlet L-series is defined by

$$L(s, \chi) = \sum_{n=1}^{\infty} \frac{\chi(n)}{n^s}$$

where χ is a character mod k, and s is a complex variable. Hence Dirichlet's proof was an extension of the Euler proof of the infinitude of primes using the real zeta series. Along the same lines both the prime number theorem and the Riemann hypothesis can be extended to primes in arithmetic progressions.

For $(a, b) = 1$ let

$$\pi(x; a, b) = \text{ numbers of primes congruent to } b \text{ mod } a \text{ and } \leq x.$$

The prime number theorem for arithmetic progressions can then be expressed as:

Theorem 4.8.5 *(The Prime Number Theorem for Arithmetic Progressions) For fixed* $a, b > 0$ *with* $(a, b) = 1$ *then*

$$\pi(x; a, b) \sim \frac{1}{\phi(a)}\pi(x) \sim \frac{1}{\phi(a)}\frac{x}{\ln x} \sim \frac{1}{\phi(a)}Li(x).$$

The result can be expressed in probabilistic terms by saying that the primes are uniformly distributed in the $\phi(a)$ residue classes relatively prime to a. In fact much of the material on the prime number theorem can be rephrased in terms of probability theory. The prime number theorem itself can be expressed as:

Theorem 4.8.6 *(The Prime Number Theorem) The probability of randomly choosing a prime less than or equal to x is asymptotically given by* $\frac{1}{\ln x}$.

Most of the ideas surrounding the use of probabilistic methods are discussed in the book **Probabilistic Number Theory** by Elliott [E].

The extension of the Riemann hypothesis to the case of arithmetic progressions is called the **generalized Riemann hypothesis** or the **extended Riemann hypothesis**. This says that the zeros of any Dirichlet L-series also lie along the critical line $\text{Re}\, s = \frac{1}{2}$.

Generalized Riemann Hypothesis: *For an integer k and any character* χ mod *k then the nontrivial zeros of the L-series*

$$L(s, \chi) = \sum_{n=1}^{\infty} \frac{\chi(n)}{n^s}$$

all lie along the critical line $\text{Re}\, s = \frac{1}{2}$.

We close this chapter with a brief discussion of primes in short intervals $[x, x + \epsilon]$ where $\epsilon > 0$ is a positive constant. Bertrand's theorem (Theorem 4.2.5) showed that for any real number $x \geq 1$ there is always a prime in the interval $[x, 2x]$. Further the proof used the same methods as the proof of Chebyshev's estimate. As an immediate consequence of the prime number theorem we can obtain the following result. We leave the proof to the exercises.

Theorem 4.8.7 *For any* $\epsilon > 0$ *there exists an* $x_0 = x_0(\epsilon)$ *such that there is always a prime in the interval* $[x, (1 + \epsilon)x]$ *for* $x > x_0$. *Equivalently* $\pi(x + y) > \pi(x)$ *for* $y = \epsilon x$.

The above theorem and its proof has the following interesting interpretation. For large x (see the exercises)

$$\pi(2x) - \pi(x) \sim \pi(x).$$

Hence for large x there are as many primes asymptotically between x and $2x$ as there are less than x, despite the fact that by the prime number theorem the density of primes tends to thin out. However, it can be shown that

$$2\pi(x) - \pi(2x) \to \infty$$

as $x \to \infty$.

The result given in Theorem 4.7.4 has been improved upon in various ways. Huxley in 1972 continuing a long line of research in this direction showed that there is always a prime in the interval $[x, x + x^c]$ if $c > \frac{7}{12}$ for large enough x. The value of c has subsequently been improved, the most recent being done by Baker and Harman who reduced c to .535 again for large enough x. Further Baker and Harman show that

$$\pi(x + x^{.535}) - \pi(x) > \frac{x^{.535}}{20 \ln x}$$

for large enough x.

Earlier Erdos, using Selberg's formula, had proved that for each $\epsilon > 0$ there exists a constant $c(\epsilon)$ such that in the interval $[x, (1 + \epsilon)x]$ there are at least $\frac{c(\epsilon)x}{\ln x}$ primes.

Finally, we mention the following remarkable result which is a consequence of Bertrand's theorem. We outline a proof in the exercises.

Theorem 4.8.8 *Given any positive integer n, the set of integers $\{1, 2, \ldots, n\}$ can be partitioned into n disjoint pairs so that the sum of each pair is a prime.*

So for example $\{1, 2, 3, 4, 5, 6, 7, 8, 9, 10\}$ can be partitioned into

$$\{1, 10\}, \{2, 9\}, \{3, 4\}, \{5, 8\}, \{6, 7\}.$$

The result is in the same spirit as the **Goldbach conjecture** which states that any even integer is the sum of two primes.

4.9 Exercises

4.1 Show that $Li(x) = \int_2^x \frac{1}{\ln t} dt$ is asymptotically equal to $\frac{x}{\ln x}$. (Hint: Take the Taylor expansion of $Li(x)$.)

4.2 If p_n is the nth prime show that $\lim_{n\to\infty} \frac{p_{n+1}}{p_n} = 1$.

Recall that the binomial coefficient $\binom{n}{k}$ (see Section 4.2) is defined by

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}$$

4.3 Prove the following facts about $\binom{n}{k}$:

(a) $\binom{n}{k}$ represents the number of ways of choosing k objects out of n without replacement and without order (Lemma 4.2.1). This is equivalent to the number of possible subsets of size k in a finite set with n elements. (Hint: Consider the number of ways of choosing k out of n with order—this would be $n(n-1)\cdots(n-k+1)$. Then consider how many ways each choice of k objects can be rearranged.)

(b) $\binom{n}{k} = \binom{n}{n-k}$.

(c) $\binom{n}{k} + \binom{n}{k-1} = \binom{n+1}{k}$. (This is the basis for Pascal's triangle).

4.4 Prove the Binomial Theorem: For any real numbers a, b, and natural number n we have

$$(a+b)^n = \sum_{k=0}^{n} \binom{n}{k} a^k b^{n-k}.$$

(Hint: Use induction and part (c) of Exercise 4.3.)

4.5 Prove: For a prime p, $(x+y)^p \equiv x^p + y^p \mod p$. (Hence the beginning algebra mistake $(x+y)^p = x^p + y^p$ is true in the field $\mathbb{Z}_p$.)

4.6 If $s > 0$ the **Gamma function** is given by

$$\Gamma(s) = \int_0^\infty x^{s-1} e^{-x} dx.$$

Show that

(a) $\Gamma(s+1) = s\Gamma(s)$. (Use integration by parts.)

(b) $\Gamma(n) = (n-1)!$ for any $n \geq 1, n \in \mathbb{N}$.

4.7 (a) Show that $\int_0^\infty e^{-x^2} dx = \frac{\sqrt{\pi}}{2}$. (Hint: Let $A = \int_0^\infty e^{-x^2} dx$. Then

$$A^2 = (\int_0^\infty e^{-x^2} dx)(\int_0^\infty e^{-y^2} dy) = \int_0^\infty \int_0^\infty e^{-(x^2+y^2)} dx dy$$

Now change to polar coordinates. Recall that $dxdy = rdrd\theta$).

(b) Use part (a) to show that $\Gamma(\frac{1}{2}) = \sqrt{\pi}$.

4.8 Recall that Stirling's Approximation is

$$n! \approxeq \sqrt{2\pi n}(\frac{n}{e})^n.$$

We outline a proof of this result.

(a) From Problem 4.6 Stirling's approximation is equivalent to

$$\Gamma(p+1) \approxeq p^p e^{-p} \sqrt{2\pi p}.$$

(b) Write the integral for $\Gamma(p+1)$ as follows

$$\Gamma(p+1) = \int_o^\infty x^p e^{-x} dx = \int_0^\infty e^{p\ln x - x} dx.$$

Now substitute the variable $x = p + y\sqrt{p}$ so that $dx = \sqrt{p}dy$. Show then that

$$\Gamma(p+1) = \int_{-\sqrt{p}}^\infty e^{p\ln(p+\sqrt{p}y)-p-\sqrt{p}y}\sqrt{p}dy.$$

(c) By looking at the Taylor series for $\ln x$ show that for large p

$$\ln(p+\sqrt{p}y) = \ln p + \ln(1+\frac{y}{\sqrt{p}}) \approxeq \ln p + \frac{y}{\sqrt{p}} - \frac{y^2}{2p}.$$

(d) By using part (c) and the integral in part (b) show that

$$\Gamma(p+1) = e^{p\ln p - p}\sqrt{p}\int_{-\sqrt{p}}^\infty e^{-\frac{1}{2}y^2} dy$$

$$= p^p e^{-p}\sqrt{p}(\int_{-\infty}^\infty e^{-\frac{1}{2}y^2} dy - \int_{-\infty}^{-\sqrt{p}} e^{-\frac{1}{2}y^2} dy).$$

(e) Evaluate the two integrals in part (d) to get Stirling's approximation. Notice that from Exercise 4.4 we have

$$\int_{-\infty}^\infty e^{-x^2} dx = \sqrt{\pi}$$

and so

$$\int_{-\infty}^\infty e^{-\frac{1}{2}x^2} dx = \sqrt{2\pi}$$

and

$$\int_{-\infty}^{-\sqrt{p}} e^{-\frac{1}{2}y^2} dy$$

goes to zero as p goes to infinity.

4.9 Use the prime number theorem to give an alternative proof that there are arbitrarily large gaps in the sequence of primes. (Hint: Suppose that there is a bound A so that there is always a prime between x and $x + A$. Then consider $\pi(nA)$ to deduce a contradiction.)

4.10 Show that $f(x) \sim g(x)$ is equivalent to $f(x) = g(x)(1 + o(1))$.

4.11 Show that $f = o(g)$ implies $f = O(g)$.

4.12 Show that:
(a) $\cos x = O(1)$,
(b) $\sin x = o(x)$,
(c) $x = o(x^d)$ if $d > 1$,
(d) If $P(x)$ is a polynomial of degree n with leading coefficient a then $P(x) \sim ax^n$.

4.13 (a) Show that if $f = O(1)$ and $g = O(1)$ then $f + g = O(1)$ or equivalently $O(1) + O(1) = O(1)$
(b) Show that $O(1) = o(x)$

4.14 Show that $\frac{\ln x}{x^\delta} \to 0$ as $x \to \infty$ for any $\delta > 0$. Equivalently $\ln x = o(x^\delta)$. Hence $\ln x$ goes to infinity more slowly than any positive power of x.

4.15 Using Bertrand's theorem show that $p_{n+1} < 2p_n$ where p_n is the nth prime.

4.16 Prove that for each $\epsilon > 0$ there exists an $x_0 = x_0(\epsilon)$ such that there is always a prime in the interval $[x, (1+\epsilon)x]$ for $x > x_0$. (Hint: Consider $\pi(x + \epsilon x) - \pi(x)$ and apply the prime number theorem.)

4.17 Show that $\pi(2x) - \pi(x) \sim \pi(x)$. Hence, asymptotically there are as many primes between x and $2x$ as are less than x.

4.18 Prove that

$$\frac{1}{\zeta(s)} = \sum_{n=1}^{\infty} \frac{\mu(n)}{n^s}$$

where $\mu(n)$ is the Möbius function.

4.19 Prove that the set of rationals of the form $\{\frac{p}{q}; p, q \text{ primes}\}$ is dense in the set of positive reals. Recall that a set S is dense in the reals if given any real number r and $\epsilon > 0$ there is an $s \in S$ with $|r - s| < \epsilon$.

4.20 Prove Theorem 4.7.6: Given any positive integer n the set of integers $\{1, 2, \ldots, 2n\}$ can be partitioned into n disjoint pairs so that the sum of each pair is a prime. (Hint: Use induction and then notice that for $n = 2k$ by Bertrand's Theorem there exists an m with $1 \leq m < 2k$ such that $2k + m$ is prime.)

4.21 Prove that the equation $n! = m^k$ has no solutions in integers with $m, n, k > 1$.

4.22 Prove that there exists real numbers a, b such that for all n

$$n^{an} < \prod_{i=1}^{n} p_i < n^{bn}$$

with p_i the ith prime.

4.23 Let $\Lambda(n)$ be the Van Mangoldt function. Prove that

$$\sum_{d|n} \Lambda(d) = \ln n$$

or equivalently $\Lambda = \mu \star L$.

4.24 Prove the following orthogonality relations among the trigonometric functions:

(a) $\int_{-\pi}^{\pi} \cos(mx)\cos(nx)dx = 0 \,\text{if}\, m \neq n; = \pi \,\text{if}\, m = n \neq 0; = 2\pi \,\text{if}\, m = n = 0.$
(b) $\int_{-\pi}^{\pi} \sin(mx)\sin(nx)dx = 0$ if $m \neq n$; $= \pi$ if $m = n \neq 0$.
(c) $\int_{-\pi}^{\pi} \cos(mx)\sin(nx)dx = 0$ for all m, n.

4.22 Use the previous problem to show that if $f(x)$ is a periodic function with period 2π and Fourier series

$$\overline{f} = a_0 + \sum_{n=1}^{\infty} (a_n \cos(\frac{n\pi x}{L}) + b_n \sin(\frac{n\pi x}{L}))$$

then, if $f(x) = \overline{f}(x)$, the coefficients a_0, a_n, b_n must be given by

$$\begin{aligned} a_0 &= \frac{1}{2L}\int_{-L}^{L} f(x)dx \\ a_n &= \frac{1}{L}\int_{-L}^{L} f(x)\cos(\frac{n\pi x}{L})dx, n = 1, 2, \ldots \\ b_n &= \frac{1}{L}\int_{-L}^{L} f(x)\sin(\frac{n\pi x}{l})dx, n = 1, 2, \ldots \end{aligned}$$

4.23 Using the formula for complements

$$\Gamma(s)\Gamma(1-s) = \frac{\pi}{\sin(\pi s)}$$

and the **duplication formula**

$$\Gamma(s)\Gamma(s+\frac{1}{2}) = \sqrt{\pi}2^{1-2s}\Gamma(2s).$$

Show that the relation

$$\pi^{-s/2}\Gamma(\frac{s}{2})\zeta(s) = \pi^{-(1-s)/2}\Gamma(\frac{1-s}{2})\zeta(1-s)$$

can be transformed into

$$\zeta(s) = 2^s\pi^{s-1}\sin(\frac{\pi s}{2})\Gamma(1-s)\zeta(1-s), s \neq 0, 1.$$

4.24 Prove Theorem 4.6.3: The set of number theoretic functions with addition defined pointwise and multiplication given by Dirichlet convolution forms a commutative ring with an identity.

4.25 Prove Euler's identity for the cotangent function

$$\pi\cot(\pi x) = \frac{1}{x} + \sum_{n=1}^{\infty}\frac{2x}{x^2-n^2} \qquad (x \in \mathbb{C}\backslash.$$

4.26 Prove that the Taylor expansion of the logarithm of the Gamma function at $z = 0$ is given by

$$\log\Gamma(1-z) = \gamma z + \sum_{k=2}^{\infty}\frac{\zeta(k)}{k}z^k$$

where γ is the Euler–Mascheroni constant

$$\gamma = \lim_{n\to\infty}\left(\sum_{k=1}^{n}\frac{1}{k} - \log(n)\right).$$

Chapter 5
Primality Testing—An Overview

5.1 Primality Testing and Factorization

In the previous two chapters we have seen that there are infinitely many primes and showed that as we move through larger and larger integers, the density of primes thins out. In particular we proved that

$$\frac{\pi(x)}{x} \sim \frac{1}{\ln x} \text{ as } x \to \infty$$

where $\pi(x)$ represents the number of primes less than the positive real number x. This result, the prime number theorem, could be interpreted as saying that the probability of randomly choosing a prime number less than or equal to a positive real number x is approximately $\frac{1}{\ln x}$ as x gets large. In this chapter we consider the question of determining whether a particular given positive integer n is prime or not prime. The methods concerning this problem are called **primality testing** and consist of algorithms to determine whether or not an inputted positive integer is prime. Primality testing has become extremely important and has been of great interest in recent years due to its close ties to **cryptography** and especially **public key cryptography**. Cryptography is the science of encoding and decoding secret messages. Many of the most powerful and secure encoding methods depend on number theory, especially on the computational difficulty of factoring large integers. It turns out, somewhat surprisingly, that relative to ease of computation, determining if a number is prime is easier than actually factoring it.

Public key cryptography is that part of cryptography that deals with sending secret (or hopefully secure) messages across public communications systems. The major algorithm in this area, called the RSA algorithm, depends directly on the difficulty of factoring large integers. We will briefly introduce cryptography and the RSA algorithm in Section 5.4. First we take a short overview look at primality testing.

At first glance, the problem of determining if a positive integer n is prime, seems like an easy one. If n is not prime it must have a divisor m with $1 < m < n$. Therefore

B. Fine and G. Rosenberger, *Number Theory*,
DOI 10.1007/978-3-319-43875-7_5

test all integers $2, \ldots, \frac{n}{2}$ to see if they divide n or not. If there is such a divisor then n is composite. If not, then n is prime.

Of course this can be improved in several ways. First of all, if $n = mk$ then one of m, k must be $\leq \sqrt{n}$. Hence we need only check integers from 2 to $\sqrt{n}$ rather than from 2 to $\frac{n}{2}$. Further if n has a divisor m with $1 < m \leq \sqrt{n}$ then n must have a prime divisor p with $1 < p \leq \sqrt{n}$. Therefore it is only necessary to check the primes $\leq \sqrt{n}$. Therefore knowing all the primes $\leq \sqrt{n}$ allows to test for primality all the integers $\leq n$. We summarize all these comments to give a general algorithm for primality testing.

General Algorithm for Primality Testing: *Given $n > 0$, test all primes p with $p \leq \sqrt{n}$. The integer n is prime if and only if none of these primes divides n.*

EXAMPLE 5.1.1 Test whether the integer 83 is prime.

Now $9 < \sqrt{83} < 10$ so we must test all the primes less than 9. Hence we must test 2, 3, 5, 7. None of these divides 83 and therefore 83 is prime.

This general algorithm is simple and always works. However, it becomes computationally infeasible for large integers. Therefore other methods become necessary to determine primality. Most of these methods rely on a number theoretic property, such as Fermat's theorem, which is true for all primes but may not be true for all composites. Recall that Fermat's theorem (see Chapter 2) says that $a^{p-1} \equiv 1 \bmod p$ for any prime p and for any a with $1 < a < p$. We will return to this in Section 5.3. In the next section we examine a series of techniques for determining primes called **sieving methods**.

5.2 Sieving Methods

In ordinary language a sieve is a device to separate or sift finer particles from coarser particles. This idea has been applied to number theory via numerical sieving methods. A **sieve** in number theory is a method or procedure to find numbers with desired properties (for example, primes) by sifting through all the positive integers up to a certain bound, successively eliminating invalid candidates until only numbers with the particular attributes desired are left. Sieving methods are quite effective for obtaining lists of primes (and numbers with other characteristics) up to a reasonably small limit.

Relative to generating lists of primes, sieving methods originated with the **Sieve of Eratosthenes**. This is a straightforward method to obtain all the primes less than or equal to a fixed bound x. It is ascribed (as the name suggests) to Eratosthenes (276–194 B.C.) who was the chief librarian of the great ancient library in Alexandria. Besides the sieve method he was an influential scientist and scholar in the ancient world, developing a chronology of ancient history (up to that point) and helping to obtain an accurate measure (within the measurement errors of his time) of the dimensions of the Earth.

The method of the Sieve of Eratosthenes is direct and works as follows. Given $x > 0$ list all the positive integers less than or equal to x. Starting with 2, which is prime, cross out all proper multiples of 2 on the list. The next number on the list, not crossed out, which is 3, is prime. Now cross out all the proper multiples of 3 not already eliminated. The next number left uneliminated, 5, is prime. Continue in this manner. As explained for the primality test described in the previous section the elimination must only be done for numbers $\leq \sqrt{x}$. Upon completion of this process, any number, except 1, not crossed out must be a prime.

Below we exhibit the Sieve of Eratosthenes for numbers $\leq$ 100. In beginning each round of elimination we must only consider numbers $\leq \sqrt{100} = 10$.

1	2	3	~~4~~	5	~~6~~	7	~~8~~	~~9~~	~~10~~
11	~~12~~	13	~~14~~	~~15~~	~~16~~	17	~~18~~	19	~~20~~
~~21~~	~~22~~	23	~~24~~	~~25~~	~~26~~	~~27~~	~~28~~	29	~~30~~
31	~~32~~	~~33~~	~~34~~	~~35~~	~~36~~	37	~~38~~	~~39~~	~~40~~
41	~~42~~	43	~~44~~	~~45~~	~~46~~	47	~~48~~	~~49~~	~~50~~
~~51~~	~~52~~	53	~~54~~	~~55~~	~~56~~	~~57~~	~~58~~	59	~~60~~
61	~~62~~	~~63~~	~~64~~	~~65~~	~~66~~	67	~~68~~	~~69~~	~~70~~
71	~~72~~	73	~~74~~	~~75~~	~~76~~	~~77~~	~~78~~	79	~~80~~
~~81~~	~~82~~	83	~~84~~	~~85~~	~~86~~	~~87~~	~~88~~	89	~~90~~
~~91~~	~~92~~	~~93~~	~~94~~	~~95~~	~~96~~	97	~~98~~	~~99~~	~~100~~

After completing the sieving operation we obtain the list

$$\{2, 3, 5, 7, 11, 13, 17, 19, 23, 29, 31, 37, 41, 43, 47, 53, 59, 61\}$$

$$\{67, 71, 73, 79, 83, 89, 97\}$$

which comprises all the primes less than or equal to 100.

Given positive integers m, x, by a slight modification, the Sieve of Eratosthenes can be used to determine all the positive integers relatively prime to m and less than or equal to x.

Here suppose we are given m and x. Let $p_1, \ldots, p_k$ be the distinct prime factors of m arranged in ascending order, that is, $p_1 < p_2 < \cdots < p_k$. Next list all the positive integers less than or equal to x as we did for the ordinary sieve. Start with p_1 and eliminate all multiples of p_1 on the list. Then successively do the same for p_2 through p_k. The numbers remaining on the list are precisely those relatively prime to m that are also less than or equal to x. If $p_i > x$ ignore this prime and all higher primes.

Below we exhibit the sieve applied to finding the numbers less than 50 and relatively prime to 180.

Since $180 = 2^2 3^2 5$ we must sieve out multiples of 2, 3 and 5.

$$\begin{array}{cccccccccc}
1 & \cancel{2} & \cancel{3} & \cancel{4} & \cancel{5} & \cancel{6} & 7 & \cancel{8} & \cancel{9} & \cancel{10} \\
11 & \cancel{12} & 13 & \cancel{14} & \cancel{15} & \cancel{16} & 17 & \cancel{18} & 19 & \cancel{20} \\
\cancel{21} & \cancel{22} & 23 & \cancel{24} & \cancel{25} & \cancel{26} & \cancel{27} & \cancel{28} & 29 & \cancel{30} \\
31 & \cancel{32} & \cancel{33} & \cancel{34} & \cancel{35} & \cancel{36} & 37 & \cancel{38} & \cancel{39} & \cancel{40} \\
41 & \cancel{42} & 43 & \cancel{44} & \cancel{45} & \cancel{46} & 47 & \cancel{48} & 49 & \cancel{50}
\end{array}$$

The remaining list is

$$\{1, 7, 11, 13, 17, 19, 23, 29, 31, 37, 41, 43, 47, 49\}.$$

These are all relatively prime to 180. Recall that these numbers then are all units modulo 180.

Legendre in 1808, in an attempt to determine the distribution, $\pi(x)$, of primes, derived a computational formula for the Sieve of Eratosthenes. Recall (see Chapter 4) that Legendre had conjectured the prime number theorem in the form

$$\pi(x) \approxeq \frac{x}{\ln x - 1.08}.$$

We first present a slightly more general form of Legendre's formula. Given a positive integer m and a positive x let

$$N_m(x) = \text{ number of positive integers } \le x \text{ and relatively prime to } m.$$

This is precisely the size of the list obtained in the modified Sieve of Eratosthenes derived above. We obtain:

Theorem 5.2.1 *(Legendre's Formula for the Sieve of Eratosthenes) Let $m \in \mathbb{N}, x \geq 0$, then*

$$N_m(x) = \sum_{d|m} \mu(d)[\frac{x}{d}]$$

where $\mu(d)$ is the Möbius function and [ ] *is the greatest integer function.*

Proof If $m = 1$ then clearly

$$N_1(x) = [x].$$

Now given $m > 1$ let $p_1 < p_2 < \cdots < p_k$ be the distinct prime factors of m and for each j with $1 \le j \le k$ let $m_j = p_1 \cdot p_2 \cdots p_j$.

For a given m_j the only integers counted by $N_{m_j}(x)$ not counted by $N_{m_{j+1}}(x)$ are those of the form $p_{j+1}n \le x$ where $(n, m_j) = 1$. It then follows that

$$N_{m_j}(x) - N_{m_{j+1}}(x) = N_{m_j}(\frac{x}{p_{j+1}}).$$

Applying this repeatedly we obtain

$$N_{m_1}(x) = N_1(x) - N_1(\frac{x}{p_1}) = [x] - [\frac{x}{p_1}]$$

$$N_{m_2}(x) = N_{m_1}(x) - N_{m_1}(\frac{x}{p_2}) = [x] - [\frac{x}{p_1}] - [\frac{x}{p_2}] + [\frac{x}{p_1 p_2}].$$

Continuing in this manner inductively we arrive at

$$N_m(x) = \sum_{d|\overline{m}} (-1)^{\omega(d)} [\frac{x}{d}] \qquad (5.2.1)$$

where $\overline{m} = p_1 p_2 \cdots p_k$ and $\omega(d)$ is the number of distinct prime factors of d. The integer $\overline{m}$ is called the **square-free kernel** of m. This can then be expressed in terms of the Möbius function. Recall (see Chapter 2 and Section 3.6) that the Möbius function is defined by

$$\mu(d) = \begin{cases} (-1)^{\omega(d)} & \text{if } d \text{ is squarefree} \\ 0 & \text{otherwise} \end{cases}$$

Substituting this in the form of Legendre's formula (5.2.1) and realizing that $\mu(d) = 0$ except for the factors of the square-free kernel we obtain

$$N_m(x) = \sum_{d|m} \mu(d) [\frac{x}{d}] \qquad (5.2.2)$$

proving the theorem. □

Now let $x \geq 2$ and let

$$m = \prod_{p \leq \sqrt{x}} p$$

where p is prime. Then $N_m(x)$ counts the number of primes in the interval $[\sqrt{x}, x]$. It follows that

$$N_m(x) = \pi(x) - \pi(\sqrt{x}) + 1.$$

Substituting Legendre's formula (5.2.2) into this expression we obtain as a corollary:

Corollary 5.2.1 *For $x \geq 2$*

$$\pi(x) = -1 + \pi(\sqrt{x}) + \sum_{\nu(d) \leq \sqrt{x}} \mu(d) [\frac{x}{d}]$$

where $\nu(d)$ is the greatest prime factor of d.

Although this gives a formula for $\pi(x)$, it is essentially useless in truly computing $\pi(x)$ for large x, or in shedding any light on the prime number theorem. First of all if we estimate $[\frac{x}{d}]$ by $\frac{x}{d} + O(1)$ and substitute in the formula we have

$$\begin{aligned}\pi(x) - \pi(\sqrt{x}) + 1 &= \sum_{\nu(d)\leq\sqrt{x}} \mu(d)(\frac{x}{d} + O(1)) \\ &= x \prod_{p\leq\sqrt{x}} (1 - \frac{1}{p}) + O(2^{\pi(\sqrt{x})})\end{aligned}$$

Hence the error term is exponentially larger than the main term. Further the number of steps in the Sieve of Eratosthenes and hence in the computation of the formula is proportional to $\sum_{p\leq x} \frac{x}{p}$. However, it can be shown that

$$\sum_{p\leq x} \frac{x}{p} = x \ln \ln x + O(x)$$

(see [CP] page 113 and [HW] Theorem 427). Therefore the number of steps is proportional to $\ln \ln x$ which goes to infinity (albeit slowly) with x. In addition, from a computer/computational point of view, one of the major computational drawbacks to implementing the Sieve of Eratosthenes (for large x) is the computer space it requires (see [CP]) which can be substantial. We mention that Brun attempted to make Legendre's formula computable. As an application he was able to prove the spectacular result that the sum of the reciprocals of the twin primes

$$\sum_{p,p+2 \text{ primes}} (\frac{1}{p} + \frac{1}{p+2})$$

converges. We will look at Brun's method and his proof of this result in the next section. We note that further slight modification of the Sieve of Eratosthenes can be utilized to obtain a complete prime factorization of a positive integer n.

Meisel in 1870 also gave an improvement to Legendre's formula and was able to use this technique to compute $\pi(x)$ correctly up to $x = 10^8$.

Theorem 5.2.2 *(Meisel's Formula) Let $p_1 < p_2 < \cdots < p_n < \cdots$ be the listing of the primes in increasing order so that p_j is the jth prime. Let $x \geq 4$, $n = \pi(\sqrt{x})$ and $m_n = p_1 \cdots p_n$. Then*

$$\pi(x) = N_{m_m}(x) + m(1+s) + \frac{1}{2}s(s-1) - 1 - \sum_{j=1}^{s} \pi(\frac{x}{p_{m+j}})$$

where $m = \pi(x^{\frac{1}{3}})$ and $s = n - m$.

Proof From the proof of Legendre's formula we have

$$N_{m_j}(x) - N_{m_{j+1}}(x) = N_{m_j}(\frac{x}{p_{j+1}}).$$

This holds for $1 \leq j \leq n$. Summing this equality for $j = m+1, \ldots, n$ we obtain

$$N_{m_n}(x) = N_{m_m}(x) - \sum_{j=1}^{s} N_{m_{m+j-1}}(\frac{x}{p_{m+j}}).$$

The inequalities

$$x^{\frac{1}{3}} < p_{m+j} \leq x^{\frac{1}{2}} < \frac{x}{p_{m+j}} < x^{\frac{2}{3}}$$

holding for $j = 1, 2, \ldots, s$, then imply that

$$N_{m_n}(x) = 1 + \pi(x) - \pi(\sqrt{x}) = \pi(x) - n + 1$$

and

$$N_{m_{m+j-1}}(\frac{x}{p_{m+j}}) = 1 + \pi(\frac{x}{p_{m+j}}) - \pi(p_{m+j-1}) = \pi(\frac{x}{p_{m+j}}) - (m+j-2).$$

Therefore

$$\pi(x) = N_{m_n}(x) + n - 1 = N_{m_m}(x) - \sum_{j=1}^{s}(\pi(\frac{x}{p_{m+j}}) - m - j + 2) + n - 1$$
$$= N_{m_m}(x) - \sum_{j=1}^{s} \pi(\frac{x}{p_{m+j}}) + m(1+s) + \frac{s(s-1)}{2} - 1$$

proving the theorem. □

Note that $N_n(n)$ is the total number of integers less than n and relatively prime to n. Hence

$$N_n(n) = \phi(n)$$

the Euler phi function introduced in Chapter 2. Applying Legendre's formula with $m = n = x$ we obtain

$$\phi(n) = \sum_{d|n} \mu(d)\frac{n}{d} = n\prod_{p|n}(1 - \frac{1}{p}).$$

This recovers the formulas given for $\phi(n)$ in Theorems 2.4.7 and 2.4.8.

A variation of Legendre's formula can be obtained in the following manner. Suppose

$$p_1 < p_2 < \cdots < p_n < \cdots$$

are the primes listed in increasing order. Let $x \geq 2$ and

$$\Phi(x,k)$$

be the number of positive integers $\leq x$ not divisible by the first k primes. Hence

$$\Phi(x,k) = N_m(x)$$

if the square-free kernel of m is $p_1 \cdots p_k$. The same counting arguments applied to this function lead us to the next result.

Theorem 5.2.3 *Let the function* Φ *be defined as above. Then*

$$\Phi(x,n) = [x] - \sum [\frac{x}{p_i}] + \sum [\frac{x}{p_i p_j}] - \sum [\frac{x}{p_i p_j p_k}] + \cdots$$

where each sum is over the set of pairwise distinct primes less than or equal to x.

Here $\Phi(x,x) = N_x(x)$ so

$$\begin{aligned}\Phi(x,x) &= \pi(x) - \pi(\sqrt{x}) + 1\\ &= [x] - \sum_{p_i \leq \sqrt{x}} [\frac{x}{p_i}] + \sum_{p_i < p_j \leq \sqrt{x}} [\frac{x}{p_i p_j}] - \sum_{p_i < p_j < p_k \leq \sqrt{x}} [\frac{x}{p_i p_j p_k}] + \cdots .\end{aligned}$$

This version of Legendre's formula satisfies a very nice recurrence relation.

Corollary 5.2.2 *Let the function* Φ *be defined as above. Then*

$$\Phi(x,k) = \Phi(x,k-1) - \Phi(\frac{x}{p_k}, k-1).$$

There is a very nice visual quadratic sieve which also generates the prime numbers. Consider the parabola $x = y^2$ and consider the points (n^2, n) lying on the parabola for $n = 2, 3, \ldots$ Now connect all pairs of such points lying on the two branches of the parabola, above and below the x-axis by straight line segments. The intersection points of these lines with the positive x-axis corresponds to composite numbers. The integer points remaining are precisely the primes (see the exercises). In Figure 5.1 we give the picture of this.

5.2.1 Brun's Sieve and Brun's Theorem

The Sieve of Eratosthenes and the extensions of it described in the last section are really just the tip of the iceberg as far as sieving methods in number theory are concerned (see [HR]). In this section we give one beautiful application by V. Brun of a refinement of Legendre's formula for the Sieve of Eratosthenes.

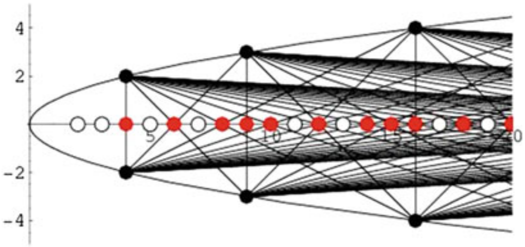


Fig. 5.1 Brun's sieve

Recall that the **twin primes** are the set $\{(p, p+2)\}$ where both p and $p+2$ are primes. There are two related still open questions concerning this set. Both are called the **twin primes conjecture**. The first is that there are infinitely many twin primes. Empirical evidence and a probabilistic argument suggests that there are infinitely many such pairs and most people working in the area feel that this part of the conjecture is almost certainly true. However, it remains still open. The second twin prime conjecture deals with the density of the twin primes and is in the same spirit as the prime number theorem.

If we let

$$\pi_2(x) = \text{ the number of pairs of twin primes } (p, p+2) \text{ with } p \leq x$$

then the second twin prime conjecture or **strong twin prime conjecture** is that

$$\pi_2(x) \sim C \int_2^x \frac{dt}{(\ln t)^2}.$$

The constant C is called the **twin primes constant** and is given by

$$C = 2\Pi_2$$

where

$$\Pi_2 = \prod_{p>2, p \text{ prime}} (1 - \frac{1}{(p-1)^2}).$$

Sometimes Π_2 is also called the twin primes constant. The value of Π_2 has been computed to a great many decimal places and has the approximate value

$$\Pi_2 \approxeq .660161815\ldots.$$

Brun proved that there exists an integer N such that

$$\pi_2(x) \leq \frac{100x}{(\ln x)^2} \text{ for } x \geq N.$$

It has further been proved that

$$\pi_2(x) \leq k\Pi_2(\frac{x}{(\ln x)^2})(1+O(\frac{\ln\ln x}{\ln x}))$$

where k is a constant. Hardy and Littlewood proposed the value of 2 in the strong twin primes conjecture.

The strong twin primes conjecture is actually the smallest case of a general conjecture called the **Hardy–Littlewood conjecture** or **k-tuple conjecture**.

Here suppose $0 < m_1 < m_2 < \cdots < m_k$ are k odd integers. Then a **prime constellation** is a set $\{p, p+2m_1, p+2m_2, \ldots, p+2m_k\}$ where all are primes. If we let

$$\pi_{m_1,\ldots,m_k}(x)$$

denote the number of prime constellations (relative to a fixed set $\{m_1, \ldots, m_k\}$) less than or equal to x then the **k-tuple conjecture** or **Hardy–Littlewood conjecture** is that

$$\pi_{m_1,\ldots,m_k}(x) \sim C(m_1, \ldots, m_k)\int_2^x \frac{dt}{(\ln t)^{k+1}}$$

where $C(m_1, \ldots, m_k)$ is a constant depending only on $m_1, \ldots, m_k$. The strong twin primes conjecture is the special case of this with $m_1 = 1$ and $k = 1$.

Although these conjectures are still open, V. Brun in 1920 was able to prove the amazing result that the sum of the reciprocals of the twin primes converges. We call this amazing since this result can be accomplished without even knowing if there are infinitely many twin primes. Brun's theorem is the following:

Theorem 5.2.4 *(Brun) If $S = \{(p, p+2)\}$ denotes the set of twin prime pairs then the series*

$$\sum_{(p,p+2)\in S}(\frac{1}{p}+\frac{1}{p+2})$$

converges. That is,

$$\frac{1}{3}+\frac{1}{5}+\frac{1}{5}+\frac{1}{7}+\frac{1}{11}+\frac{1}{13}+\cdots$$

converges.

Of course if there are only finitely many twin prime pairs the series would trivially converge.

The value of the series

$$B = \sum_{(p,p+2)\in S}(\frac{1}{p}+\frac{1}{p+2})$$

is called **Brun's constant**. A great deal of work has gone into determining the exact value of B. Empirically the value of B has been computed as (see [CP])

$$B \approxeq 1.902160583104\ldots.$$

Brun's theorem has been extended to further pairs of primes separated by a constant $d > 2$. For example, if $d = 4$ the pairs of primes of the form $(p, p+4)$ are called **cousin primes**. Again it is open whether there are infinitely many of these (for each d or for any fixed d) but Segal [S] proved that for any given d the sums of the reciprocals of the pairs is also convergent.

In 2014 Y. Zhang [Zh] proved that there is a positive constant with the property that infinitely many pairs of primes differ by less than that constant. In 2015 J. Maynard [Ma] gave a numerical extension.

Brun's proof of Theorem 5.2.4 is technical and involves attempting to improve computationally on Legendre's formula for the Sieve of Eratosthenes. His proof depends on the following technical results. After giving the proof of Brun's theorem we will give the proofs of the lemmas.

Lemma 5.2.1 *If $n \geq 0$ and $m \geq 0$ then*

$$\sum_{i=0}^{m}(-1)^i\binom{n}{i} = (-1)^m\binom{n-1}{m}.$$

In particular if m is odd then

$$\sum_{i=0}^{m-1}(-1)^i\binom{n}{i} \geq 0.$$

The next lemma depends on **symmetric polynomials** and **symmetric functions**. In Chapter 6 we will look at these in detail here we just introduce what is needed for the next result.

Suppose $y_1, \ldots, y_n$ are n distinct real numbers. (Later we will look at a more general situation). Form the polynomial

$$p(x, y_1, \ldots, y_n) = (x - y_1)\cdots(x - y_n).$$

The **ith elementary symmetric polynomial** or **ith elementary symmetric function** s_i in $y_1, \ldots, y_n$ for $i = 1, \ldots, n$, is $(-1)^i a_i$, where a_i is the coefficient of x^{n-i} in $p(x, y_1, \ldots, y_n)$.

To be more specific consider y_1, y_2, y_3. Then

$$\begin{aligned} p(x, y_1, y_2, y_3) &= (x - y_1)(x - y_2)(x - y_3) \\ &= x^3 - (y_1 + y_2 + y_3)x^2 + (y_1y_2 + y_1y_3 + y_2y_3)x - y_1y_2y_3. \end{aligned}$$

Therefore, the three elementary symmetric polynomials in y_1, y_2, y_3 are

1. $s_1 = y_1 + y_2 + y_3$.

2. $s_2 = y_1 y_2 + y_1 y_3 + y_2 y_3$.
3. $s_3 = y_1 y_2 y_3$.

In general, the pattern of the last example holds for $y_1, \ldots, y_n$. That is,

$$s_1 = y_1 + y_2 + \cdots + y_n$$

$$s_2 = y_1 y_2 + y_1 y_3 + \cdots + y_{n-1} y_n$$

$$s_3 = y_1 y_2 y_3 + y_1 y_2 y_4 + \cdots + y_{n-2} y_{n-1} y_n$$

$$\vdots$$

$$s_n = y_1 \cdots y_n.$$

We now state the lemma we need.

Lemma 5.2.2 *If S_n is the nth elementary symmetric function of s positive numbers $a_1, \ldots, a_s$, $1 \le n \le s$, then*

$$S_n \le \frac{S_1^n}{n!}.$$

Lemma 5.2.3 *Let $d > 0, n > 0$. Then the number of positive integers $m \le n$ which belong to any given residue class mod d differs from $\frac{n}{d}$ by less than 1.*

The following is the crucial lemma.

Lemma 5.2.4 *Let $P(x)$ denote the number of primes $p \le x$ for which $p+2$ is prime. Then for $x \ge 3$ we have*

$$P(x) < c \frac{x}{(\ln x)^2} (\ln \ln x)^2$$

where c is a constant.

We can now give a proof of Brun's theorem.

Proof (Theorem 5.2.4) As in the statement of Lemma 5.2.4 let $P(x)$ denote the number of primes $p \le x$ for which $p+2$ is prime. It follows then from Lemma 5.2.4 that for $x \ge 3$ (see the exercises)

$$P(x) \le k \frac{x}{(\ln x)^{\frac{3}{2}}}$$

where k is a constant. Let $(p_r, p_r + 2)$ denote the rth twin prime pair. Then for all $r \ge 1$ we have

$$r = P(p_r) < k \frac{p_r}{(\ln p_r)^{\frac{3}{2}}} < k \frac{p_r}{(\ln(r+1))^{\frac{3}{2}}}$$

$$\implies \frac{1}{p_r} < \frac{k}{r(\ln(r+1))^{\frac{3}{2}}}.$$

Now it follows easily from the integral test for infinite series (see the exercises) that the series

$$\sum_{r=1}^{\infty} \frac{1}{r(\ln(r+1))^{\frac{3}{2}}}$$

converges. Therefore by the comparison test

$$2\sum_{r=1}^{\infty} \frac{1}{p_r} \geq \sum_{r=1}^{\infty} (\frac{1}{p_r} + \frac{1}{p_r+2})$$

converges. □

We now give the proofs of the four technical lemmas. The first three are very straightforward. The real difficulty lies in Lemma 5.2.4.

Proof (Lemma 5.2.1) We wish to prove that if $n, m \geq 0$ then

$$\sum_{i=0}^{m} (-1)^i \binom{n}{i} = (-1)^m \binom{n-1}{m}.$$

The second assertion, that if m is odd then

$$\sum_{i=0}^{m-1} \binom{n}{i} \geq 0,$$

follows directly from the first.

We prove the first assertion by induction on m. If $m = 0$ then

$$\sum_{i=0}^{m} (-1)^i \binom{n}{i} = (-1)^0 \binom{n}{0} = 1 = (-1)^0 \binom{n-1}{0} = 1$$

so it is true for $m = 0$. Suppose that

$$\sum_{i=0}^{m} (-1)^i \binom{n}{i} = (-1)^m \binom{n-1}{m}.$$

Then

$$\begin{aligned}\sum_{i=0}^{m+1}(-1)^i\binom{n}{i} &= (-1)^{m+1}\binom{n}{m+1} + \sum_{i=0}^{m}(-1)^i\binom{n}{i}\\ &= (-1)^{m+1}\binom{n}{m+1} + (-1)^m\binom{n-1}{m}\\ &= (-1)^{m+1}\binom{n-1}{m+1} \text{ (see the exercises).}\end{aligned}$$

Therefore the first statement is true by induction. □

Proof (Lemma 5.2.2) Here we wish to show that

$$S_n \leq \frac{S_1^n}{n!}$$

where S_n is the *nth* elementary symmetric function of s positive numbers $a_1, \dots, a_s$, $1 \leq n \leq s$. Notice that S_n consists of the sum of all n-fold products taken from $a_1, \dots, a_s$. Now consider

$$S_1^n = (a_1 + \cdots + a_s)^n.$$

There are $\binom{s}{n}$ n-fold products $a_{i_1} \cdots a_{i_n}$ in the binomial expansion and each has coefficient $n!$. Hence the result follows. □

Proof (Lemma 5.2.3) Let $d > 0$, $n > 0$. We wish to show that the number of positive integers $m \leq n$ which belong to any given residue class mod d differs from $\frac{n}{d}$ by less than 1.

On each set of d consecutive integers there is only one number counted for a given residue class mod d. Up to a given positive n there are $[\frac{n}{d}]$ complete sets of residues mod d and, if $\frac{n}{d}$ is not integral, an additional partial set of residues. Hence the number counted in the statement of the lemma is either $[\frac{n}{d}]$ or possibly $[\frac{n}{d}] + 1$ depending on whether $\frac{n}{d}$ is integral or not. Therefore the number m in the lemma always satisfies

$$\frac{n}{d} - 1 < m < \frac{n}{d} + 1.$$

□

Proof (Lemma 5.2.4) Let $P(x)$ denote the number of primes $p \leq x$ for which $p + 2$ is prime. Then we wish to show that for $x \geq 3$

$$P(x) < c\frac{x}{(\ln x)^2}(\ln \ln x)^2$$

where c is a constant. First, suppose that $x > 5$ and y is chosen so that $5 \leq y < x$. Let $Q(x)$ be the number of integers n in the interval $y \leq n < x$ for which both n and $n + 2$ are primes. Clearly then

$$P(x) \leq y + Q(x). \tag{5.2.3}$$

Let $p_1 < p_2 < \cdots < p_n < \cdots$ denote the sequence of primes and suppose that $\pi(y) = r$. Let $A(x)$ denote the number of integers n for which $0 < n \leq x$ and n is not congruent to either 0 or -2 mod p_i for $i = 2, \ldots, r$. Then

$$Q(x) \leq A(x) \tag{5.2.4}$$

for every n, counted in $Q(x)$, is greater than y and therefore greater than p_h for $h \leq r$ since $\pi(y) = r$. Combining (5.2.3) and (5.2.4) we get

$$P(x) \leq y + A(x).$$

Let $\Omega(d)$ denote the number of distinct primes factors of $d > 0$. If d is odd and square-free let $B(d, x)$ be the number of positive integers $n \leq x$ for which for every prime factor p of d either $n \equiv 0 \bmod p$ or $n \equiv -2 \bmod p$. From Lemma 5.2.3 we have

$$|B(d, x) - 2^{\Omega(d)}\frac{x}{d}| < 2^{\Omega(d)} \tag{5.2.5}$$

for if $0 < n \leq x$ then n belongs to $2^{\Omega(d)}$ residue classes mod d. (Two classes for each of the $\Omega(d)$ prime factors of $d = \prod_{p|d} p$.)

We next claim that

$$A(x) \leq \sum_{d|p_2\cdots p_r, \Omega(d)<m} \mu(d)B(d, x) \tag{5.2.6}$$

where m is an arbitrary positive odd integer.

Every n with $0 < n \leq x$ which is not counted in $A(x)$ satisfies $n \equiv 0 \bmod p_{t_i}$ or $n \equiv -2 \bmod p_{t_i}$ for b primes $p_{t_1}, \ldots, p_{t_b}$ with $2 \leq t_1 < \cdots < t_b \leq r$. Hence those n not counted in $A(x)$ are counted in the sum precisely for those terms $B(d, x)$ for which $d|(p_2 \cdots p_r)$ and $d|(p_{t_1} \cdots p_{t_b})$ and further $\Omega(d) < m$.

Since $p_2 \cdots p_r$ is square-free it follows that every n with $0 < n \leq x$ which is counted in $A(x)$ is counted exactly once in the sum since $\mu(d) = 0$ unless $d = 1$ or d is squarefree. Combining these two observations we get that the complete count in the sum is then

$$\sum_{d|p_2\cdots p_r, \Omega(d)<m} \mu(d)B(d, x) = \sum_{i=1}^{m-1}(-1)^i\binom{n}{i} \geq 0$$

by Lemmas 5.2.1 and 5.2.2. Hence the inequality (5.2.6) is proved.

Combining this inequality with inequality (5.2.5) we have

$$A(x) < x \sum_{\substack{d|p_2\cdots p_r \\ \Omega(d)<m}} \frac{\mu(d)2^{\Omega(d)}}{d} + \sum_{i=0}^{m-1} 2^i \binom{r-1}{i}.$$

First we have

$$\sum_{i=1}^{m-1} 2^i \binom{r-1}{i} \leq 2^m \sum_{i=1}^{m-1} \binom{r-1}{i} \leq 2^m \sum_{i=1}^{m-1} r^i$$

since

$$\binom{r-1}{i} = \frac{(r-1)\cdots(r-i)}{i!} \leq r^i.$$

But this last sum satisfies

$$2^m \sum_{i=1}^{m-1} \leq 2^m \frac{r^m - 1}{r-1} < 2^m r^m \leq (2y)^m$$

since $r - 1 \geq 2$, $r \leq y$.

For the second part of the sum

$$\sum_{\substack{d|p_2\cdots p_r, \\ \Omega(d)<m}} \frac{\mu(d)2^{\Omega(d)}}{d} = \sum_{d|p_2\cdots p_r} \frac{\mu(d)2^{\Omega(d)}}{d} - \sum_{n=m}^{r-1} \sum_{\substack{d|p_2\cdots p_r, \\ \Omega(d)=n}} \frac{\mu(d)2^{\Omega(d)}}{d}.$$

If $m \geq r$ the last term is zero. We have by Euler expansion

$$\sum_{\substack{d|p_2..p_r, \\ \Omega(d)<m}} \frac{\mu(d)2^{\Omega(d)}}{d} = \prod_{2<p\leq p_r} (1 - \frac{2}{p}) - \sum_{n=m}^{r-1} (-1)^n 2^n \sum_{\substack{d|p_2\cdots p_r, \\ \Omega(d)=n}} \frac{1}{d}$$

$$= \prod_{2<p\leq n} (1 - \frac{2}{p}) - \sum_{n=m}^{r-1} (-1)^n 2^n S_n$$

where S_n is the nth elementary symmetric polynomial in

$$\frac{1}{p_2}, \ldots, \frac{1}{p_r}.$$

From Lemma 5.2.2 and since $n!e^n > n^n$ (see the exercises) it follows that

$$S_n \leq \frac{S_1^n}{n!} \leq \frac{(eS_1)^n}{n^n} < (\frac{3c \ln \ln y}{n})^n$$

where c is a constant. Then

$$|\sum_{n=m}^{r-1}(-1)^n 2^n S_n| \leq \sum_{n=m}^{r-1}(\frac{6c\ln\ln y}{m})^n \leq \sum_{n=m}^{r-1}(\frac{c_1\ln\ln y}{m})^n$$

with c_1 another constant. It follows that if

$$m > 2c_1 \ln\ln y$$

then

$$|\sum_{n=m}^{r-1}(-1)^n 2^n S_n| < \sum_{n=m}^{\infty}\frac{1}{2^n} = \frac{1}{2^{m-1}}.$$

Combining this with the earlier inequalities we obtain

$$|\sum_{\substack{d|p_2\cdots p_r,\\ \Omega(d)<m}} \frac{\mu(d)2^{\Omega(d)}}{d}| < \frac{c_2}{(\ln y)^2} + \frac{1}{2^{m-1}}$$

with c_2 another constant. Therefore

$$P(x) < y + \frac{c_2}{(\ln y)^2} + \frac{x}{2^{m-1}} + (2y)^m.$$

These inequalities are true if $5 \leq y < x$ and $m > 2c_1 \ln\ln y$. If we choose

$$y = x^{\frac{1}{3c_1 \ln\ln x}} \text{ and } m = 2[c_1 \ln\ln x] - 1$$

then these conditions are met and so the derived inequalities hold. Therefore

$$P(x) \leq c_4(y + \frac{x}{(\ln y)^2} + \frac{x}{2^{2c_1 \ln\ln x}} + (2y)^{2c_1 \ln\ln x})$$

for $x > c_5$ with c_5 some large enough constant.

Each of the terms in the parenthesis satisfies

$$< c_6 \frac{x}{(\ln x)^2}(\ln\ln x)^2$$

for some constant c_6 holding for all of them. To see this we have first

$$y \leq k_1\sqrt{x} \text{ for some constant } k_1.$$

Further

$$\frac{x}{(\ln y)^2} \leq \frac{x}{(\ln x)^2}(k_2 \ln\ln x)^2$$

and

$$\frac{x}{2^{2c_1 \ln\ln x}} = \frac{x}{(\ln x)^{2c_1 \ln 2}} < \frac{x}{(\ln x)^2}$$

since $c_1 > 2$ and $2 \ln 2 > 1$. Finally

$$(2y)^{2c_1 \ln\ln x} = e^{2c_1 \ln\ln x(\frac{\ln x}{3c_1 \ln\ln x} + \ln 2)} < e^{\frac{2}{3} \ln x + c_1 \ln\ln x}$$

$$< c_7 e^{\frac{3}{4} \ln x} = c_7 x^{\frac{3}{4}} \text{ for some constant } c_7.$$

Therefore for $x > c_5$, large enough, we have

$$P(x) < c_6 \frac{x}{(\ln x)^2} (\ln\ln x)^2$$

To obtain the result for $x \geq 3$ we combine the first terms into a new constant C and get that for $x \geq 3$,

$$P(x) < C \frac{x}{(\ln x)^2} (\ln\ln x)^2$$

proving the lemma. □

5.3 Primality Testing and Prime Records

As we have seen in the previous two sections it is theoretically very straightforward, using either the direct method of trial division or the Sieve of Eratosthenes, to test an integer for primality. The problem is that for large integers n these methods become computationally intractable if not almost impossible. Hence direct trial division and the Sieve of Eratosthenes can only be used for small integers and therefore for large integers other methods must be employed. We should note before going further that the concepts of *small* and *large* are very relative in number theory to the type of computing machinery one is using. Numbers as large as $10,000,000,000$ can be tested very easily, even on small computers, using the Sieve of Eratosthenes. In terms of computational asymptotic number theory, 10^9 is still small. Similarly, for human computation the total number of atoms in the universes is massive. This number is estimated as being of the order of 10^{79}. However, 79 digit integers are only considered moderate in asymptotic computational number theory which may want to handle integers with hundreds or even thousands of digits. Therefore what is needed are tests for testing primality which will handle some of these gigantic integers.

A **primality test** is then an algorithm which inputs a positive integer n and outputs whether it is prime or not. These tests can be subclassified as either **deterministic primality tests** or **probabilistic primality tests**. In a deterministic test an integer n is inputted and the output is, yes the integer is prime, or, no, the integer is not prime.

Hence both the direct method of trial division, and the Sieve of Eratosthenes, are deterministic tests.

A nondeterministic primality test takes an inputted integer n and returns either no, it is not prime, or it may be a prime. A **probabilistic primality test** is a nondeterministic test that returns either the inputted integer is not a prime or is probably a prime to some given degree of accuracy. There are various tests (that we will look at in the next section) which can give this accuracy to as high a probability as desired. Numbers that pass a probabilistic primality test are called **probable primes**. For use in cryptography, knowing if an integer is prime to a high probability, is often just as good as knowing if it is definitely prime. For this reason probable primes with a high degree of probability are called **industrial grade primes**, a term originally coined by M. Cohen.

The majority of nondeterminsitic tests are based on either Fermat's theorem or some variation of it. Recall from Chapter 2, Fermat's (Little) theorem, (Corollary 2.4.2):

Theorem 5.3.1 *(Fermat's Theorem) If p is a prime and $p \nmid a$ then*

$$a^{p-1} \equiv 1 \; mod \, p.$$

This was a special case of the more general Euler's theorem, which we will also need.

Theorem 5.3.2 *(Euler's Theorem) If $(a, n) = 1$ then*

$$a^{\phi(n)} \equiv 1 \; mod \, n.$$

Hence if n is an integer and a is relatively prime to n with a^{n-1} not congruent to 1 mod n then n cannot be prime. This is usually called the **Fermat Probable Prime Test** and was introduced briefly in Chapter 2. Basically given n we find an a with $(a, n) = 1$ and compute a^{n-1} mod n. If this value is not 1 mod n then n is not prime. If it is congruent to 1 mod n then n may be prime. In the latter case, by trying different values for a we can assign a probability value. We will make this precise in the next section. For now we will state the basic Fermat Probable Prime Test and present an example.

The Fermat Probable Prime Test: *Suppose n is an inputted integer. Find an a with $(a, n) = 1$. Compute a^{n-1} mod n. If this value is not 1 mod n then n is not prime. If this value is 1 mod n then n may be prime.*

EXAMPLE 5.3.1 Test whether 11387 is prime.

This integer is relatively small so even by trial division determining whether it is prime is easy. We use the Fermat method just to illustrate the technique.

Start with $a = 2$ and we test 2^{11386} mod 11387. The basic idea is to use repeated squarings to reduce the congruence. All the equivalences are modulo 11387.

$$2^{13} = 8192 \equiv -3195 \implies 2^{26} \equiv 10208025 \equiv 5273$$

$$\implies 2^{52} \equiv 8862 \equiv 2525 \implies 2^{104} \equiv 10292 \equiv -1095$$

$$\implies 2^{208} \equiv 3390 \implies 2^{416} \equiv 2617 \implies 2^{832} \equiv 5102$$

Continuing in this manner we eventually get

$$2^{11388} \equiv 8642 \implies 2^{11387} \equiv 4321.$$

From Fermat's theorem, if n is prime we would have $a^{n-1} \equiv 1 \bmod n$ and therefore $a^n \equiv a \bmod n$. Here 4321 is not congruent to 2 mod 11387. Therefore 11387 is not prime.

For this integer using trial division it is easy to obtain the factorization

$$11387 = (59)(193).$$

However even with an integer this size at least a calculator is necessary.

In 1891 Lucas gave the following extension of Fermat's theorem which actually makes the Fermat Test deterministic.

Theorem 5.3.3 *(Lucas) Let $n > 1$. If for every prime factor p of $n - 1$ there exists an integer a such that*

1. $a^{n-1} \equiv 1$ *mod n and*
2. $a^{\frac{n-1}{p}} \not\equiv 1$ *mod n*

then n is prime.

Proof Suppose n satisfies the conditions of the theorem. To show that n is prime we will show that $\phi(n) = n - 1$ where ϕ is the Euler phi function. Since in general $\phi(n) < n - 1$, to show equality we will show that under the above conditions $n - 1$ divides $\phi(n)$. Suppose not. Then there exists a prime p such that p^r divides $n - 1$, but p^r does not divide $\phi(n)$ for some exponent $r \geq 1$. For this prime p, there exists an integer a satisfying the conditions of the theorem. Let m be the order of a modulo n. Then m divides $n - 1$ since the order of an element divides any power which equals 1 (see Chapter 2). However by the second condition in the theorem and for the same reason, m does not divide $\frac{n-1}{p}$. Therefore p^r divides m which divides $\phi(n)$ contradicting our assumption. Hence $n - 1 = \phi(n)$ and therefore n is prime. □

Although this Lucas test is deterministic, it is, in most cases, no more computationally feasible than trial division or sieving since it depends on the factorization of $n-1$. In general, factorization is even more difficult than solely testing for primality. Therefore even here further methods are necessary. We note that the idea in the Lucas test has been quite effective in developing methods for testing Fermat and Mersenne numbers for primality. We will return to these in Section 5.3.2.

The majority of probabilistic primality tests are based on the Fermat test or some variation of it. The basic idea is that if an integer passes the test for a base b (so that its a probable prime) then try another base. Doing this there is then a technique to attach a probability tied to the number of bases attempted. We will make this precise in the next section. For now we would like to look at a new (2003) deterministic algorithm which answered a major open problem in both number theory and computer science.

Primality testing is essentially a computational problem. Therefore a primality test raises questions about the accompanying algorithm's computational speed and computational complexity. For these types of number theoretic algorithms the computational complexity is measured in terms of functions of the input length, which here is roughly the number of digits of the inputted integer. The Sieve of Eratosthenes requires, for an inputted integer n, roughly the same order n of operations. If n has $\log_{10} n$ digits then the Sieve requires $O(10^{\log_{10} n})$ operations to prove primality. We say that this algorithm is of **exponential time** in terms of the input length. The big open question was whether there exists a deterministic algorithm which is of **polynomial time** in the input length. This means that for this algorithm there is a positive integer d such that the number of operations in the algorithm to prove primality is $O((\ln n)^d)$. Earlier, Miller, and Rabin had shown that the Miller–Rabin test, which we will describe in the next section, can be made deterministic. Further it is of polynomial time **if** one accepts as true the extended Riemann hypothesis (see Chapter 4). However prior to 2003 it was an open question whether there was a deterministic algorithm for primality which could be shown to be of polynomial time without using any unproved conjectures.

In 2003, M. Agrawal and two of his students, N. Kayal and N. Saxena, developed an algorithm, now called the **AKS Algorithm**, which was deterministic and could be proved to be of polynomial time. The result was even more spectacular since it was accomplished with relatively elementary methods. The basic algorithm depends on two rather straightforward extensions of Fermat's theorem. This result has of course generated a great deal of attention and much has already been written about it. We refer the reader to the articles [Bo] and [Be] for a more complete discussion of the algorithm and its development. Because of the timeliness and excitement this result has generated we will present the basic arguments in the paper of [AKS]. This will be done in Section 5.5 at the conclusion of this chapter.

The first result needed is the following which was well known in the theory of finite fields.

Theorem 5.3.4 *Suppose $(a, n) = 1$ with $n > 1$. Then n is a prime if and only if*

$$(x - a)^n \equiv x^n - a \bmod n$$

in the ring of polynomials $\mathbb{Z}[x]$.

Proof Suppose n is prime. If $n = 2$ the statement holds. Now we assume that n is an odd prime. From the binomial theorem

$$(x - a)^n = \sum_{k=0}^{n} \binom{n}{k} x^{n-k}(-a)^k.$$

If n is prime and $k \neq 0$, n then $\binom{n}{k} \equiv 0 \bmod n$ (see the exercises). Therefore

$$(x-a)^n = x^n - a^n \text{ in } \mathbb{Z}_n[x].$$

But from Fermat's theorem $a^n \equiv a \bmod n$ and so the result follows.

Conversely, if n is composite then it has a prime divisor p. Suppose p^k is the highest power of p dividing n. Then p^k does not divide $\binom{n}{p}$. Therefore in the binomial expansion of $(x-a)^n$ the coefficient of the x^p term is not zero mod n and hence

$$(x-a)^n \neq x^n - a \text{ in } \mathbb{Z}_n[x].$$

□

This theorem is computationally just as difficult to use as Fermat's theorem in proving primality. Agrawal, Kayal, and Saxena then proved the following extension of the above result which leads to the AKS algorithm. To state the theorem we need the following notation. If $p(x)$, $q(x)$ are integral polynomials then we say

$$p(x) \equiv q(x) \bmod (x^r - 1, n)$$

if the remainders of $p(x)$ and $q(x)$ after division by $x^r - 1$ are equal (equal coefficients) modulo n.

Theorem 5.3.5 *(AKS) Suppose that n is a natural number and $s \leq n$. Suppose that q, r are primes satisfying $q|(r-1)$, $n^{\frac{r-1}{q}}$ is not congruent to 0 or 1 modulo r and $\binom{q+s-1}{s} \geq n^{2[\sqrt{r}]}$. If for all a with $1 \leq a < s$*

1. *$(a, n) = 1$, and*
2. *$(x-a)^n \equiv x^n - a$ mod $(x^r - 1, n)$.*

Then n is a prime power.

The proof of this theorem is not difficult but requires some results from the theory of cyclotomic fields which are outside the scope of this book. Hence at this point we omit the proof. However as mentioned, the basic arguments in the paper of [AKS] will be presented in Section 5.5. The most difficult part of the proof is showing that given n there do exist primes q, r satisfying the conditions in the theorem.

From Theorem 5.3.4 we get the following algorithm (the AKS algorithm), which is deterministic.

The AKS Algorithm: Input an integer $n > 1$

Step (1): Determine if $n = a^b$ for some integers a, b. If so and $b > 1$ output **composite** and done.

Step (2): Choose q, r, s satisfying the hypotheses of Theorem 5.3.5

Step (3): For $a = 1, 2, \ldots, s-1$ do the following

If a is a divisor of n output **composite** and done

If $(x-a)^n$ is not congruent to $x^n - a$ mod $(x^r - 1, n)$ output **composite** and done

Step (4): Output **prime**.

Although the algorithm is deterministic, it is not clear that it can be accomplished in polynomial time. What is necessary is to show that polynomial bounds can be placed on determining q, r, s. This can be done. The following is a program written in pseudocode which can be implemented even on a relatively small computer which places the appropriate bounds. It is also necessary to have an algorithm to implement the first step. This can be done in linear time.

AKS Algorithm Program: Input an integer $n > 1$.

```
If $n = a^b$ for some natural numbers $a, b$ with $b > 1$ then output COMPOSITE.
$r = 2$
while $(r < n)$ do {
    if $((n, r) \neq 1)$ output COMPOSITE
    if ($r$ is prime )
        let $q$ be the largest prime factor of $r - 1$
        if $(q \geq 4\sqrt{r}\log_2 n)$ and $(n^{\frac{r-1}{q}} \neq 1 \bmod r)$
            break;
    $r \leftarrow r + 1$
}
for $a = 1$ to $2\sqrt{r}\log_2 n$
    If $(x - a)^n$ is not congruent to $x^n - a \bmod (x^r - 1, n)$ output
        COMPOSITE;
output PRIME;
```

The crucial thing is that by determining these bounds it makes the algorithm run in polynomial time.

Theorem 5.3.6 *(AKS) The AKS algorithm runs in* $O((\log_2 n)^{12} f(\log_2 \log_2 n))$ *time. That is, the time to run this algorithm is bounded by a constant times the number of digits to the 12th power times a polynomial in the log of the number of digits.*

The proof of the AKS algorithm has been refined by several people (see [Be]) and it has been conjectured that it actually has polynomial running time $O((\log_2 n)^6)$.

In theory the AKS algorithm should be the fastest running primality tester. However, computational complexity is only a theoretical statement as $n \to \infty$. In practice, at the present time, several of the existing algorithms actually run faster. However, the implementation of the AKS algorithm will probably improve. As mentioned, in Section 5.5 we will give the proof of this theorem. In the next section we introduce the ideas behind the probabilistic primality tests.

5.3.1 Pseudo-Primes and Probabilistic Testing

In this section we present two probabilistic primality tests; the **Solovay–Strassen test** and the **Miller–Rabin test**. The basic idea in both of these is to test, for an inputted integer n, a sequence of bases in the Fermat test. The hope is that a base will be located for which the test fails. In this case the number is not prime. If no such base is found a probability can be assigned, determined by the number of bases tested, that the number is prime. First we introduce some necessary concepts.

Definition 5.3.1 *Let n be a composite integer. If $b > 1$ with $(n, b) = 1$ then n is a* **pseudoprime** *to the base b if $b^{n-1} \equiv 1 \bmod n$.*

Hence n is a pseudoprime to the base b if it passes the Fermat test and hence is a probable prime.

EXAMPLE 5.3.1.1 25 is a pseudoprime to the base 7. To see this notice that

$$7^2 = 49 \equiv -1 \bmod 25.$$

This implies that $7^4 \equiv 1 \bmod 25$ and hence $7^{24} \equiv 1^6 \equiv 1 \bmod 25$.

Notice that 25 is not a pseudoprime mod 2 or 3.

Theorem 5.3.7 *For each base $b > 1$ there exists infinitely many pseudoprimes to the base b.*

Proof Suppose $b > 1$. We show that if p is any odd prime not dividing $b^2 - 1$ then the integer $n = \frac{b^{2p}-1}{b^2-1}$ is a pseudoprime to the base b. Note that for this n we have

$$n = \frac{b^{2p} - 1}{b^2 - 1} = \frac{b^p - 1}{b - 1} \cdot \frac{b^p + 1}{b + 1}$$

so that n is composite.

Given b from Fermat's theorem we have $b^p \equiv b \bmod p$ and hence $b^{2p} \equiv b^2 \bmod p$. Now $n - 1 = \frac{b^{2p}-b^2}{b^2-1}$ and since p does not divide $b^2 - 1$ by assumption it follows that p divides $n - 1$.

Further

$$n - 1 = b^{2p-2} + b^{2p-4} + \cdots + b^{2p}.$$

Therefore $n - 1$ is a sum of an even number of terms of the same parity so $n - 1$ must be even. It follows that $2p$ divides $n - 1$. Hence $b^{2p} - 1$ is a divisor of $b^{n-1} - 1$. However

$$b^{2p} - 1 \equiv 0 \bmod n \implies b^{n-1} - 1 \equiv 0 \bmod n.$$

Therefore n is a pseudoprime to the base b proving the theorem. □

Although there are infinitely many pseudoprimes they are not that common. It has been shown, for example, that there are only 21,853 pseudoprimes to the base 2 among the first 25,000,000,000 integers. Hence there is a good chance that if a number, especially a large number, passes a test as a pseudoprime, then it is really a prime. The question becomes how to make this chance or probability precise. List of many pseudoprimes can be found on various Internet websites (see [PP]).

From simple congruences the following is clear.

Lemma 5.3.1 *If n is a pseudoprime to the base b_1 and also a pseudoprime to the base b_2 then it is a pseudoprime to the base b_1b_2.*

Probabilistic methods proceed by testing n to a base b_1. If it is not a pseudoprime then it is composite and we are done. If it is a pseudoprime, test a second base b_2 and so on, in the hope of finding a base where it is not a pseudoprime. However there do exist numbers which are pseudoprimes to every possible base.

Definition 5.3.2 *A composite integer n is a* **Carmichael number** *if n is a pseudoprime to each base $b > 1$ with $(n, b) = 1$.*

If $n > 3$ is a Carmichael number then n must be odd. To see this suppose that n is even. We have $(n - 1, n) = 1$ and since n is a Carmichael number $(n - 1)^{n-1} \equiv 1 \bmod n$. However $(n - 1)^{n-1} \equiv -1 \bmod n$ since n is even. Hence $n|2$ which is a contradiction since $n > 3$. It follows that n must be odd.

The Carmichael numbers can be completely classified. Interestingly this was done even before the existence of Carmichael numbers was shown. The following is called the **Korselt criterion** after A. Korselt.

Theorem 5.3.8 *An odd composite number n is a Carmichael number if and only if n is squarefree and $(p - 1)|(n - 1)$ for every prime p dividing n.*

Proof Suppose that n is odd and composite. We first show that if a number n is not squarefree then it cannot be a Carmichael number.

Suppose that n is not squarefree. Then there exists a prime p with $p^2|n$. From Theorem 2.4.14 the multiplicative group in $\mathbb{Z}_{p^2}$ is cyclic (that is there exists a primitive element) and hence there is a multiplicative generator $g \bmod p^2$. Since $\phi(p^2) = p(p-1)$ we have $g^{p(p-1)} \equiv 1 \bmod p^2$ and this is the least power of g that is congruent to $1 \bmod p^2$. Now let $m = p_1 \cdots p_k$ where $p_1, \ldots, p_k$ are the other primes besides p dividing n. Notice that p^k is not a Carmichael number so these primes exist. Choose a solution b to the pair of congruences

$$b \equiv g \bmod p^2$$

$$b \equiv 1 \bmod m$$

which exists from the Chinese remainder theorem. Since $b \equiv g \bmod p^2$ it follows that b also has multiplicative order $p(p - 1) \bmod p^2$. Suppose n was a Carmichael number. Then n would be a pseudoprime to the base b and hence

$$b^{n-1} \equiv 1 \bmod n.$$

This implies that $p(p - 1)|n$ from the multiplicative order of b. However since $p|n$ we have $n - 1 \equiv -1 \bmod p$. On the other hand if $p(p - 1)|(n - 1)$ we have $n - 1 \equiv 0 \bmod p$ a contradiction. Therefore n cannot be a pseudoprime to the base b and hence is not a Carmichael number.

Now suppose that n is squarefree so that $n = p_1p_2\cdots p_k$ with $k \geq 2$ and the p_i distinct primes. Suppose first that $(p_i - 1)|(n - 1)$ for $i = 1, \dots, k$ and suppose that $(b, n) = 1$. Then

$$b^{n-1} \equiv b^{(p_i-1)k} \equiv 1^k \equiv 1 \bmod p_i, i = 1, \dots, k$$

Hence

$$b^{n-1} \equiv 1 \bmod p_1 \cdots p_k.$$

That is,

$$b^{n-1} \equiv 1 \bmod n.$$

Therefore n is a pseudoprime to the base b and since b was arbitrary with $(b, n) = 1$ it follows that n is a Carmichael number.

Conversely suppose that $n = p_1 \cdots p_k$ is a Carmichael number. Let p_i be one of these primes and suppose that g is a generator of the multiplicative group of $\mathbb{Z}_{p_i}$. Recall as in the proof of the squarefree property that this group is cyclic. Hence g has multiplicative order $p_i - 1 \bmod p_i$. Now let b be a solution to the pair of congruences

$$b \equiv g \bmod p_i$$

$$b \equiv 1 \bmod \frac{n}{p_i}.$$

Then b also has multiplicative order $p_i - 1 \bmod p_i$. Further since $(b, p_i) = 1$ and $(b, \frac{n}{p_i}) = 1$ it follows that $(b, n) = 1$. Since n is a Carmichael number it is a pseudoprime to the base b and hence

$$b^{n-1} \equiv 1 \bmod \text{n} \implies b^{n-1} \equiv 1 \bmod p_i.$$

It follows that $(p_i - 1)|(n - 1)$ proving the theorem.

□

Corollary 5.3.1 *A Carmichael number must be divisible by at least 3 primes.*

Proof Suppose that n is a Carmichael number. Then from the proof of the previous theorem $n = p_1 \cdots p_k$ with $k \geq 2$ and the p_1 distinct primes. We must show that $k > 2$. Suppose that $n = pq$ with $p < q$, p and q primes. Since n is a Carmichael number from the previous theorem $(q - 1)|(n - 1)$. However

$$n - 1 = pq - 1 = p(q - 1 + 1) - 1 \equiv (p - 1) \bmod (q - 1).$$

Since $(q-1)|(n-1)$ this would imply that $(q-1)|(p-1)$ which is impossible since $p < q$. Therefore if $n = pq$ it cannot be a Carmichael number and hence $k > 2$ so that n must be divisible by at least 3 distinct primes. □

Using the Korselt criterion we can present an example of a Carmichael number.

EXAMPLE 5.3.1.2 The integer $n = 561 = 3 \cdot 11 \cdot 17$ is a Carmichael number. Here $n - 1 = 560$ which is divisible by $2, 10$ and 16 and hence by the Korselt criterion it is a Carmichael number. This is well known as the smallest Carmichael number (see the exercises).

Carmichael numbers are relatively infrequent. It has been shown for example that there are only 2163 Carmichael numbers among the first 25 billion integers. However it has been proved by Alford, Granville and Pomerance that there do exist infinitely many Carmichael numbers. There is a list of Carmichael numbers up to 10^{16} (see [CP]).

Theorem 5.3.9 *(Alford, Granville, Pomerance) There are infinitely many Carmichael numbers. In particular if $C(x)$ denotes the number of Carmichael numbers less than or equal to x then $C(x) > x^{\frac{2}{7}}$ for x sufficiently large.*

We note that there are conjectured theorems on the distribution of $C(x)$ analogous to the Prime Number Theorem (see [CP]).

To proceed further we define several stronger types of pseudoprimes. Recall that if $n = p$ is a prime then $\mathbb{Z}_p$ is a field. Hence the polynomial equation

$$x^2 \equiv 1 \bmod p$$

has only the solutions $x \equiv 1 \bmod p$ and $x \equiv -1 \bmod p$. Therefore if $(a, p) = 1$ we must have

$$a^{\frac{p-1}{2}} \equiv \pm 1 \bmod p. \tag{5.3.1}$$

Recall that for a prime p the Legendre symbol $(a/p) = \pm 1$ whether or not a is a quadratic residue mod p (see Section 2.6). We need an extension of the Legendre symbol.

Definition 5.3.3 *If n is a positive odd integer with prime factorization $n = p_1^{e_1} \cdots p_k^{e_k}$ and a is a positive integer then the* **Jacobi symbol** *is*

$$(a/n) = (a/p_1)^{e_1} \cdots (a/p_k)^{e_k}.$$

Several of the results concerning the Legendre symbol, including quadratic reciprocity, can be extended to the Jacobi symbol.

Theorem 5.3.10 *If m, n are odd positive integers then:*

(1) $(2/n) = (-1)^{\frac{n^2-1}{8}}$

(2) (Jacobi Quadratic Reciprocity)

$$(m/n) = (-1)^{\frac{(m-1)(n-1)}{4}}(n/m).$$

The proofs of both of these assertions follow easily from the corresponding results on the Legendre symbol and we leave them to the exercises.

Note that if p is a prime then the Jacobi symbol and the Legendre symbol are identical. Hence for any prime p and integer a with $(a, p) = 1$

$$a^{\frac{p-1}{2}} \equiv (a/p) \bmod p$$

where on the right hand side we consider (a/p) as the Jacobi symbol.

Definition 5.3.4 *An odd composite integer n is an* **Euler pseudoprime** *to the base b if*

$$b^{\frac{n-1}{2}} \equiv (b/n)\ mod\, n$$

where (b/n) is the Jacobi symbol.

Since $(b/n) = \pm 1$ it follows easily that an Euler pseudoprime to the base b must also be a pseudoprime to the base b (see the exercises). However, the converse is not true—there exists pseudoprimes to a base b which are not Euler pseudoprimes to that base.

EXAMPLE 5.3.1.3 91 is a pseudoprime to the base 3 since $3^{90} \equiv 1 \bmod 91$. However, $3^{45} \equiv 27 \bmod 91$ so 91 is not an Euler pseudoprime to the base 3.

What is crucial in describing our first probabilistic primality test is that there are no "Carmichael type" numbers for Euler pseudoprimes. In fact, if n is composite it will fail to be an Euler pseudoprime for at least $\frac{1}{2}$ of the bases b with $(b, n) = 1$.

Theorem 5.3.11 *(Solovay, Strassen) If n is an odd composite integer then n is an Euler pseudoprime for at most $\frac{1}{2}$ of the bases b with $1 < b < n$ and $(b, n) = 1$.*

Proof Suppose that n is odd and composite. We first show that in this case if n is not an Euler pseudoprime for at least one base b then it is not an Euler pseudoprime for at least half of the bases b with $1 < b < n$, $(b, n) = 1$. We then show that if n is odd and composite there is a base b for which n is not an Euler pseudoprime.

Suppose that n is odd and composite and suppose that n is not an Euler pseudoprime to the base b. That is

$$b^{\frac{n-1}{2}} \not\equiv \pm 1 \bmod n.$$

If n is not an Euler pseudoprime to any base then certainly it is not an Euler pseudoprime for at least half of the possible bases. Suppose then that n is an Euler pseudoprime to the base b_1 so that

$$b_1^{\frac{n-1}{2}} \equiv \pm 1 \bmod n.$$

Then

$$(bb_1)^{\frac{n-1}{2}} \equiv b^{\frac{n-1}{2}} b_1^{\frac{n-1}{2}} \equiv b^{\frac{n-1}{2}} \not\equiv \pm 1 \bmod n.$$

Hence n is not an Euler pseudoprime to the base bb_1. Therefore for every base b_i for which n is an Euler pseudoprime, n is not an Euler pseudoprime for the base bb_i.

Further if b_i, b_j are distinct (mod n) bases for which n is an Euler pseudoprime bb_i is not congruent to bb_j mod n. It follows that if $\{b_1, \ldots, b_k\}$ are the distinct bases for which n is an Euler pseudoprime then $\{bb_i, \ldots, bb_k\}$ are distinct bases for which n is not an Euler pseudoprime. Therefore there are at least as many bases for which n is not an Euler pseudoprime as there are bases for which it is. We conclude then that if there exists at least one base b for which n is an Euler pseudoprime then n is an Euler pseudoprime for at most $\frac{1}{2}$ of the possible bases.

We now show that there must exist a base b for which n is not an Euler pseudoprime. Suppose first that n is not square-free so that there exists a prime p with $p^2|n$. Let g be a generator of the multiplicative group of $\mathbb{Z}_{p^2}$. Then as in the proof of the Korselt criterion, g has exact multiplicative order $\phi(p^2) = p(p-1)$. Let b solve the pair of congruences

$$b \equiv g \bmod p^2$$

$$b \equiv 1 \bmod \frac{n}{p^2}.$$

Then suppose that $b^{\frac{n-1}{2}} \equiv 1 \bmod n$. It follows that $p(p-1)|(n-1)$ which is impossible since $p^2|n$. Next suppose that $b^{\frac{n-1}{2}} \equiv -1 \bmod n$. Then $b^{n-1} \equiv 1 \bmod n$ so $b^{n-1} \equiv 1 \bmod p^2$. It follows that $p(p-1)|(n-1)$. But then again $p|(n-1)$ a contradiction. Hence if n is not squarefree, b as chosen above, is a base for which n is not an Euler pseudoprime.

Now suppose that n is square-free with $n = p_1 \cdots p_k$ with p_i distinct primes. Let g be a nonsquare mod p_1. Recall that there are only $\frac{p-1}{2}$ squares mod p_1 so such nonsquares exist. Hence $(g/p_1) = -1$. Choose a base b satisfying the simultaneous congruences

$$b \equiv g \bmod p_1$$

$$b \equiv 1 \bmod p_i, i = 2, \ldots, k$$

which exists by the Chinese remainder theorem. We then have for the Jacobi symbol

$$(b/n) = (b/p_1)(b/p_2) \cdots (b/p_k).$$

But $(b/p_1) = -1$ since $b \equiv g \bmod p_1$ and $(b/p_i) = (1/p_i) = 1$. Hence

$$(b/n) = -1.$$

If n were an Euler pseudoprime to the base b then

$$b^{\frac{n-1}{2}} \equiv (b/n) \bmod n$$

so that

$$b^{\frac{n-1}{2}} \equiv -1 \bmod n.$$

But then

$$b^{\frac{n-1}{2}} \equiv -1 \bmod p_2$$

which is a contradiction since $b \equiv 1 \bmod p_2$. Therefore n cannot be an Euler pseudoprime to the base b. Hence in each case there does exist a base for which n is not an Euler pseudoprime, proving the theorem. □

Theorem 5.3.11 is the basis for the **Solovay–Strassen Primality Test**. Suppose that we are given an odd integer n. Choose k integers $b_1, b_2, \dots, b_k$ at random with $1 < b_i < n$. If for some i we have $(b_i, n) > 1$ then n is composite. If all b_i are relatively prime to n then for each b_i compute

(1) $b_i^{(n-1)/2} \bmod n$ and

(2) $(b_i/n) \bmod n$.

If (1) does not equal (2) for some b_i then n is composite. Finally if

$$b_i^{(n-1)/2} \equiv (b_i/n) \bmod n$$

for all $i = 1, \dots, k$ then the probability that n is not prime is less than $(\frac{1}{2})^k$.

To see this notice that if n passes the conditions for b_1 then the probability of being composite from the Solovay–Strassen result is less than $\frac{1}{2}$. But b_2 is chosen randomly so the events that n passes the conditions for b_1 and b_2 are independent. Hence the probability that n passes the conditions for both b_1 and b_2 is $\frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}$ and so on.

```
Solovay–Strassen Primality Test: Input an odd integer $n$
    1: Choose $k$ random integers $b_1, \dots, b_k$ with $1 < b_i < n$
    2: For $i = 1, \dots, k$
        a: Compute $(b_i, n)$ (by the Euclidean algorithm)
            i: If $(b_i, n) > 1$ then $n$ is composite and stop.
        b: Compute (1) $b_i^{(n-1)/2}$ mod $n$ and (2) $(b_i/n)$ mod $n$
            i: If (1) $\neq$ (2) then $n$ is composite and stop
    3: The probability that $n$ is prime is greater then $1 - \frac{1}{2^k}$.
```

Miller and Rabin determined an even stronger test than the above by extending the idea of an Euler pseudoprime.

Definition 5.3.5 *Let n be an n composite integer with $n - 1 = 2^s t$ with t odd. If $b > 1$ and $(n, b) = 1$ then n is a* **strong pseudoprime** *to the base b if either*

(1) $b^t \equiv 1$ mod n or

(2) there exists r with $0 \le r < s$ such that $b^{2^r t} \equiv -1$ mod n.

The Miller–Rabin test is based on the following theorem analogous to the Solovay–Strassen result. It was proved independently by Monier and Rabin.

Theorem 5.3.12 *For each composite integer $n > 9$ the number of bases b with $0 < b < n$ for which n is a strong pseudoprime is less than $\frac{1}{4}$.*

If n is not a strong pseudoprime to the base b we say that b is a **witness** for n (a witness that n is composite). Hence if n is composite, Theorem 5.3.11, says that at least $\frac{3}{4}$ of all the integers in $[1, n-1]$ are witnesses for n. The Miller–Rabin test now proceeds exactly as the Solovay–Strassen test, except that the probability now that n is prime is greater than $1 - \frac{1}{4^k}$.

Miller–Rabin Primality test: Input an odd integer n and suppose that $n-1 = 2^s t$ with t odd

```
1: Choose $k$ random integers $b_1, \dots, b_k$ with $1 < b_i < n$
2: For $i = 1, \dots, k$
    a: Compute $(b_i, n)$ (by the Euclidean algorithm)
        i: If $(b_i, n) > 1$ then $n$ is composite and stop
    b: For $i = 1, \dots, k$
        i.Compute $m_i = b_i^t \bmod n$
            j: If $m_i = \pm 1$ then $n$ is a strong pseudoprime to the base $b_i$ and go on to the next $i$. Else
                k: For $j = 1, \dots, s-1$ compute $k_j = b_i^{2^j t} \bmod n$
                    l: If $k_j \equiv -1 \bmod n$ then $n$ is a strong pseudoprime to the base $b_i$ and go on to the next $i$. If not then go to the next $j$.
                    m. If $k_j \not\equiv -1 \bmod n$ for all $j$ then $n$ is composite and stop
3: The probability that $n$ is prime is greater then $1 - \frac{1}{4^k}$.
```

The Miller–Rabin test can be made deterministic under the assumption that the Extended Riemann Hypothesis holds (see Chapter 4). In particular Bach proved the following.

Theorem 5.3.13 *Assuming that the Extended Riemann Hypothesis holds then for any odd composite integer n there is a witness less than* $2(\ln n)^2$.

Hence based on the theorem we would only have to test for witnesses less than $2(\ln n)^2$. If there are none, then n is prime. This is then a deterministic polynomial time algorithm. However it depends on the unproved Extended Riemann Hypothesis.

5.3.2 The Lucas–Lehmer Test and Prime Records

A large portion of primality testing has centered on the Mersenne primes. In fact most of the prime *records*, that is, the determination of a largest known prime involves finding larger and larger Mersenne primes.

Recall from Section 3.1.3 that a Mersenne number is a positive integer of the form $M_n = 2^n - 1$, $n = 1, 2, \dots$. If M_n is prime then M_n is a **Mersenne prime**. Recall that it is not known whether or not there are infinitely many Mersenne primes. However it is conjectured, and believed, that there are infinitely many Mersenne primes.

Testing Mersenne numbers for primality has been particularly fruitful because of the **Lucas–Lehmer test**. This is a straightforward deterministic primality test specific to the Mersenne numbers. It is relatively easy to implement on a computer and has been quite successful in finding larger and larger Mersenne primes. For the most part historically, the largest known Mersenne prime, is also the largest known prime or current prime record. From Theorem 3.1.6 if $M_n = 2^n - 1$ is prime then n must be prime. Finding Mersenne primes then is often an experimental procedure with random prime exponents being tested by using the Lucas–Lehmer Test. In Table 5.1 we list the known Mersenne primes as of the writing of this book. Note that because the choice of prime exponents to test is random there may be other Mersenne primes between those on the list.

When looking at this table it should be mentioned how enormous the recent Mersenne primes are. In particular the most recent (in 2016) has close to 22.3 million decimal digits. We should also point out that although there may be intermediate Mersenne primes between those on the list, as of 2015, all Mersenne primes less than or equal to number 48 have been checked. It has been conjectured that there is a prime number type theorem for Mersenne primes. It particular it has been conjectured that if $M(x)$ is the number of primes $p \leq x$ with M_p prime then $M(x) \sim c \ln x$. Further $c = \frac{e^\gamma}{\ln 2}$ where γ is Euler's constant (see [CP]).

Before giving the Lucas–Lehmer Test we review some facts about the Mersenne numbers. Recall that the Mersenne numbers are closely tied to the **perfect numbers**. A natural number n is a perfect number if it is equal to the sum of its proper divisors. That is,

$$n = \sum_{d|n, d \geq 1, d \neq n} d$$

For example the number 6 is perfect since its proper divisors are 1, 2, 3 which add up to 6. We then have the following concerning Mersenne numbers, Mersenne primes and the ties to perfect numbers.

Theorem 5.3.14 *(1) If $M_n = 2^n - 1$ is prime then n is prime (Theorem 3.1.6).*

(2) If $M_p = 2^p - 1$ is a Mersenne prime then $n = 2^{p-1}(2^p - 1)$ is a perfect number. (Due to Euclid and given in Theorem 3.1.7).

(3) Conversely if $n \geq 2$ is a perfect number and even then $n = 2^{p-1}(2^p - 1)$ and $M_p = 2^p - 1$ is a Mersenne prime. (Due to Euler and given in Theorem 3.1.7).

Notice that from the theorem in searching for Mersenne primes only prime exponents must be considered. We now state the Lucas–Lehmer Test (note this was presented also in Section 3.1.3).

Theorem 5.3.15 *(Lucas–Lehmer Test). Let p be an odd prime and define the sequence (S_n) inductively by*

$$S_1 = 4 \text{ and } S_n = S_{n-1}^2 - 2.$$

Then the Mersenne number $M_p = 2^p - 1$ is a Mersenne prime if and only if M_p divides S_{p-1}.

Table 5.1 The known Mersenne primes M_p with p prime

Number	p	Discoverer and year
1	2	Unknown - pre 1500
2	3	Unknown - pre 1500
3	5	Unknown - pre 1500
4	7	Unknown - pre 1500
5	13	Anonymous - 1456
6	17	Cataldi - 1588
7	19	Cataldi - 1588
8	31	Euler - 1772
9	61	Pervushin - 1883
10	89	Powers - 1911
11	107	Powers - 1914
12	127	Lucas - 1876
13	521	Robinson - 1952
14	607	Robinson - 1952
15	1279	Robinson - 1952
16	2203	Robinson - 1952
17	2281	Robinson - 1952
18	3217	Riesel - 1957
19	4253	Hurwitz and Selfridge -1961
20	4423	Hurwitz and Selfridge -1961
21	9689	Gillies -1963
22	9941	Gillies -1963
23	11213	Gillies -1963
24	19937	Tuckerman -1971
25	21701	Noll and Nickel - 1978
26	23209	Noll -1979
27	44497	Slowinski and Nelson - 1979
28	86243	Slowinski - 1982
29	110503	Colquitt and Welsh - 1988
30	132049	Slowinski - 1983
31	216091	Slowinski - 1985
32	756839	Slowinski and Gage - 1992
33	859433	Slowinski and Gage - 1994
34	1257787	Slowinski and Gage - 1996
35	1398269	Armengaud, Woltman et. al - 1996
36	2976221	Spence, Woltman et.al - 1996
37	3021377	Clarkson,Woltman, Kurowski et.al - 1998
38	6972593	Hajratwala,Woltman and Kurowski - 2000

(continued)

Table 5.1 (continued)

Number	p	Discoverer and year
39	13466917	Cameron,Waltman,Kurowski - 2001
40	20996011	Shafer,Woltman and Kurowksi - 2003
41	24036583	Findley,Woltman and Kurowksi - 2004
42	25964951	Nowak,Woltman and Kurowksi - 2005
43	30402457	Cooper, Boone - 2005
44	32582657	Cooper, Boone - 2006
45	37156667	Elvenich,Woltman and Kurowksi - 2008
46	43112609	Smith,Woltman and Kurowksi - 2008
47	42643801	Odd Magnar Strimo,Melhus - 2009
48	57885161	Cooper - 2013
49	74207281	Cooper - 2016

Proof We first show that if M_p divides S_{p-1} then M_p is prime. We follow the proof given in [Br] and redone in [Tu] and [PP].

Let $u = 2 - \sqrt{3}$, $v = 2 + \sqrt{3}$. Then $u + v = 4 = S_1$ and $uv = 1$. An easy induction (see the exercises) shows that

$$S_n = u^{2^{n-1}} + v^{2^{n-1}}.$$

Suppose that $M_p | S_{p-1}$. We show that M_p must be a prime. Suppose not and let q be a prime dividing M_p with $q < \sqrt{M_p}$. Since $M_p | S_{p-1}$ we also have $q | S_{p-1}$.

Consider the finite field $\mathbb{Z}_q$. If 3 is a square mod q, that is, $(\frac{3}{q}) = 1$, let $F = \mathbb{Z}_q$. If 3 is not a square mod q let F be the extension field of $\mathbb{Z}_q$ adjoining a $\sqrt{3}$. That is $F = \mathbb{Z}_q(w)$ where $w^2 = 3$ (see Chapter 6). In either case F is a finite field, of order q in the former case and order q^2 in the latter. Recall that the multiplicative group of a finite field is cyclic (see Chapter 2). Hence if $g \in F$ with $g \neq 0$ then g has multiplicative order d with either $d|(q-1)$ or $d|(q^2-1)$. Since $(q-1)|(q^2-1)$ we can assume without loss of generality that $d|(q^2-1)$.

From $uv = 1$ and the induction we have

$$S_{p-1} = u^{2^{p-2}} + v^{2^{p-2}} = u^{2^{p-2}}(1 + v^{2 \cdot 2^{p-2}}).$$

Since $q | S_{p-1}$ we then obtain

$$u^{2^{p-2}}(1 + v^{2 \cdot 2^{p-2}}) \equiv 0 \bmod q.$$

Now $u = 2 - \sqrt{3}$ is not congruent to 0 mod q for if it were

$$2 \equiv \sqrt{3} \bmod q \implies 4 \equiv 3 \bmod q$$

which is possible only if $q = 1$. Hence mod q

$$1 + v^{2 \cdot 2^{p-2}} \equiv 1 + v^{2^{p-1}} \equiv 0 \implies v^{2^{p-1}} \equiv -1.$$

Therefore $v^{2^p} \neq 1$ in F_q. It follows that the multiplicative order of v mod q must divide 2^p and therefore the multiplicative order of v as an element of F must also divide 2^p. This then must be a power of 2 say 2^m. If $m \leq p - 1$ then $2^m | 2^{p-1}$ from which it follows that $v^{2^{p-1}} = 1$ in F_q and not -1. Therefore m must equal p and the order of v in F must be exactly 2^p.

However as explained earlier the order of any nonzero element in F must divide $q^2 - 1$ and so $2^p | (q^2 - 1)$ which implies that $2^p < q^2 - 1$. On the other hand we have $2^p = M_p + 1$ and $q < \sqrt{M_p}$ and so we have the inequality

$$M_p + 1 = 2^p < q^2 - 1 < M_p - 1$$

which is a contradiction. Therefore no such q can exist and therefore M_p must be prime proving the Lucas–Lehmer Theorem in one direction.

Conversely we show that if M_p is prime then $M_p | S_{p-1}$.

Let $q = M_p$ and let $u = 2 - \sqrt{3}$, $v = 2 + \sqrt{3}$ as in the first part of the proof. We will show that

$$v^{2^{p-1}} \equiv -1 \bmod q$$

and hence

$$S_{p-1} = u^{2^{p-2}} + v^{2^{p-2}} = u^{2^{p-2}}(1 + v^{2 \cdot 2^{p-2}}) \equiv 0 \bmod q.$$

This then shows that $M_p | S_{p-1}$.

To show that v has this order notice first that $q - 1 = 2^p - 2 = 2(2^{p-1} - 1)$. It follows that $\frac{q-1}{2}$ is odd so that $(-1)^{\frac{q-1}{2}} = -1$ so that -1 is not a square mod q.

Next, notice that since q is prime $2^q \equiv 2 \bmod q$ from Fermat's theorem. Hence $2^{q+1} \equiv 4 \bmod q$ which implies that $2^{2^p} \equiv 4 \bmod q$. Since p is a prime ≥ 3 it follows that mod q, 2 has both a square root ($2^{1/2} = 2^{(q+1)/4}$) and a fourth root ($2^{1/4} = 2^{\frac{q+1}{8}}$) mod q.

Finally as a preliminary we show that 3 is not a square mod q. One of the three consecutive integers $q - 1, q, q + 1$ must be divisible by 3. $q + 1 = 2^p$ is a power of 2 and q is a prime > 3. Hence $3 | (q - 1)$. Let g be a generator of the multiplicative group of $\mathbb{Z}_q$. It follows that $w = g^{\frac{q-1}{3}}$ satisfies $w^3 \equiv 1 \bmod q$ and $w \not\equiv 1 \bmod q$. Since

$$w^3 - 1 = (w - 1)(w^2 + w + 1)$$

it follows that

$$w^2 + w + 1 \equiv 0 \bmod q.$$

Let $z = w - w^2$. Then mod q,

$$z^2 \equiv (w - w^2)^2 \equiv w^2 - 2w^3 + w^4 \equiv w^2 - 2 + w \equiv -3.$$

Therefore -3 is a square mod q. Since -1 is not a square mod q it follows that 3 is also not a square mod q.

Since 3 is not a square mod q let F be the extension field of $\mathbb{Z}_q$ adjoining a $\sqrt{3}$. That is $F = \mathbb{Z}_q(r)$ where $r^2 = 3$. F is then a finite field of order q^2.

Let $v = 2 + r = 2 + \sqrt{3}$ in F. Since 3 is not a square mod q we have $3^{\frac{q-1}{2}} = -1$ mod q. Hence in F,

$$v^q = (2 + r)^q = 2^q + r^q = 2 + (\sqrt{3})^q = 2 + 3^{\frac{q}{2}};$$

$$\Longrightarrow\ v^q = 2 + 3^{\frac{q-1}{2}} \cdot 3^{\frac{1}{2}} = 2 - 3^{\frac{1}{2}} = 2 - \sqrt{3} = u.$$

Since 2 is a square mod q, 2^{-1} is also a square mod q. Here 2^{-1} is the multiplicative inverse of 2 mod q which exists since q is an odd prime. Let $2^{-\frac{1}{2}}$ be a squareroot of 2^{-1} mod q. Let $t \in F$ be given by

$$t = (1 + r)2^{-\frac{1}{2}}.$$

Then in F we have

$$t^2 = (1 + r)^2(2^{-\frac{1}{2}})^2 = (1 + 2r + r^2)2^{-1} = (1 + 2r + 3)2^{-1} = 2 + r = v.$$

Therefore t is a squareroot of v in F. We show that v does not have a fourth root in F.

Suppose v had a fourth root. Then t would have to be a square and since $2^{-\frac{1}{2}}$ is a square this would imply that $1 + r$ would have to be a square also. Hence we show that $1 + r$ is not a square in F. This is done by computation in F. The elements of F are of the form $a + bw$ with $a, b \in \mathbb{Z}_q$. Suppose that $(a + bw)^2 = 1 + r$. Then

$$a^2 + 2abw + b^2w^2 = (a^2 + 3b^2) + (2ab)w = 1 + r.$$

This would imply that

$$a^2 + 3b^2 = 1 \text{ and } 2ab = 1 \implies a^2 + 3b^2 \equiv 2ab \bmod q$$

$$\Longrightarrow\ a^2 - 2ab + 3b^2 = (a - b)^2 + 2b^2 \equiv 0 \bmod q$$

$$\Longrightarrow\ \frac{(a-b)^2}{b^2} = (\frac{a-b}{b})^2 \equiv -2 \bmod q.$$

Hence -2 must be a square mod q. However 2 is a square mod q and -1 is not a square mod q and therefore -2 cannot be a square. Therefore $1 + r$ is not a square in F and hence v has no fourth root in F.

Now $v^q = u$ so $v^{q+1} = uv \equiv 1 \bmod q$. Since v has no fourth root it follows that in F the order of t is precisely $2(q+1)$. Since this must divide $q^2 - 1 = (q+1)(q-1)$ it follows that the order of v must be exactly $q + 1$. But then

$$v^{\frac{q+1}{2}} = v^{2^{p-1}} \equiv -1 \bmod q$$

completing the proof. □

Based on the theorem the algorithm for testing a Mersenne prime is particularly simple.

Lucas–Lehmer Algorithm

```
Input a prime $p$.
        a: Let $u = 4$.
        b: For $i = 3$ to $p$
           (1): Let $u \equiv u^2 - 2 \bmod (2^p - 1)$.
              (a): If $u = 0$ output **prime** and finish
              (b): else next $i$.
        c: output **composite**.
```

5.3.3 Some Additional Primality Tests

The Lucas–Lehmer test is called an **n+1-test** since it requires knowledge of a complete factorization of $n + 1$. (Recall $M_n = 2^n - 1$ so $M_n + 1 = 2^n$.) Other tests have been developed to handle the situation where there is knowledge of a complete factorization of $n - 1$. These are known as **(n − 1) tests** and handle, in particular, testing for Fermat primes. Recall (see Chapter 3) that the Fermat numbers are the sequence (F_n) of positive integers defined by

$$F_n = 2^{2^n} + 1, n = 1, 2, 3, \ldots..$$

If F_m is prime it is called a **Fermat prime**. As discussed in Chapter 3, Fermat believed that all the numbers in this sequence were primes. In fact F_1, F_2, F_3, F_4 are all prime but F_5 is composite. It is still an open question whether or not there are infinitely many Fermat primes, however, it has been conjectured that there are only finitely many. On the other hand if a number of the form $2^n + 1$ is a prime for some integer n then it must be a Fermat prime (see Theorem 3.1.5). Lucas' primality test (Theorem 5.3.3) can be considered an $(n - 1)$-test.

Lucas' result was strengthened by Pocklington in the following form:

Theorem 5.3.16 *(Pocklington's Theorem) Suppose* $(n-1) = fr$ *with* $(f, r) = 1$ *and suppose that a complete factorization of* f *is known. Suppose that there exists an* a *such that*

$$a^{n-1} \equiv 1 \textit{ mod } n \textit{ and } (a^{\frac{n-1}{q}}, n) = 1$$

for every prime factor q of f. Then every prime factor of n is congruent to 1 *mod* f.

Proof Let p be a prime factor of n. Since $a^{n-1} \equiv 1 \bmod n$ the multiplicative order d of a^r in the finite field $\mathbb{Z}_p$ is a divisor of $\frac{n-1}{r} = f$. However from $(a^{\frac{n-1}{q}}, n) = 1$ it follows that d cannot be a proper divisor of f and hence $d = f$. Therefore $f|(p-1)$ since the multiplicative group in $\mathbb{Z}_p$ has order $p-1$. □

Pocklington's theorem can then be fashioned into a primality test.

Corollary 5.3.2 *Suppose $n - 1 = fr$ with $(f, r) = 1$ and suppose that a complete factorization of f is known. Suppose that there exists an a such that*

$$a^{n-1} \equiv 1 \text{ mod } n \text{ and } (a^{\frac{n-1}{q}}, n) = 1$$

for every prime factor q of f. Then if $f \geq \sqrt{n}$ it follows that n is prime.

Proof From Theorem 5.3.16 it follows that each prime factor p of n is congruent to 1 mod f. Hence $p > f$. But $f \geq \sqrt{n}$ so each $p > \sqrt{n}$. Therefore n cannot have a prime factor $\leq \sqrt{n}$ and so $n = p$ and n is prime. □

Pocklington's theorem which was done in 1914 actually extended several earlier results that were specific to the testing of Fermat numbers for primality. Pepin' theorem (Theorem 5.3.17) was done in 1877 and Proth's theorem in 1878.

Theorem 5.3.17 *(Pepin's Theorem) Let $F_n = 2^{2^n} + 1$ be the nth Fermat number. Then F_n is prime if and only if $3^{\frac{F_n-1}{2}} \equiv -1$ mod F_n.*

Proof If $3^{\frac{F_n-1}{2}} \equiv -1 \bmod F_n$ then the argument used in proving Pocklington's theorem with $a = 3$ can be used to show that F_n is prime. Conversely suppose F_n is prime. Then $3^{\frac{F_n-1}{2}} \equiv (\frac{3}{F_n}) \bmod F_n$ where $(\frac{3}{F_n})$ is the Jacobi symbol. It is straightforward to check (see the exercises) that $(\frac{3}{F_n}) = -1$. □

Theorem 5.3.18 *(Proth's Theorem) Let $n = f \cdot 2^k + 1$ with $2^k > f$. If there exists an integer a with $a^{\frac{n-1}{2}} \equiv -1$ mod n then n is prime.*

Proof The same arguments as in the proof of Pocklington's theorem can be applied. □

These results, together with the Lucas–Lehmer test, just begin to scratch the surface of primality testing. A complete discussion of primality testing together with discussions of computational complexity of both primality testing and factorization algorithms can be found in the excellent and comprehensive book by Crandall and Pomerance [CP]. There are also many suggestions given in [CP] for research problems.

Recent work, leading eventually to the polynomial time algorithm (AKS), has concentrated on improving both the running time and computational complexity of primality testing algorithms. The major breakthrough from a computational point of view came with the development in 1983 by Adelman, Pomerance, and Rumely of a deterministic algorithm (the APR algorithm) based on Jacobi sums (see [CP]) which ran in subexponential time. The fact that this could be done was in essence the first step toward the eventual polynomial time algorithm. The approach of the APR algorithm extended a line of research that considered testing for primality via Gauss sums (see [CP]).

5.3.4 Elliptic Curve Methods

There have been many additional approaches to primality testing. A very fruitful approach which has had wide ranging applications both in number theory and cryptography is to use **elliptic curves**. In this section we define and explains elliptic curves and their utilization in primality testing. Then in Section 5.6 we will discuss elliptic curve cryptography.

If F is a field of characteristic not equal to 2 or 3 then an **elliptic curve** over F is the locus of points $(x, y) \in F \times F$ satisfying the equation

$$y^2 = x^3 + ax + b \text{ with } 4a^3 + 27b^2 \neq 0.$$

We denote by 0 a single point at infinity and let

$$E(F) = \{(x, y) \in F \times F; y^2 = x^3 + ax + b\} \cup \{0\}.$$

We also call $E(F)$ an elliptic curve over F.

The important thing about elliptic curves from the viewpoint of number theory and primality testing is that a group structure can be placed on $E(F)$. In particular we define the operation $+$ on $E(F)$ by the following rules that for future reference we will denote by (ECR):

1. $0 + P = P$ for any point $P \in E(F)$
2. If $P = (x, y)$ then $-P = (x, -y)$ and $-0 = 0$
3. $P + (-P) = 0$ for any point $P \in E(F)$
4. If $P_1 = (x_1, y_1), P_2 = (x_2, y_2)$ with $P_1 \neq -P_2$ then

$$P_1 + P_2 = (x_3, y_3) \text{ with}$$

$$x_3 = m^2 - (x_1 + x_2), y_3 = -m(x_3 - x_1) - y_1$$

and

$$m = \frac{y_2 - y_1}{x_2 - x_1} \text{ if } x_2 \neq x_1 \text{ and}$$

$$m = \frac{3x_1^2 + a}{2y_1} \text{ if } x_2 = x_1.$$

This operation has a very nice geometric interpretation if $F = \mathbb{R}$ the real numbers. It is known as the chord and tangent method. If $P_1 \neq P_2$ are two points on the curve then the line through P_1, P_2 intersects the curve at another point P_3. If we reflect P_3 through the x-axis we get $P_1 + P_2$. If $P_1 = P_2$ we take the tangent line at P_1.

With this operation $E(F)$ becomes an abelian group (due to Cassels) whose structure can be worked out (see [CP]).

Theorem 5.3.19 *$E(F)$ together with the operations defined above forms an abelian group. If F is a finite field of order p^k then $E(F)$ is either cyclic or has the structure*

$$E(F) = \mathbb{Z}_{m_1} \times \mathbb{Z}_{m_2}$$

with $m_1|m_2$ and $m_1|(p^k - 1)$.

By considering the order of the group $E(F)$ over finite fields, Lenstra developed a factorization algorithm (ECM) (see [CP]). His method, as well as elliptic curve primality testing, depends on the concept of an **elliptic pseudocurve**. An important fact in forming the elliptic curve group is that F is a field. An **elliptic pseudocurve** is the set of points satisfying an elliptic curve equation over a modular ring $\mathbb{Z}_n$ not necessarily a field. If n is not a prime then we cannot expect the total validity of the group laws. Even the combination of two points is not necessarily defined in all cases. This is the idea behind **Lenstra's Factorization Algorithm** (ECM).

In particular if n is a positive integer with $(n, 6) = 1$, $a, b \in \mathbb{Z}_n$ and $(4a^3 + 27b^2, n) = 1$ if a, b are considered as integers, then an elliptic pseudocurve over $\mathbb{Z}_n$ is a set

$$E(\mathbb{Z}_n) = \{(x, y) \in \mathbb{Z}_n \times \mathbb{Z}_n; y^2 = x^3 + ax + b\} \cup \{0\}$$

with 0 a point at infinity. As usual we identify $\mathbb{Z}_n$ with $\{0, 1, \dots, n - 1\}$. The name **pseudocurve** indicates that $\mathbb{Z}_n$ need not be a field.

Using the concept of a pseudocurve, Goldwater and Killian developed an elliptic curve analog of Pocklington's theorem (Theorem 5.3.16) which ushered in elliptic curve primality proving (ECPP). We state the theorem and then discuss pseudocurves in more detail.

Theorem 5.3.20 *(ECPP) Let $n > 1$ with $(n, 6) = 1$, $E(\mathbb{Z}_n)$ an elliptic pseudocurve over $\mathbb{Z}_n$ and s, m positive integers with $s|m$. Let $[m]$ denote the residue class of m and assume that there exists a point $P \in E$ such that $[m]P = 0$ and $[\frac{m}{q}]P \neq 0$ for every prime divisor q of s. Then for every prime p dividing n we have*

$$|E(\mathbb{Z}_p)| \equiv 0 \; mod \; s.$$

Further if $s > (n^{\frac{1}{4}} + 1)^2$ then n is prime.

The Goldwater–Killian theorem was improved upon by Atkin and Morain who developed a very efficient elliptic curve primality testing algorithm. In practice this algorithm seems to be at present the fastest computationally. However, it is felt that ultimately an implementation of the theoretically faster AKS algorithm will be developed that will be computationally faster.

To handle pseudocurves we first transfer the addition operation from elliptic curves to pseudocurves and assume the restriction that the results are not always defined. Let us consider the rules (ECR) for the definition of P_1+P_2. If $P_1 \neq P_2$ then P_1+P_2 is defined only if x_2-x_1 is invertible in $\mathbb{Z}_n$, that is, if $(x_2-x_1, n)=1$. If $P_1=P_2$ then $P_1+P_2=2P_1$ is only defined if $2y_1$ is invertible in $\mathbb{Z}_n$, that is, if $(y_1, n)=1$ because $2 \nmid n$. We remark that if $P=(x, y)$ then $-P=(x,-y)$ with $P+(-P)=P-P=0$ always exists.

The idea of using this for factorization is due to H.W. Lenstra Jr. We now describe the details of the method. As above let $n \in \mathbb{N}$ with $n \geq 2$ and $(n, 6)=1$. First, we randomly determine $a, x, y \in \{0, 1, \ldots, n-1\}$ and then compute b according to the rule

$$b \equiv y^2-x^3-ax \bmod n.$$

If $(4a^3+27b^2, n) \neq 1$ then we either found a nontrivial divisor of n or we repeat the process. If we determined a and b first it would be more difficult to find a (random) point of the curve.

The pseudocurve $E(\mathbb{Z}_n)$ is now given by the equation

$$y^2=x^3+ax+b \text{ with } (4a^3+27b^2, n)=1$$

and the point on $E(\mathbb{Z}_n)$ is $P=(x, y)$.

Lemma 5.3.2 *Let $m>1$ be a divisor of n and let $P_1, P_2 \in E(\mathbb{Z}_n)$ be points such that P_1+P_2 is defined. Then*

$$(P_1 \textit{ mod } m)+(P_2 \textit{ mod } m) \equiv (P_1+P_2) \textit{ mod } m.$$

Proof We may assume that $P_1 \neq 0$ and $P_2 \neq 0$ and let $P_1=(x_1, y_1), P_2=(x_2, y_2)$. The ring homomorphism mod m: $\mathbb{Z}_n \rightarrow \mathbb{Z}_m$ is compatible with forming inverses, that is, for $(x, n)=1$ we have

$$x^{-1} \bmod m=(x \bmod m)^{-1}$$

where on the left-hand side of the equation, the inverse is meant modulo n, while on the right side it is done in $\mathbb{Z}_m$. If for computing the points in $E(\mathbb{Z}_n)$ and $E(\mathbb{Z}_m)$ the same computation rules are applied, then it is true that we can draw mod m into the terms. In particular, we then have that

$$(P_1 \bmod m)+(P_2 \bmod m)$$

is defined.

The critical cases to consider are:

(a) $x_1 \not\equiv x_2 \bmod n$ and $x_1 \equiv x_2 \bmod m$,

(b) $x_1 \equiv x_2 \bmod n$, $y_1 \equiv y_2 \not\equiv 0 \bmod n$, and $y_1 \equiv -y_2 \bmod m$,

(c) $x_1 \equiv x_2 \bmod n$ and $y_1 \not\equiv \pm y_2 \bmod n$.

In case (a), $P_1 + P_2$ is not defined because $x_2 - x_1$ is divisible by m and therefore not invertible modulo n. Thus the conditions of the statement are not met.

In case (b), $P_1 + P_2$ is also not defined because $2y_1 \equiv y_1 + y_2 \bmod n$ and $m|(y_1 + y_2)$. Thus $2y_1$ is not invertible.

If case (c) occurs again $P_1 + P_2$ is not defined. □

Let $E(\mathbb{Z}_n)$, $n \in \mathbb{N}$ with $n \geq 2$ and $(6, n) = 1$ be a (pseudo)curve. If n is not a prime then there are $P_1, P_2 \in E(\mathbb{Z}_n)$ such that $P_1 + P_2$ is not defined. It follows that if n is not a prime then there exist

$$P_1 = (x_1, y_1) \in E(\mathbb{Z}_n), P_1 \neq 0 \text{ and } P_2 = (x_2, y_2) \in E(\mathbb{Z}_n), P_2 \neq 0$$

such that one of the following holds:

(a) $x_1 \not\equiv x_2 \bmod n$ and $x_1 \equiv x_2 \bmod m$ for a divisor $m > 1$ of n,

(b) $x_1 \equiv x_2 \bmod n$, $y_1 \equiv y_2 \not\equiv 0 \bmod n$ and $y_1 \equiv -y_2 \bmod m$ for a divisor $m > 1$ of n,

(c) $x_1 \equiv x_2 \bmod n$ and $y_1 \not\equiv \pm y_2 \bmod n$.

Theorem 5.3.21 *Suppose that $P_1 + P_2$ for two points $P_1, P_2 \in E(\mathbb{Z}_n)$ is not defined. Then this yields a nontrivial divisor of n.*

Proof Let $P_1 = (x_1, y_1)$ and $P_2 = (x_2, y_2)$. If $P_1 + P_2$ if undefined then there are three possibilities.

The first is that $x_1 \not\equiv x_2 \bmod n$ but $x_2 - x_1$ is not invertible modulo n. Then $x_2 - x_1$ is not a multiple of n but also not relatively prime to n. Thus $(x_2 - x_1, n)$ is a nontrivial divisor of n.

The second possibility is $x_1 \equiv x_2 \bmod n$ and $y_1 \equiv y_2 \not\equiv 0 \bmod n$ but $2y_1$ is not invertible modulo n. Since n is odd the greatest common divisor (y_1, n) yields a nontrivial divisor of n.

The final possibility is $x_1 \equiv x_2 \bmod n$ but $y_1 \not\equiv \pm y_2 \bmod n$. Then we have

$$y_2^2 - y_1^2 = (x_2^3 + ax_2 + b) - (x_1^3 + ax_1 + b) \equiv 0 \bmod n.$$

Therefore

$$y_2^2 - y_1^2 = (y_2 + y_1)(y_2 - y_1)$$

is a multiple of n but neither $y_2 + y_1$ nor $y_2 - y_1$ are multiples of n. This implies that both $(y_1 + y_2, n)$ and $(y_2 - y_1, n)$ are nontrivial divisors of n. □

We give an example of using this theorem to factorize n.

EXAMPLE 5.3.4.1 Let $n = 1715761513$ and C a curve given by $y^2 = x^3 + ax + b$ with $a = 42$ and $b = -91$. We first check that C is an elliptic pseudocurve for n. To do this we calculate that $\gcd(6, n) = \gcd(4a^3 + 27b^2, n) = 1$.

The point $P = (2, 1)$ is an element of C so we may test with P. Take

$$k_1 = 2^4 + 2^6 + \cdots + 2^{21}$$

$$k_2 = 2^{23}.$$

Let

$$P_1 = k_1 P = (x_1, y_1)$$

$$P_2 = k_2 P = (x_2, y_2).$$

Then we calculate

$$P_1 = (1115004543, 1676196055)$$

and

$$P_2 = (1267572925, 848156341) \bmod n$$

Then $x_2 - x_1 = 152568382$ and $\gcd(152568382, n) = 26927$. Hence 26927 divides n and we find that

$$n = 26927 \cdot 63719.$$

We now consider the **prime number certification** using the method of Goldwasser–Killian. The idea behind certificates of primality for a number is to provide an efficiently verifiable proof that n is a prime number. One approach goes back to H. Pocklington which was Theorem 5.3.16 and Corollary 5.3.2. We restate it below and then present an example using it.

Theorem 5.3.22 *(Pocklington) Let $a, k, n, q \in \mathbb{N}$ with $n - 1 = qk$ and $q > k$, and let the following properties be satisfied:*

1. *q is a prime number,*
2. *$a^{n-1} \equiv 1 \bmod n$,*
3. *$(a^k - 1, n) = 1$.*

Then n is a prime number.

EXAMPLE 5.3.4.2 We present a certification using Pocklington that 2922529 is a prime number. Note that this number is small enough that it can be checked directly. We have $2922259 - 1 = 1721 \cdot 1698$ and

$$2^{2922259-1} \equiv 1 \bmod 2922259$$

$$(2^{1698} - 1, 2922259) = 1.$$

Both facts can be checked efficiently using modular exponentiation and the Euclidean algorithm. If we knew that 1721 is a prime number than Pocklington's theorem certifies that 2922259 is also a prime number.

We use the same approach for 1721. We have $1721 - 1 = 43 \cdot 40$ and

$$2^{1721-1} \equiv 1 \bmod 1721$$

$$(2^{40} - 1, 43) = 1.$$

Since 43 is a prime number it follows that 1721 is prime and finally 2922259 is prime. The certificate for primality now consists of all the numbers involved in the proof above.

$$\begin{array}{lll} n_1 = 43 & q_1 = 7 & a_1 = 2 \\ n_2 = 1721 & q_2 = 43 & a_2 = 2. \\ n_3 = 2922259 & q_3 = 1721 & a_3 = 2 \end{array}$$

The problem with primality certification using Pocklington's method in this form is that it only works for numbers n where $n - 1$ has a large prime factor. The method of Goldwasser and Kilian carries Pocklington's idea over to elliptic curves. Here, by choosing different curves, very many groups are available. Similar to the situation with the factorization problem, here this method would apply only if one can with high probability count on n being a prime number.

If p is a prime there are many results concerning bounds on the order of the elliptic curve group $|E(\mathbb{Z}_p)|$. We need the following theorem of Hasse, a proof of which can be found in the book by J.H. Silverman [Si].

Theorem 5.3.23 *(Hasse's Theorem) Let $F = F_q$ with $q = p^n$, p prime and $n \geq 1$ and let $E(F)$ be the elliptic curve group for the ellipitc curve $y^2 = x^3 + ax + b$. Then*

$$q + 1 - 2\sqrt{q} \leq |E(F)| \leq q + 1 + 2\sqrt{q}.$$

With this we give another primality test using elliptic pseudocurves.

Theorem 5.3.24 *Let $n \in \mathbb{N}$ and let $E(\mathbb{Z}_n)$ be a pseudocurve over $\mathbb{Z}_n$. Suppose that $0 \neq P \in E(\mathbb{Z}_n)$ and $q > (n^{\frac{1}{4}} + 1)^2$ a prime number. If $q \cdot P = 0$ in $E(\mathbb{Z}_n)$, then n is a prime number.*

Proof Suppose that n is not a prime number. Then there exists a prime factor p of n with $p \leq \sqrt{n}$. Let d be the order of p in the elliptic curve $E(\mathbb{Z}_p)$. It is clear from the lemma that we have $q \cdot P = 0$ in $E(\mathbb{Z}_p)$. This implies that $d|q$. Since q is a prime number and $d \neq 1$ we obtain $d = q$. Thus $q \leq |\overline{E}(\mathbb{Z}_p)|$. However from Hasse's theorem we have

$$|E(\mathbb{Z}_p)| \leq (n^{\frac{1}{4}} + 1)^2$$

providing a contradiction. Therefore n must be prime. □

We note that this is a special case of Theorem 5.3.20.

We now describe the algorithm of Goldwasser and Kilian. Let $n \in \mathbb{N}$ and we want to prove that n is prime and further the test used provides a certificate for the primality. For small n this can be solved by the direct primality testing so we assume that n is sufficiently large. First, by a probabilistic procedure we convince ourselves that with high probability n is a prime number. Then we choose a random (pseudo)curve E over $\mathbb{Z}_n$ and compute the number $|E(\mathbb{Z}_n)|$ under the assumption that n is prime (see [BFKR]). If this calculation is not possible, for example due division by zero, then n is not a prime. We keep searching until $|E(\mathbb{Z}_n)| = kq$ for a number q that is very likely a prime and satisfies

$$(n^{\frac{1}{4}} + 1)^2 < q < \frac{n}{2}$$

for n sufficiently small and k in some sense is small.

Before we certify that q is a prime, we choose a random point $P = (x, y) \in E(\mathbb{Z}_n)$. To do this we repeatedly choose an $x \in \mathbb{Z}_n$ at random, until a $y \in \mathbb{Z}_n$ is found such that $y^2 \equiv x^3 + ax + b \bmod n$. To determine y we use one of the randomized methods for extracting roots in finite fields. Should the process fail, then the chances are good to find a proper divisor of n and thus prove that n is not prime.

In the next step we compute $P' = kP \in E(\mathbb{Z}_n)$. If kP is not defined, then n is not prime.

If $kP = 0$, we search for a new point $P \in E(\mathbb{Z}_n)$.

If $P' \neq 0$ then P' must have order q, or n is not a prime number.

If the computation of $qP' = 0$ in $E(\mathbb{Z}_n)$ is successful then from Theorem 5.3.24 we obtain that n is a prime number unless q is not a prime.

Therefore, finally we apply the method recursively on input q to certify that q is a prime. The certificate to the primality of n consists of the parameters of E, the point P, the value q together with a certificate for the primality of q. If this algorithm yields a result then this algorithm and certificate is correct.

A comprehensive description and discussion of elliptic curve methods can be found in Crandall and Pomerance [CP].

5.4 Cryptography and Primes

Cryptography refers to the science and/or art of sending and receiving coded messages. Coding and hidden ciphering is an old endeavor used by governments and militaries and between private individuals from ancient times. Recently, it has become even more prominent because of the necessity of sending secure and private information, such as credit card numbers, over essentially open communication systems.

In general both the **plaintext message** (uncoded message) and the **ciphertext message** (coded message) are written in some N-letter alphabet which is usually the same for both plaintext and code. The method of coding or the encoding algorithm is then a transformation of the N-letters. The most common way to perform this transformation is to consider the N letters as N integers modulo N and then perform a number theoretical function on them. Therefore most encoding algorithms use modular arithmetic and hence cryptography is closely tied to number theory. In this section we give a brief overview of cryptography and some number theoretic algorithms used in encryption. The subject is very broad, and as mentioned above, very current, due to the need for publically viewed but coded messages. There are many references to the subject. The book by Koblitz [Ko] gives an outstanding introduction to the interaction between number theory and cryptography. It also includes many references to other sources. The book by Baumslag, Fine, Kreuzer and Rosenberger [BFKR] provides a further comprehensive look at mathematical cryptography while the book by Stinson [St] describes the whole area.

Modern cryptography is usually separated into **classical cryptography** and **public key cryptography**. In the former, both the encoding and decoding algorithms are supposedly known only to the sender and receiver, usually referred to as Bob and Alice. In the latter, the encryption method is public knowledge but only the receiver knows how to decode. We make this more precise in Section 5.5 when we introduce public key methods. Here we present first the basic terminology used in classical cryptography.

The message that one wants to send is written in **plaintext** and then converted into code. The coded message is written in **ciphertext**. The plaintext message and ciphertext message are written in some **alphabets** that are usually the same. The process of putting the plaintext message into code is called **enciphering** or **encryption** while the reverse process is called **deciphering** or **decryption**. Encryption algorithms break the plaintext and ciphertext message into **message units**. These are single letters or pairs of letters or more generally k-vectors of letters. The transformations are done on these message units and the encryption algorithm is a mapping from the set of plaintext message units to the set of ciphertext message units. Putting this into a mathematical formulation we let

$$\mathcal{M} = \text{ set of all plaintext message units and}$$

$$\mathcal{C} = \text{ set of all ciphertext message units.}$$

The encryption algorithm is then the application of a left invertible function

$$f : \mathcal{M} \to \mathcal{C}.$$

The function f is the **encryption map**. The left inverse

$$g : \mathcal{C} \to \mathcal{M}$$

is the **decryption** or **deciphering map**. The triple $\{\mathcal{M}, \mathcal{C}, f\}$, consisting of a set of plaintext message units, a set of ciphertext message units and an encryption map is called a **cryptosystem**.

Breaking a code is called **cryptanalysis**. An attempt to break a code is called an **attack**. Often cryptanalysis depends on a statistical frequency analysis of the plaintext language used (see the exercises). Cryptanalysis depends also on a knowledge of the form of the code, that is, the type of cryptosystem used.

We now give some examples of cryptosystems and cryptanalysis.

EXAMPLE 5.4.1 The simplest type of encryption algorithm is a **permutation cipher**. Here the letters of the plaintext alphabet are permuted and the plaintext message is sent in the permuted letters. Mathematically if the alphabet has N letters and σ is a permutation on $1, \ldots, N$, the letter i in each message unit is replaced by $\sigma(i)$. For example suppose the plaintext language is English and the plaintext word is *BOB* and the permutation algorithm is

$$\begin{array}{ccccccccccccc} a & b & c & d & e & f & g & h & i & j & k & l & m \\ b & c & d & f & g & h & j & k & l & n & o & p & r \end{array}$$

$$\begin{array}{ccccccccccccc} n & o & p & q & r & s & t & u & v & w & x & y & z \\ s & t & v & w & x & a & e & i & z & m & q & y & u \end{array}$$

then $bob \rightarrow ctc$.

EXAMPLE 5.4.2 A very straightforward example of a permutation encryption algorithm is a **shift algorithm**. Here we consider the plaintext alphabet as the integers $0, 1, \ldots, N-1 \bmod N$. We choose a fixed integer k and the encryption algorithm is

$$f : m \rightarrow m + k \bmod N.$$

This is often known as a **Caesar code** after Julius Caesar who supposedly invented it. It was used by the Union Army during the American Civil War. For example if both the plaintext and ciphertext alphabets were English and each message unit was a single letter then the number of letters N is $N = 26$. Suppose $k = 5$ and we wish to send the message *ATTACK*. If $A = 0$ then *ATTACK* is the numerical sequence 0, 19, 19, 0, 2, 10. The encoded message would then be *FYYFIP*.

Any permutation encryption algorithm which goes letter to letter is very simple to attack using a statistical analysis. If enough messages are intercepted and the plaintext language is guessed then a frequency analysis of the letters will suffice to crack the code. For example in the English language the three most commonly occurring letters are E, T and A with a frequency of occurrence of approximately 13 % and 9 % and 8 % respectively. By examining the frequency of occurrences of letters in the ciphertext the letters corresponding to E, T and A can be uncovered (see the exercises).

EXAMPLE 5.4.3 A variation on the Caesar code is the **Vigenére code** Here message units are considered as k-vectors of integers mod N from an N letter alphabet. Let $B = (b_1, \ldots, b_k)$ be a fixed k-vector in $\mathbb{Z}_N^k$. The Vigenére code then takes a message unit

$$(a_1, \ldots, a_k) \rightarrow (a_1 + b_1, \ldots, a_k + b_k) \bmod N.$$

From a cryptanalysis point of view a Vigenére code is no more secure than a Caesar code and is susceptible to the same type of statistical attack.

The **Alberti Code** is a **polyalphabetic cipher** and can often be used to thwart a statistical frequency attack. Originally this type of polyalphabetic cipher was developed by Leon Alberti about 1470. Nowadays it more commonly referred to as a Vigenére code after Blaise Vigenére who worked a century after Alberti. A version of a polyalphabetic cipher is described it in the next example.

EXAMPLE 5.4.4 Suppose we have an N-letter alphabet. We then form an $N \times N$ matrix P where each row and column is a distinct permutation of the plaintext alphabet. Hence P is a permutation matrix on the integers $0, \ldots, N-1$. Bob and Alice decide on a **keyword**. The keyword is placed above the plaintext message and the intersection of the keyword letter and plaintext letter below it will determine which cipher alphabet to use. We will make this precise with an 9 letter alphabet $A, B, C, D, E, O, S, T, U$. Here for simplicity we will assume that each row is just a shift of the previous row, but any permutation can be used.

Key Letters

$$\begin{array}{cccccccccccc}
 & & A & B & C & D & E & O & S & T & U \\
a & A & a & b & c & d & e & o & s & t & u \\
l & B & b & c & d & e & o & s & t & u & a \\
p & C & c & d & e & o & s & t & u & a & b \\
h & D & d & e & o & s & t & u & a & b & c \\
a & E & e & o & s & t & u & a & b & c & d \\
b & O & o & s & t & u & a & b & c & d & e \\
e & S & s & t & u & a & b & c & d & e & o \\
t & T & t & u & a & b & c & d & e & o & s \\
s & U & u & a & b & c & d & e & o & s & t
\end{array}$$

Suppose the plaintext message is STAB DOC and Bob and Alice have chosen the keyword BET. We place the keyword repeatedly over the message

$$\begin{array}{ccccccc}
B & E & T & B & E & T & B \\
S & T & A & B & D & O & C
\end{array}$$

To encode we look at B which lies over S. The intersection of the B key letter and the S alphabet is a t so we encrypt the S with T. The next key letter is E which lies over T. The intersection of the E keyletter with the T alphabet is c. Continuing in this manner and ignoring the space we get the encryption

$$\text{STAB DOC} \rightarrow \text{TCTCTDD}$$

EXAMPLE 5.4.5 A final example, which is not number theory based, is the so-called **Beale Cipher**. This has a very interesting history which is related in the popular book *Archimedes Revenge* by P. Hoffman (see [Ho]). Here letters are encrypted by numbering the first letters of each word in some document like the Declaration of Independence or the Bible. There will then be several choices for each letter and a Beale cipher is quite difficult to attack.

5.4.1 Some Number Theoretic Cryptosystems

Here we describe some basic number theoretically derived crytosystems. In applying a cryptosystem to an N letter alphabet we consider the letters as integers mod N. The encryption algorithms then apply number theoretic functions and use modular arithmetic on these integers. One example of this was the shift or Caesar cipher described in Example 5.4.2. In this encryption method a fixed integer k is chosen and the encryption map is given

$$f : m \to m + k \bmod N.$$

The shift algorithm is a special case of an **affine cipher**. Recall that an **affine map** on a ring R is a function $f(x) = ax + b$ with $a, b, x \in R$. We apply such a map to the ring $R = \mathbb{Z}_N$ as the encryption map. Specifically again suppose we have an N letter alphabet and we consider the letters as the integers $0, 1, \dots, (N - 1) \bmod N$, that is in the ring $\mathbb{Z}_N$. We choose integers $a, b \in \mathbb{Z}_N$ with $(a, N) = 1$ and $b \neq 0$. a, b are called the **keys** of the cryptosystem. The encryption map is then given by

$$f : m \to am + b \bmod N$$

EXAMPLE 5.4.1.1 Using an affine cipher with the English language and keys $a = 3, b = 5$ encode the message EAT AT JOE'S. Ignore spaces and punctuation.

The numerical sequence for the message ignoring the spaces and punctuation is

$$4, 0, 19, 0, 19, 9, 14, 4, 18$$

Applying the map $f(m) = 3m + 5 \bmod 26$ we get

$$17, 5, 62, 5, 62, 32, 47, 17, 59 \to 17, 5, 10, 5, 10, 6, 21, 17, 7 \bmod 26.$$

Now rewriting these as letters we get

$$\text{EAT AT JOE'S} \to \text{RFKFKGVRH}.$$

Since $(a, N) = 1$ the integer a has a multiplicative inverse mod N. The decryption map for an affine cipher with keys a, b is then

$$f^{-1} : m \to a^{-1}(m - b) \bmod N.$$

Since an affine cipher, as given above, goes letter to letter it is easy to attack using a statistical frequency approach. Further if an attacker can determine two letters and knows that it is an affine cipher the keys can often easily be determined and the code broken (see the exercises). To give better security it is preferable to use k-vectors of letters as message units. The form then of an affine cipher becomes

$$f : v \to Av + B$$

where here v and B are k-vectors from $\mathbb{Z}_N^k$ and A is an invertible $k \times k$ matrix with entries from the ring $\mathbb{Z}_N$. The computations are then done modulo N. Since v is a k-vector and A is a $k \times k$ matrix the matrix product Av produces another k-vector from $\mathbb{Z}_N^k$. Adding the k-vector B again produces a k-vector so the ciphertext message unit is again a k-vector. The keys for this affine cryptosystem are the enciphering matrix A and the shift vector B. The matrix A is chosen to be invertible over $\mathbb{Z}_N$ (equivalent to the determinant of A being a unit in the ring $\mathbb{Z}_N$) so the decryption map is given by

$$v \to A^{-1}(v - B).$$

Here A^{-1} is the matrix inverse over $\mathbb{Z}_N$ and v is a k-vector. The **enciphering matrix** A and the shift vector B are now the keys of the cryptosystem.

A statistical frequency attack on such a cryptosystem requires knowledge, within a given language, of the statistical frequency of k-strings of letters. This is more difficult to determine than the statistical frequency of single letters. As for a letter to letter affine cipher, if $k + 1$ message units, where k is the message block length, are discovered, then the code can often easily be broken.

EXAMPLE 5.4.1.2 Using an affine cipher with message units of length 2 in the English language and keys

$$A = \begin{pmatrix} 5 & 1 \\ 8 & 7 \end{pmatrix}, B = (5, 3)$$

encode the message EAT AT JOE'S. Again ignore spaces and punctuation.

Message units of length 2, that is 2-vectors of letters are called **digraphs**. We first must place the plaintext message in terms of these message units. The numerical sequence for the message EAT AT JOE's ignoring the spaces and punctuation is as before

$$4, 0, 19, 0, 19, 9, 14, 4, 18.$$

Therefore the message units are

$$(4, 0), (19, 0), (19, 9), (14, 4), (18, 18)$$

repeating the last letter to end the message.

The enciphering matrix A has determinant 27 which is 1 modulo 26 and hence is a unit mod 26. Therefore A is invertible and so it is a valid key.

Now we must apply the map $f(v) = Av + B$ mod 26 to each digraph. For example

$$A\begin{pmatrix} 4 \\ 0 \end{pmatrix} + B = \begin{pmatrix} 5 & 1 \\ 8 & 7 \end{pmatrix}\begin{pmatrix} 4 \\ 0 \end{pmatrix} + \begin{pmatrix} 5 \\ 3 \end{pmatrix} = \begin{pmatrix} 20 \\ 32 \end{pmatrix} + \begin{pmatrix} 5 \\ 3 \end{pmatrix} = \begin{pmatrix} 25 \\ 9 \end{pmatrix}.$$

Doing this to the other message units we obtain

$$(25, 9), (22, 25), (5, 10), (1, 13), (9, 13).$$

Now rewriting these as digraphs of letters we get

$$(Z, J), (W, Z), (F, K), (B, N), (J, N).$$

Therefore the coded message is

EAT AT JOE'S → ZJWZFKBNJN.

EXAMPLE 5.4.1.3 Suppose we receive the message ZJWZFKBNJN and we wish to decode it. We know that an affine cipher with message units of length 2 in the English language and keys

$$A = \begin{pmatrix} 5 & 1 \\ 8 & 7 \end{pmatrix}, B = (5, 3)$$

is being used.

The decryption map is given by

$$v \rightarrow A^{-1}(v - B).$$

so we must find the inverse matrix for A. For a 2×2 invertible matrix $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$ we have

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}^{-1} = \frac{1}{ad - bc} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix},$$

Therefore, in this case, recalling that multiplication is mod 26,

$$A = \begin{pmatrix} 5 & 1 \\ 8 & 7 \end{pmatrix} \implies A^{-1} = \begin{pmatrix} 7 & -1 \\ -8 & 5 \end{pmatrix}.$$

The message ZJWZFKBNJN in terms of message units is

$$(25, 9), (22, 25), (5, 10), (1, 13), (9, 13).$$

We apply the decryption map to each digraph. For example

$$A^{-1}\left(\begin{pmatrix} 25 \\ 9 \end{pmatrix} - B\right) = \left(\begin{pmatrix} 7 & -1 \\ -8 & 5 \end{pmatrix}\begin{pmatrix} 25 \\ 9 \end{pmatrix} - \begin{pmatrix} 5 \\ 3 \end{pmatrix}\right) = (4, 0).$$

Doing this to each we obtain

$$(4, 0), (19, 0), (19, 9), (14, 4), (18, 18)$$

and rewriting in terms of letters

$$(E,A),\ (T,A),\ (T,J),\ (O,E),\ (S,S).$$

This gives us

$$\text{ZJWZFKBNJN} \rightarrow \text{EATATJOESS}$$

5.5 Public Key Cryptography and the RSA Algorithm

Presently there are many instances where secure information must be sent over open communication lines. These include for example banking and financial transactions, purchasing items via credit cards over the internet and similar things. This led to the development of **public key cryptography**. Roughly, in classical cryptography only the sender and receiver know the encoding and decoding methods. Further it is a feature of such cryptosystems, such as the ones that we have looked at, that if the encrypting method is known the decrypting can be carried out. In public key cryptography the encryption method is public knowledge but only the receiver knows how to decode. More precisely in a classical cryptosystem once the encrypting algorithm is known the decryption algorithm can be implemented in approximately the same order of magnitude of time. In a public key cryptosystem, developed first by Diffie and Hellman, the decryption algorithm is much more difficult to implement. This difficulty depends on the type of computing machinery used (much as primality testing) and as computers get better, new and more secure pulic key cryptosystems become necessary.

The basic idea in a public key cryptosystem is to have a **one-way function**. That is a function which is easy to implement but very hard to left invert. Hence it becomes simple to encrypt a message but very hard, unless you know the left inverse, to decrypt. The standard model for public key systems is the following. We assume that the set $\mathcal{M}$ of plaintext message units is the same as the set $\mathcal{C}$ of ciphertext message units and that the decrypting map is the inverse of the encrypting map. Alice wants to send a message to Bob. The encrypting map f_A for Alice is public knowledge as well as the encrypting map f_B for Bob. On the other hand the decryption algorithms g_A and g_B are secret and known only to Alice and Bob respectively. Let m be the message Alice wants to send to Bob. She sends $f_B g_A(m)$. To decode Bob applies first g_B, which only he knows. This gives him $g_B(f_B g_A(m)) = g_A(m)$. He then looks up f_A which is publically available and applies this $f_A(g_A(m)) = m$ to obtain the message. Why not just send $f_B(m)$. Bob is the only one who can decode this. The idea is **authentication**, that is being certain from Bob's point of view that the message really came from Alice. Suppose p is Alice's verification; for example her signature, social security number etc.. If Bob receives $f_B(p)$ it could be sent by anyone since f_B is public. On the other hand since only Alice supposedly knows g_A getting a reasonable message from $g_A(g_B f_B f_A(m))$ would verify that it is from Alice. Applying g_B alone should result in nonsense.

Getting a reasonable one-way function can be a formidable task. The most widely used (at present) public key systems are based on difficult to invert number theoretic functions. Diffie–Hellman in 1976 developed the original public key idea using the **discrete log problem**. In modular arithmetic it is easy to raise an element to a power but difficult to determine, given an element, if it is a power of another element. Specifically if G is a finite group, such as the cyclic multiplicative group of $\mathbb{Z}_p$ where p is a prime, and $h = g^k$ for some k then the **discrete log** of h to the base g is any integer t with $h = g^t$. The rough form of the Diffie–Helman public key system is as follows. Bob and Alice will use a classical cryptosystem based on a key k with $1 < k < q - 1$ where q is a prime. It is the key k that Alice must send to Bob. Let g be a multiplicative generator of $\mathbb{Z}_q^\star$ Alice chooses an $a \in \mathbb{Z}_q$ with $1 < a < q - 1$. She makes public g^a. Bob chooses an element $b \in \mathbb{Z}_q^\star$ and makes public g^b. The secret key is g^{ab}. Both Bob and Alice, but presumably no one else, can discover this key. Alice knows her secret power a and the value g^b is public from Bob. Hence she can compute the key $g^{ab} = (g^b)^a$. The analogous situation holds for Bob. An attacker however only knows g^a and g^b. Unless the attacker can solve the discrete log problem, that is finding the base g, the key exchange is secure.

In 1977 Rivest, Adelman, and Shamir developed the **RSA Algorithm** which is presently one of the most widely used public key cryptosystems. It is based on the difficulty of factoring large integers and in particular on the fact that it is easier to test for primality (hence the inclusion in this chapter) than to factor. It works as follows. Alice chooses two large primes p_A, q_A and an integer e_A relatively prime to $\phi(p_A q_A) = (p_A - 1)(q_A - 1)$. It is assumed that these integers are chosen randomly to minimize attack. The primality tests arise in the following manner. Alice first randomly chooses a large odd integer m and tests it for primality. If its prime it is used. If not, she tests $m+2, m+4, \ldots$ and so on until she gets her first prime p_A. She then repeats the process to get q_A. Similarly she chooses another odd integer m and tests until she gets an e_A relatively prime to $\phi(p_A q_A)$. The primes she chooses should be quite large. Originally RSA used primes of approximately 100 decimal digits, but as computing and attack have become more sophisticated, larger primes have had to be utilized. We will say more of this shortly. Once Alice has obtained p_A, q_A, e_A she lets $n_A = p_A q_A$ and computes d_A, the multiplicative inverse of e_A modulo $\phi(n_A)$. That is d_A satisfies $e_A d_A \equiv 1 \bmod (p_A - 1)(q_A - 1)$. She makes public the enciphering key $K_A = (n_A, e_A)$ and the encryption algorithm known to all is

$$f_A(m) = m^{e_A} \bmod n_A$$

where $m \in \mathbb{Z}_{n_A}$ is a message unit. It can be shown that if

$$(e_A, (p_A - 1)(q_A - 1)) = 1 \text{ and } e_A d_A \equiv 1 \bmod (p_A - 1)(q_A - 1)$$

then $m^{e_A d_A} \equiv m \bmod n_A$ (see the exercises). Therefore the decryption algorithm is

$$g_A(c) = c^{d_A} \bmod n_A.$$

Notice then that $g_A(f_A(m)) = m^{e_A d_A} \equiv m \bmod n_A$ so it is the inverse.

Now Bob makes the same type of choices to obtain p_B, q_B, e_B. He lets $n_B = p_B q_B$ and makes public his key $K_B = (n_B, e_B)$.

If Alice wants to send a message to Bob that can be authenticated to be from Alice she sends $f_B(g_A(m))$, where here we assume that $n_A > n_B$. An attack then requires factoring n_A or n_B which is much more difficult than obtaining the primes p_A, q_A, p_B, q_B.

In practice suppose there is an N letter alphabet which is to be used for both plaintext and ciphertext. The plaintext message is to consist of k vectors of letters and the ciphertext message of l vectors of letters with $k < l$. Each of the k plaintext letters in a message unit m are then considered as integers mod N and the whole plaintext message is considered as a k digit integer written to the base N (see the example below). The transformed message is then written as an l digit integer mod N and then the digits are then considered integers mod N from which encrypted letters are found. To ensure that the range of plaintext messages and ciphertext messages are the same, $k < l$, are chosen so that

$$N^k < n_U < N^l$$

for each user U, that is $n_U = p_U q_U$. In this case any plaintext message m is an integer less than N^k considered as an element of $\mathbb{Z}_{n_U}$. Since $n_U < N^l$ the image under the power transformation corresponds to an l digit integer written to the base N and hence to an l letter block. We give an example with relatively small primes. In real world applications the primes would be chosen to have over a hundred digits and the computations and choices must be done using good computing machinery.

EXAMPLE 5.4.2.1 Suppose $N = 26, k = 2$ and $l = 3$. Suppose further that Alice chooses $p_A = 29, q_A = 41, e_A = 13$. Here $n_A = 29 \cdot 41 = 1189$ so she makes public the key $K_A = (1189, 13)$. She then computes the multiplicative inverse d_A of 13 mod 1120 where $1120 = 28 \cdot 40$. Now suppose we want to send her the message TABU. Since $k = 2$ the message units in plaintext are 2 vectors of letters so we separate the message into TA BU. We show how to send TA. First the numerical sequence for the letters TA mod 26 is (19,0). We then use these as the digits of a 2-digit number to the base 26. Hence

$$\text{TA} \mathrel{\hat{=}} 19 \cdot 26 + 0 \cdot 1 = 494.$$

We now compute the power transformation using her $e_A = 13$ to evaluate

$$f(19, 0) = 494^{13} \bmod 1189.$$

This is evaluated as 320. Now we write 320 to the base 26. By our choices of k, l this can be written with a maximum of 3 digits to this base. Then

$$320 = 0 \cdot 26^2 + 12 \cdot 26 + 8.$$

The letters in the encoded message then correspond to (0, 12, 8) and therefore the encryption of TA is AMI.

To decode the message Alice knows d_A and applies the inverse transformation.

Since we have assumed that $k < l$ this seems to restrict the direction in which messages can be sent. In practice to allow messages to go between any two users the following is done. Suppose Alice is sending an authenticated message to Bob. The keys $k_A = (n_A, e_A), k_B = (n_B, e_B)$ are public. If $n_A \leq n_B$ Alice sends $f_B g_A(m)$. On the other hand if $n_A > n_B$ she sends $g_A f_B(m)$.

The computations and choices used in real-world implementations of the RSA algorithm must be done with computers. Similarly, attacks on RSA are done via computers. As computing machinery gets stronger and factoring algorithms get faster, RSA becomes less secure and larger and larger primes must be used. In order to combat this, other public key methods are in various stages of ongoing development. RSA and Diffie–Hellman and many related public key cryptosystems use properties in abelian groups. In recent years a great deal of work has been done to encrypt and decrypt using certain nonabelian groups such as linear groups or braid groups (see [AG] or [BFX]). Complete treatments of group based cryptography can be found in the books [St], [MSU] and [BFKR].

5.6 Elliptic Curve Cryptography

In Section 5.3.4 we discussed elliptic curves and how they can be utilized in primality testing. Here we show how they can be used effectively in public key cryptography.

The standard public key systems that we have described so far, the Diffie–Hellman and RSA systems, require very large key spaces. In an attempt to use the same ideas but reduce the key space size it was suggested that the Diffie–Hellman method be applied to other abelian groups. To accomplish this, algebraic geometry was introduced into cryptography. In 1985, Neil Koblitz, and independently Victor Miller, suggested the use of elliptic curves over finite fields, and their corresponding groups, as possible cryptographic platforms. These methods have been quite successful and result, in many cases, in faster encryption and smaller key spaces than standard RSA methods. First we describe a basic encryption method developed by ElGamal.

In 1984, T. ElGamal devised a method to turn the Diffie–Hellman key exchange protocol into a public key encryption protocol. This is now known as **ElGamal encryption**. The basic scheme for an ElGamal encryption system is the following. Given a large prime p there is a fixed efficiently invertible procedure to encrypt a plaintext into residue classes within $\mathbb{Z}_p^*$, the unit group within $\mathbb{Z}_p$. Let g be a generator for the cyclic group $\mathbb{Z}_p^*$.

For each message transmission the user's public key is (p, g, A) where $A = g^a$ for some integer a.

The encryption and decryption works as follows. Suppose that Bob wants to send a message to Alice. Alice's public key, which is public knowledge, is (p, g, A) as above. The message is m and, as above, is encrypted in some workable efficient manner within $\mathbb{Z}_p^*$, that is, the message is encrypted in a manner known to all users (once p is given) as an integer in 0, 1, ..., $p-1$. Bob now randomly chooses an integer b and computes $B = g^b$. He now sends (B, mC) to Alice where $C = g^{ab}$. Notice that C is the common shared key in the Diffie–Hellman key exchange and in the encryption this is multiplied by the message m.

To decrypt Alice first uses B to determine the common shared key C. Since $B = g^b$ and she knows $A = g^a$ she knows $C = g^{ab}$ for the same reasons as the Diffie–Hellman key exchange works. Since she knows $C = g^{ab}$ and she knows the modulus p she can compute the inverse $g^{-(ab)}$. This is efficient since it only requires one exponentiation modulo p. She then multiplies $mC = mg^{ab}$ by g^{-ab} to obtain the message m.

Although ElGamal proposed using the cyclic groups $\mathbb{Z}_p^*$ for large primes p, this type of encryption can be used in any cyclic group where the discrete log problem is assumed hard. If the group is a cyclic subgroup within the group of an elliptic curve, ElGamal encryption becomes the basis for **elliptic curve cryptography**.

We assume the material on elliptic curves described in Section 5.3.4 and apply the ElGamal method to the group of an elliptic curve to obtain the **elliptic curve cryptosystem**. We restrict ourselves to odd prime numbers $p \geq 5$ and the corresponding finite fields $\mathbb{Z}_p$.

Consider the elliptic curve (in Weierstrass form) over $\mathbb{Z}_p$ given by

$$y^2 = x^3 + ax + b, a, b \in \mathbb{Z}_p$$

with $\Delta = -4a^3 - 27b^2 \neq 0$ in $\mathbb{Z}_p$.

Now let

$$E(\mathbb{Z}_p) = \{(x, y) \in \mathbb{Z}_p \times \mathbb{Z}_p; y^2 = x^3 + ax + b\} \cup \{0\}$$

be the elliptic curve group of $E(\mathbb{Z}_p)$. The basic idea is to use the ElGamal method, and its dependence on the corresponding discrete log problem, in $E(\mathbb{Z}_p)$, this is, given $P \in E(\mathbb{Z}_p)$, $P \neq 0$, and $nP \in E(\mathbb{Z}_p)$ find n.

We now define the **elliptic curve encryption scheme** which we will abbreviate by ECES. This is also known as the **Elliptic Curve ElGamal Cryptosystem** or the **Meneses-Vanstone Cryptosystem**. This is in general, efficient to encrypt, and requires a smaller keyspace than the RSA method.

ECES Preparation

1. Choose a large odd prime p with $p \geq 5$ and $a, b \in \mathbb{Z}_p$ such that
$$y^2 = x^3 + ax + b$$
is an elliptic curve.
2. Choose an injective efficiently invertible (on the image) map $\rho : \mathcal{M} \to E(\mathbb{Z}_p)\backslash\{0\}$ from the set $\mathcal{M}$ of plain text units to $E(\mathbb{Z}_p)$. We describe such a choice below.
3. Choose a point $P \neq 0$ in $E(\mathbb{Z}_p)$
4. Choose a secret integer $d \in \mathbb{Z}$ and calculate $dP \in E(\mathbb{Z}_p)$

Encryption and Decryption in ECES

1. The public key is (P, dP) with $P \neq 0, P \in E(\mathbb{Z}_p) \times E(\mathbb{Z}_p)$ and the elliptic curve itself. The secret key is d.
2. **Encryption**: Let $m \in \mathcal{M}$ be a plain text message unit. Calculate $Q = \rho(m)$. Choose a random integer $k \in \mathbb{Z}$ and define

$$c = (kP, Q + k(dP)) \in E(\mathbb{Z}_p))^2 = \mathcal{C}.$$

   This is the encrypted message unit.
3. **Decryption**: Let $c = (c_1, c_2) \in \mathcal{C}$ be a ciphertext unit. Calculate $Q = c_2 - dc_1$ and $m = \rho^{-1}(Q)$ the preimage of Q.
   Recall that $Q \in E(\mathbb{Z}_p)$ if $Q = \rho(m)$ and $(c_1, c_2) = (kP, Q + k(dP))$.

Theorem 5.6.1 *ECEC provides a valid cryptosystem.*

Proof Let $(c_1, c_2) = (kP, Q + k(dP))$. Then $c_2 - dc_1 = Q = \rho(m)$. □

Notice that if the discrete log problem for $E(\mathbb{Z}_p)$ is solvable, that is, if we can calculate d from (P, dP) then the ECES is broken.

We now show how to construct an injective, efficiently invertible map $\mathcal{M} \to E(\mathbb{Z}_p) \setminus \{0\}$.

Let $y^2 = x^3 + ax + b$ be an elliptic curve over $\mathbb{Z}_p$ with $p \geq 5$. We have by Hasse's Theorem (see [Sil])

$$|E(\mathbb{Z}_p)| \in [p + 1 - 2\sqrt{p}, p + 1 + 2\sqrt{p}] \cap \mathbb{N}.$$

There are efficient probabilistic algorithms to generate points of $E(\mathbb{Z}_p)$ (see [BFKR]). We need many points in $E(\mathbb{Z}_p)$.

1. Choose $k \in \mathbb{N}$ such that the permitted probability of an error is $< \frac{1}{2^k}$
2. Let $\mathcal{M} = \{0, 1, ..., M\}$. We should have $p > (M + 2)k$.
3. Define an injective map:

$$\Psi : \mathcal{M} \times \{1, ..., k\} \to \mathbb{Z}_p \text{ by } (m, j) \to mk + j.$$

   Recall that $0 \leq mk + j < p$ because $p > (M + 2)k$.
4. Let $x = \Psi(m, 1)$. Calculate $f(x) = x^3 + ax + b$ and check if there exists $y \in \mathbb{Z}_p$ with $y^2 = f(x)$. If this is the case then choose y so that $y \in \{0, 1, ..., \frac{p-1}{2}\}$ and define $\rho(m) = (x, y)$.
   We note that $f(x)$ is a quadratic residue modulo p for about half of the $f(x)$, with $f(x) \neq 0$, and $x \in \mathbb{Z}_p$ gives 0, 1 or 2 points on the elliptic curve.

5. If $x = \Psi(m, 1)$ and there is no $y \in \mathbb{Z}_p$ with $y^2 = f(x)$ then try $x = \Psi(m, 2)$, $x = \Psi(m, 3)$ and so on.
   With probability $> 1 - \frac{1}{2^k}$ we find an element $x \in \{\Psi(m, 1), ..., \Psi(m, k)\}$ with $f(x) = y^2$ for some $y \in \mathbb{Z}_p$.
   If j with $1 \leq j \leq k$ is the smallest integer j such that $x = \Psi(m, j)$ and $f(x) = y^2$ for some $y \in \mathbb{Z}_p$ - such a j exists with probability $> 1 - \frac{1}{2^k}$ - then choose $y \in \{0, 1, ..., \frac{p-1}{2}\}$ and define $\rho(m) = (x, y)$.
6. If $(x, y) \in Im(\rho) \subset E(\mathbb{Z}_p)$ then we may recover m efficiently. If $x = mk + j$ then $m = \frac{x-j}{k}$ because $k \in \mathbb{N}$ and $p > (M + 2)k$.

There has been extensive work on the cryptanalysis of ECES. We mention some general ideas and refer to [BFKR] for more information.

Recall that $E(\mathbb{Z}_p)$ and P are public. An attacker has to calculate $|E(\mathbb{Z}_p)|$ or $|P|$ the order of P in $E(\mathbb{Z}_p)$.

1. ECES is not secure if $|E(\mathbb{Z}_p)|$ has only small prime factors. Hence $|E(\mathbb{Z}_p)|$ should have at least one large prime factor (see[BFKR])
2. Analogously $|P|$ should have at least one large prime factor.
3. ECES is not secure if $|P| = p$. Here we can determine $|E(\mathbb{Z}_p)|$ effectively via the trace t,$t = q + 1 - |E(\mathbb{Z}_p)|$, of the Frobenius map using what is called Schoof's algorithm (see [Sc 1,2]).

Elliptic curves that have passed all known attacks so far can be found at the website http://www.ecc-brainpool.org/ecc-standards.htm.

5.7 The AKS Algorithm

The development of the AKS algorithm and the fact that it is of polynomial time is the major most recent theoretical breakthrough in primality testing. Because of the timeliness and relative simplicity of the proof we here reproduce the arguments in the original paper of Agrawal, Kayena, and Saxena [AKS]. There have already been substantial improvements (see [Bo], [Be]), however, the elegance of the original stands out. For the most part, this section, with some explanatory material, is taken directly from their paper. We first need the following notation. If $p(x), q(x)$ are integral polynomials then we say

$$p(x) \equiv q(x) \bmod (x^r - 1, n)$$

if the remainders of $p(x)$ and $q(x)$ after division by $x^r - 1$ are equal (equal coefficients) modulo n. Further if p is a prime $o_p(r)$ is the multiplicative order of r mod p. Two further number theoretic results are needed.

Lemma 5.7.1 *([Fou85], [BH96]) Let $P(n)$ denote the greatest prime divisor of n. Then there exist constants $c > 0$ and n_0 such that for all $x \geq n_0$*

$$|\{p; p \textit{ prime } p \leq x \textit{ and } P(p-1) > x^{\frac{2}{3}}\}| \geq c\frac{x}{\log_2 x}$$

Lemma 5.7.2 *([A]) If $\pi(x)$ is the standard prime number function then for $n \geq 1$,*

$$\frac{n}{6\log_2 n} \leq \pi(n) \leq \frac{8n}{\log_2 n}$$

We now restate the AKS algorithm as given in [AKS].

AKS Algorithm Program: Input an integer $n > 1$.

```
If $n = a^b$ for some natural numbers $a, b$ with $b > 1$ then output COMPOSITE.
$r = 2$
while $(r < n)$ do {
    if $((n, r) \neq 1)$ output COMPOSITE
    if ($r$ is prime )
        let $q$ be the largest prime factor of $r - 1$
        if $(q \geq 4\sqrt{r}\log_2 n)$ and $(n^{\frac{r-1}{q}} \neq 1) \bmod r$
            break;
    $r \leftarrow r + 1$
}
for $a = 1$ to $[2\sqrt{r}\log_2 n]$
    if $(x - a)^n$ is not congruent to $x^n - a \bmod (x^r - 1, n)$ output COMPOSITE;
output PRIME;
```

The proof by Agrawal, Kayena, and Saxena is in two parts. The first establishes that the algorithm is deterministic. That is the algorithm will return PRIME if and only if the inputted integer is a prime. The second part shows that the algorithm is polynomial in $\log_2 n$ the number of binary digits of n. The remainder of this section is taken from the original paper [AKS].

Theorem 5.7.1 (AKS) *The AKS algorithm returns PRIME if and only if n is prime.*

The proof is established by a series of lemmas. The first lemma bounds the number of iterations in the **while** loop. This loop attempts to find a prime r such that $r - 1$ has a large prime factor $q \geq 4\sqrt{r}\log_2 n$ and $q | o_r(n)$ where $o_r(n)$ is the multiplicative order of n mod r.

Lemma 5.7.3 *There exist positive constants c_1, c_2 for which there is a prime r in the interval $[c_1(\log_2 n)^6, c_2(\log_2 n)^6]$ such that $r - 1$ has a prime factor q with $q \geq 4\sqrt{r}\log_2 n$ and $q | o_r(n)$.*

Proof Let c and $P(n)$ be as in Lemma 5.5.1. For any c_1, c_2 call the primes r in the interval $[c_1(\log_2 n)^6, c_2(\log_2 n)^6]$ that satisfy $P(r-1) > (c_2 \log_2 n)^6)^{\frac{2}{3}} > r^{\frac{2}{3}}$ special primes. Then for n large enough the number of special primes is greater than or equal to

number of special primes in $[1, c_2(\log_2 n)^6]-$ number of primes in $[1, c_1(\log_2 n)^6]$.

Using Lemmas 5.7.1 and 5.7.2 then this value is greater than or equal to

$$\frac{cc_2(\log_2 n)^6}{7\log_2\log_2 n} - \frac{8c_1(\log_2 n)^6}{6\log_2\log_2 n} = \frac{(\log_2 n)^6}{\log_2\log_2 n}(\frac{cc_2}{7} - \frac{8c_1}{6}).$$

Now choose the constants $c_1 \geq 4^6$ and c_2 so that $\frac{cc_2}{7} - \frac{8c_1}{6} > 0$. Call this positive value c_3.

Let $x = c_3(\log_2 n)^6$. Consider the product

$$P = (n-1)(n^2-1)\cdots(n^{[x^{\frac{1}{3}}]}-1).$$

This product has at most $x^{\frac{2}{3}}\log_2 n$ different prime factors. Note that

$$x^{\frac{2}{3}\log_2 n} < \frac{c_3(\log_2 n)^6}{\log_2\log_2 n}.$$

It follows that there is at least one special prime, say r, that does not divide the product P. This is the required prime in the Lemma 5.7.3. $r-1$ has a large prime factor $q \geq r^{\frac{2}{3}} \geq 4\sqrt{r}\log_2 n$ since $c_1 \geq 4^6$ and $q|o_r(n)$. □

Lemma 5.7.4 *If n is prime the AKS algorithm returns PRIME.*

Proof Suppose that n is a prime. Then the **while** loop in the algorithm cannot return COMPOSITE since $(n, r) = 1$ for all $r \leq c_2(\log_2 n)^6$ where c_2 is the constant from Lemma 5.7.3. Since $f(x)^p \equiv f(x^p) \bmod p$ for any integral polynomial, the **for** loop in the algorithm also cannot return COMPOSITE. Hence the algorithm will identify n as PRIME. □

It must be shown now that if n is composite then the algorithm will return COMPOSITE. Suppose that n is composite with the different prime factors $p_1, \ldots, p_k$. Let r be the prime found in the while loop as in Lemma 5.7.3. Then in this case $o_r(n)|lcm(o_r(p_1), \ldots, o_r(p_k))$ and hence there exists a prime factor p of n such that $q|o_r(p)$ with q the largest prime factor of $r-1$. Let p be such a prime factor of n.

The bottom loop in the program uses the value of r to do polynomial computations on the $t = [2\sqrt{r}\log_2 n]$ polynomials $x - a$ for $1 \leq a \leq t$. In the finite field $\mathbb{Z}_p$ the polynomial $x^r - 1$ has an irreducible factor $h(x)$ of degree $d = o_r(p)$. Now

$$(x-a)^n \equiv (x^n - a) \bmod (x^r - 1, n)$$

implies that

$$(x-a)^n \equiv (x^n - a) \bmod (h(x), p).$$

It follows that the polynomial identities on the set of $(x-a)$ hold in the quotient field $\mathbb{Z}_p[x]/(h(x))$. The set of $(x-a)$ form a large cyclic group in this field.

Lemma 5.7.5 *In the field $F = \mathbb{Z}_p[x]/(h(x))$ the group G generated by the t polynomials $(x-a)$ with $1 \le a \le t$ is cyclic and of size $> (\frac{d}{t})^t$ where d is the degree of $h(x)$.*

Proof Recall that the multiplicative group of a finite field is cyclic. Since F is finite and G is a multiplicative subgroup of F it follows that G is also cyclic. What must be shown is the size.

Consider the set

$$S = \{ \prod_{1 \le a \le t} (x-a)^{\alpha_a};\ \sum_{1 \le a \le t} \alpha_a \le d-1, \alpha_a \ge 0,\ \text{for all } 1 \le a \le t\}.$$

The **while** loop ensures that the final r on halting satisfies $r > q > 4\sqrt{r}\log_2 n > t$. If any of the a's are congruent mod p then $p < t < r$ and step 4 of the algorithm identifies n as composite. Therefore any two elements of S are distinct modulo p. This implies that all elements of S are distinct in the field $F = \mathbb{Z}_p[x]/(h(x))$ since the degree of an element of S is less than d, the degree of $h(x)$.

The cardinality of S is then

$$\binom{t+d-1}{t} = \frac{(t+d-1)(t+d-2)\cdots d}{t!} > (\frac{d}{t})^t.$$

Since S is a subset of G this gives the desired result.

□

Since $d > 2t$ the size of G is $> 2^t = n^{2\sqrt{r}}$. From the previous lemma G is cyclic. Let $g(x)$ be a generator of G. The order of $g(x)$ in F is then $> n^{2\sqrt{r}}$. Let

$$I_{g(x)} = \{m;\ g(x)^m \equiv g(x^m) \bmod (x^r - 1, p)\}.$$

Lemma 5.7.6 *The set $I_{g(x)}$ is closed under multiplication.*

Proof Let $m_1, m_2 \in I_{g(x)}$. Then

$$g(x)^{m_1} \equiv g(x^{m_1}) \bmod (x^r - 1, p)$$

and

$$g(x)^{m_2} \equiv g(x^{m_2}) \bmod (x^r - 1, p).$$

Substituting x^{m_1} for x in the second congruence we get

$$g(x^{m_1})^{m_2} \equiv g(x^{m_1 m_2}) \bmod (x^r - 1, p).$$

From these it follows that

$$g(x)^{m_1 m_2} \equiv g(x^{m_1 m_2}) \bmod (x^r - 1, p)$$

and hence $m_1 m_2 \in I_{g(x)}$. □

Lemma 5.7.7 *Let o_g be the order of $g(x)$ in F. Let $m_1, m_2 \in I_{g(x)}$. Then $m_1 \equiv m_2$ mod r implies that $m_1 \equiv m_2$ mod o_g.*

Proof Since $m_1 \equiv m_2 \bmod r$ we have $m_2 = m_1 + kr$ for some $k \geq 0$. Since $m_2 \in I_{g(x)}$, taking congruences in $F = \mathbb{Z}_p[x]/(h(x))$, we get

$$\begin{aligned} g(x)^{m_2} &\equiv g(x^{m_2}) \bmod (x^r - 1, p) \\ \implies g(x)^{m_2} &\equiv g(x^{m_2}) \\ \implies g(x)^{m_1+kr} &\equiv g(x^{m_1+kr}) \\ \implies g(x)^{m_1} g(x)^{kr} &\equiv g(x)^{m_1}. \end{aligned}$$

Now $g(x)$ not congruent to 0 implies that $g(x)^{m_1}$ is not congruent to 0 and hence it has a multiplicative inverse in F. Canceling it from both sides of the congruence above gives

$$g(x)^{kr} \equiv 1.$$

Therefore

$$kr \equiv 0 \bmod o_g \implies m_1 \equiv m_2 \bmod o_g.$$

□

Lemma 5.7.8 *If n is composite the AKS algorithm will return COMPOSITE.*

Proof Suppose that n is composite and suppose that the algorithm returns PRIME. We show a contradiction. The **for** loop ensures that for all $1 \leq a \leq 2\sqrt{r}\log_2 n$,

$$(x - a)^n \equiv (x^n - a) \bmod (x^r - 1, p).$$

The polynomial $g(x)$, the generator of G, is a product of powers of t polynomials $(x - a)$ with $1 \leq a \leq t$ all of which satisfy the above equation. Thus

$$g(x)^n \equiv g(x^n) \bmod (x^r - 1, p).$$

Therefore $n \in I_{g(x)}$. Further $p \in I_{g(x)}$ and $1 \in I_{g(x)}$. We show that $I_{g(x)}$ has too many numbers less than o_g contradicting Lemma 5.7.7.

Consider the set

$$E = \{n^i p^j; 0 \leq i, j \leq [\sqrt{r}]\}.$$

By Lemma 5.7.6, $E \subset I_{g(x)}$. Since $|E| = (1 + [\sqrt{r}])^2 > r$ there are two elements $n^{i_1}p^{j_1}$ and $n^{i_2}p^{j_2}$ in E with $i_1 \neq i_2$ or $j_1 \neq j_2$ such that

$$n^{i_1}p^{j_1} \equiv n^{i_2}p^{j_2} \bmod r$$

by the pigeonhole principle. Then from Lemma 5.7.7

$$n^{i_1}p^{j_1} \equiv n^{i_2}p^{j_2} \bmod o_g.$$

This implies

$$n^{i_1-i_2} \equiv p^{j_2-j_1} \bmod o_g.$$

Since $o_g \geq n^{2\sqrt{r}}$ and $n^{|i_1-i_2|} < n^{2\sqrt{r}}$ and $p^{|j_2-j_1|} < n^{2\sqrt{r}}$ the above congruence becomes an equality. Since p is prime this equality implies $n = p^k$ for some $k \geq 1$. However in step 1 of the algorithm composite numbers of the form p^k for $k \geq 2$ have already been detected. Therefore $n = p$ a contradiction. □

This establishes that the AKS algorithm is deterministic and completes the proof of Theorem 5.7.1.

The final theorem calculates the time complexity of the algorithm. For further details see [AKS].

Theorem 5.7.2 *The asymptotic time complexity of the AKS algorithm is* $O((\log_2 n)^{12} f(\log_2 \log_2 n))$ *where* f *is a polynomial.*

Proof Let $\tilde{O}(t(n))$ stand for $O(t(n)\,\text{poly}(\log_2(t(n)))$ where $t(n)$ is some function of n and poly means polynomial in the argument. In this notation the theorem says that the time complexity is $\tilde{O}((\log_2 n)^{12})$. The first step in the algorithm has asymptotic time complexity $O(\log_2 n)^3$ while the **while** loop makes $O(\log_2 n)^6$ iterations.

The first step in the **while** loop, the gcd computation, takes poly$(\log_2 \log_2 r)$ asymptotic time. The next two steps in the **while** loop would take at most $r^{\frac{2}{2}\,\text{poly}\,(\log_2 \log_2 n)}$ in brute-force implementation. The next three steps take at most poly$(\log_2 \log_2 n)$ steps. Thus the total asymptotic time taken by the **while** loop is $\tilde{O}((r^{\frac{2}{2}(\log_2 n)^6})) = \tilde{O}((\log_2 n)^9)$

The **for** loop does modular computation over polynomials. If repeated squaring and Fast-Fourier Multiplication is used then one iteration of the **for** loop takes $\tilde{O}(\log_2 n \cdot r \log_2 n)$ steps. Thus the **for** loop takes asymptotic time $\tilde{O}(r^{\frac{3}{2}}(\log_2 n)^3) = \tilde{O}((\log_2 n)^{12})$. □

As pointed out in [AKS] in practice the algorithm should actually work much faster. This is due to the relationship to an older conjecture involving what are called Sophie Germain primes. If both r and $\frac{r-1}{2}$ are primes then $\frac{r-1}{2}$ is a **Sophie Germain**

prime and r is a **co-Sophie Germain prime**. In this case $P(r-1) = \frac{r-1}{2}$. It has been conjectured that the number of co-Sophie Germain primes is asymptotic to $\frac{Dx}{(\log_2 x)^2}$ where D is the twin prime constant (see Section 5.2.1). It has been verified for $r \leq 10^{10}$. If the conjecture is true then the **while** loop exits with an r of size $O((\log_2 n)^2)$ taking the overall complexity to $\tilde{O}(\log_2 n)^6)$.

5.8 Exercises

5.1 Use trial division to determine which if any of the following integers are prime.
(a) 10387 (b) 269 (c) 46411

5.2 Use the Sieve of Eratosthenes to develop a list of primes less than 300. (Note this list could be used for Problem 5.1).

5.3 Use the modified Sieve of Eratosthenes to find the integers less than 100 and relatively prime to 891.

5.4 Apply Legendre's formula to evaluate
(a) $N_{655}(200)$ (b) $N_{891}(100)$

5.5 Let $P(x)$ denote the number of primes $p \leq x$ for which $p+2$ is prime. Then by Lemma 5.2.4 for $x \geq 3$ we have

$$P(x) < c\frac{x}{(\ln x)^2}(\ln\ln x)^2$$

where c is a constant. Show that this implies that for $x \geq 3$

$$P(x) \leq k\frac{x}{(\ln x)^{\frac{3}{2}}}$$

where k is a constant.

5.6 Use the integral test for infinite series to show that

$$\sum_{r=1}^{\infty} \frac{1}{r(\ln(r+1))^{\frac{3}{2}}}$$

converges.

5.7 Prove that

$$(-1)^{m+1}\binom{n}{m+1} + (-1)^m\binom{n-1}{m} = (-1)^{m+1}\binom{n-1}{m+1}.$$

5.8 Use the Fermat probable prime test to determine if 42671 is prime or not.

5.9 Use the Lucas test to establish that 271 is prime

5.10 Show that if n is prime and $k \neq 0$, $n \neq 0$ then the binomial coefficient $\binom{n}{k}$ is congruent to 0 mod n.

5.11 Use Problem 5.10 to show that if p is prime then

$$(x - a)^p = x^p - a \text{ in } \mathbb{Z}_p.$$

5.12 Determine the bases b (if any) $0 < b < 14$ for which 14 is a pseudoprime to the base b.

5.13 Prove Lemma 5.3.1: If n is a pseudoprime to the base b_1 and also a pseudoprime to the base b_2 then it is a pseudoprime to the base $b_1 b_2$.

5.14 Show that $561 = 3 \cdot 11 \cdot 17$ is the smallest Carmichael number. (Use the Korselt criterion together with Corollary 5.3.1).

5.15 Define the sequence (S_n) inductively by

$$S_1 = 4 \text{ and } S_n = S_{n-1}^2 - 2.$$

Let $u = 2 - \sqrt{3}$, $v = 2 + \sqrt{3}$. Show that $u + v = 4 = S_1$ and $uv = 1$. Then use induction to show that

$$S_n = u^{2^{n-1}} + v^{2^{n-1}}.$$

5.16 Let $F_n = 2^{2^n} + 1$ be the nth Fermat number. Show that $(\frac{3}{F_n}) = -1$ where $(\frac{3}{F_n})$ is the Jacobi symbol.

5.17 Show that if p, q are primes and e, d are positive integers with $(e, (p-1)(q-1)) = 1$ and $ed \equiv 1 \bmod (p-1)(q-1)$ then $a^{ed} \equiv a \bmod pq$ for any integer a. (This is the basis of the decryption function used in the RSA algorithm.)

5.18 The following table gives the approximate statistical frequency of occurrence of letters in the English language. The passage below is encrypted with a simple permutation cipher without punctuation. Use a frequency analysis to try to decode it.

letter	frequency	letter	frequency	letter	frequency
A	.082	B	.015	C	.027
D	.043	E	.127	F	.022
G	.020	H	.061	I	.070
J	.002	K	.008	L	.040
M	.024	N	.067	O	.075
P	.019	Q	.001	R	.060
S	.063	T	.091	U	.028
V	.010	W	.023	X	.001
Y	.020	Z	.001		

ZKIRNVMFNYVIRHZKLHRGREVRMGVTVIDSR
XSSZHZHGHLMOBKLHRGREVWRERHLIHLMVZ
MWRGHVOUKIRNVMFNYVIHKOZBZXIFXRZOI
LOVRMMFNYVIGSVLIBZMWZIVGSVYZHRHUL
IGHSHVMLGVHGSVIVZIVRMURMRGVOBNZMB
KIRNVHZMWGSVBHVIEVZHYFROWRMTYOLXP
HULIZOOGSVKLHRGREVRMGVTVIH

5.19 Encrypt the message NO MORE WAR using an affine cipher with single letters keys $a = 7, b = 5$.

5.20 Encrypt the message NO MORE WAR using an affine cipher on 2-vectors of letters and an encrypting keys

$$A = \begin{pmatrix} 5 & 2 \\ 1 & 1 \end{pmatrix}, B = (3, 7).$$

5.21 What is the decryption algorithm for the affine cipher given in the last problem.

5.22 How many different affine enciphering transformations are there on single letters with an N letter alphabet.

5.23 Let $N \in \mathbb{N}$ with $N \geq 2$ and $n \rightarrow an + b$ with $b \neq 0$, $(a, N) = 1$ and $(a - 1, N) = 1$. Show that there is always a unique fixed letter. (This can be used in cryptanalysis.)

5.24 Let $N \in \mathbb{N}$ with $N \geq 2$ and $n \rightarrow an + b$ with $(a, N) = 1$ is an affine cipher on an N letter alphabet. Show that if any two letters are guessed $n_1 \rightarrow m_1, n_2 \rightarrow m_2$ with $(n_1 - n_2, N) = 1$ then the code can be broken.

Chapter 6
Primes and Algebraic Number Theory

6.1 Algebraic Number Theory

The final major area within the theory of numbers is **algebraic number theory**. In this chapter we present an overview of the major ideas in this discipline. In line with the theme of these notes, we will concentrate on primes and prime decompositions.

Algebraic number theory is roughly the study of **algebraic number fields**, which are finite extensions of the rationals, and their rings of **algebraic integers**. We will define each of these concepts formally in Section 6.3. Algebraic number theory lies between pure abstract algebra and (elementary) number theory. It originated in methods to solve classical problems in number theory, such as proving Fermat's Big Theorem, but evolved into an independent discipline. It is a true melding of algebra and number theory. Whereas in many places in these notes we used abstract algebra to simplify a proof or clarify an idea in elementary number theory, in algebraic number theory the algebraic concepts are crucial to what is being studied. In fact, the basic terminology and format of modern abstract algebra come from algebraic number theory. While the concepts of rings and fields were implicit in the work of Galois and Abel, it was Kronecker and Dedekind working in number theory who formally defined them in the modern manner.

The starting off point for algebraic number theory was the observation, first made by Gauss, that unique factorization into primes is not unique to the integers. That is, there are other algebraic systems which also permit such unique factorizations. Gauss, in attempting to extend the quadratic reciprocity law, investigated the complex integers $\mathbb{Z}[i] = \{a + bi; a, b \in \mathbb{Z}\}$. They are now called the **Gaussian integers** in his honor. He discovered that he could define divisibility and primes in $\mathbb{Z}[i]$ and that there was a division algorithm analogous to the division algorithm in the ordinary integers $\mathbb{Z}$. From this he derived that in $\mathbb{Z}[i]$ there was unique factorization into primes—of course primes in $\mathbb{Z}[i]$. We will discuss the Gaussian integers in detail in Sections 6.2 and 6.3.

B. Fine and G. Rosenberger, *Number Theory*,
DOI 10.1007/978-3-319-43875-7_6

Kummer, who studied with Gauss, extended these investigations to **complex integers**, which was Kummer's terminology, of the form

$$a_0 + a_1\omega + \cdots + a_{p-1}\omega^{p-1},$$

where $a_i \in \mathbb{Z}$ and ω is a primitive pth root of unity where p is a prime. That is ω is a root of the polynomial equation $x^p - 1 = 0$ with $x \neq 1$. His original motivation was an attempt to prove Fermat's Last Theorem for prime exponents. Kummer's idea was to take $x^p + y^p$ and factor it into

$$x^p + y^p = (x + y)(x + \omega y) \cdots (x + \omega^{p-1} y).$$

Kummer defined divisibility and primes for the sets of complex integers. However it became clear that for some primes p they did not satisfy unique factorization. We will give an example to show this in the next section. To alleviate this problem, the lack of unique factorization, Kummer adjoined to his sets of complex integers certain other complex numbers which he called **ideal numbers**. By allowing these ideal numbers, there was unique factorization. This allowed him to actually settle many cases of Fermat's Last Theorem for prime exponents.

Dedekind, another student of Gauss, extended both Gauss' work on the Gaussian integers and Kummer's ideal numbers. Dedekind introduced the idea of an **algebraic integer** which is defined as a complex number that is a root of a monic polynomial with integral coefficients. That is $\theta \in \mathbb{C}$ is an algebraic integer if $p(\theta) = 0$, where

$$p(x) = x^n + a_{n-1}x^{n-1} + \cdots + a_0, \quad , n \geq 1, a_i \in \mathbb{Z}.$$

Each integer m is of course an algebraic integer satisfying the polynomial $p(x) = x - m$. In this context the ordinary integers are called the **rational integers**. Dedekind introduced the definition of a ring and showed that the set of algebraic integers forms a ring. Further he showed that the algebraic integers within each algebraic number field form a ring within that number field. We will discuss algebraic integers in Section 6.4.

To handle unique factorization, Dedekind worked not with the algebraic integers themselves, but with special subrings of algebraic integers that he called **ideals** in honor of Kummer's ideal numbers. He then showed that he could define divisibility and primes for ideals and then that there was unique factorization of ideals. The concept of an ideal in a ring is now fundamental in abstract algebra. We will discuss general ideals in the next section and then ideals in algebraic number rings in Section 6.5.

Finally Kronecker, a student of Kummer, developed a general theory of fields and algebraic numbers over a field. By considering polynomial rings over a general field he showed, given an irreducible polynomial, that it was always possible to construct a field where this polynomial has a root. This is done by **adjoining** the root to the original field. This is now known as **Kronecker's Theorem**. It was implied in the work of Abel and Galois done earlier but Kronecker's Theorem is now the cornerstone of Galois Theory.

We begin our overview of algebraic number theory by looking at unique factorization.

6.2 Unique Factorization Domains

The true beginning point for the theory of numbers was the Fundamental Theorem of Arithmetic which said that any rational integer could be factored into primes and that this factorization is unique up to ordering and unit factors. Algebraic number theory begins with the observation that this property is not unique to $\mathbb{Z}$ but actually holds in many other integral domains. We start by reviewing some basic concepts from abstract algebra that were introduced in Chapter 2.

Recall that an **integral domain** R is a commutative ring R with an identity and with no **zero divisors**. That is, R has the property that if $ab = 0$ with $a, b \in R$ then either $a = 0$ or $b = 0$. It is clear that the integers $\mathbb{Z}$ form an integral domain. A **unit** in an integral domain is an element u with a multiplicative inverse, that is, there exists an element, u_1, which we denote by u^{-1} such that $u \cdot u^{-1} = 1$. It is easy to show that the product of two units is again a unit and hence the set of units in an integral domain forms a group under multiplication (see Chapter 2 and the exercises). A **field** F is an integral domain where every nonzero element is a unit. The rationals $\mathbb{Q}$, the reals $\mathbb{R}$, and the complex numbers $\mathbb{C}$ all form fields.

Two elements r_1, r_2 in an integral domain R are **associates** if there exists a unit u such that $r_1 = ur_2$. We now extend to any integral domain the ideas of divisibility and primes.

Definition 6.2.1 *Let R be an integral domain. If $r_1, r_2 \in R$ then r_1* **divides** *r_2, denoted $r_1|r_2$, if there exists an $r_3 \in R$ such that $r_2 = r_1r_3$. In analogy with the integers, the elements r_1, r_3 are* **factors** *of r_2 and r_1r_3 is a* **factorization** *of r_2. An element $r \in R$ is a* **prime** *if r is not a unit and whenever $r = r_1r_2$ one factor must be a unit.*

We now use the statement of the Fundamental Theorem of Arithmetic to define a unique factorization domain.

Definition 6.2.2 *An integral domain R is a* **unique factorization domain** *or* **UFD** *if for each $r \in R$ then either $r = 0$, r is a unit or r has a factorization into primes which is unique up to ordering and unit factors. This means that if*

$$r = p_1 \cdots p_m = q_1 \cdots q_k,$$

where the p_i and q_j are primes, then $m = k$ and each p_i is an associate of some q_j.

Hence in this more general algebraic language the Fundamental Theorem of Arithmetic states that the integers $\mathbb{Z}$ are a unique factorization domain. However they are the only one. The complex integers, $\mathbb{Z}[i]$, are also a UFD. We will look at these in the next section. As a first example we show that the ring of polynomials over any field F (which we define below) forms a UFD.

If F is a field and n is a nonnegative integer, then a **polynomial of degree** n **over** F is a formal sum of the form

$$P(x) = a_0 + a_1x + \cdots + a_nx^n \tag{6.2.1}$$

with $a_i \in F$ for $i = 0, \ldots, n$, $a_n \neq 0$, and x an indeterminate. A **polynomial** $P(x)$ over F is either a polynomial of some degree or the expression $P(x) = 0$, which is called the **zero polynomial** and has no degree. We denote the degree of $P(x)$ by **deg** $P(x)$. A polynomial of zero degree has the form $P(x) = a_0$ and is called a **constant polynomial** and can be identified with the corresponding element of F. We also call the zero polynomial a constant polynomial and identify it with the zero element of F. The elements $a_i \in F$ are called the **coefficients of** $P(x)$; a_n is the **leading coefficient**. If $a_n = 1$, $P(x)$ is called a **monic polynomial.** Two nonzero polynomials are equal if and only if they have the same degree and exactly the same coefficients. A polynomial of degree 1 is called a **linear polynomial** while one of degree two is a **quadratic polynomial**.

We denote by $F[x]$ the set of all polynomials over F and we will show that $F[x]$ becomes a unique factorization domain. We first define addition, subtraction, and multiplication on $F[x]$ by algebraic manipulation. That is, suppose $P(x) = a_0 + a_1x + \cdots + a_nx^n$, $Q(x) = b_0 + b_1x + \cdots + b_mx^m$ then

$$P(x) \pm Q(x) = (a_0 \pm b_0) + (a_1 \pm b_1)x + \cdots$$

that is, the coefficient of x^i in $P(x) \pm Q(x)$ is $a_i \pm b_i$, where $a_i = 0$ for $i > n$ and $b_j = 0$ for $j > m$. Multiplication is given by:

$$P(x)Q(x) = (a_0b_0) + (a_1b_0 + a_0b_1)x + (a_0b_2 + a_1b_1 + a_2b_0)x^2 + \cdots + (a_nb_m)x^{n+m}$$

that is, the coefficient of x^i in $P(x)Q(x)$ is $(a_0b_i + a_1b_{i-1} + \cdots + a_ib_0)$.

EXAMPLE 6.2.1 Let $P(x) = 3x^2 + 4x - 6$ and $Q(x) = 2x + 7$ be in $\mathbb{Q}[x]$. Then

$$P(x) + Q(x) = 3x^2 + 6x + 1$$

and

$$P(x)Q(x) = (3x^2 + 4x - 6)(2x + 7) = 6x^3 + 29x^2 + 16x - 42.$$

From the definitions the following degree relationships are clear. The proofs are in the exercises.

Lemma 6.2.1 *Let* $P(x) \neq 0$, $Q(x) \neq 0 \in F[x]$. *Then:*

1. *deg* $P(x)Q(x)$ = *deg* $P(x)$ + *deg* $Q(x)$.

2. *deg* $(P(x) \pm Q(x)) \leq \max($*deg* $P(x),$ *deg* $Q(x))$ *if* $P(x) \pm Q(x) \neq 0$.

We next obtain the following.

Theorem 6.2.1 *If F is a field, then $F[x]$ forms an integral domain. F can be naturally embedded into $F[x]$ by identifying each element of F with the corresponding constant polynomial. The only units in $F[x]$ are the nonzero elements of F.*

Proof Verification of the basic ring properties is solely computational and is left to the exercises. Since deg $P(x)Q(x) =$ deg $P(x) +$ deg $Q(x)$ for nonzero polynomials, it follows that if neither $P(x) \neq 0$ nor $Q(x) \neq 0$ then $P(x)Q(x) \neq 0$ and therefore $F[x]$ is an integral domain.

If $G(x)$ is a unit in $F[x]$, then there exists an $H(x) \in F[x]$ with

$$G(x)H(x) = 1.$$

From the degrees we have deg $G(x)$ + deg $H(x) = 0$ and since deg $G(x) \geq 0$, deg $H(x) \geq 0$. This is possible only if deg $G(x)$ = deg $H(x) = 0$. Therefore $G(x) \in F$. □

Now that we have $F[x]$ as an integral domain we proceed to show that there is unique factorization into primes. We first repeat the definition of a prime in $F[x]$. If $0 \neq f(x)$ has no nontrivial, nonunit factors (it cannot be factorized into polynomials of lower degree) then $f(x)$ is a **prime** in $F[x]$ or a **prime polynomial**. A prime polynomial is also called an **irreducible polynomial**. Clearly, if deg $g(x) = 1$ then $g(x)$ is irreducible.

The fact that $F[x]$ is a UFD follows from the division algorithm for polynomials, which is entirely analogous to the division algorithm for integers.

Lemma 6.2.2 *(Division Algorithm in $F[x]$) If $0 \neq f(x), 0 \neq g(x) \in F[x]$ then there exist unique polynomials $q(x), r(x) \in F[x]$ such that $f(x) = q(x)g(x) + r(x)$, where $r(x) = 0$ or deg $r(x) <$ deg $g(x)$. (The polynomials $q(x)$ and $r(x)$ are called, respectively, the quotient and remainder.)*

This theorem is essentially long division of polynomials. A formal proof is based on induction on the degree of $g(x)$. We omit this but give some examples from $\mathbb{Q}[x]$.

EXAMPLE 6.2.2

(a) Let $f(x) = 3x^4 - 6x^2 + 8x - 6$, $g(x) = 2x^2 + 4$. Then

$$\frac{3x^4 - 6x^2 + 8x - 6}{2x^2 + 4} = \frac{3}{2}x^2 - 6 \text{ with remainder } 8x + 18.$$

Thus here $q(x) = \frac{3}{2}x^2 - 6$, $r(x) = 8x + 18$.

(b) Let $f(x) = 2x^5 + 2x^4 + 6x^3 + 10x^2 + 4x$, $g(x) = x^2 + x$. Then

$$\frac{2x^5 + 2x^4 + 6x^3 + 10x^2 + 4x}{x^2 + x} = 2x^3 + 6x + 4.$$

Thus here $q(x) = 2x^3 + 6x + 4$ and $r(x) = 0$.

Using the division algorithm, the development of unique factorization follows in exactly the same manner as in $\mathbb{Z}$. We need the idea of a **greatest common divisor**, or **gcd**, and the lemmas following the definition.

Definition 6.2.3 *(1) If $f(x), g(x) \in F[x]$ with $g(x) \neq 0$ then a polynomial $d(x) \in F[x]$ is the* **greatest common divisor**, *or* **gcd**, *of $f(x), g(x)$ if $d(x)$ is monic, $d(x)$ divides both $g(x)$ and $f(x)$, and if $d_1(x)$ divides both $g(x)$ and $f(x)$ then $d_1(x)$ divides $d(x)$. We write $d(x) = (g(x), f(x))$. If $(f(x), g(x)) = 1$, then we say that $f(x)$ and $g(x)$ are* **relatively prime**. *If $f(x) = g(x) = 0$ then $d(x) = 0$ is the gcd of $f(x)$ and $g(x)$.*

(2) An expression of the form $f(x)h(x) + g(x)k(x)$ is called a **linear combination** *of $f(x), g(x)$.*

Lemma 6.2.3 *Given $f(x), g(x) \in F[x]$ with $g(x) \neq 0$ then the gcd exists, is unique, and equals the monic polynomial of least degree that is expressible as a linear combination of $f(x), g(x)$.*

Finding the gcd of two polynomials is done in the same manner as finding the gcd of two integers. That is, we use the **Euclidean algorithm**. Recall from Chapter 2 that this is done in the following manner. Suppose $0 \neq f(x), 0 \neq g(x) \in F[x]$. Use repeated applications of the division algorithm to obtain the sequence:

$$f(x) = q(x)g(x) + r(x)$$

$$g(x) = q_1(x)r(x) + r_1(x)$$

$$r(x) = q_2(x)r_1(x) + r_2(x)$$

$$\cdots\cdots$$

$$r_{k-1}(x) = q_{k+1}(x)r_k(x).$$

Since each division reduces the degree, and the degree is finite, this process will ultimately end. Let $r_k(x)$ be the last nonzero remainder polynomial and suppose c is the leading coefficient of $r_k(x)$. Then $c^{-1}r_k(x)$ is the gcd. If there does not exist a last nonzero remainder polynomial then $r(x) = 0$ and $g(x)$ is a divisor of $f(x)$. In this case $(f(x), g(x)) = c^{-1}g(x)$, where c is the leading coefficient of $g(x)$. We give an example.

EXAMPLE 6.2.3 In $\mathbb{Q}[x]$ find the gcd of the polynomials

$$f(x) = x^3 - 1 \text{ and } g(x) = x^2 - 2x + 1$$

and express it as a linear combination of the two.

Using the Euclidean algorithm we obtain

$$x^3 - 1 = (x^2 - 2x + 1)(x + 2) + (3x - 3),$$

$$x^2 - 2x + 1 = (3x - 3)(\frac{1}{3}x - \frac{1}{3}).$$

Therefore the last nonzero remainder is $3x - 3$. Since the gcd must be a monic polynomial we divide through by 3 and hence the gcd is $x - 1$.

Working backwards we have

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

so

$$x - 1 = \frac{1}{3}(x^3 - 1) - \frac{1}{3}(x^2 - 2x + 1)(x + 2)$$

expressing the gcd as a linear combination of the two given polynomials.

The next component is Euclid's Lemma applied to the polynomial ring.

Lemma 6.2.4 *(Euclid's Lemma) If $p(x)$ is an irreducible polynomial and $p(x)$ divides $f(x)g(x)$, then $p(x)$ divides $f(x)$ or $p(x)$ divides $g(x)$.*

Proof The proof is identical to the proof in $\mathbb{Z}$. Suppose $p(x)$ does not divide $f(x)$. Then since $p(x)$ is irreducible, $p(x)$ and $f(x)$ must be relatively prime. Therefore, there exist $h(x), k(x)$ such that

$$f(x)h(x) + p(x)k(x) = 1.$$

Multiply through by $g(x)$ to obtain

$$g(x)f(x)h(x) + g(x)p(x)k(x) = g(x).$$

Now, $p(x)$ divides each term on the left-hand side since $p(x)|g(x)f(x)$ and therefore $p(x)|g(x)$. □

Theorem 6.2.2 *If $0 \neq f(x) \in F[x]$ and $f(x)$ is nonconstant, then $f(x)$ has a factorization into irreducible polynomials that is unique up to ordering and unit factors. In other words $F[x]$ is a UFD.*

Proof The proof is almost identical to the proof for $\mathbb{Z}$, and we sketch it. We outlined this sketch in the exercises to Chapter 2. First we use induction on the degree of $f(x)$ to obtain a prime factorization. If deg $f(x)$ =1, then $f(x)$ is irreducible, so suppose deg $f(x) = n > 1$. If $f(x)$ is irreducible, then it has such a prime factorization. If $f(x)$ is not irreducible, then $f(x) = h(x)g(x)$ with deg $g(x) < n$ and deg $h(x) < n$. By the inductive hypothesis, both $g(x)$ and $h(x)$ have prime factorizations, and $f(x)$ does as well.

Now suppose that $f(x)$ has two prime factorizations

$$f(x) = p_1(x)^{n_1} \cdots p_k(x)^{n_k} = q_1(x)^{m_1} \cdots q_t(x)^{m_t},$$

where $p_i(x), i = 1, \dots, n, q_j(x), j = 1, \dots, t$ are prime polynomials and the $p_i(x)$ also the $q_j(x)$ are pairwise relatively prime. Consider $p_i(x)$. Then $p_i(x) | q_1(x)^{m_1} \cdots q_t(x)^{m_t}$, and hence from Euclid's lemma, $p_i(x) | q_j(x)$ for some j. Since both are irreducible, $p_i(x) = cq_j(x)$ for some unit c. By repeated application of this argument we get that $k = t$ and $n_i = m_j$. Thus we have the same primes with the same multiplicities but perhaps unit factors, proving the theorem. □

A polynomial $P(x) \in F[x]$ can also be considered as a function

$$P : F \to F$$

via the **substitution process**. If $P(x) = a_0 + a_1x + \cdots + a_nx^n \in F[x]$ and $t \in F$, then

$$P(t) = a_0 + a_1t + \cdots + a_nt^n \in F$$

since F is closed under all the operations used in the polynomial. If $r \in F$, $P(x) \in F[x]$, and $P(r) = 0$ under the substitution process, we say that r is a **root** of $P(x)$ or a **zero** of $P(x)$. Synonymously we say that r **satisfies** $P(x)$.

Before closing this section, we further review some properties of roots of polynomials which will be essential when we deal with algebraic number fields. First we have an important divisibility property.

Lemma 6.2.5 *If $P(x) \neq 0$ and c is a root of $P(x)$ then $(x - c)$ divides $P(x)$, that is, $P(x) = (x - c)Q(x)$ with deg $Q(x)$ = deg $P(x) - 1$.*

Proof Suppose $P(c) = 0$. Then from the division algorithm

$$P(x) = (x - c)Q(x) + r(x),$$

where $r(x) = 0$ or $r(x) = f \in F$, since deg $r(x) <$ deg $(x - c) = 1$. Therefore

$$P(x) = (x - c)Q(x) + f.$$

Substituting, we have $P(c) = 0 + f = 0$, and $f = 0$. Hence

$$P(x) = (x - c)Q(x).$$

□

Corollary 6.2.1 *An irreducible polynomial of degree greater than one over a field F has no roots in F.*

From this we obtain the following result which bounds the number of roots of a polynomial over a field.

Lemma 6.2.6 *A polynomial of degree n in $F[x]$ can have at most n distinct roots.*

Proof Suppose $P(x)$ has degree n and suppose $c_1, \ldots, c_n$ are n distinct roots. From repeated application of Lemma 6.2.4,

$$P(x) = k(x - c_1) \cdots (x - c_n),$$

where $k \in F$. Suppose c is any other root. Then

$$P(c) = 0 = k(c - c_1) \cdots (c - c_n).$$

Since a field F has no zero divisors, one of these terms must be zero: $c - c_i = 0$ for some i, and hence $c = c_i$. □

Besides having a maximum of n roots (with n the degree) the roots of a polynomial are unique. Suppose $P(x)$ has degree n and distinct roots $c_1, .., c_k$ with $k \leq n$. Then from the unique factorization in $F[x]$ we have

$$P(x) = (x - c_1)^{m_1} \cdots (x - c_k)^{m_k} Q_1(x) \cdots Q_t(x),$$

where $Q_i(x), i = 1, \ldots, t$ are irreducible and of degree greater than 1. The exponents m_i are called the **multiplicities** of the roots c_i. Let c be a root. Then as above,

$$(c - c_1)^{m_1} \cdots (c - c_k)^{m_k} Q_1(c) \cdots Q_t(c) = 0.$$

Now $Q_i(c) \neq 0$ for $i = 1, .., t$ since $Q_i(x)$ are irreducible of degree > 1. Therefore, $(c - c_i) = 0$ for some i, and hence $c = c_i$.

Finally the famous Fundamental Theorem of Algebra (see [FR 2]) says that any nonconstant complex polynomial must have a root. As a consequence of this and the divisibility property it follows that a complex polynomial of degree n must have n roots counting multiplicities.

Theorem 6.2.3 *(Fundamental Theorem of Algebra) If $p(x)$ is a nonconstant complex polynomial, $p(x) \in \mathbb{C}[x]$, then $p(x)$ has a complex root.*

6.2.1 Euclidean Domains and the Gaussian Integers

In analyzing the proof of unique factorization in both $\mathbb{Z}$ and $F[x]$ it is clear that it depends primarily on the division algorithm. In $\mathbb{Z}$ the division algorithm depended on the fact that the positive integers could be ordered and in $F[x]$ on the fact the degrees of nonzero polynomials are nonnegative integers and hence could be ordered. This basic idea can be generalized in the following way.

Definition 6.2.4 *Let R be an integral domain. Then R is a* **Euclidean domain** *if there exists a function N from $R^\star = R \backslash \{0\}$ to the nonnegative integers such that*

1. $N(r_1) \leq N(r_1 r_2)$ *for any* $r_1, r_2 \in R^\star$.
2. *For all* $r_1, r_2 \in R$ *with* $r_2 \neq 0$ *there exists* $q, r \in R$ *such that*

$$r_2 = qr_1 + r,$$

where either $r = 0$ *or* $N(r) < N(r_1)$.

The function N is called a **Euclidean norm** *on R.*

Therefore Euclidean domains are precisely those integral domains which allow division algorithms. In the integers $\mathbb{Z}$ define $N(z) = |z|$. Then N is a Euclidean norm on $\mathbb{Z}$ and hence $\mathbb{Z}$ is a Euclidean domain. On $F[x]$ define $N(p(x)) = deg(p(x))$ if $p(x) \neq 0$. Then N is also a Euclidean norm on $F[x]$ so that $F[x]$ is also a Euclidean domain. In any Euclidean domain we can mimic the proofs of unique factorization in both $\mathbb{Z}$ and $F[x]$ to obtain the following:

Theorem 6.2.4 *Every Euclidean domain is a unique factorization domain.*

Before proving this theorem we must develop some results on the *number theory* of general Euclidean domains. First some properties of the norm.

Lemma 6.2.7 *If R is a Euclidean domain then*

1. $N(1)$ *is minimal among* $\{N(r); r \in R^\star\}$.
2. $N(u) = N(1)$ *if and only if u is a unit.*
3. $N(a) = N(b)$ *for* $a, b \in R^\star$ *if* a, b *are associates.*
4. $N(a) < N(ab)$ *unless b is a unit.*

Proof (1) From property (1) of Euclidean norms we have

$$N(1) \leq N(1 \cdot r) = N(r) \text{ for any } r \in R^\star.$$

(2) Suppose u is a unit. Then there exists u^{-1} with $u \cdot u^{-1} = 1$. Then

$$N(u) \leq N(u \cdot u^{-1}) = N(1)$$

From the minimality of $N(1)$ it follows that $N(u) = N(1)$.

Conversely suppose $N(u) = N(1)$. Apply the division algorithm to get

$$1 = qu + r.$$

If $r \neq 0$ then $N(r) < N(u) = N(1)$ contradicting the minimality of $N(1)$. Therefore $r = 0$ and $1 = qu$. Then u has a multiplicative inverse and hence is a unit.
(3) Suppose $a, b \in \mathbb{R}^\star$ are associates. Then $a = ub$ with u a unit. Then

$$N(b) \leq N(ub) = N(a).$$

On the other hand $b = u^{-1}a$ so

$$N(a) \leq N(u^{-1}a) = N(b).$$

Since $N(a) \leq N(b)$ and $N(b) \leq N(a)$ it follows that $N(a) = N(b)$.
(4) Suppose $N(a) = N(ab)$. Apply the division algorithm

$$a = q(ab) + r,$$

where $r = 0$ or $N(r) < N(ab)$. If $r \neq 0$ then

$$r = a - qab = a(1 - qb) \implies N(ab) = N(a) \leq N(a(1 - qb)) = N(r)$$

contradicting that $N(r) < N(ab)$. Hence $r = 0$ and $a = q(ab) = (qb)a$. Then

$$a = (qb)a = 1 \cdot a \implies qb = 1$$

since there are no zero divisors in an integral domain. Hence b is a unit. Since $N(a) \leq N(ab)$ it follows that if b is not a unit we must have $N(a) < N(ab)$. □

We next need the concept of a gcd.

Definition 6.2.5 *Let R be a Euclidean domain and let $r_1, r_2 \in R$. If $r_2 \neq 0$ then $d \in R$ is a* **GCD** *for r_1, r_2 if $d|r_1$ and $d|r_2$ and if $d_1|r_1$ and $d_1|r_2$ then $d_1|d$. If $r_1 = r_2 = 0$ then $d = 0$ is the gcd of r_1, r_2.*

In $\mathbb{Z}$ GCD's are unique if we choose d to be positive. In general they are only unique up to associates.

Lemma 6.2.8 *Any two gcds of $r_1, r_2 \in R$ are associates. Further an associate of a gcd of r_1, r_2 is also a gcd.*

The proof is straightforward and we leave it to the exercises.

Lemma 6.2.9 *Suppose R is a Euclidean domain and $r_1, r_2 \in R$ with $r_2 \neq 0$. Then a gcd d for r_1, r_2 exists and is expressible as a linear combination with minimal norm. That is there exists $x, y \in R$ with*

$$d = r_1x + r_2y$$

and $N(d) \leq N(d_1)$ for any other linear combination $d_1 = r_1u + r_2v$ of r_1, r_2.

Further if $r_1 \neq 0, r_2 \neq 0$ then a gcd can be found by the Euclidean algorithm exactly as in $\mathbb{Z}$ and $F[x]$.

The proof of this lemma, except for uniqueness which from Lemma 6.2.8 is only true up to associates, is identical to the proof in $\mathbb{Z}$ and we leave it to the exercises (see Chapter 2).

Unique factorization will follow from the analog of Euclid's lemma.

Lemma 6.2.10 *(Euclid's Lemma) Suppose R is a Euclidean domain and $r \in R$ is a prime. If $r|r_1r_2$ then $r|r_1$ or $r|r_2$.*

Proof Suppose $r|r_1r_2$. If r does not divide r_1 then the gcd of r and r_1 must be a unit u since the only factors of r are units and associates of r. Then from Lemma 6.2.8, 1 is also a gcd since 1 is an associate of any unit. Therefore there exists $x, y \in R$ with

$$1 = r_1x + ry.$$

Multiplying through by r_2 we obtain

$$r_2 = (r_1r_2)x + r_2ry.$$

Since $r|r_1r_2$ and $r|r$ it follows that $r|r_2$. □

We can now prove Theorem 6.2.4. Suppose that R is a Euclidean domain. We must show that R is a UFD. First let $r \in R$ with $r \neq 0$. To show that r either is a unit or has a prime factorization we use induction on the norm. If $N(r)$ is minimal then $N(r) = N(1)$ and r is a unit. Suppose that $N(r)$ is the minimal norm greater than $N(1)$. We claim that r must be a prime. If $r = r_1r_2$ and neither r_1 nor r_2 were units from Lemma 6.2.7 then both $N(r_1) < N(r)$, $N(r_2) < N(r)$ contradicting the minimality of $N(r)$ among nonunits. Therefore r is a prime and the beginning of the induction is correct. Assume that if $N(r) < k$ then r has a prime factorization and suppose then that $N(r) = k$. If r is prime then it certainly has a prime factorization. If r is not prime then $r = r_1r_2$ with both r_1, r_2 nonunits. Then $N(r_1) < N(r)$ and $N(r_2) < N(r)$ and from the inductive hypothesis both r_1 and r_2 have prime factorizations and hence so does r.

The uniqueness of the factorization, at least up to units and ordering follows almost identically to what was done in $\mathbb{Z}$. Notice that if r, s are both primes in R and $r|s$ then r, s are associates. Then, as in $\mathbb{Z}$, assume that r has two prime factorizations

$$r = r_1 \cdots r_k = s_1 \cdots s_t$$

with $r_1, \ldots, r_k, s_1, \ldots, s_t$ all primes in R. We now apply Euclid's Lemma repeatedly to get that each r_i is an associate of some s_j and $k = t$. We leave the details to the exercises.

We now apply these ideas to the **Gaussian integers**

$$\mathbb{Z}[i] = \{a + bi; a, b \in \mathbb{Z}\}.$$

It was first observed by Gauss that this set permits unique factorization. To show this we need a Euclidean norm on $\mathbb{Z}[i]$.

Definition 6.2.6 *If $z = a + bi \in \mathbb{Z}[i]$ then its* **norm** *$N(z)$ is defined by*

$$N(a + bi) = a^2 + b^2$$

The basic properties of this norm follow directly from the definition (see exercises).

Lemma 6.2.11 *If $\alpha, \beta \in \mathbb{Z}[i]$ then:*

1. *$N(\alpha)$ is an integer for all $\alpha \in \mathbb{Z}[i]$.*
2. *$N(\alpha) \geq 0$ for all $\alpha \in \mathbb{Z}[i]$.*
3. *$N(\alpha) = 0$ if and only if $\alpha = 0$.*
4. *$N(\alpha) \geq 1$ for all $\alpha \neq 0$.*
5. *$N(\alpha\beta) = N(\alpha)N(\beta)$ that is the norm is multiplicative.*

From the multiplicativity of the norm we have the following concerning primes and units in $\mathbb{Z}[i]$.

Lemma 6.2.12 *(1) $u \in \mathbb{Z}[i]$ is a unit if and only if $N(u) = 1$.*
(2) If $\pi \in \mathbb{Z}[i]$ and $N(\pi) = p$, where p is an ordinary prime in $\mathbb{Z}$ then π is a prime in $\mathbb{Z}[i]$.

Proof Certainly u is a unit if and only if $N(u) = N(1)$. But in $\mathbb{Z}[i]$ we have $N(1) = 1$ so the first part follows.

Suppose next that $\pi \in \mathbb{Z}[i]$ with $N(\pi) = p$ for some $p \in \mathbb{Z}$. Suppose that $\pi = \pi_1\pi_2$. From the multiplicativity of the norm we have

$$N(\pi) = p = N(\pi_1)N(\pi_2).$$

Since each norm is a positive ordinary integer and p is a prime it follows that either $N(\pi_1) = 1$ or $N(\pi_2) = 1$. Hence either π_1 or π_2 is a unit. Therefore π is a prime in $\mathbb{Z}[i]$. □

Armed with this norm we can show that $\mathbb{Z}[i]$ is a Euclidean domain.

Theorem 6.2.5 *The Gaussian integers $\mathbb{Z}[i]$ form a Euclidean domain.*

Proof That $\mathbb{Z}[i]$ forms a commutative ring with an identity can be verified directly and easily. If $\alpha\beta = 0$ then $N(\alpha)N(\beta) = 0$ and since there are no zero divisors in $\mathbb{Z}$ we must have $N(\alpha) = 0$ or $N(\beta) = 0$. But then either $\alpha = 0$ or $\beta = 0$ and hence $\mathbb{Z}[i]$ is an integral domain. To complete the proof we show that the norm N is a Euclidean norm.

From the multiplicativity of the norm we have if $\alpha, \beta \neq 0$,

$$N(\alpha\beta) = N(\alpha)N(\beta) \geq N(\alpha) \text{ since } N(\beta) \geq 1.$$

Therefore property (1) of Euclidean norms is satisfied. We must now show that the division algorithm holds.

Let $\alpha = a + bi$ and $\beta = c + di$ be Gaussian integers. Recall that for a nonzero complex number $z = x + iy$ its inverse is

$$\frac{1}{z} = \frac{\overline{z}}{|z|^2} = \frac{x - iy}{x^2 + y^2}.$$

Therefore as a complex number

$$\frac{\alpha}{\beta} = \alpha\frac{\overline{\beta}}{|\beta|^2} = (a + bi)\frac{c - di}{c^2 + d^2}$$
$$= \frac{ac + bd}{c^2 + d^2} + \frac{bc - ad}{c^2 + d^2}i = u + iv.$$

Now since a, b, c, d are integers u, v must be rationals. The set

$$\{u + iv; u, v \in \mathbb{Q}\}$$

is called the **Gaussian rationals**.

If $u, v \in \mathbb{Z}$ then $u + iv \in \mathbb{Z}[i]$, $\alpha = q\beta$ with $q = u + iv$ and we are done. Otherwise choose ordinary integers m, n satisfying $|u - m| \le \frac{1}{2}$ and $|v - n| \le \frac{1}{2}$ and let $q = m + in$. Then $q \in \mathbb{Z}[i]$. Let $r = \alpha - q\beta$. We must show that $N(r) < N(\beta)$.

Working with complex absolute value we get

$$|r| = |\alpha - q\beta| = |\beta||\frac{\alpha}{\beta} - q|.$$

Now

$$|\frac{\alpha}{\beta} - q| = |(u - m) + i(v - n)| = \sqrt{(u - m)^2 + (v - n)^2} \le \sqrt{(\frac{1}{2})^2 + (\frac{1}{2})^2} < 1.$$

Therefore

$$|r| < |\beta| \implies |r|^2 < |\beta|^2 \implies N(r) < N(\beta)$$

completing the proof. □

Since $\mathbb{Z}[i]$ forms a Euclidean domain it follows from our previous results that $\mathbb{Z}[i]$ must be a UFD.

Corollary 6.2.2 *The Gaussian integers are a UFD.*

Since we will now be dealing with many kinds of **integers** we will refer to the ordinary integers $\mathbb{Z}$ as the **rational integers** and the ordinary primes p as the **rational primes**. It is clear that $\mathbb{Z}$ can be embedded into $\mathbb{Z}[i]$. However not every rational prime is also prime in $\mathbb{Z}[i]$. The primes in $\mathbb{Z}[i]$ are called the **Gaussian primes**. For example we can show that both $1 + i$ and $1 - i$ are Gaussian primes, that is primes in $\mathbb{Z}[i]$.

However $(1+i)(1-i)=2$ so that the rational prime 2 is not a prime in $\mathbb{Z}[i]$. Using the multiplicativity of the Euclidean norm in $\mathbb{Z}[i]$ we can describe all the units and primes in $\mathbb{Z}[i]$.

Theorem 6.2.6 *Consider the Gaussian integers* $\mathbb{Z}[i]$.

1. *The only units in* $\mathbb{Z}[i]$ *are* $\pm 1, \pm i$.
2. *Suppose* π *is a Gaussian prime. Then* π *is either:*
   - *(a)* *a positive rational prime* $p \equiv 3 \bmod 4$ *or an associate of such a rational prime.*
   - *(b)* $1+i$ *or an associate of* $1+i$.
   - *(c)* $a+bi$ *or* $a-bi$, *where* $a>0, b>0$, *a is even and* $N(\pi)=a^2+b^2=p$ *with p a rational prime congruent to 1 mod 4 or an associate of* $a+bi$ *or* $a-bi$.

Proof (1) Suppose $u=x+iy\in\mathbb{Z}[i]$ is a unit. Then from Lemma 6.2.12 we have $N(u)=x^2+y^2=1$ implying that $(x,y)=(0,\pm 1)$ or $(x,y)=(\pm 1,0)$. Hence $u=\pm 1$ or $u=\pm i$.

(2) Now suppose that π is a Gaussian prime. Since $N(\pi)=\pi\overline{\pi}$ and $\overline{\pi}\in\mathbb{Z}[i]$ it follows that $\pi|N(\pi)$. $N(\pi)$ is a rational integer so $N(\pi)=p_1\cdots p_k$, where the p_i's are rational primes. By Euclid's lemma $\pi|p_i$ for some p_i and hence a Gaussian prime must divide at least one rational prime. On the other hand suppose $\pi|p$ and $\pi|q$, where p,q are different primes. Then $(p,q)=1$ and hence there exist $x,y\in\mathbb{Z}$ such that $1=px+qy$. It follows that $\pi|1$ a contradiction. Therefore a Gaussian prime divides one and only one rational prime.

Let p be the rational prime that π divides. Then $N(\pi)|N(p)=p^2$. Since $N(\pi)$ is a rational integer it follows that $N(\pi)=p$ or $N(\pi)=p^2$. If $\pi=a+bi$ then $a^2+b^2=p$ or $a^2+b^2=p^2$.

If $p=2$ then $a^2+b^2=2$ or $a^2+b^2=4$. It follows that $\pi=\pm 2,\pm 2i$ or $\pi=1+i$ or an associate of $1+i$. Since $(1+i)(1-i)=2$ and neither $1+i$ nor $1-i$ are units it follows that neither 2 nor any of its associates are primes. Then $\pi=1+i$ or an associate of $1+i$. To see that $1+i$ is prime suppose $1+i=\alpha\beta$. Then $N(1+i)=2=N(\alpha)N(\beta)$. It follows that either $N(\alpha)=1$ or $N(\beta)=1$ and either α or β is a unit.

If $p\neq 2$ then either $p\equiv 3 \bmod 4$ or $p\equiv 1 \bmod 4$. Suppose first that $p\equiv 3 \bmod 4$. Then $a^2+b^2=p$ would imply from Fermat's two-square theorem (see Chapter 2) that $p\equiv 1 \bmod 4$. Therefore from the remarks above $a^2+b^2=p^2$ and $N(\pi)=N(p)$. Since $\pi|p$ we have $\pi=\alpha p$ with $\alpha\in\mathbb{Z}[i]$. From $N(\pi)=N(p)$ we get that $N(\alpha)=1$ and α is a unit. Therefore π and p are associates. Hence in this case π is an associate of a rational prime congruent to 3 mod 4.

Finally suppose $p\equiv 1 \bmod 4$. From the remarks above either $N(\pi)=p$ or $N(\pi)=p^2$. If $N(\pi)=p^2$ then $a^2+b^2=p^2$. Since $p\equiv 1 \bmod 4$ from Fermat's two-square theorem there exist $m,n\in\mathbb{Z}$ with $m^2+n^2=p$. Let $u=m+in$ then the norm $N(u)=p$. Since p is a rational prime it follows from Lemma 6.2.12 that u is a Gaussian prime. Similarly its conjugate $\overline{u}$ is also a Gaussian prime. Now $u\overline{u}|p^2$

and $p^2 = N(\pi)$. Since $\pi | N(\pi)$ it follows that $\pi | u\overline{u}$ and from Euclid's Lemma either $\pi | u$ or $\pi | \overline{u}$. If $\pi | u$ they are associates since both are primes. But this is a contradiction since $N(\pi) \neq N(u)$. The same is true if $\pi | \overline{u}$. It follows that if $p \equiv 1 \bmod 4$ that $N(\pi) \neq p^2$. Therefore in this case $\mathbb{N}(\pi) = p = a^2 + b^2$. An associate of π has both $a, b > 0$ (see exercises). Further since $a^2 + b^2 = p$ one of a or b must be even. If a is odd then b is even and then $i\pi$ is an associate of π with a even, completing the proof. □

In the proof above we used Fermat's two-square theorem. Gauss's original motivation in investigating the complex integers was to prove results in elementary number theory. As an application of unique factorization in $\mathbb{Z}[i]$ we give another proof of the Fermat two-square theorem in the following form.

Theorem 6.2.7 *Let p be an odd rational prime. Then $p = a^2 + b^2$ for $a, b \in \mathbb{Z}$ if and only if $p \equiv 1 \bmod 4$.*

Proof Suppose first that $p = a^2 + b^2$. Since p is odd one of a, b is even and the other is odd. Suppose $a = 2n$ and $b = 2m + 1$ then

$$p = a^2 + b^2 = (2n)^2 + (2m+1)^2 = 4n^2 + 4m^2 + 4m + 1 = 4(n^2 + m^2 + m) + 1$$

and therefore $p \equiv 1 \bmod 4$.

Conversely suppose that $p \equiv 1 \bmod 4$. From Chapter 3 we then have that -1 is a quadratic residue mod p that is there exists an integer x such that $x^2 + 1 \equiv 0 \bmod p$. Then $p | (x^2 + 1)$ so $p | (x + i)(x - i)$. If p were prime, (we cannot use the characterization of primes in $\mathbb{Z}[i]$ since we used the two-square theorem in that proof), then $p | (x + i)$ or $p | (x - i)$. If $p | (x + i)$ then $x + i = p(a + bi)$ for some integers a, b. This would imply that $pb = 1$ which is impossible. Hence p cannot divide $x + i$. An identical argument shows that p cannot divide $x - i$. Therefore p cannot be a Gaussian prime.

Since p is not a Gaussian prime we have a factorization $p = (a + bi)(c + di)$, where neither factor is a unit. Then

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

Since p is prime this implies that $a^2 + b^2 = p$ or $a^2 + b^2 = p^2$. If $a^2 + b^2 = p^2$ then $c^2 + d^2 = 1$ and $c + di$ is a unit contradicting that it is not a unit. Therefore $a^2 + b^2 = p$ and we are done. □

Finally we show that the methods used in $\mathbb{Z}[i]$ cannot be applied to all quadratic integers. Kummer, as mentioned in Section 6.1, considered rings of the form

$$\mathbb{Z}[\sqrt{-p}] = \{a + ib\sqrt{p};\, a, b \in \mathbb{Z},\, p \text{ a prime}\}.$$

One can then define the norm as $N(a + ib\sqrt{p}) = a^2 + pb^2$. This norm is multiplicative $N(\alpha\beta) = N(\alpha)N(\beta)$. However not all of these rings are UFD's. We show for example that there is not unique factorization in $\mathbb{Z}[\sqrt{-5}]$.

By using the multiplicativity of the norm in $\mathbb{Z}[\sqrt{-5}]$ it can be shown that $3, 7, 1 + 2i\sqrt{5}, 1 - 2i\sqrt{5}$ are all primes and not associates (see the exercises). However

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

Therefore factorization into primes in $\mathbb{Z}[\sqrt{-5}]$ is not unique and hence this set is not a UFD. We will examine these rings of quadratic integers more closely in Section 6.4 and consider the question of exactly which ones are UFD's.

6.2.2 *Principal Ideal Domains*

We now take a slightly different approach to UFD's which will eventually lead us to Dedekind's theory of ideals.

Definition 6.2.7 *An* **ideal** *I in an integral domain R is a subring with the property that $RI \subset I$, that is $ri \in I$ for all $r \in R$ and $i \in I$. An ideal is thus a subring closed under multiplication from the whole ring.*

In the rational integers $\mathbb{Z}$ the set $n\mathbb{Z}$ consisting of all multiples of n is an ideal. We will see shortly that every ideal in $\mathbb{Z}$ has this form.

Theorem 6.2.8 *Let R be an integral domain and $\alpha_1, \ldots, \alpha_n$ fixed elements of R. Let $I = \{r_1\alpha_1 + \cdots + r_n\alpha_n; r_i \in R\}$. Then I forms an ideal in R called the* **ideal generated by** $\{\alpha_1, \ldots, \alpha_n\}$. *We will denote this by*

$$< \alpha_1, \ldots, \alpha_n > .$$

If $n = 1$ so that $I =< \alpha >$ with $\alpha \in R$ then I consists of all R-multiples of α. An ideal of this form $< \alpha >$ is called a **principal ideal**.

Proof The proof is straightforward. If $I = \{r_1\alpha_1 + \cdots + r_n\alpha_n; r_i \in R\}$ and $i_1 = r_1\alpha_1 + \cdots + r_n\alpha_n$, $i_2 = s_1\alpha_1 + \cdots + s_n\alpha_n$ are two elements of I then

$$i_1 \pm i_2 = (r_1 \pm s_1)\alpha_1 + \cdots + (r_n \pm s_n)\alpha_n \in I$$

and hence I is closed under addition and additive inverses. If $r \in R$ then

$$ri_1 = (rr_1)\alpha_1 + \cdots + (rr_n)\alpha_n \in I$$

so that I is closed under multiplication from R. Therefore $RI \subset I$ and in particular $I \cdot I \subset I$ so I is closed under multiplication. Therefore I is an ideal. □

Notice that $n\mathbb{Z} =< n >$ is a principal ideal. In the rational integers $\mathbb{Z}$ we have the following.

Theorem 6.2.9 *Every ideal in $\mathbb{Z}$ has the form $n\mathbb{Z}$ for some $n \in \mathbb{Z}$. In particular every ideal in $\mathbb{Z}$ is a principal ideal.*

Proof Let I be an ideal in $\mathbb{Z}$. If $I = \{0\}$ then $I = 0\mathbb{Z}$. If $I \neq \{0\}$ then there exists $z \in I$ with $z \neq 0$. Since I is a subring $-z$ is also in I. Since either z or $-z$ is positive it follows that I must contain positive elements. Let n be the least positive element of I. We show that $I = n\mathbb{Z}$.

Let a be a positive element of I. Then by the division algorithm

$$a = nq + r,$$

where $r = 0$ or $0 < r < n$. If $r \neq 0$ then $0 < r = a - nq < n$. Now $a \in I, n \in I$ and hence nq and $a - nq \in I$ since I is an ideal. This contradicts the minimality of n as the least positive element of I. Therefore $r = 0$ and $a = nq$. If a is a negative element of I then $-a > 0$ and $-a = nq$. Then $a = n(-q)$. Hence every element of I is a multiple of n and therefore $I = n\mathbb{Z}$. □

Definition 6.2.8 *A* **principal ideal domain**, *abbreviated as PID, is an integral domain where every ideal is a principal ideal.*

In this language, Theorem 6.2.9 says that the rational integers $\mathbb{Z}$ are a PID. The same proof using degrees of polynomials would show that the polynomial ring $F[x]$ over a field F is also a PID. This is no accident since both are Euclidean domains and the following is true.

Theorem 6.2.10 *Any Euclidean domain R is a PID.*

The proof is entirely analogous to the proof of Theorem 6.2.3 using the Euclidean norm. We leave the details to the exercises. Euclidean domains are PID's and UFD's. This will follow also from the next result although we proved unique factorization in Euclidean domains directly.

Theorem 6.2.11 *Every PID R is a UFD.*

We use a series of lemmas to obtain a proof of the above result. As for Euclidean domains, uniqueness of prime factorization depends on an analog of Euclid's Lemma. The existence of a prime factorization depends on a property in PID's called the **ascending chain condition**.

Lemma 6.2.13 *Let R be an integral domain and $I_1 \subset I_2 \subset \cdots$, an ascending chain of ideals of R. Then $I = \cup_i I_i$ is also an ideal.*

Proof Let $r_1, r_2 \in I$. Then since $\{I_i\}$ is an ascending chain there exists an I_n with both $r_1, r_2 \in I_n$. Then $r_1 \pm r_2$ and rr_1 with $r \in R$ are all in I_n since I_n is an ideal. But $I_n \subset I$ so all are in I and hence I is an ideal. □

We next show that in a PID every strictly increasing sequence of ideals must terminate. We call this the ascending chain condition or ACC on ideals.

Definition 6.2.9 *An integral domain R satisfies the* **ascending chain condition** *or* **ACC** *on ideals if for every ascending chain of ideals $I_1 \subset I_2 \subset \cdots$ there exists a positive integer n such that $I_i = I_n$ for all $i \geq n$. Equivalently every strictly increasing ascending chain, that is all inclusions proper, must have finite length.*

Lemma 6.2.14 *Every PID satisfies the ACC.*

Proof Let $I_1 \subset I_2 \subset \cdots$ be an ascending chain of ideals in the PID R. Then $I = \cup_i I_i$ is an ideal in R. Since R is a PID we have $I =< r >$ for some $r \in R$. Now $r \in I$ so $r \in I_n$ for some I_n. Then for all $i \geq n$

$$< r > \subset I_n \subset I_i \subset I =< r > .$$

It follows that $I_i = I_n$ for all $i \geq n$ and R satisfies the ACC. □

Finally we need the analog of Euclid's Lemma.

Lemma 6.2.15 *(Euclid's Lemma for PID's) Suppose R is a PID and $p \in R$ is a prime. If $p|ab$ then $p|a$ or $p|b$.*

Proof Notice first the following relationships between divisibility and principal ideals in a PID.

(i) $a|b$ if and only if $< b > \subset < a >$

(ii) $< b > = < c >$ if and only if b and c are associates.

(iii) $< a > = R$ if and only if a is a unit.

The proofs of these properties follow directly from the definitions (see exercises).

Now suppose that p is a prime in R and $p|ab$. Suppose p does not divide a. Then $< a >$ is not contained in $< p >$. It follows that $I =< a, p >$ the ideal generated by a and p is not equal to $< p >$. Since R is a PID we have an element $c \in R$ with $< a, p > = < c >$. Therefore $< p > \subset < c >$ so $p = cr$. Since p is a prime either c or r is a unit. If c is not a unit then p and c are associates and $< p > = < c >$ and hence $< a, p > = < p >$ a contradiction. Therefore c is a unit and $< c > = < a, p > = R$ the whole integral domain. In the next subsection we will see that what we have actually proved is that if p is a prime in a PID then $< p >$ is a maximal ideal. Then since $< a, p > = R$ we must have $1 \in < a, p >$, where 1 is the multiplicative identity.

$$1 \in < a, p > \Longrightarrow ar + ps = 1 \text{ for some } r, s \in R.$$

As in the proof for rational integers multiply through by b to obtain

$$abr + pbs = b.$$

Since $p|ab$ and $p|p$ it follows then that $p|b$. □

We can now prove Theorem 6.2.10.

Proof (of Theorem 6.2.10). We show first that each non-unit in R can be expressed as a product of primes. Let $r \in R$ with $r \neq 0$ and r a nonunit. We show that there is a prime $p \in R$ which divides it. If r is a prime we are done. If not $r = r_1 s$ with neither r_1 nor s a unit. It follows that

$$< r > \subset < r_1 > .$$

If r_1 is prime then r is an associate of r_1 and we are done. If not continue in this manner to obtain an ascending chain of ideals

$$< r > \subset < r_1 > \subset < r_2 > \cdots .$$

By the ACC this chain must terminate at some $< r_n >$ and hence r_n must be a prime, Hence r must be divisible by at least one prime p_1. Therefore $r = p_1 s_1$. By the same argument there is a prime $p_2 | s_1$ so that $r = p_1 p_2 s_2$. We cannot get an infinite factorization by the ACC so it follows that there must be a finite factorization $r = p_1 \cdots p_k$ with p_i all primes. Therefore there must be a prime factorization.

The uniqueness of this factorization up to ordering and units follows analogously to all the previous cases from Euclid's Lemma. If $r = p_1 \cdots p_k = q_1 \cdots q_t$ with p_i, q_j all primes in R then $p_1 | q_j$ for some j and $k = t$. Since both are primes p_1 and q_j are associates. It now goes through as before. □

Hence every PID is a UFD. Are there UFD's which are not PID's? The answer is yes. To give an example we state the following theorem. This is not directly relevant to our subsequent work on algebraic numbers so we omit the proof (and sketch an outline of it in the exercises).

Theorem 6.2.12 *If R is a UFD then the polynomial ring $R[x]$ is also a UFD.*

From this result we have

Corollary 6.2.3 *$\mathbb{Z}[x]$ is a UFD.*

Corollary 6.2.4 *If F is a field then $F[x_1, \ldots, x_n]$, the ring of polynomials in n variables over F, is a UFD.*

From this second corollary we get the example. $F[x, y]$ is a UFD for any field F. Let I be the set of polynomials in $F[x, y]$ with constant term 0. This forms an ideal but it is not principal (see exercises).

6.2.3 Prime and Maximal Ideals

Certain ideas arose in the proof of Theorem 6.2.11 which we look at a bit more closely.

Definition 6.2.10 *An ideal I in an integral domain R is a* **prime ideal** *if whenever $r_1r_2 \in I$ then either $r_1 \in I$ or $r_2 \in I$. I is a* **maximal ideal** *if whenever $I \subset I_1$ with I_1 an ideal then either $I_1 = I$ or $I_1 = R$.*

Hence a maximal ideal is an ideal which is contained in no larger ideal other than the whole integral domain. This is equivalent to $< I, r > = R$ if $r \notin I$. In the proof of Euclid's Lemma for PID's we actually showed that if p is a prime then $< p >$ is a maximal ideal. The general relationship between primes and the principal ideals they generate in PID's is given in the next theorem.

Theorem 6.2.13 *Let R be a PID and let $r \in R$ with $r \neq 0$. Then the following are equivalent:*

1. $r \in R$ is prime.
2. $< r >$ is a prime ideal.
3. $< r >$ is a maximal ideal.

In particular in a PID a nonzero ideal is maximal if and only if it is prime.

Proof We show first that (1) is equivalent to (2). Suppose r is a prime and $r_1r_2 \in < r >$. Then $r|r_1r_2$ so by Euclid's Lemma $r|r_1$ or $r|r_2$. If $r|r_1$ then $r_1 \in < r >$ while if $r|r_2$ then $r_2 \in < r >$. It follows that $< r >$ is a prime ideal.

Conversely suppose that $< r >$ is a prime ideal and $r = r_1r_2$. Since $r_1r_2 \in < r >$ we have either $r_1 \in < r >$ or $r_2 \in < r >$. If $r_1 \in < r >$ then $r_1 = r_3r$ and then

$$r = r_1r_2 = (r_2r_3)r \implies r_3r_2 = 1.$$

Hence r_2 is a unit. Similarly if $r_2 \in < r >$ then r_1 is a unit. It follows that r is prime.

The proof about maximality is essentially the proof of Euclid's Lemma.

We now show that (1) is equivalent to (3). Suppose r is a prime and $< r > \subset I$. If $< r > \neq I$ then there exists an $r_1 \in I$ with $r_1 \notin < r >$. Hence $< r, r_1 > \neq < r >$. Since R is a PID $< r, r_1 > = < r_2 >$ so $r \in < r_2 >$. Then $r_2|r$ and hence r_2 is either a unit or an associate of r. If r_2 is a unit then $< r_2 > = R$ and hence $I = R$. If $< r_2 >$ is not a unit then r_2 is an associate of r and hence

$$< r, r_1 > = < r_2 > = < r >$$

a contradiction since $r_1 \notin < r >$. Hence r_2 is a unit, $I = R$ and $< r >$ is a maximal ideal.

Conversely suppose that $< r >$ is maximal and $r_1r_2 = r$. Suppose first that $r|r_1$. Since $r_1|r$ then r and r_1 are associates. Now if r does not divide r_1 then $r_1 \notin < r >$ so that $< r, r_1 > \neq < r >$. It follows from the maximality of $< r >$ that $< r, r_1 > = R$. Hence $1 \in < r, r_1 >$ and there exists $x, y \in R$ with

$$rx + r_1y = 1.$$

Multiplying through by r_2 we have

$$rr_2x + r_1r_2y = r_2.$$

Then $r|r_2$. Therefore $r_2 = r_3r$ and we have $r = (r_1r_3)r$. Hence $r_1r_3 = 1$ and r_1 is a unit. Hence either r_1 is an associate of r or a unit. In either case r_2 is either an associate of r or a unit. Therefore r is prime. □

In an integral domain R we can use ideals to build **factor rings**. This is a fundamental concept in abstract algebra and will also play a role in algebraic number theory. We define this in general.

Definition 6.2.11 *If R is an integral domain and I is an ideal in R then a* **coset** *of I is a subset of the form*

$$r + I = \{r + i; i \in I\}.$$

The set of cosets of I in R is denoted R/I.

Lemma 6.2.16 *(1) The set of cosets R/I partition R and $r \in I$ if and only if $r + I = 0 + I$.*

Proof On R define $r_1 \sim r_2$ if $r_1 - r_2 \in I$. This is an equivalence relation (see exercises) and therefore the equivalence classes partition R. If $r \in R$ its equivalence class $[r]$ is precisely the coset $r + I$. □

Next we define operations on R/I. If $[r_1] = r_1 + I$ and $[r_2] = r_2 + I$ then

$$[r_1] + [r_2] = (r_1 + r_2) + I = [r_1 + r_2]$$

$$[r_2][r_2] = (r_1r_2) + I = [r_1r_2].$$

Lemma 6.2.17 *The operations defined on R/I are well-defined.*

Proof Well-defined means that if $[r_1] = [r_2]$ and $[r_3] = [r_4]$ then $[r_1] + [r_3] = [r_2] + [r_4]$ and $[r_1][r_3] = [r_2][r_4]$. We show this is true for addition and leave multiplication to the exercises.

Suppose $[r_1] = [r_2]$ then $r_1 \sim r_2 \implies r_1 - r_2 \in I$. Similarly if $[r_3] = [r_4]$ then $r_3 - r_4 \in I$. Then $(r_1 - r_2) + (r_3 - r_4) \in I$ which implies $(r_1 + r_3) - (r_2 + r_4) \in I$. Therefore $[r_1 + r_3] = [r_2 + r_4]$ and addition is well-defined. □

If $r_1 + I = r_2 + I$ we will also write $r_1 \equiv r_2 \bmod I$.

Theorem 6.2.14 *Let R be an integral domain and $I \subset R$ an ideal. Then*

1. *R/I forms a commutative ring with an identity under the operations defined above.*
2. *R/I is an integral domain if and only if I is a prime ideal.*
3. *R/I is a field if and only if I is a maximal ideal.*

The ring R/I is called the **factor ring** *or* **quotient ring** *of R modulo I.*

Proof The proof that R/I is a commutative ring with an identity is a routine exercise. We show (2) and (3). We need that the elements of R/I are the cosets which we will now denote as $[r]$ and that the additive identity is $[0]$ which we will just write as 0 in R/I. Further the multiplicative identity of R/I is $[1]$ which we will write as 1 in R/I.

Suppose I is a prime ideal and suppose $[r_1][r_2] = [0] = 0$ in R/I. Then $r_1r_2 \in I$ and then either $r_1 \in I$ or $r_2 \in I$. If $r_1 \in I$ then $[r_1] = 0$ in R/I and if $r_2 \in I$ then $[r_2] = 0$ in R/I. Therefore there are no zero divisors in R/I and hence its an integral domain.

Conversely suppose R/I is an integral domain and suppose $r_1r_2 \in I$. Then $[r_1][r_2] = 0$ and since R/I is an integral domain either $[r_1] = 0$ or $[r_2] = 0$. In the former case $r_1 \in I$ and in the latter $r_2 \in I$. Therefore I is a prime ideal.

Next suppose that I is maximal. If $[r] \neq 0$ in R/I then $r \notin I$. From the maximality of I it follows that $< I, r > = R$ and then $1 \in < I, r >$. This implies that there exist $x, y \in R$ with

$$rx + iy = 1 \text{ for some } i \in I.$$

But then in R/I we have $[r][x] = [1] = 1$ since $[iy] = [0] = 0$. Hence in the factor ring $[r]$ is a unit. Since $[r]$ was an arbitrary nonzero element of R/I it follows that R/I is a field.

Conversely suppose R/I is a field. If $r \notin I$ then $[r] \neq 0$ in R/I and hence there exists an inverse $[x]$ with $[r][x] = 1$. Hence there is an $i \in I$, $y \in R$ with

$$rx + iy = 1.$$

It follows that $1 \in < I, r >$ which implies that $< I, r > = R$. Therefore I is maximal. □

Now a field F is always an integral domain. Therefore if R/I is a field it follows that R/I is an integral domain. Translating this into statements about the ideal I we have:

Corollary 6.2.5 *In any integral domain a maximal ideal is a prime ideal.*

Note that the converse of this corollary is not necessarily true in general but it is true in a PID for nonzero prime ideals.

Finally we sketch a beautiful application of these ideas called Kronecker's Theorem. Although it was proved by Kronecker well after the work of Galois, from a modern perspective it is really the starting off point for Galois Theory. We will look more carefully at this in the next section.

Theorem 6.2.15 *Let F be a field and $p(x) \in F[x]$ an irreducible polynomial. Then there exists a field F' with $F \subset F'$ in which $p(x)$ has a root.*

Proof Since $p(x)$ is irreducible and $F[x]$ is a PID the ideal $< p(x) >$ is a maximal ideal. Then the factor ring

$$F' = F[x]/ < p(x) >$$

is a field. The elements of F' are cosets $g(x) + < p(x) >$. If we identify $f \in F$ with the coset $f + < p(x) > = [f]$ this gives an embedding of F into F'. Therefore F can be considered as a subfield of F'.

Now consider $[x] = x + < p(x) >$. Then by considering the operations in F' it is clear that $p([x]) = [p(x)]$ (see exercises). But $[p(x)] = p(x) + < p(x) > = < p(x) > = [0]$. Therefore in F' we have $p([x]) = 0$ and $[x]$ is a root of $p(x)$ in F'. □

We will give a well-known example to clarify the theorem. Let $F = \mathbb{R}$ and $p(x) = x^2 + 1$. Then $p(x)$ is irreducible in $\mathbb{R}[x]$. Let $\mathbb{R}' = \mathbb{R}[x]/ < x^2 + 1 >$. Since $x^2 + 1$ is prime the ideal $< x^2 + 1 >$ is a maximal ideal and hence $\mathbb{R}'$ is a field.

Each element of $\mathbb{R}'$ is a polynomial in $\mathbb{R}[x]$ modulo $< x^2 + 1 >$. By the division algorithm if $h(x) \in \mathbb{R}[x]$ with $h(x) \neq 0$ then

$$h(x) = q(x)(x^2 + 1) + h_1(x) \text{ with } \deg(h_1(x)) < \deg(x^2 + 1) = 2.$$

Therefore $h_1(x) = a + bx$ with $a, b \in \mathbb{R}$. However

$$h(x) \equiv h_1(x) \bmod < x^2 + 1 > .$$

It follows that every element of $\mathbb{R}'$ can be expressed as $a + bx$ with $a, b \in \mathbb{R}$. Therefore

$$\mathbb{R}' = \{a + bx; a, b \in \mathbb{R}\}.$$

Further in $\mathbb{R}'$ we have $x^2 + 1 = 0$ and hence $x^2 = -1$. Then

$$\mathbb{R}' = \{a + bx; a, b \in \mathbb{R}, x^2 = -1\}.$$

Mapping $\mathbb{R}'$ onto $\mathbb{C}$ the complex numbers by $1 \to 1, x \to i$ gives an isomorphism. Therefore $\mathbb{R}'$ is precisely $\mathbb{C}$ the complex numbers.

6.3 Algebraic Number Fields

An algebraic number field is a finite field extension of the rational numbers $\mathbb{Q}$ within the complex numbers $\mathbb{C}$. As before we must first look at some essential definitions from abstract algebra.

If F and F' are fields with F a subfield of F', then F' is an **extension field**, or simply an **extension**, of F. If we have a chain of fields and extension fields

$$F \subset E \subset E' \subset F'$$

then F is called the **ground field** and E and E' are **intermediate fields**.

Recall that if F is a field then a **vector space** V over F consists of an abelian group V together with scalar multiplication from F satisfying:

1. $fv \in V$ if $f \in F, v \in V$.
2. $f(u+v) = fu + fv$ for $f \in F, u, v \in V$.
3. $(f+g)v = fv + gv$ for $f, g \in F, v \in V$.
4. $(fg)v = f(gv)$ for $f, g \in F, v \in V$.
5. $1v = v$ for $v \in V$.

A set of elements in a vector space V, $\{v_1, \ldots, v_n\}$ is **independent**, over F if whenever $f_1v_1 + \cdots + f_nv_n = 0$ then each scalar $f_i = 0$. If a set is not independent then it is called **dependent**. For a subset $U \subset V$ the set

$$\{f_1v_1 + \cdots + f_nv_n; n \geq 1, v_i \in U, f_i \in F\}$$

of linear combinations of elements of U forms a subspace of V called the **span** of U or the subspace **spanned** by U. This is denoted by $< U >$. If $U = \{v_1, \ldots, v_n\}$ is finite then we write $< U > = < v_1, \ldots, v_n >$. An independent set which spans the whole vector space V is called a **basis** for V. The number of elements in a basis is unique and is called the **dimension** of V over F denoted $\dim_F V$ or just $\dim V$ if F is understood. If there is a finite basis then V is **finite-dimensional** over F.

If $v_1, \ldots, v_n$ is a basis for V and $w_1, \ldots, w_n$ is another set of vectors in V then

$$w_1 = f_{11}v_1 + \cdots + f_{1n}v_n$$

$$w_2 = f_{21}v_1 + \cdots + f_{2n}v_n$$

$$\ldots\ldots$$

$$w_n = f_{n1}v_1 + \cdots + f_{nn}v_n$$

for some scalars $f_{ij} \in F$. Then $w_1, .., w_n$ is also a basis if and only if the transition matrix

$$\begin{pmatrix} f_{11} & \cdots & f_{1,n} \\ f_{21} & \cdots & f_{2n} \\ \cdots & & \\ f_{n1} & \cdots & f_{nn} \end{pmatrix}$$

has nonzero determinant.

If F' is an extension field of F then multiplication of elements of F' by elements of F are still in F'. Since F' is an abelian group under addition, F' can be considered as a vector space over F. Thus any extension field is a vector space over any of its subfields. The **degree of the extension** is the dimension of F' as a vector space over F. We denote the degree by $|F' : F|$. If the degree is finite, that is, $|F' : F| < \infty$, so that F' is a finite-dimensional vector space over F, then F' is called a **finite extension** of F.

From vector space theory we easily obtain that the degrees are multiplicative. Specifically:

Lemma 6.3.1 *If $F \subset F' \subset F''$ are fields with F'' a finite extension of F, then $|F' : F|$ and $|F'' : F'|$ are also finite, and $|F'' : F| = |F'' : F'||F' : F|$.*

Proof The fact that $|F' : F|$ and $|F'' : F'|$ are also finite follows easily from linear algebra since the dimension of a subspace must be less than the dimension of the whole vector space.

If $|F' : F| = n$ with $\alpha_1, \dots, \alpha_n$ a basis for F' over F, and $|F'' : F'| = m$ with $\beta_1, \dots, \beta_m$ a basis for F'' over F' then the mn products $\{\alpha_i \beta_j\}$ form a basis for F'' over F (see the exercises). Then

$$|F'' : F| = mn = |F'' : F'||F' : F|.$$

□

This last argument also shows that if $F \subset F' \subset F''$ are fields, with $|F' : F|$ and $|F'' : F'|$ finite, then F'' is a finite extension of F.

EXAMPLE 6.3.1 $\mathbb{C}$ is a finite extension of $\mathbb{R}$, but $\mathbb{R}$ is an infinite extension of $\mathbb{Q}$.

The complex numbers $1, i$ form a basis for $\mathbb{C}$ over $\mathbb{R}$. It follows that the degree of $\mathbb{C}$ over $\mathbb{R}$ is 2, that is, $|\mathbb{C} : \mathbb{R}| = 2$.

That $\mathbb{R}$ is infinite dimensional over $\mathbb{Q}$ depends on the existence of **transcendental numbers**. An element $r \in \mathbb{R}$ is **algebraic** (over $\mathbb{Q}$) if it satisfies some nonzero polynomial with coefficients from $\mathbb{Q}$. That is, $P(r) = 0$, where

$$0 \neq P(x) = a_0 + a_1 x + \cdots + a_n x^n \text{ with } a_i \in \mathbb{Q}.$$

An element $r \in \mathbb{R}$ is **transcendental** if it is not algebraic.

In general it is very difficult to show that a particular element is transcendental. However there are uncountably many transcendental elements as we will show in Section 6.3.2. Specific examples are our old friends e and π. We give a proof of their transcendence later in this chapter.

Since e is transcendental, for any natural number n the set of vectors $\{1, e, e^2, \dots, e^n\}$ must be independent over $\mathbb{Q}$, for otherwise there would be a polynomial that e would satisfy. Therefore, we have infinitely many independent vectors in $\mathbb{R}$ over $\mathbb{Q}$ which would be impossible if $\mathbb{R}$ had finite degree over $\mathbb{Q}$.

We are interested in special types of field extensions called **algebraic extensions**. We present the definitions in general and then specialize to extensions of the rationals $\mathbb{Q}$ within $\mathbb{C}$.

Definition 6.3.1 *Suppose F' is an extension field of F and $\alpha \in F'$. Then α is* **algebraic over F** *if there exists a nonzero polynomial $p(x)$ in $F[x]$ with $p(\alpha) = 0$. (α*

is a root of a polynomial with coefficients in F.) If every element of F' is algebraic over F, then F' is an **algebraic extension** *of F.*

If $\alpha \in F'$ is nonalgebraic over F then α is called **transcendental** *over F. A nonalgebraic extension is called a* **transcendental extension**.

Lemma 6.3.2 *Every element of F is algebraic over F.*

Proof If $f \in F$ then $p(x) = x - f \in F[x]$ and $p(f) = 0$. □

The tie-in to finite extensions is via the following theorem.

Theorem 6.3.1 *If F' is a finite extension of F, then F' is an algebraic extension.*

Proof Suppose $\alpha \in F'$. We must show that there exists a nonzero polynomial $0 \neq p(x) \in F[x]$ with $p(\alpha) = 0$.

Since F' is a finite extension, $|F' : F| = n < \infty$. This implies that there are n elements in a basis for F' over F, and hence any set of $(n+1)$ elements in F' must be linearly dependent over F.

Consider then $1, \alpha, \alpha^2, \ldots, \alpha^n$. These are $(n+1)$ elements in F' and therefore must be linearly dependent. Then there must exist elements

$$f_0, f_1, \ldots, f_n \in F,$$

not all zero, such that

$$f_0 + f_1\alpha + \cdots + f_n\alpha^n = 0. \tag{6.3.1}$$

Let $p(x) = f_0 + f_1x + \cdots + f_nx^n$. Then $p(x) \in F[x]$ and $p(\alpha) = 0$ from (6.3.1). Therefore any $\alpha \in F'$ is algebraic over F and hence F' is an algebraic extension of F. □

EXAMPLE 6.3.2 $\mathbb{C}$ is algebraic over $\mathbb{R}$, but $\mathbb{R}$ is transcendental over $\mathbb{Q}$.

Since $|\mathbb{C} : \mathbb{R}| = 2$, $\mathbb{C}$ being algebraic over $\mathbb{R}$ follows from Theorem 6.3.1. More directly, if $z \in \mathbb{C}$ then $p(x) = (x - z)(x - \overline{z}) \in \mathbb{R}[x]$ and $p(z) = 0$.

$\mathbb{R}$ (and thus $\mathbb{C}$) being transcendental over $\mathbb{Q}$ follows from the existence of transcendental numbers such as e and π.

If α is algebraic over F, it satisfies a polynomial over F. It follows that it must then also satisfy an irreducible polynomial over F. Since F is a field, if $f \in F$ with $f \neq 0$ and $p(x) \in F[x]$, then $f^{-1}p(x) \in F[x]$ also. This implies that if $p(\alpha) = 0$ with $a_n \neq 0$ the leading coefficient of $p(x)$, then $p_1(x) = a_n^{-1}p(x)$ is a monic polynomial in $F[x]$ that α also satisfies. Thus if α is algebraic over F there is a monic irreducible polynomial that α satisfies. The next result says that this polynomial is unique.

Lemma 6.3.3 *If $\alpha \in F'$ is algebraic over F, then there exists a unique monic irreducible polynomial $p(x) \in F[x]$ such that $p(\alpha) = 0$.*

This unique monic irreducible polynomial is denoted by $irr(\alpha, F)$.

Proof Suppose $f(\alpha) = 0$ with $0 \neq f(x) \in F[x]$. Then $f(x)$ factors into irreducible polynomials. Since there are no zero divisors in a field, one of these factors, say $p_1(x)$ must also have α as a root. If the leading coefficient of $p_1(x)$ is a_n then $p(x) = a_n^{-1} p_1(x)$ is a monic irreducible polynomial in $F[x]$ that also has α as a root.

Therefore, there exist monic irreducible polynomials that have α as a root. Let $p(x)$ be one such polynomial of minimal degree. It remains to show that $p(x)$ is unique.

Suppose $g(x)$ is another monic irreducible polynomial with $g(\alpha) = 0$. Since $p(x)$ has minimal degree, deg $p(x) \leq$ deg $g(x)$. By the division algorithm

$$g(x) = q(x)p(x) + r(x), \tag{6.3.2}$$

where $r(x) = 0$ or deg $r(x) <$ deg $p(x)$. Substituting α into (6.3.2) we get

$$g(\alpha) = q(\alpha)p(\alpha) + r(\alpha),$$

which implies that $r(\alpha) = 0$ since $g(\alpha) = p(\alpha) = 0$. But then if $r(x)$ is not identically 0, α is a root of $r(x)$, which contradicts the minimality of the degree of $p(x)$. Therefore, $r(x) = 0$ and $g(x) = q(x)p(x)$. The polynomial $q(x)$ must be a constant (unit factor) since $g(x)$ is irreducible, but then $q(x) = 1$ since both $g(x)$, $p(x)$ are monic. This says that $g(x) = p(x)$, and hence $p(x)$ is unique. □

We say that an algebraic element has **degree** n if the degree of $irr(\alpha, F)$ is n.

Embedded in the proof of Lemma 6.3.3 is the following important corollary.

Corollary 6.3.1 *If α is algebraic over F and $f(\alpha) = 0$ for $f(x) \in F[x]$ then $irr(\alpha, F) | f(x)$. That is $irr(\alpha, F)$ divides any polynomial over F which has α as a root.*

Suppose $\alpha \in F'$ is algebraic over F and $p(x) = irr(\alpha, F)$. Then there exists a smallest intermediate field E with $F \subset E \subset F'$ such that $\alpha \in E$. By smallest we mean that if E' is another intermediate field with $\alpha \in E'$ then $E \subset E'$. To see that this smallest field exists, notice that there are subfields E' in F' in which $\alpha \in E'$ (namely F' itself). Let E be the intersection of all subfields of F' containing α and F. E is a subfield of F' (see the exercises) and E contains both α and F. Further, this intersection is contained in any other subfield containing α and F.

This smallest subfield has a very special form.

Definition 6.3.2 *Suppose $\alpha \in F'$ is algebraic over F and*

$$p(x) = irr(\alpha, F) = a_0 + a_1 x + \cdots + a_{n-1}x^{n-1} + x^n.$$

Let

$$F(\alpha) = \{f_0 + f_1\alpha + \cdots + f_{n-1}\alpha^{n-1};\ f_i \in F\}.$$

On $F(\alpha)$ define addition and subtraction componentwise and define multiplication by algebraic manipulation, replacing powers of α higher than α^{n-1} by using

$$\alpha^n = -a_0 - a_1\alpha - \cdots - a_{n-1}\alpha^{n-1}.$$

Theorem 6.3.2 *$F(\alpha)$ forms a finite algebraic extension of F with*

$$|F(\alpha) : F| = \ deg(irr(\alpha, F)).$$

The field, $F(\alpha)$, is the smallest subfield of F' that contains the root α. A field extension of the form $F(\alpha)$ for some α is called a **simple extension** *of F.*

Proof Recall that $F_{n-1}[x]$ is the set of all polynomials over F of degree $\leq n-1$ together with the zero polynomial. This set forms a vector space of dimension n over F. As defined in Definition 6.3.2, relative to addition and subtraction $F(\alpha)$ is the same as $F_{n-1}[x]$, and thus $F(\alpha)$ is a vector space of dimension deg $irr(\alpha, F)$ over F and hence an abelian group.

Multiplication is done via multiplication of polynomials, so it is straightforward then that $F(\alpha)$ forms a commutative ring with an identity. We must show that it forms a field. To do this we must show that every nonzero element of $F(\alpha)$ has a multiplicative inverse.

Suppose $0 \neq g(x) \in F[x]$. If deg $g(x) \ < n =$ deg $irr(\alpha, F)$, then $g(\alpha) \neq 0$ since $irr(\alpha, F)$ is the irreducible polynomial of minimal degree that has α as a root.

If $h(x) \in F[x]$ with deg $h(x) \geq n$, then $h(\alpha) = h_1(\alpha)$, where $h_1(x)$ is a polynomial of degree $\leq n-1$, obtained by replacing powers of α higher than α^{n-1} by combinations of lower powers using

$$\alpha^n = -a_0 - a_1\alpha - \cdots - a_{n-1}\alpha^{n-1}.$$

Now suppose $g(\alpha) \in F(\alpha)$, $g(\alpha) \neq 0$. Consider the corresponding polynomial $g(x) \in F[x]$ of degree $\leq n-1$. Since $p(x) = irr(\alpha, F)$ is irreducible, it follows that $g(x)$ and $p(x)$ must be relatively prime, that is, $(g(x), p(x)) = 1$. Therefore, there exist $h(x), k(x) \in F[x]$ such that

$$g(x)h(x) + p(x)k(x) = 1.$$

Substituting α into the above we obtain:

$$g(\alpha)h(\alpha) + p(\alpha)k(\alpha) = 1.$$

However, $p(\alpha) = 0$ and $h(\alpha) = h_1(\alpha) \in F(\alpha)$, so that

$$g(\alpha)h_1(\alpha) = 1.$$

It follows then that in $F(\alpha)$, $h_1(\alpha)$ is the multiplicative inverse of $g(\alpha)$. Since every nonzero element of $F(\alpha)$ has such an inverse $F(\alpha)$ forms a field.

F is contained in $F(\alpha)$ by identifying F with the constant polynomials. Therefore, $F(\alpha)$ is an extension field of F. From the definition of $F(\alpha)$, we have that $\{1, \alpha, \alpha^2, \dots, \alpha^{n-1}\}$ form a basis, so $F(\alpha)$ has degree n over F. Therefore, $F(\alpha)$ is a finite extension and hence an algebraic extension.

If $F \subset E \subset F'$ and E contains α, then clearly E contains all powers of α since E is a subfield. E then contains $F(\alpha)$, and hence $F(\alpha)$ is the smallest subfield containing both F and α. □

EXAMPLE 6.3.3 Consider $p(x) = x^3 - 2$ over $\mathbb{Q}$. This is irreducible over $\mathbb{Q}$ but has the root $\alpha = 2^{1/3} \in \mathbb{R}$. The field $\mathbb{Q}(\alpha) = \mathbb{Q}(2^{1/3})$ is then the smallest subfield of $\mathbb{R}$ that contains $\mathbb{Q}$ and $2^{1/3}$.

Here

$$\mathbb{Q}(\alpha) = \{q_0 + q_1\alpha + q_2\alpha^2;\ q_i \in \mathbb{Q} \text{ and } \alpha^3 = 2\}.$$

We first give examples of addition and multiplication in $\mathbb{Q}(\alpha)$.

Let $g = 3 + 4\alpha + 5\alpha^2$, $h = 2 - \alpha + \alpha^2$. Then

$$g + h = 5 + 3\alpha + 6\alpha^2$$

and

$$\begin{aligned} gh &= 6 - 3\alpha + 3\alpha^2 + 8\alpha - 4\alpha^2 + 4\alpha^3 + 10\alpha^2 - 5\alpha^3 + 5\alpha^4 \\ &= 6 + 5\alpha + 9\alpha^2 - \alpha^3 + 5\alpha^4. \end{aligned}$$

But $\alpha^3 = 2$, so $\alpha^4 = 2\alpha$, and then

$$gh = 6 + 5\alpha + 9\alpha^2 - 2 + 5(2\alpha) = 4 + 15\alpha + 9\alpha^2.$$

We now show how to find the inverse of h in $\mathbb{Q}(\alpha)$.

Let $h(x) = 2 - x + x^2$, $p(x) = x^3 - 2$. Use the Euclidean algorithm as in Chapter 3 to express 1 as a linear combination of $h(x)$, $p(x)$.

$$x^3 - 2 = (x^2 - x + 2)(x + 1) + (-x - 4),$$

$$x^2 - x + 2 = (-x - 4)(-x + 5) + 22.$$

This implies that

$$22 = (x^2 - x + 2)(1 + (x + 1)(-x + 5)) - ((x^3 - 2)(-x + 5))$$

or

$$1 = \frac{1}{22}[(x^2 - x + 2)(-x^2 + 4x + 6)] - [(x^3 - 2)(-x + 5)].$$

Now substituting α and using that $\alpha^3 = 2$, we have

$$1 = \frac{1}{22}[(\alpha^2 - \alpha + 2)(-\alpha^2 + 4\alpha + 6)],$$

and hence

$$h^{-1} = \frac{1}{22}(-\alpha^2 + 4\alpha + 6).$$

Now suppose $\alpha, \beta \in F'$ with both elements algebraic over F and suppose $irr(\alpha, F) = irr(\beta, F)$. From the construction of $F(\alpha)$ we can see that it would be essentially the same as $F(\beta)$. We now make this idea precise.

Definition 6.3.3 *Let F', F'' be extension fields of F. An* **F-isomorphism** *is an isomorphism $\sigma : F' \to F''$ such that $\sigma(f) = f$ for all $f \in F$. That is, an F-isomorphism is an isomorphism of the extension fields that* **fixes each element of the ground field.** *If F', F'' are F-isomorphic, we denote this relationship by $F' \underset{F}{\cong} F''$.*

Lemma 6.3.4 *Suppose $\alpha, \beta \in F'$ are both algebraic over F and suppose $irr(\alpha, F) = irr(\beta, F)$. Then $F(\alpha)$ is F-isomorphic to $F(\beta)$.*

Proof Define the map $\sigma : F(\alpha) \to F(\beta)$ by $\sigma(\alpha) = \beta$ and $\sigma(f) = f$ for all $f \in F$. Allow σ to be a homomorphism, that is, preserve addition and multiplication. It follows then that σ maps

$$f_0 + f_1\alpha + \cdots + f_n\alpha^{n-1} \in F(\alpha) \text{ to } f_0 + f_1\beta + \cdots + f_n\beta^{n-1} \in F(\beta).$$

From this it is straightforward that σ is an F-isomorphism. □

Further we note that if $\alpha, \beta \in F'$ with both algebraic over F and suppose that $F(\alpha)$ is F-isomorphic to $F(\beta)$. Then there is a $\gamma \in F(\beta)$ with $irr(\alpha, F) = irr(\gamma, F)$. We can take for γ the image of α under the F-isomorphism.

If $\alpha, \beta \in F'$ are two algebraic elements over F, we use $F(\alpha, \beta)$ to denote $(F(\alpha))(\beta)$. $F(\alpha, \beta)$ and $F(\beta, \alpha)$ are F-isomorphic so we treat them as the same. We now show that the set of algebraic elements over a ground field is closed under the arithmetic operations and from this obtained that the algebraic elements form a subfield.

Lemma 6.3.5 *If $\alpha, \beta \in F'$, $\beta \neq 0$ are two algebraic elements over F, then $\alpha \pm \beta$, $\alpha\beta$, and α/β are also algebraic over F.*

Proof Since α, β are algebraic, the subfield $F(\alpha, \beta)$ will be of finite degree over F and therefore algebraic over F. Now, $\alpha, \beta \in F(\alpha, \beta)$ and since $F(\alpha, \beta)$ is a subfield, it follows that $\alpha \pm \beta$, $\alpha\beta$ and α/β are also elements of $F(\alpha, \beta)$. Since $F(\alpha, \beta)$ is an algebraic extension of F, each of these elements is algebraic over F. □

Theorem 6.3.3 *If F' is an extension field of F, then the set of elements of F' that are algebraic over F forms a subfield. This subfield is called the* **algebraic closure of F in F'**.

Proof Let $A_F(F')$ be the set of algebraic elements over F in F'. $A_F(F') \neq \emptyset$ since it contains F. From the previous lemma it is closed under addition, subtraction, multiplication, and division, and therefore it forms a subfield. □

We close this subsection with a final result, that says that every finite extension is formed by taking successive simple extensions.

Theorem 6.3.4 *If F' is a finite extension of F, then there exists a finite set of algebraic elements $\alpha_1, \dots, \alpha_n$ such that $F' = F(\alpha_1, \dots, \alpha_n)$.*

Proof Suppose $|F' : F| = k < \infty$. Then F' is algebraic over F. Choose an $\alpha_1 \in F'$, $\alpha_1 \notin F$. Then $F \subset F(\alpha_1) \subset F'$ and $|F' : F(\alpha_1)| < k$. If the degree of this extension is 1, then $F' = F(\alpha_1)$, and we are done. If not, choose an $\alpha_2 \in F'$, $\alpha_2 \notin F(\alpha_1)$. Then as above

$$F \subset F(\alpha_1) \subset F(\alpha_1, \alpha_2) \subset F' \text{ with } |F' : F(\alpha_1, \alpha_2)| < |F' : F(\alpha_1)|.$$

As before, if this degree is one we are done; if not, continue. Since k is finite this process must terminate in a finite number of steps. □

6.3.1 Algebraic Extensions of $\mathbb{Q}$

We now specialize to the case where the ground field is the rationals $\mathbb{Q}$. An **algebraic number field** is a finite and hence algebraic extension field of $\mathbb{Q}$ within $\mathbb{C}$. Hence an algebraic number field is a field K such that

$$\mathbb{Q} \subset K \subset \mathbb{C}$$

with $|K : \mathbb{Q}| < \infty$. We will prove shortly that K is actually a simple extension of $\mathbb{Q}$.

Definition 6.3.4 *An* **algebraic number** *α is an element of $\mathbb{C}$ which is algebraic over $\mathbb{Q}$. Hence an algebraic number is an $\alpha \in \mathbb{C}$ such that $f(\alpha) = 0$ for some $f(x) \in \mathbb{Q}[x]$. If $\alpha \in \mathbb{C}$ is not algebraic it is* **transcendental**.

We will let $\mathcal{A}$ denote the totality of algebraic numbers within the complex numbers $\mathbb{C}$, and $\mathcal{T}$ the set of transcendentals so that $\mathbb{C} = \mathcal{A} \cup \mathcal{T}$. In the language of the last subsection, $\mathcal{A}$ is the algebraic closure of $\mathbb{Q}$ within $\mathbb{C}$. As in the general case, if $\alpha \in \mathbb{C}$ is algebraic we will let $irr(\alpha, \mathbb{Q})$ denote the unique monic irreducible polynomial of minimal degree that α satisfies over $\mathbb{Q}$. Then $irr(\alpha, \mathbb{Q})$ divides any rational polynomial $p(x)$ which satisfies $p(\alpha) = 0$.

If $\alpha \notin \mathbb{Q}$ then $\mathbb{Q}(\alpha)$ is the smallest subfield containing both $\mathbb{Q}$ and α. Since $|Q(\alpha) : Q| = \deg(irr(\alpha, Q))$ it follows that $K = \mathbb{Q}(\alpha)$ is an algebraic number field. It then follows trivially that an algebraic number is any element of $\mathbb{C}$ which falls in an algebraic number field and $\mathcal{A}$ is the union of all algebraic number fields.

We next need the following.

Lemma 6.3.6 *If $p(x) \in \mathbb{Q}[x]$ is irreducible of degree n then $p(x)$ has n* **pairwise distinct** *roots in $\mathbb{C}$.*

Proof That $p(x)$ has n roots is a consequence of the Fundamental Theorem of Algebra. What is important here is that if $p(x)$ is irreducible over $\mathbb{Q}$ then its roots in $\mathbb{C}$ are distinct.

Let c be a root of $p(x)$. Then c is an algebraic number and then $irr(c, \mathbb{Q}) | p(x)$. Since $p(x)$ is irreducible it follows that $p(x)$ is just a constant multiple of $irr(c, \mathbb{Q})$ and hence they have the same degree which is minimal among the degrees of all rational polynomials which have c as a root.

Suppose that c is a double root. Then

$$p(x) = (x - c)^2 h(x) \text{ where } h(x) \in \mathbb{C}[x].$$

Now the formal derivative of a rational polynomial is also a rational polynomial. Therefore $p'(x) \in \mathbb{Q}[x]$. However from above using the product rule

$$p'(x) = 2(x - c)h(x) + (x - c)^2 h'(x).$$

Therefore $p'(c) = 0$. This is a contradiction since $\deg(p'(x)) < \deg(p(x))$. Therefore a root cannot be a double root and hence all the n roots are pairwise distinct. □

It follows that if α is an algebraic number of degree n then its minimal polynomial $irr(\alpha, \mathbb{Q})$ has n distinct roots in $\mathbb{C}$.

Definition 6.3.5 *If α is an algebraic number, then its* **conjugates** *over $\mathbb{Q}$, consist of the set, $\{\alpha_1 = \alpha, \dots, \alpha_n\}$, of distinct roots of $irr(\alpha, \mathbb{Q})$ in $\mathbb{C}$.*

Since distinct monic irreducible polynomials cannot have a root in common it follows that if α_i is conjugate to α then $irr(\alpha_i, \mathbb{Q}) = irr(\alpha, \mathbb{Q})$ (see exercises). It follows that $\mathbb{Q}(\alpha_i)$ is $\mathbb{Q}$-isomorphic (see last section) to $\mathbb{Q}(\alpha)$ with the $\mathbb{Q}$-isomorphism being given by $\sigma_i : 1 \to 1, \alpha \to \alpha_i$.

We now get that any algebraic number field is actually a simple extension of $\mathbb{Q}$.

Theorem 6.3.5 *Any algebraic number field K is a simple extension of $\mathbb{Q}$, that is, $K = \mathbb{Q}(\alpha)$ for some algebraic number α. α is called a* **primitive element**.

Proof Since K is a finite extension, $K = \mathbb{Q}(\alpha_1, \ldots, \alpha_n)$ for some algebraic numbers $\alpha_1, \ldots, \alpha_n$. Then to show that K is a simple extension it is sufficient to show that $\mathbb{Q}(\alpha, \beta) = \mathbb{Q}(\gamma)$ for algebraic numbers α, β.

Let $\alpha_1 = \alpha, \ldots, \alpha_n$ be the conjugates of α over $\mathbb{Q}$ and let $\beta_1 = \beta, \ldots, \beta_m$ be the conjugates of β over $\mathbb{Q}$. If $j \neq 1$ then $\beta_i \neq \beta$ since the conjugates are distinct. It follows that for each $i = 1, \ldots, n$ and each $j \neq 1$, $j = 2, \ldots, m$ the equation

$$\alpha_i + \beta_j x = \alpha + \beta x$$

has exactly one complex solution and hence at most one rational solution. Since there are only finitely many such equations there are only finitely many rational solutions x and therefore there exists a rational number q with $q \neq 0$ and q differing from all the solutions. That is

$$\alpha_i + \beta_j q \neq \alpha + \beta q$$

for all i and all $j \neq 1$.

Let $\gamma = \alpha + q\beta$. We claim that $\mathbb{Q}(\alpha, \beta) = \mathbb{Q}(\gamma)$. Since $\mathbb{Q}(\alpha, \beta)$ contains all of $\mathbb{Q}$ as well as α and β it is clear that $\gamma \in \mathbb{Q}(\alpha, \beta)$ and hence $\mathbb{Q}(\gamma) \subset \mathbb{Q}(\alpha, \beta)$. We show that $\mathbb{Q}(\alpha, \beta) \subset \mathbb{Q}(\gamma)$. Here it suffices to show that each of $\alpha, \beta \in \mathbb{Q}(\gamma)$.

Let $f(x) = irr(\alpha, Q)$ and $g(x) = irr(\beta, \mathbb{Q})$. Then $f(\gamma - q\beta) = f(\alpha) = 0$. Therefore β is a root of the polynomials $g(x)$ and $h(x) = f(\gamma - qx)$. If $h(\beta_i) = f(\gamma - q\beta_i) = 0$ for some conjugate $\beta_i \neq \beta$ then $\gamma - \beta_i q = \alpha_j$ for some α_j contradicting the choice of q. Therefore $g(x)$ and $h(x)$ have only β as a common root.

Now $g(x)$ and $h(x) = f(\gamma - qx)$ are polynomials in $K[x]$, where $K = \mathbb{Q}(\gamma)$. Since $Q(\alpha, \beta)$ has finite degree over $\mathbb{Q}$ then $\mathbb{Q}(\beta)$ has finite degree over $\mathbb{Q}(\alpha)$ and β is algebraic over K. Let $h_1(x) = irr(\beta, K)$. Since $g(\beta) = 0$ and $h(\beta) = 0$ it follows that $h_1(x) | g(x)$ and $h_1(x) | h(x)$ in $K[x]$. Since then every root of $h_1(x)$ is then a root of both $g(x)$ and $h(x)$ and β is the only common root of $g(x)$ and $h(x)$ it follows that $h_1(x)$ must have degree one. Therefore

$$h_1(x) = ax + b \text{ for some } a, b \in K.$$

But $h_1(\beta) = 0$ so $\beta = \frac{-b}{a} \in K$. Therefore $\beta \in K = \mathbb{Q}(\gamma)$. An analogous argument shows that $\alpha \in K$. Hence $\mathbb{Q}(\alpha, \beta) \subset \mathbb{Q}(\gamma)$ and so

$$\mathbb{Q}(\alpha, \beta) = \mathbb{Q}(\gamma).$$

□

Let K be an algebraic number field and α a primitive element so that $K = \mathbb{Q}(\alpha)$. It follows that K must have at least one basis (as a vector space over $\mathbb{Q}$) of the form

$$1, \alpha, \alpha^2, \ldots, \alpha^{n-1},$$

where $n = |K : \mathbb{Q}|$. We will use this observation in Section 6.3.4 to define an invariant of a number field called its discriminant.

6.3.2 Algebraic and Transcendental Numbers

In this section we examine the sets $\mathcal{A}$ and $\mathcal{T}$ more closely. Since $\mathcal{A}$ is precisely the algebraic closure of $\mathbb{Q}$ in $\mathbb{C}$ we have from our general result that $\mathcal{A}$ actually forms a subfield of $\mathbb{C}$. Further since the intersection of subfields is again a subfield it follows that $\mathcal{A}' = \mathcal{A} \cap \mathbb{R}$ the real algebraic numbers form a subfield of the reals.

Theorem 6.3.6 *The set $\mathcal{A}$ of algebraic numbers forms a subfield of $\mathbb{C}$. The subset $\mathcal{A}' = \mathcal{A} \cap \mathbb{R}$ of real algebraic numbers forms a subfield of $\mathbb{R}$.*

Since each rational is algebraic it is clear that there are algebraic numbers. Further there are irrational algebraic numbers, $\sqrt{2}$ for example, since it satisfies the irreducible polynomial $x^2 - 2 = 0$ over $\mathbb{Q}$. On the other hand we have not examined the question of whether transcendental numbers really exist. To show that any particular complex number is transcendental is in general quite difficult. However it is relatively easy to show that there are uncountably infinitely many transcendentals.

Theorem 6.3.7 *The set $\mathcal{A}$ of algebraic numbers is countably infinite. Therefore, $\mathcal{T}$, the set of transcendental numbers, and $\mathcal{T}' = \mathcal{T} \cap \mathbb{R}$, the real transcendental numbers are uncountably infinite.*

Proof Let

$$\mathcal{P}_n = \{f(x) \in \mathbb{Q}[x]; \deg(f(x)) \le n\}.$$

Since if $f(x) \in \mathcal{P}_n$, $f(x) = q_o + q_1 x + \cdots + q_n x^n$ with $q_i \in \mathbb{Q}$ we can identify a polynomial of degree $\le n$ with an $(n+1)$-tuple $(q_0, q_1, \ldots, q_n)$ of rational numbers. Therefore the set $\mathcal{P}_n$ has the same size as the $(n+1)$-fold Cartesian product of $\mathbb{Q}$:

$$\mathbb{Q}^{n+1} = \mathbb{Q} \times \mathbb{Q} \times \cdots \times \mathbb{Q}.$$

Since a finite Cartesian product of countable sets is still countable it follows that $\mathcal{P}_n$ is a countable set.

Now let

$$\mathcal{B}_n = \bigcup_{p(x) \in \mathcal{P}_n} \{\text{roots of } p(x)\},$$

that is $\mathcal{B}_n$ is the union of all roots in $\mathbb{C}$ of all rational polynomials of degree $\le n$. Since each such $p(x)$ has a maximum of n roots and since $\mathcal{P}_n$ is countable it follows that $\mathcal{B}_n$ is a countable union of finite sets and hence is still countable. Now

$$\mathcal{A} = \bigcup_{n=1}^{\infty} \mathcal{B}_n$$

so $\mathcal{A}$ is a countable union of countable sets and is therefore countable.

Since both $\mathbb{R}$ and $\mathbb{C}$ are uncountably infinite, the second assertions follow directly from the countability of $\mathcal{A}$. If say $\mathcal{T}$ were countable then $\mathbb{C} = \mathcal{A} \cup \mathcal{T}$ would also be countable which is a contradiction. □

Therefore we now know that there exist infinitely many transcendental numbers. Liouville in 1851 gave the first proof of the existence of transcendentals by exhibiting a few. He gave as one the following example.

Theorem 6.3.8 *The real number*

$$c = \sum_{j=1}^{\infty} \frac{1}{10^{j!}}$$

is transcendental.

Proof First of all since $\frac{1}{10^{j!}} < \frac{1}{10^j}$ and $\sum_{j=1}^{\infty} \frac{1}{10^j}$ is a convergent geometric series it follows from the comparison test that the infinite series defining c converges and defines a real number. Further since $\sum_{j=1}^{\infty} \frac{1}{10^j} = \frac{1}{9}$. It follows that $c < \frac{1}{9} < 1$.

Suppose that c is algebraic so that $g(c) = 0$ for some rational nonzero polynomial $g(x)$. Multiplying through by the least common multiple of all the denominators in $g(x)$ we may suppose that $f(c) = 0$ for some integral polynomial $f(x) = \sum_{j=0}^{n} m_j x^j$. Then c satisfies

$$\sum_{j=0}^{n} m_j c^j = 0$$

for some integers $m_0, \ldots, m_n$.

If $0 < x < 1$ then by the triangle inequality

$$|f'(x)| = |\sum_{j=1}^{n} j m_j x^{j-1}| \leq \sum_{j=1}^{n} |j m_j| = B,$$

where B is a real constant depending only on the coefficients of $f(x)$.

Now let

$$c_k = \sum_{j=1}^{k} \frac{1}{10^{j!}}$$

be the kth partial sum for c. Then

$$|c - c_k| = \sum_{j=k+1}^{\infty} \frac{1}{10^{j!}} < 2 \cdot \frac{1}{10^{(k+1)!}}.$$

Apply the Mean Value Theorem to $f(x)$ at c and c_k to obtain

$$|f(c) - f(c_k)| = |c - c_k||f'(\zeta)|$$

for some ζ with $c_k < \zeta < c < 1$. Now since $0 < \zeta < 1$ we have

$$|c - c_k||f'(\zeta)| < 2B\frac{1}{10^{(k+1)!}}.$$

On the other hand, since $f(x)$ can have at most n roots, it follows that for all k large enough we would have $f(c_k) \neq 0$. Since $f(c) = 0$ we have

$$|f(c) - f(c_k)| = |f(c_k)| = |\sum_{j=1}^{n} m_j c_k^j| > \frac{1}{10^{nk!}}$$

since for each j, $m_j c_k^j$ is a rational number with denominator $10^{jk!}$. However if k is chosen sufficiently large and n is fixed we have

$$\frac{1}{10^{nk!}} > \frac{2B}{10^{(k+1)!}}$$

contradicting the equality from the Mean Value Theorem. Therefore c is transcendental. □

After we discuss algebraic integers we will show that both e and π are transcendental. The transcendence of e was proved first by Hermite in 1873 while Lindemann in 1881 proved the transcendence of π.

6.3.3 Symmetric Polynomials

Many results on algebraic number fields and algebraic integers depend on the properties of **symmetric polynomials**. These were briefly introduced and used in Section 5.2.1. Here we look at them more carefully and present a fundamental result concerning them.

Definition 6.3.6 *Let $y_1, \ldots, y_n$ be (independent) indeterminates over a field F. A polynomial $f(y_1, \ldots, y_n) \in F[y_1, \ldots, y_n]$ is a* **symmetric polynomial** *in $y_1, \ldots, y_n$ if $f(y_1, \ldots, y_n)$ is unchanged by any permutation σ of $\{y_1, \ldots, y_n\}$, that is, $f(y_1, \ldots, y_n) = f(\sigma(y_1), \ldots, \sigma(y_n))$.*

If $F \subset F'$ are fields and $\alpha_1, \ldots, \alpha_n$ are in F', then we call a polynomial $f(\alpha_1, \ldots, \alpha_n)$ with coefficients in F **symmetric** *in $\alpha_1, \ldots, \alpha_n$ if $f(\alpha_1, \ldots, \alpha_n)$ is unchanged by any permutation σ of $\{\alpha_1, \ldots, \alpha_n\}$.*

EXAMPLE 6.3.3.1 Let F be a field and $f_0, f_1 \in F$. Let $h(y_1, y_2) = f_0(y_1 + y_2) + f_1(y_1 y_2)$.

There are two permutations on $\{y_1, y_2\}$, namely $\sigma_1 : y_1 \to y_1, y_2 \to y_2$ and $\sigma_2 : y_1 \to y_2, y_2 \to y_1$.

Applying either one of these two to $\{y_1, y_2\}$ leaves $h(y_1, y_2)$ invariant. Therefore, $h(y_1, y_2)$ is a symmetric polynomial.

Definition 6.3.7 *Let $x, y_1, \dots, y_n$ be indeterminates over a field F(or elements of an extension field F' over F). Form the polynomial*

$$p(x, y_1, \dots, y_n) = (x - y_1) \cdots (x - y_n).$$

The **ith elementary symmetric polynomial** *s_i in $y_1, \dots, y_n$ for $i = 1, \dots, n$, is $(-1)^i a_i$, where a_i is the coefficient of x^{n-i} in $p(x, y_1, \dots, y_n)$ as a polynomial in x with coefficients from $F(y_1, \dots, y_n)$.*

EXAMPLE 6.3.3.2 Consider y_1, y_2, y_3. Then

$$p(x, y_1, y_2, y_3) = (x - y_1)(x - y_2)(x - y_3)$$

$$= x^3 - (y_1 + y_2 + y_3)x^2 + (y_1 y_2 + y_1 y_3 + y_2 y_3)x - y_1 y_2 y_3.$$

Therefore, the three elementary symmetric polynomials in y_1, y_2, y_3 over any field are

1. $s_1 = y_1 + y_2 + y_3$.
2. $s_2 = y_1 y_2 + y_1 y_3 + y_2 y_3$.
3. $s_3 = y_1 y_2 y_3$.

In general, the pattern of the last example holds for $y_1, \dots, y_n$. That is,

$$s_1 = y_1 + y_2 + \cdots + y_n$$

$$s_2 = y_1 y_2 + y_1 y_3 + \cdots + y_{n-1} y_n$$

$$s_3 = y_1 y_2 y_3 + y_1 y_2 y_4 + \cdots + y_{n-2} y_{n-1} y_n$$

$$\vdots$$

$$s_n = y_1 \cdots y_n.$$

The importance of the elementary symmetric polynomials is that any symmetric polynomial can be built up from the elementary symmetric polynomials. We make this precise in the next theorem called the **fundamental theorem of symmetric polynomials**. We will use this important result several times in our study of algebraic numbers and algebraic integers.

Theorem 6.3.9 *(Fundamental Theorem of Symmetric Polynomials) If P is a symmetric polynomial in the indeterminates $y_1, .., y_n$ over a field F, that is, $P \in F[y_1, .., y_n]$ and P is symmetric, then there exists a unique $g \in F[y_1, .., y_n]$ such that $P(y_1, \dots, y_n) = g(s_1, .., s_n)$. That is, any symmetric polynomial in $y_1, \dots, y_n$ is a polynomial expression in the elementary symmetric polynomials in $y_1, .., y_n$.*

In order to prove this result we need the concept of a **piece**. Any polynomial $f(x_1, \dots, x_n) \in F[x_1, .., x_n]$ is composed of a sum of **pieces** of the form $ax_1^{i_1} \cdots x_n^{i_n}$ with $a \in F$. We first put an order on these pieces of a polynomial.

The piece $ax_1^{i_1} \cdots x_n^{i_n}$ with $a \neq 0$ is called **higher** than the piece $bx_1^{j_1} \cdots x_n^{j_n}$ with $b \neq 0$ if the first one of the differences

$$i_1 - j_1, i_2 - j_2, \dots, i_n - j_n$$

that differs from zero is in fact positive. The highest piece of a polynomial $f(x_1, \dots, x_n)$ is denoted by $HG(f)$.

Lemma 6.3.7 *For $f(x_1, \dots, x_n), g(x_1, .., x_n) \in F[x_1, \dots, x_n]$ we have*

$$HG(fg) = HG(f)HG(g).$$

Proof We use an induction on n, the number of indeterminates. It is clearly true for $n = 1$, and now assume that the statement holds for all polynomials in k variables with $k < n$ and $n \geq 2$. Order the polynomials via exponents on the first variable x_1 so that

$$f(x_1, \dots, x_n) = x_1^r \phi_r(x_2, \dots, x_n) + x_1^{r-1}\phi_{r-1}(x_2, \dots, x_n) + \cdots + \phi_0(x_2, \dots, x_n),$$

$$g(x_1, \dots, x_n) = x_1^s \psi_s(x_2, \dots, x_n) + x_1^{s-1}\psi_{s-1}(x_2, \dots, x_n) + \cdots + \psi_0(x_2, \dots, x_n).$$

Then

$$HG(fg) = x_1^{r+s} HG(\phi_r \psi_s).$$

By the inductive hypothesis

$$HG(\phi_r \psi_s) = HG(\phi_r)HG(\psi_s).$$

Hence

$$\begin{aligned} HG(fg) &= x_1^{r+s} HG(\phi_r)HG(\psi_s) \\ &= (x_1^r HG(\phi_r))(x_1^s HG(\psi_s)) \\ &= HG(f)HG(g). \end{aligned}$$

□

In general the kth elementary symmetric polynomial is given by

$$s_k = \sum_{i_1<i_2<\cdots<i_k} x_{i_1}x_{i_2}\cdots x_{i_k},$$

where the sum is taken over all the $\binom{n}{k}$ different systems of indices $i_1, \ldots, i_k$ with $i_1 < i_2 < \cdots < i_k$. We need the following concerning the pieces of s_k.

Lemma 6.3.8 *In the highest piece $ax_1^{k_1}\cdots x_n^{k_n}$, $a \neq 0$, of a symmetric polynomial $s(x_1, \ldots, x_n)$ we have $k_1 \geq k_2 \geq \cdots \geq k_n$.*

Proof Assume that $k_i < k_j$ for some $i < j$. As a symmetric polynomial, $s(x_1, \ldots, x_n)$ also must then contain the piece $ax_1^{k_1}\cdots x_i^{k_j}\cdots x_j^{k_i}\cdots x_n^{k_n}$, which is higher than $ax_1^{k_1}\cdots x_i^{k_i}\cdots x_j^{k_j}\cdots x_n^{k_n}$, giving a contradiction. □

Lemma 6.3.9 *The product $s_1^{k_1-k_2}s_2^{k_2-k_3}\cdots s_{n-1}^{k_{n-1}-k_n}s_n^{k_n}$ with $k_1 \geq k_2 \geq \cdots \geq k_n$ has the highest piece $x_1^{k_1}x_2^{k_2}\cdots x_n^{k_n}$.*

Proof From the definition of the elementary symmetric polynomials we have that

$$HG(s_k^t) = (x_1x_2\cdots x_k)^t,\ 1 \leq k \leq n, t \geq 1.$$

From Lemma 6.3.7,

$$\begin{aligned} &HG(s_1^{k_1-k_2}s_2^{k_2-k_3}\cdots s_{n-1}^{k_{n-1}-k_n}s_n^{k_n}) \\ &\quad = x_1^{k_1-k_2}(x_1x_2)^{k_2-k_3}\cdots(x_1\cdots x_{n-1})^{k_{n-1}-k_n}(x_1\cdots x_n)^{k_n} \\ &\quad = x_1^{k_1}x_2^{k_2}\cdots x_n^{k_n}. \end{aligned}$$

□

We can now prove the fundamental theorem of symmetric polynomials.

Proof (Theorem 6.3.7) Let $s(x_1, \ldots, x_n) \in F[x_1, \ldots, x_n]$ be a symmetric polynomial. We must show that $s(x_1, \ldots, x_n)$ can be uniquely expressed as a polynomial $f(s_1, \ldots, s_n)$ in the elementary symmetric polynomials $s_1, \ldots, s_n$ with coefficients from F. We prove the existence of the polynomial f by induction on the size of the highest piece. If in the highest piece of a symmetric polynomial all exponents are zero, then it is constant, that is, an element of F and there is nothing to prove.

Now we assume that each symmetric polynomial with highest piece smaller than that of $s(x_1, \ldots, x_n)$ can be written as a polynomial in the elementary symmetric polynomials. Let $ax_1^{k_1} \cdots x_n^{k_n}$, $a \neq 0$, be the highest piece of $s(x_1, \ldots, x_n)$. Let

$$t(x_1, \ldots, x_n) = s(x_1, \ldots, x_n) - as_1^{k_1-k_2} \cdots s_{n-1}^{k_{n-1}-k_n} s_n^{k_n}.$$

Clearly, $t(x_1, .., x_n)$ is another symmetric polynomial, and from Lemma 6.3.9 the highest piece of $t(x_1, \ldots, x_n)$ is smaller than that of $s(x_1, \ldots, x_n)$. Therefore, $t(x_1, \ldots, x_n)$ and hence $s(x_1, \ldots, x_n) = t(x_1, .., x_n) + as_1^{k_1-k_2} \cdots s_{n-1}^{k_{n-1}-k_n} s_n^{k_n}$ can be written as a polynomial in $s_1, \ldots, s_n$.

To prove the uniqueness of this expression assume that $s(x_1, \ldots, x_n) = f(s_1, \ldots, s_n) = r(s_1, \ldots, s_n)$. Then $f(s_1, \ldots, s_n) - r(s_1, .., s_n) = h(s_1, \ldots, s_n) = \phi(x_1, \ldots, x_n)$ is the zero polynomial in $x_1, \ldots, x_n$. Hence, if we write $h(s_1, \ldots, s_n)$ as a sum of products of powers of the $s_1, \ldots, s_n$, all coefficients disappear because two different products of powers in the $s_1, .., s_n$ have different highest pieces. This follows from Lemma 6.3.9. Therefore, f and r are the same, proving the theorem. □

From this theorem we obtain the following theorem, which is crucial in our study of both algebraic numbers in general and algebraic integers.

Theorem 6.3.10 *Let α be an algebraic number and $\alpha_1, \ldots, \alpha_n$ be its set of conjugates in $\mathbb{C}$. Then any symmetric polynomial in $\alpha_1, \ldots, \alpha_n$ over $\mathbb{Q}$ is a rational number.*

Proof Since α is algebraic we have $irr(\alpha, \mathbb{Q}) \in \mathbb{Q}[x]$. Since $\alpha_1, \ldots, \alpha_n$ are the conjugates of α we have that $irr(\alpha, \mathbb{Q})$ splits in $\mathbb{C}$ as

$$irr(\alpha, \mathbb{Q}) = (x - \alpha_1)(x - \alpha_2) \cdots (x - \alpha_n).$$

Therefore the coefficients of $irr(\alpha, \mathbb{Q})$ are up to ±1 precisely the elementary symmetric polynomials in the conjugates. Since $irr(\alpha, \mathbb{Q}) \in \mathbb{Q}[x]$ it follows then that any elementary symmetric polynomial in the conjugates of α is a rational number and then Theorem 6.3.10 follows from the fundamental theorem of symmetric polynomials. □

6.3.4 Discriminant and Norm

We introduce certain complex numbers that will be used to further describe both algebraic numbers and algebraic number fields. We first must extend our definition of conjugate.

Let $K = \mathbb{Q}(\theta)$ be an algebraic number field of degree n. Then K has precisely n embeddings $\sigma_i : K \to \mathbb{C}$ which fix $\mathbb{Q}$. These can be defined by $\sigma_i : 1 \to 1, \theta \to \theta_i$, where θ_i is a conjugate of θ. Now let $\alpha \in K$ of degree m. Since $|K : \mathbb{Q}(\alpha)||\mathbb{Q}(\alpha) : \mathbb{Q}| = |K : \mathbb{Q}|$ it follows that $m|n$. Let $d = \frac{n}{m}$.

Definition 6.3.8 *Let K be an algebraic number field of degree n and $\alpha \in K$ of degree m. Then the set of* **conjugates of** α **for** K *is the set $\{\sigma_i(\alpha)\}$, where σ_i are the n embeddings of K into $\mathbb{C}$.*

Lemma 6.3.10 *Let K be an algebraic number field of degree n and $\alpha \in K$ of degree m. Then the set of* **conjugates of** α **for** K *consists of the m distinct conjugates of α in $\mathbb{C}$ each repeated $d = \frac{n}{m}$ times.*

Proof On the set of n embeddings $K \to \mathbb{C}$ fixing $\mathbb{Q}$ define the relation $\sigma \sim \tau$ if $\sigma(\alpha) = \tau(\alpha)$. This is an equivalence relation (see exercises). Each equivalence class has size $|K : \mathbb{Q}(\alpha)| = d$ and hence there are m of them. Since each $\sigma(\alpha)$ is a conjugate of α in $\mathbb{C}$ it follows that the set $\{\sigma_i(\alpha)\}$ consists of the m conjugates of α in $\mathbb{C}$ each repeated d times. □

Hence an $\alpha \in K$ always has n conjugates for K. By looking at degrees it follows that these conjugates will be distinct if and only if $K = \mathbb{Q}(\alpha)$. Next we define the discriminant of a basis.

Definition 6.3.9 *Let K be an algebraic number field of degree n and let $\alpha_1, \ldots, \alpha_n$ be a basis for K over $\mathbb{Q}$. For each α_i let α_{ij}, $j = 1, \ldots, n$ be the n conjugates of α_i for K. Then the* **discriminant** *of the basis $\alpha_1, \ldots, \alpha_n$ is*

$$\Delta(\alpha_1, \ldots, \alpha_n) = (det(\alpha_{ij}))^2 = \begin{vmatrix} \alpha_{11} & \alpha_{12} & \ldots & \alpha_{1n} \\ \alpha_{21} & \alpha_{22} & \ldots & \alpha_{2n} \\ \ldots & \ldots & \ldots & \ldots \\ \alpha_{n1} & \alpha_{n2} & \ldots & \alpha_{nn} \end{vmatrix}^2 .$$

Notice that if we change the ordering of the basis we interchange a column of the matrix (α_{ij}) and thus multiply the determinant by ±1. Hence by squaring the determinant the value remains the same. Therefore the discriminant of a basis is independent of the ordering. Second, notice that if $\beta_1, \ldots, \beta_n$ is another basis then

$$\Delta(\beta_1, \ldots, \beta_n) = |(c_{ij})|^2 \Delta(\alpha_1, \ldots, \alpha_n),$$

where (c_{ij}) is the transition matrix. Therefore the discriminant of any basis has the same sign. We show below that the discriminant is a rational number.

Theorem 6.3.11 *Let $K = \mathbb{Q}(\alpha)$ be an algebraic number field. Then the discriminant of any basis is rational and nonzero.*

Proof Now $\Delta(\alpha_1, \ldots, \alpha_n)$ is a symmetric function of $\alpha_1, \ldots, \alpha_n$ and their conjugates so by the results of the last section it follows that the discriminant is rational.

Since $K = \mathbb{Q}(\alpha)$ it has a basis of the form $1, \alpha, \dots, \alpha^{n-1}$. If α_i is a conjugate of α then α_i^j is a conjugate of α^j. Therefore if $\alpha_1 = \alpha, \dots, \alpha_n$ are the conjugates of α for K we have

$$\Delta(1, \alpha, \dots, \alpha^{n-1}) = \begin{vmatrix} 1 & \alpha_1 & \alpha_1^2 & \dots & \alpha_1^{n-1} \\ 1 & \alpha_2 & \alpha_2^2 & \dots & \alpha_2^{n-1} \\ \dots \\ 1 & \alpha_n & \alpha_n^2 & \dots & \alpha_n^{n-1} \end{vmatrix}^2 .$$

This determinant is called the **Vandermonde determinant** and can be shown to have the value (see exercises)

$$V(\alpha) = \begin{vmatrix} 1 & \alpha_1 & \alpha_1^2 & \dots & \alpha_1^{n-1} \\ 1 & \alpha_2 & \alpha_2^2 & \dots & \alpha_2^{n-1} \\ \dots \\ 1 & \alpha_n & \alpha_n^2 & \dots & \alpha_n^{n-1} \end{vmatrix} = \prod_{i<j} (\alpha_j - \alpha_i).$$

Since the elements of a basis are all distinct it follows that $V(\alpha) \neq 0$ so that $\Delta(1, \alpha, \dots, \alpha^{n-1}) \neq 0$. Since the discriminant of one basis is nonzero the discriminant of any basis is nonzero completing the theorem. □

As part of our discussion of algebraic integers in the next section we will look at bases which have minimal discriminant and from these define the discriminant not only of a particular basis but as an invariant of the whole field K.

We next define two further concepts.

Definition 6.3.10 *Suppose $\alpha \in K$, where K is an algebraic number field of degree n. Let*

$$\alpha_1 = \sigma_1(\alpha), \dots, \alpha_n = \sigma_n(\alpha)$$

be the conjugates of α for K, where the σ_i are the n embeddings of K into $\mathbb{C}$. Then the **norm** *of α in K is*

$$N_K(\alpha) = \alpha_1 \alpha_2 \cdots \alpha_n.$$

This definition agrees with our previous definition of norm in $\mathbb{Z}[i]$. If $\alpha \in \mathbb{Z}[i] \subset \mathbb{Q}(i) = K$ then its conjugate for K is precisely its complex conjugate $\overline{\alpha}$. To see this notice that if $\alpha = a + bi \in \mathbb{Z}[i]$ then $p(\alpha) = 0$, where $p(x) = (x - \alpha)(x - \overline{\alpha}) \in \mathbb{Q}[x]$. If $\alpha \notin \mathbb{Z}$ then $p(x) = irr(\alpha, \mathbb{Q})$. Hence $N_K(\alpha) = \alpha\overline{\alpha} = a^2 + b^2$ which agrees with the previous definition. We will discuss quadratic integers and their norms more completely in the next section. In $\mathbb{Z}[i]$ the norm was multiplicative and always had rational value. In general:

Lemma 6.3.11 *(1) $N_K(\alpha)$ is a rational number for $\alpha \in K$.*

(2) If α, β are in the algebraic number field K then

$$N_K(\alpha\beta) = N_K(\alpha) N_K(\beta).$$

Proof If $\alpha_1, \ldots, \alpha_n$ are the conjugates of α for K then the norm $N_K(\alpha)$ is a symmetric function of $\alpha_1, \ldots, \alpha_n$ and hence rational.

If $\beta_1, \ldots, \beta_n$ are the conjugates of β for K then $\alpha_1\beta_1, \ldots, \alpha_n\beta_n$ are the conjugates of $\alpha\beta$ for K. It follows that $N_K(\alpha\beta) = N_K(\alpha)N_K(\beta)$. □

Finally if $\alpha \in K$ for an algebraic number field K we define the **trace** of α in K as $tr_K(\alpha) = \alpha_1 + \cdots + \alpha_n$, where $\alpha_1 = \sigma_1(\alpha), \ldots, \alpha_n = \sigma_n(\alpha)$ are the conjugates of α for K.

Now let $K = \mathbb{Q}(\theta)$ be an algebraic number field of degree n. For $\alpha \in K$ define the mapping $T_\alpha : K \to K$ by

$$T_\alpha(x) = \alpha x.$$

This is a linear transformation of the n-dimensional $\mathbb{Q}$-vector space K (see exercises) and therefore is given by an $n \times n$ matrix. This matrix is related to the trace and norm in the following manner.

Theorem 6.3.12 *Let $K = \mathbb{Q}(\theta)$ be an algebraic number field of degree n and let $\alpha \in K$. Then if T_α is the linear transformation defined above*

1. $N_K(\alpha) = det(T_\alpha)$
2. $tr_K(\alpha) = tr(T_\alpha)$

Let $f_\alpha(t) = \det(tI - T_\alpha)$ be the characteristic polynomial of T_α and let $p_\alpha(t) = irr(\alpha, \mathbb{Q})$. Theorem 6.3.12 will then follow from the next two lemmas. Notice that the multiplicativity of the norm and the additivity of the trace follow directly from this matrix formulation.

Lemma 6.3.12 *Let K be an algebraic number field of degree n and $\alpha \in K$ of degree m. Let $d = \frac{n}{m}$ and suppose that $f_\alpha(t)$ and $p_\alpha(t)$ are as above. Then*

$$f_\alpha(t) = (p_\alpha(t))^d.$$

Proof Let $p_\alpha(t) = t^m + c_{m-1}t^{m-1} + \cdots + c_0$. Now $\{1, \alpha, \alpha^2, \ldots, \alpha^{m-1}\}$ is a basis for $\mathbb{Q}(\alpha)$ over $\mathbb{Q}$. Let $\alpha_1, \ldots, \alpha_d$ be a basis for K over $\mathbb{Q}(\alpha)$. Then

$$\{\alpha_1, \alpha_1\alpha, \ldots, \alpha_1\alpha^{m-1}, \ldots, \alpha_d\alpha^{m-1}\}$$

is a basis of K over $\mathbb{Q}$. The matrix of the linear transformation T_α with respect to this basis has the form

$$\begin{pmatrix} M & 0 & \ldots \\ 0 & M & \ldots \\ \ldots & \ldots & 0 \\ \ldots & 0 & M \end{pmatrix},$$

where

$$M = \begin{pmatrix} 0 \ 0 \dots 0 & -c_0 \\ 1 \ 0 \dots 0 & -c_1 \\ 0 \ 1 \dots 0 & -c_2 \\ \dots & \\ 0 \ 0 \dots 1 & -c_{n-1} \end{pmatrix}.$$

The characteristic polynomial of M is

$$\det(tI - M) = t^m + c_{m-1}t^{m-1} + \cdots + c_0 = p_\alpha(t).$$

Then from the form of the matrix for T_α we have $f_\alpha(t) = (p_\alpha(t))^d$. □

Lemma 6.3.13 *Let σ run through all the embeddings of K into $\mathbb{C}$ which fix $\mathbb{Q}$. Then:*

1. $f_\alpha(t) = \prod_\sigma (t - \sigma(\alpha))$
2. $tr_K(\alpha) = \sum_\sigma \sigma(\alpha)$
3. $N_K(\alpha) = \prod_\sigma \sigma(\alpha)$

Proof As before the embeddings of K into $\mathbb{C}$ fall into m equivalence classes. Let $\sigma_1, \dots, \sigma_m$ be a set of representatives. Then

$$p_\alpha(t) = \prod_{i=1}^{m} (t - \sigma_i(\alpha))$$

and from the previous lemma

$$\begin{aligned} f_\alpha(t) &= (\prod_{i=1}^{m} (t - \sigma_i(\alpha)))^d \\ &= \prod_{i=1}^{m} \prod_{\sigma \sim \sigma_i} (t - \sigma(\alpha)) = \prod_\sigma (t - \sigma(\alpha)) \end{aligned}$$

This proves part (1) and the other two parts follow directly from the definitions trace and norm in terms of α. □

6.4 Algebraic Integers

We now look at **integers** in an algebraic number field.

Definition 6.4.1 *An* **algebraic integer** *is a complex number α that is a root of a* **monic** *integral polynomial. That is, $\alpha \in \mathbb{C}$ is an algebraic integer if there exists $f(x) \in \mathbb{Z}[x]$ with $f(x) = x^n + b_{n-1}x^{n-1} + \cdots + b_0, b_i \in \mathbb{Z}, n \geq 1$, and $f(\alpha) = 0$.*

An algebraic integer is clearly an algebraic number. Hence there exists $p(x) = irr(\alpha, \mathbb{Q})$.

Lemma 6.4.1 *If $\alpha \in \mathbb{C}$ is an algebraic integer, then all its conjugates, $\alpha_1, \ldots, \alpha_n$, over $\mathbb{Q}$ are also algebraic integers.*

Proof Let $f(x) \in \mathbb{Z}[x]$ be a monic polynomial with $f(\alpha) = 0$. Let $p(x) = irr(\alpha, \mathbb{Q})$. Let $\alpha_1, \ldots, \alpha_n$ be the conjugates of α. Since $p(x) = irr(\alpha, \mathbb{Q}) = irr(\alpha_i, \mathbb{Q}) = p_{\alpha_i}(x)$, for $i = 1, \ldots, n$ we have $p_{\alpha_i}(x)|f(x)$ for $i = 1, \ldots, n$. Hence $f(\alpha_i) = 0$ for $i = 1, \ldots, n$. □

Lemma 6.4.2 *$\alpha \in \mathbb{C}$ is an algebraic integer if and only if $irr(\alpha, \mathbb{Q}) \in \mathbb{Z}[x]$.*

Proof If $irr(\alpha, \mathbb{Q}) \in \mathbb{Z}[x]$ then α is an algebraic integer directly from the definition.

To prove the converse we need the concept of a **primitive integral polynomial**. This is a polynomial $p(x) \in \mathbb{Z}[x]$ such that the gcd of all its coefficients is 1. The following can be proved (see exercises):
(1) If $f(x)$ and $g(x)$ are primitive then so is $f(x)g(x)$.
(2) If $f(x) \in \mathbb{Z}[x]$ is monic then it is primitive.
(3) If $f(x) \in \mathbb{Q}[x]$ then there exists a rational number c such that $f(x) = cf_1(x)$ with $f_1(x)$ primitive.

Now suppose $f(x) \in \mathbb{Z}[x]$ is a monic polynomial with $f(\alpha) = 0$. Let $p(x) = irr(\alpha, \mathbb{Q})$. Then $p(x)$ divides $f(x)$ so $f(x) = p(x)q(x)$.

Let $p(x) = c_1 p_1(x)$ with $p_1(x)$ primitive and let $q(x) = c_2 q_2(x)$ with $q_2(x)$ primitive. Then

$$f(x) = cp_1(x)q_1(x).$$

Since $f(x)$ is monic it is primitive and hence $c = 1$ so $f(x) = p_1(x)q_1(x)$.

Since $p_1(x)$ and $q_1(x)$ are integral and their product is monic they both must be monic. Since $p(x) = c_1 p_1(x)$ and they are both monic it follows that $c_1 = 1$ and hence $p(x) = p_1(x)$. Therefore $p(x) = irr(\alpha, \mathbb{Q})$ is integral. □

We now show the close ties between algebraic integers and rational integers.

Lemma 6.4.3 *If α is an algebraic integer and also rational then it is a rational integer.*

Proof If $\alpha \in \mathbb{Q}$ then $irr(\alpha, \mathbb{Q}) = x - \alpha$. But if α is also an algebraic integer than $irr(\alpha, \mathbb{Q}) \in \mathbb{Z}[x]$. Hence $x - \alpha \in \mathbb{Z}[x]$ and $\alpha \in \mathbb{Z}$. □

The following ties algebraic numbers in general to corresponding algebraic integers. Notice that if $q \in \mathbb{Q}$ then there exists a rational integer n such that $nq \in \mathbb{Z}$. This result generalizes this simple idea.

Theorem 6.4.1 *If θ is an algebraic number then there exists a rational integer $r \neq 0$ such that $r\theta$ is an algebraic integer.*

Proof Since θ is an algebraic number there exists a $p(x) \in \mathbb{Z}[x]$ with $p(\theta) = 0$. Suppose $p(x) = a_n x^n + a_{n-1}x^{n-1} + \cdots + a_0$ with $a_i \in \mathbb{Z}$. Then

$$a_n\theta^n + a_{n-1}\theta^{n-1} + \cdots + a_0 = 0.$$

Let $\zeta = a_n\theta$. Then

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

Let $p(x) = x^n + a_{n-1}x^{n-1} + a_n a_{n-2}x^{n-2} + \cdots + a_n^{n-1}a_0$. Then from the above $p(\zeta) = 0$ and therefore $\zeta = a_n\theta$ is an algebraic integer. □

6.4.1 *The Ring of Algebraic Integers*

We saw that the set $\mathcal{A}$ of all algebraic numbers is a subfield of $\mathbb{C}$. We now show that the set $\mathcal{I}$ of all algebraic integers forms a subring of $\mathcal{A}$. First an extension of the following result on algebraic numbers.

Lemma 6.4.4 *Suppose $\alpha_1, \ldots, \alpha_n$ are the set of conjugates over $\mathbb{Q}$ of an algebraic integer α. Then any integral symmetric function of $\alpha_1, \ldots, \alpha_n$ is a rational integer.*

Proof We have $irr(\alpha, \mathbb{Q}) = (x - \alpha_1)\cdots(x - \alpha_n) \in \mathbb{Z}[x]$. Hence the elementary symmetric functions are rational integers. It follows from the fundamental theorem of symmetric polynomials that any integral symmetric function is also a rational integer. □

Theorem 6.4.2 *The set $\mathcal{I}$ of all algebraic integers forms a subring of $\mathcal{A}$.*

Proof Clearly it suffices to show that if α, β are algebraic integers then so are $\alpha \pm \beta$ and $\alpha\beta$. Let $\alpha_1 = \alpha, \ldots, \alpha_n$ be the conjugates of α and $\beta_1 = \beta, \ldots, \beta_m$ the conjugates of β. Let

$$f(x) = \prod_{i=1}^{n}\prod_{j=1}^{m}(x - (\alpha_i + \beta_j)) = x^{n+m} + d_{n+m-1}x^{n+m-1} + \cdots + d_0.$$

The coefficients d_k are symmetric functions in α_i, β_j, and therefore from the remarks above we have $d_k \in \mathbb{Z}$. It follows that $f(x) \in \mathbb{Z}[x]$ and further $f(\alpha + \beta) = 0$. Therefore, $\alpha + \beta$ is an algebraic integer. We treat $\alpha - \beta$ and $\alpha\beta$ analogously. □

We note that $\mathcal{A}$, the field of algebraic numbers, is precisely the quotient field of the ring of algebraic integers.

Now let $K = \mathbb{Q}(\theta)$ be an algebraic number field and let $\mathcal{O}_K = K \cap \mathcal{I}$. Then $\mathcal{O}_K$ forms a subring of K called the algebraic integers or just integers of K. Further analysis of the proof of Theorem 6.4.1 shows that each $\beta \in K$ can be written as

$$\beta = \frac{\alpha}{r}$$

with $\alpha \in \mathcal{O}_K$ and $r \in \mathbb{Z}$.

We now look at the norms of algebraic integers.

Lemma 6.4.5 *If α is an algebraic integer then $N(\alpha)$ is a rational integer.*

Proof $N(\alpha) = \alpha_1 \cdots \alpha_n$, where $\alpha_1 = \sigma_1(\alpha), \dots, \alpha_n = \sigma_n(\alpha)$ are the conjugates of α for K. But this is an integral symmetric function of the conjugates and by Lemma 6.4.4 it is a rational integer. □

Lemma 6.4.6 *Let $K = \mathbb{Q}(\theta)$ be an algebraic number field. Then α is a unit in $\mathcal{O}_K$ if and only if $N(\alpha) = \pm 1$.*

Proof If $\alpha\beta = 1$ then $1 = N(\alpha\beta) = N(\alpha)N(\beta)$. But $N(\alpha), N(\beta)$ are rational integers so $|N(\alpha)| = |N(\beta)| = 1$.

Conversely suppose $N(\alpha) = \pm 1$. If $\alpha = \alpha_1$, and $\alpha_2, \dots \alpha_n$ are the conjugates of α in K then

$$\alpha_1 \cdots \alpha_n = 1 \implies \alpha_1(\alpha_2 \cdots \alpha_n) = 1.$$

Since K is a field $\alpha_1^{-1} = \alpha_2 \cdots \alpha_n \in K$. But $\alpha_2 \cdots \alpha_n$ is an algebraic integer so $\alpha_2 \cdots \alpha_n \in \mathcal{O}_K$. Hence α is a unit in $\mathcal{O}_K$. □

Based on the multiplicativity of the norm we obtain prime factorizations (not necessarily unique) in any algebraic number ring $\mathcal{O}_K$. Notice first that there are no primes at all in $\mathcal{I}$ the set of all algebraic integers. If $\alpha \in \mathcal{I}$ then $\alpha = \sqrt{\alpha}\sqrt{\alpha}$, where $\sqrt{\alpha} \in \mathbb{C}$. However if $p(\alpha) = 0$ for $p(x) \in \mathbb{Z}[x]$ then $p_1(\sqrt{\alpha}) = 0$, where $p_1(x) = p(x^2)$. Hence $\sqrt{\alpha}$ is also an algebraic integer. Since this is true for any $\alpha \in \mathcal{I}$ there is always a nontrivial factorization and hence α cannot be prime.

From now on K will denote an algebraic number field and $\mathcal{O}_K$ its ring of integers.

Lemma 6.4.7 *If $\alpha \in \mathcal{O}_K$ and $N(\alpha) = p$, where p is a rational prime then α is a prime in $\mathcal{O}_K$.*

Proof Suppose $\alpha = \beta\gamma$. Then $N(\alpha) = N(\beta)N(\gamma)$. Since all are rational integers and $N(\alpha)$ is prime we must have either $|N(\beta)| = 1$ or $|N(\gamma)| = 1$ from which it follows that either β or γ is a unit. □

Theorem 6.4.3 *Let K be an algebraic number field and $\mathcal{O}_K$ its ring of integers. Then each $\alpha \in \mathcal{O}_K$ is either 0, a unit or can be factored into a product of primes.*

Proof Suppose $\alpha \neq 0$ is not a unit. Then $N(\alpha) \neq 1$. We do an induction on $|N(\alpha)|$. If $|N(\alpha)| = 2$ then α is prime from Lemma 6.4.7. Suppose $|N(\alpha)| > 2$. If $\alpha = \beta\gamma$ then if neither β or γ is a unit it follows that $|N(\beta)| < |N(\alpha)|$ and $|N(\gamma)| < |N(\alpha)|$. From the inductive hypothesis it follows that both β and γ have prime factorizations and hence so does α. □

We stress again that the prime factorization need not be unique. However from the existence of a prime factorization we can mimic Euclid's original proof (see Chapter 2) to obtain:

Corollary 6.4.1 *There exist infinitely many primes in $\mathcal{O}_K$ for any algebraic number ring $\mathcal{O}_K$.*

6.4.2 Integral Bases

If K has degree n over $\mathbb{Q}$ we show that there exists $\omega_1, \dots, \omega_n$ in $\mathcal{O}_K$ such that each $\alpha \in \mathcal{O}_K$ is expressible as

$$\alpha = m_1\omega_1 + \cdots + m_n\omega_n,$$

where $m_1, \dots, m_n \in \mathbb{Z}$.

Definition 6.4.2 *An* **integral basis** *for $\mathcal{O}_K$ is a set of integers*

$$\omega_1, \dots, \omega_t \in \mathcal{O}_K$$

such that each $\alpha \in \mathcal{O}_K$ can be expressed uniquely as

$$\alpha = m_1\omega_1 + \cdots + m_t\omega_t,$$

where $m_1, \dots, m_t \in \mathbb{Z}$.

We show first that there must exist an integral basis.

Theorem 6.4.4 *Let $\mathcal{O}_K$ be the ring of integers in the algebraic number field K of degree n over $\mathbb{Q}$. Then there exists at least one integral basis for $\mathcal{O}_K$.*

Proof Since K has degree n there is a basis $\omega_1, \dots, \omega_n$ for K over $\mathbb{Q}$. Each ω_i is algebraic, so by Theorem 6.4.1 for each i there is a rational integer r_i such that $r_i\omega_i \in \mathcal{O}_K$. Multiplying through by a large enough rational integer r we would have $r\omega_1, \dots, r\omega_n$ all in $\mathcal{O}_K$, These are clearly still independent so they still constitute a vector space basis of K over $\mathbb{Q}$. It follows that K has bases (as a vector space) which are all integers in $\mathcal{O}_K$. Further if $\omega_1, \dots, \omega_n$ is such a basis for K all in $\mathcal{O}_K$ then the discriminant of this basis $\Delta(\omega_1, \dots, \omega_n)$ must be a rational integer since the discriminant is a symmetric polynomial over $\mathbb{Z}$ of its arguments.

Among all bases of K which are in $\mathcal{O}_K$ choose one, say $\omega_1, \dots, \omega_n$, with $|\Delta(\omega_1, \dots, \omega_n)|$ minimal. This exists since these values are positive rational integral. We claim that this is an integral basis for $\mathcal{O}_K$.

Let $\alpha \in \mathcal{O}_K$. Since $\alpha \in K$ and $\omega_1, \dots, \omega_n$ is a basis over $\mathbb{Q}$,

$$\alpha = q_1\omega_1 + \cdots + q_n\omega_n$$

with $q_i \in \mathbb{Q}$. We show that each q_i must be a rational integer. Suppose that q_1 is not rational. Then $q_1 = m_1 + r_1$ with $m_1 \in \mathbb{Z}$ and $0 < r_1 < 1$. Consider now the set $\omega_1^\star, \dots, \omega_n^\star$, where

$$\omega_1^\star = (q_1 - m_1)\omega_1 + q_2\omega_2 + \cdots + q_n\omega_n$$

$$\omega_i^\star = \omega_i \text{ if } i \neq 1.$$

The transition matrix from $\omega_1, \dots, \omega_n$ to $\omega_1^\star, \dots, \omega_n^\star$ is

$$C = \begin{pmatrix} q_1 - m_1 & q_2 & \dots q_n \\ 0 & \dots & \dots \\ \dots & & \\ \dots & 1 & \end{pmatrix}.$$

This has determinant $q_1 - m_1 = r_1 > 0$ so $\omega_1^\star, \dots, \omega_n^\star$ is another basis consisting solely of integers. Its discriminant is

$$\Delta(\omega_1^\star, \dots, \omega_n^\star) = r_1^2 \Delta(\omega_1, \dots, \omega_n).$$

Since $r_1 < 1$ this implies that

$$|\Delta(\omega_1^\star, \dots, \omega_n^\star)| < |\Delta(\omega_1, \dots, \omega_n)|$$

contradicting the minimality of $|\Delta(\omega_1, \dots, \omega_n)|$. Therefore $r = 0$ and $q_1 = m_1 \in \mathbb{Z}$. The other coefficients follow in the same manner.

□

Therefore $\mathcal{O}_K$ has at least one integral basis. We next show that the cardinality of any integral basis is the same as the degree of K.

Theorem 6.4.5 *Let $\mathcal{O}_K$ be the ring of integers in the algebraic number field K of degree n over $\mathbb{Q}$. Then any integral basis for $\mathcal{O}_K$ is also a basis for K over $\mathbb{Q}$. Hence the cardinality of any integral basis is the same as the degree of K. Further all integral bases have the same discriminant.*

Proof Let $\omega_1, \dots, \omega_t$ be an integral basis and suppose $\alpha \in K$. Then there exists an $r \in \mathbb{Z}$, $r \neq 0$, with $r\alpha \in \mathcal{O}_K$. Hence

$$r\alpha = m_1\omega_1 + \cdots + m_t\omega_t \text{ with } m_i \in \mathbb{Z}.$$

Then

$$\alpha = \frac{m_1}{r}\omega_1 + \cdots + \frac{m_t}{r}\omega_t.$$

Therefore $\omega_1, \dots, \omega_t$ span K as a vector space over $\mathbb{Q}$. We must show that they are independent over $\mathbb{Q}$.

Suppose $q_1\omega_1 + \cdots + q_t\omega_t = 0$. Then multiplying through by the lcm of the denominators of the q_i we obtain $m_1\omega_1 + \cdots + m_t\omega_t = 0$ for some $m_i \in \mathbb{Z}$. Since $\omega_1, \dots, \omega_t$ is an integral basis it follows that each $m_i = 0$. But then each $q_i = 0$ and therefore $\omega_1, \dots, \omega_t$ are independent and hence form a basis.

It then follows that $t = n$, where $n = |K : \mathbb{Q}|$.

Now let $\omega_1, \dots, \omega_n$ and $\zeta_1, \dots, \zeta_n$ be two integral bases. Their transition matrix $C = (c_{ij})$ is rational integral and

$$\Delta(\omega_1, \dots, \omega_n) = |(c_{ij})|^2 \Delta(\zeta_1, \dots, \zeta_n).$$

It follows that $\Delta(\omega_1, \dots, \omega_n)$ divides $\Delta(\zeta_1, \dots, \zeta_n)$. Reversing the roles we get that $\Delta(\zeta_1, \dots, \zeta_n)$ divides $\Delta(\omega_1, \dots, \omega_n)$ and therefore $\Delta(\omega_1, \dots, \omega_n) = \Delta(\zeta_1, \dots, \zeta_n)$. □

Definition 6.4.3 *The* **discriminant** d_K *of an algebraic number field K is the common value of the discriminants of all integral bases of its ring of integers $\mathcal{O}_K$.*

For some later work in this section we need the following result whose proof we will give in Section 6.5 after we introduce some material on ideals.

Theorem 6.4.6 *If K has degree n over $\mathbb{Q}$ then each ideal $I \subset \mathcal{O}_K$ has an integral basis of rank n. That is there exists $\omega_1, \dots, \omega_n \in I$ such that any $\alpha \in I$ can be expressed uniquely as*

$$\alpha = m_1\omega_1 + \cdots + m_n\omega_n$$

with $m_i \in \mathbb{Z}$. In particular any ideal in I is finitely generated of rank $\leq n$.

In particular this implies that the index $[\mathcal{O}_K : I]$ is finite. Then for an ideal I in $\mathcal{O}_K$ we define the discriminant $d(I)$ of I analogously via an integral basis of I. This certainly exists and the value $d(I)$ is independent of the chosen integral basis of I. Since the index $[\mathcal{O}_K : I]$ is finite we have $d(I) = [\mathcal{O}_K : I]^2 d_K$.

6.4.3 Quadratic Fields and Quadratic Integers

We now look more closely at **quadratic fields**. These are algebraic number fields K of degree 2. The Gaussian rationals $\mathbb{Q}(i)$ are an example. Let $K = \mathbb{Q}(\theta)$ with $|K : \mathbb{Q}| = 2$. Then θ satisfies a degree 2 integral polynomial $p(x) = ax^2 + bx + c$. Let $d = b^2 - 4ac$ be the discriminant of this polynomial. Then clearly $\mathbb{Q}(\sqrt{d}) \subset \mathbb{Q}(\theta)$ and hence if d is not a perfect square it follows by degrees that $\mathbb{Q}(\sqrt{d}) = \mathbb{Q}(\theta)$. Further if $d = m^2 d_1$ then $\mathbb{Q}(\sqrt{d}) = \mathbb{Q}(\sqrt{d_1})$. It follows from these comments that any quadratic field K has the form $\mathbb{Q}(\sqrt{d})$ for some squarefree integer d. In the following we always consider d to be squarefree. If $d > 0$ then K is called a **real quadratic field** while if $d < 0$ it is an **imaginary quadratic field**. In both cases $\{1, \sqrt{d}\}$ is a basis for K over $\mathbb{Q}$.

The integers in $\mathbb{Q}(\sqrt{d})$ are called **quadratic integers** and we characterize them. Suppose $\alpha \in \mathcal{O}_K$ is a quadratic integer. Since $\alpha \in K$ we have $\alpha = q_1 + q_2\sqrt{d}$. Since $irr(\alpha, \mathbb{Q})$ is a monic rational integral polynomial of degree 2 we have

$$irr(\alpha, \mathbb{Q}) = (x - \alpha)(x - \overline{\alpha}) = x^2 - (\alpha + \overline{\alpha})x + \alpha\overline{\alpha} \in \mathbb{Z}[x],$$

where $\overline{\alpha} = q_1 - q_2\sqrt{d}$. It follows that $\alpha \in \mathcal{O}_K$ if and only if its trace and norm are both rational integers:

$$tr_K(\alpha) = \alpha + \overline{\alpha} = 2q_1 \in \mathbb{Z}$$

$$N_K(\alpha) = \alpha\overline{\alpha} = q_1^2 - dq_2^2 \in \mathbb{Z}$$

since d is squarefree.

Now

$$(2q_2)^2 d = (2q_1)^2 - 4(q_1^2 - q_2^2 d) \in \mathbb{Z} \implies 2q_2 \in \mathbb{Z}.$$

Therefore $q_1 = \frac{m}{2}, q_2 = \frac{n}{2}$ for rational integers m, n and

$$\alpha = \frac{m + n\sqrt{d}}{2} \text{ with } m, n \in \mathbb{Z}.$$

Further

$$m^2 - n^2 d \equiv 0 \bmod 4.$$

If $d \equiv 2 \bmod 4$ or $d \equiv 3 \bmod 4$ this congruence is solved only if m, n are even or equivalently $q_1, q_2 \in \mathbb{Z}$.

If $d \equiv 1 \bmod 4$ then $m^2 - dn^2 \equiv 0 \bmod 4$ is equivalent to $m \equiv n \bmod 2$.

It follows that the integers in $\mathcal{O}_K$ can be described by:

(1) $m + n\sqrt{d}$ with $m, n \in \mathbb{Z}$.

(2) If $d \equiv 1 \bmod 4$ but not otherwise, also $\frac{m+n\sqrt{d}}{2}$ with m, n odd rational integers.

From this characterization it follows that if d is not congruent to 1 mod 4, every integer in $\mathcal{O}_K$ can be written as $m + n\sqrt{d}$ with $m, n \in \mathbb{Z}$. In other words $\{1, \sqrt{d}\}$ is an integral basis.

If $d \equiv 1 \bmod 4$ let $\omega = \frac{1+\sqrt{d}}{2}$. Then from the characterization every integer in $\mathcal{O}_K$ is uniquely of the form $m + n\omega, m, n \in \mathbb{Z}$ and $\{1, \omega\}$ is an integral basis (see exercises). We summarize all this discussion in the next theorem.

Theorem 6.4.7 *Let K be a quadratic field. Then:*

(1) $K = \mathbb{Q}(\sqrt{d})$ for some squarefree rational integer d.

(2) The integers in K can be characterized as

(a) $m + n\sqrt{d}$ with $m, n \in \mathbb{Z}$;

(b) If $d \equiv 1 \bmod 4$ but not otherwise, also $\frac{m+n\sqrt{d}}{2}$ with m, n odd rational integers.

(3) An integral basis for $\mathcal{O}_K$ is given by

(a) $\{1, \sqrt{d}\}$ *if* $d \equiv 2$ *mod* 4 *or* $d \equiv 3$ *mod* 4*;*
(b) $\{1, \omega\}$, *where* $\omega = \frac{1+\sqrt{d}}{2}$ *if* $d \equiv 1$ *mod* 4.

(4) The discriminant of $K = \mathbb{Q}(\sqrt{d})$ *is*

(a) $4d$ *if* $d \equiv 2, 3$ *mod* 4,
(b) d *if* $d \equiv 1$ *mod* 4.

Proof Everything was explained prior to the theorem except part (4). If $d \equiv 2, 3$ mod 4 then $\{1, \sqrt{d}\}$ is an integral basis. Then

$$\Delta(1, \sqrt{d}) = \begin{vmatrix} 1 & \sqrt{d} \\ 1 & -\sqrt{d} \end{vmatrix}^2 = 4d.$$

If $d \equiv 1$ mod 4 then $\{1, \omega\}$ is an integral basis and

$$\Delta(1, \omega) = \begin{vmatrix} 1 & \frac{1+\sqrt{d}}{2} \\ 1 & \frac{1-\sqrt{d}}{2} \end{vmatrix}^2 = d.$$

□

Theorem 6.4.8 *Suppose that* $K = \mathbb{Q}(\sqrt{d})$ *with* $d < 0$ *and* d *squarefree is a quadratic imaginary number field. If* $d \neq -1, -3$ *then the only units in* $\mathcal{O}_K$ *are* ± 1*. If* $d = -1$ *the units are* $\pm 1, \pm i$ *while if* $d = -3$ *the units are* $\pm 1, \pm\omega, \pm\overline{\omega}$*, where* $\omega = \frac{1+i\sqrt{3}}{2}$.

Proof As we have seen $\alpha \in \mathcal{O}_K$ is a unit if and only if $|N(\alpha)| = 1$. Let α be a unit in $\mathcal{O}_K$. Then $\alpha = x + y\sqrt{d}$ or $\alpha = \frac{x+y\sqrt{d}}{2}$ and then $N(\alpha) = x^2 - dy^2$ or $N(\alpha) = \frac{x^2-dy^2}{4}$.

Since $d < 0$, $x^2 - dy^2 \geq 0$. If $d < -1$ and d is not congruent to 1 mod 4 the only solutions to $x^2 - dy^2 = 1$ is $x = \pm 1$, $y = 0$.

Our analysis of the Gaussian integers showed that if $d = -1$ then $\pm i$ are also units.

If $d < -3$ then the only solutions to $x^2 - dy^2 = 4$ are $x = \pm 2$ again giving the result.

Finally if $d = -3$ we see by computation that $\pm\omega$ and $\pm\overline{\omega}$ are also units (see exercises and note that $\omega^3 = 1$).

□

Theorem 6.4.9 *In any real quadratic field there are infinitely many units.*

Proof The equation $x^2 - dy^2 = 1$ for $d > 0$ and $x, y \in \mathbb{Z}$ is called **Pell's equation**. If $d > 1$ in Section 6.4.6 we will show that this equation has infinitely many solutions. Since $\alpha = x + y\sqrt{d}$ is an integer in $\mathcal{O}_K$ with $N(\alpha) = 1$ it follows that $\mathcal{O}_K$ has infinitely many units.

□

In the real quadratic case the units can be built up from one special unit called a fundamental unit.

Theorem 6.4.10 *Suppose $K = \mathbb{Q}(\sqrt{d})$ with $d > 0$ and squarefree, Then in $\mathcal{O}_K$ there exists a special unit, ϵ_d, called the* **fundamental unit** *such that all units in $\mathcal{O}_K$ are given by*

$$\mu = \pm\epsilon_d^n, n = 0, \pm 1, \pm 2, \cdots + \pm \cdots \quad .$$

This is a special case of a general result called Dirichlet's unit theorem which we will present in Section 6.4.6.

Now what can be said about primes and prime factorization for quadratic integers? We saw in Section 6.4.2 that there is always a prime factorization. However our example in $\mathbb{Q}(\sqrt{-5})$ shows that this is not always unique. Since there is a norm in every $\mathcal{O}_K$ the first question to ask is when this is a Euclidean norm or equivalently which $\mathcal{O}_K$ are Euclidean domains. From the results in Section 6.2 this would imply unique factorization. We have already seen that the Gaussian integers are Euclidean. We state several results concerning these questions (see [Ri]).

Theorem 6.4.11 *Suppose $K = \mathbb{Q}(\sqrt{d})$ with $d < 0$ and squarefree is a quadratic imaginary number field. Then $\mathcal{O}_K$ is Euclidean if and only if*

$$d = -1, -2, -3, -7, -11.$$

We let $\tilde{O}_d$ stand for $\mathcal{O}_K$ when $K = \mathbb{Q}(\sqrt{d})$. The rings

$$\tilde{O}_{-1}, \tilde{O}_{-2}, \tilde{O}_{-3}, \tilde{O}_{-7}, \tilde{O}_{-11}$$

are called the **Euclidean quadratic imaginary number rings**. They and matrix groups with entries from them have been investigated extensively (see [F] and [FR 1]).

In the real case we have the following.

Theorem 6.4.12 *The real quadratic fields $K = \mathbb{Q}(\sqrt{d})$ for which $\mathcal{O}_K$ is Euclidean are for*

$$d = 2, 3, 5, 6, 7, 11, 13, 17, 19, 21, 29, 33, 37, 41, 57, 73.$$

Recall from Section 6.2.3 that being a principal ideal domain always implies unique factorization. It was conjectured by Gauss and finally proven in several results by Heegner, Baker, and Stark that there are only finitely many quadratic imaginary number fields whose integer rings are principal ideal domains.

Theorem 6.4.13 *Suppose $K = \mathbb{Q}(\sqrt{d})$ with $d < 0$ is a quadratic imaginary number field. Then $\mathcal{O}_K$ is a principal ideal domain if and only if*

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

It has been conjectured that there are infinitely many real quadratic fields whose integral rings are principal ideal domains.

In the case where $\mathcal{O}_K$ does have unique factorization we can analyze the primes exactly as we analyzed the Gaussian primes in Theorem 6.2.6. We state the following and leave the proof to the exercises.

Theorem 6.4.14 *Suppose K is a quadratic field and suppose $\mathcal{O}_K$ is a unique factorization domain. Then*

(1) To each prime $\pi \in \mathcal{O}_K$ there corresponds one and only one rational prime p such that $\pi | p$.

(2) Any rational prime p is either a prime in $\mathcal{O}_K$ or a product $\pi_1\pi_2$ of two primes (not necessarily distinct) from $\mathcal{O}_K$. In this case if $\pi_1 \neq \pi_2$, we say p is **decomposed**. *If $\pi_1 = \pi_2$, so that $p = \pi^2$, we say the rational prime is* **ramified**.

(3) All primes in $\mathcal{O}_K$ are either rational primes or the two factors of rational primes (and their associates).

6.4.4 The Transcendence of e and π

There are infinitely many transcendental numbers (see Section 6.3.2) however the only particular number that we have exhibited as transcendental was

$$c = \sum_{j=1}^{\infty} \frac{1}{10^{j!}}.$$

Here we show that the fundamental constants e and π are also transcendental. The transcendence of e was established first by Hermite in 1873 while Lindemann in 1881 proved the transcendence of π.

Theorem 6.4.15 *e is a transcendental number, that is, transcendental over $\mathbb{Q}$.*

Proof We use some complex analysis. Let $f(x) \in \mathbb{R}[x]$ with the degree of $f(x) = m \geq 1$. Let $z_1 \in \mathbb{C}$, $z_1 \neq 0$, and $\gamma : [0, 1] \to \mathbb{C}$, $\gamma(t) = tz_1$. Let

$$I(z_1) = \int_\gamma e^{z_1 - z} f(z)dz = (\int_0^{z_1})_\gamma e^{z_1 - z} f(z)dz.$$

By $(\int_0^{z_1})_\gamma$ we mean the integral from 0 to z_1 along γ. Recall that

$$(\int_0^{z_1})_\gamma e^{z_1 - z} f(z)dz = -f(z_1) + e^{z_1} f(0) + (\int_0^{z_1})_\gamma e^{z_1 - z} f'(z)dz.$$

It follows then by repeated partial integration that

(1) $I(z_1) = e^{z_1} \sum_{j=0}^m f^{(j)}(0) - \sum_{j=0}^m f^{(j)}(z_1)$.

Let $|f|(x)$ be the polynomial that we get if we replace the coefficients of $f(x)$ by their absolute values. Since $|e^{z_1-z}| \le e^{|z_1-z|} \le e^{|z_1|}$, we get

(2) $|I(z_1)| \le |z_1| e^{|z_1|} |f|(|z_1|)$.

Now assume that e is an algebraic number, that is,

(3) $q_0 + q_1 e + \cdots + q_n e^n = 0$ for some $n \ge 1$ and integers $q_0 \ne 0, q_1, \ldots, q_n$, and the greatest common divisor of $q_0, q_1, \ldots, q_n$, is equal to 1.

We consider now the polynomial $f(x) = x^{p-1}(x-1)^p \cdots (x-n)^p$ with p a sufficiently large prime number, and we consider $I(z_1)$ with respect to this polynomial. Let

$$J = q_0 I(0) + q_1 I(1) + \cdots + q_n I(n).$$

From (1) and (3) we get that

$$J = -\sum_{j=0}^{m} \sum_{k=0}^{n} q_k f^{(j)}(k),$$

where $m = (n+1)p - 1$ since $(q_0 + q_1 e + \cdots + q_n e^n)(\sum_{j=0}^{m} f^{(j)}(0)) = 0$.

Now, $f^{(j)}(k) = 0$ if $j < p, k > 0$, and if $j < p-1$ then also for $k = 0$, and hence $f^{(j)}(k)$ is an integer that is divisible by $p!$ for all j, k except for $j = p-1$, $k = 0$. Further, $f^{(p-1)}(0) = (p-1)!(-1)^{np}(n!)^p$, and hence, if $p > n$, then $f^{(p-1)}(0)$ is an integer divisible by $(p-1)!$ but not by $p!$.

It follows that J is a nonzero integer that is divisible by $(p-1)!$ if $p > |q_0|$ and $p > n$. So let $p > n$, $p > |q_0|$, so that $|J| \ge (p-1)!$.

Now, $|f|(k) \le (2n)^m$. Together with (2) we then get that

$$|J| \le |q_1| e |f|(1) + \cdots + |q_n| n e^n |f|(n) \le c^p$$

for a number c independent of n. It follows that

$$(p-1)! \le |J| \le c^p,$$

that is,

$$1 \le \frac{|J|}{(p-1)!} \le c \frac{c^{p-1}}{(p-1)!}.$$

This gives a contradiction, since $\frac{c^{p-1}}{(p-1)!} \to 0$ as $p \to \infty$. Therefore, e is transcendental. □

We now move on to the transcendence of π. Recall first from the proof of Theorem 6.4.1 that if $\alpha \in \mathbb{C}$ is an algebraic number and

$$f(x) = a_n x^n + \cdots + a_0, n \geq 1, a_n \neq 0,$$

and all $a_i \in \mathbb{Z}$ with $f(\alpha) = 0$. then $a_n \alpha$ is an algebraic integer.

Theorem 6.4.16 *π is a transcendental number, that is, transcendental over $\mathbb{Q}$.*

Proof Assume that π is an algebraic number. Then $\theta = i\pi$ is also algebraic. Let $\theta_1 = \theta, \theta_2, \ldots, \theta_d$ be the conjugates of θ. Suppose

$$p(x) = q_0 + q_1 x + \cdots + q_d x^d \in \mathbb{Z}[x], q_d > 0, \text{and } \gcd(q_0, \ldots, q_d) = 1$$

is the entire minimal polynomial of θ over $\mathbb{Q}$. Then $\theta_1 = \theta, \theta_2, \ldots, \theta_d$ are the zeros of this polynomial. Let $t = q_d$. Then from the discussion above $t\theta_i$ is an algebraic integer for all i. From $e^{i\pi} + 1 = 0$ and from $\theta_1 = i\pi$ we get that

$$(1 + e^{\theta_1})(1 + e^{\theta_2}) \cdots (1 + e^{\theta_d}) = 0.$$

The product on the left side can be written as a sum of 2^d terms e^{ϕ}, where $\phi = \epsilon_1\theta_1 + \cdots + \epsilon_d\theta_d, \epsilon_j = 0$ or 1. Let n be the number of terms $\epsilon_1\theta_1 + \cdots + \epsilon_d\theta_d$ that are nonzero. Call these $\alpha_1, \ldots, \alpha_n$. We then have an equation

$$q + e^{\alpha_1} + \cdots + e^{\alpha_n} = 0 \tag{6.4.1}$$

with $q = 2^d - n > 0$. Recall that all $t\alpha_i$ are algebraic integers. We consider the polynomial

$$f(x) = t^{np} x^{p-1} (x - \alpha_1)^p \cdots (x - \alpha_n)^p$$

with p a sufficiently large prime integer. We have $f(x) \in \mathbb{R}[x]$, since the α_i are algebraic numbers and the elementary symmetric polynomials in $\alpha_1, \ldots, \alpha_n$ are rational numbers.

Let $I(z_1)$ be defined as in the proof of Theorem 6.4.15, and now let

$$J = I(\alpha_1) + \cdots + I(\alpha_n).$$

From (1) in the proof of Theorem 6.4.15 and (6.4.1) we get

$$J = -q \sum_{j=0}^{m} f^{(j)}(0) - \sum_{j=0}^{m} \sum_{k=1}^{n} f^{(j)}(\alpha_k),$$

with $m = (n + 1)p - 1$.

Now, $\sum_{k=1}^{n} f^{(j)}(\alpha_k)$ is a symmetric polynomial in $t\alpha_1, \ldots, t\alpha_n$ with integer coefficients since the $t\alpha_i$ are algebraic integers. It follows from the main theorem on symmetric polynomials that $\sum_{j=0}^{m} \sum_{k=1}^{n} f^{(j)}(\alpha_k)$ is an integer. Further, $f^{(j)}(\alpha_k) = 0$ for $j < p$. Hence $\sum_{j=0}^{m} \sum_{k=1}^{n} f^{(j)}(\alpha_k)$ is an integer divisible by $p!$.

Now, $f^{(j)}(0)$ is an integer divisible by $p!$ if $j \neq p-1$, and $f^{(p-1)}(0) = (p-1)!(-t)^{np}(\alpha_1 \cdots \alpha_n)^p$ is an integer divisible by $(p-1)!$ but not divisible by $p!$ if p is sufficiently large. In particular, this is true if $p > |t^n(\alpha_1 \cdots \alpha_n)|$ and also $p > q$.

From (2) in the proof of Theorem 6.4.15 we get that

$$|J| \leq |\alpha_1|e^{|\alpha_1|}|f|(|\alpha_1|) + \cdots + |\alpha_n|e^{|\alpha_n|}|f|(|\alpha_n|) \leq c^p$$

for some number c independent of n.

As in the proof of Theorem 6.4.15, this gives us

$$(p-1)! \leq |J| \leq c^p,$$

that is,

$$1 \leq \frac{|J|}{(p-1)!} \leq c\frac{c^{p-1}}{(p-1)!}.$$

This as before gives a contradiction, since $\frac{c^{p-1}}{(p-1)!} \to 0$ as $p \to \infty$. Therefore, π is transcendental. □

6.4.5 The Geometry of Numbers—Minkowski Theory

We consider some ties between algebraic integers and the geometry of real n-space.

Definition 6.4.4 *Let V be an n-dimensional vector space over the real numbers $\mathbb{R}$. A* **lattice** *in V is a subgroup of the form*

$$\Gamma = \{m_1 v_1 + \cdots + m_k v_k; m_i \in \mathbb{Z}\}$$

with $v_1, \ldots, v_k$ linearly independent vectors of V.

The k-tuple $\{v_1, \ldots, v_k\}$ is called a **basis** *and the set*

$$\phi = \{x_1 v_1 + \cdots + x_k v_k; x_i \in \mathbb{R}, 0 \leq x_1 \ < 1\}$$

is a **fundamental mesh** *of the lattice.*

The lattice is **complete** *if $k = n$.*

As an example consider the lattice given by the Gaussian integers in real 2-space. Here $V = \mathbb{R}^2$, $\Gamma = \mathbb{Z} + \mathbb{Z}i = \mathbb{Z}[i]$ and the fundamental mesh is

$$\phi = \{x + iy; 0 \leq x \ < 1, 0 \leq y \ < 1\}.$$

Now suppose V is a real Euclidean space, that is a finite-dimensional $\mathbb{R}$-vector space with an inner product, that is a symmetric, positive definite bilinear form

$$< \, , \, >: V \times V \to \mathbb{R}.$$

On such a V we can define a volume. The cube spanned by the standard orthonormal basis $e_1, \dots, e_n$ has volume 1 and more generally the parallelopiped

$$\phi = \{x_1 v_1 + \cdots + x_n v_n ; x_i \in \mathbb{R}, 0 \leq x_i \ < 1\}$$

spanned by the independent set of vectors $v_1, \dots, v_n$ has a volume given by

$$vol(\phi) = |\det(A)|,$$

where $A = (a_{ij})$ is the transition matrix from the basis $e_1, \dots, e_n$ to the basis $v_1, \dots, v_n$, that is

$$v_i = \sum_{i=1}^{n} a_{ij} e_j.$$

As an example if we use the ordinary Euclidean inner product on $\mathbb{R}^n$ then

$$vol(\phi) = \lambda(\phi),$$

where λ is the Lebesgue measure.

Further we have $vol(\phi) = |\det(< v_i, v_j >)|^{\frac{1}{2}}$ since

$$< v_i, v_j > = \sum_{k,l} a_{ik} a_{jl} \ < e_k, e_j > = \sum_{k} a_{ik} a_{jk} = AA^t$$

Let Γ be the lattice spanned by $v_1, \dots, v_n$. If ϕ is the fundamental mesh then we define

$$vol(\Gamma) = vol(\phi).$$

This definition is independent of the choice of basis $v_1, \dots, v_n$ for the lattice because the transition matrix to another basis for the lattice is from $GL(n, \mathbb{Z})$.

Now let K be an algebraic number field with $|K : \mathbb{Q}| = n$. Then there are n different embeddings of K into $\mathbb{C}$ which fix $\mathbb{Q}$. Call these $\tau_1, \dots, \tau_n$. From these some are real and some are non-real. Let $\rho_1, \dots, \rho_r$ be the real embeddings $K \to \mathbb{R}$. The non-real complex embeddings $K \to \mathbb{C}$ are given in pairs $\sigma_1, \overline{\sigma_1}, \dots, \sigma_s, \overline{\sigma_s}$, where $\overline{\sigma_i}$ is the complex conjugate of the mapping σ_i. Altogether we have $n = r + 2s$.

For each pair $\sigma_i, \overline{\sigma_i}$ we choose a fixed non-real embedding and call this just σ_i. We define for $a \in K$ the map $f : K \to \mathbb{R}^n$ by

$$f(a) = (\rho_1(a), \dots, \rho_r(a), \text{Re}\,(\sigma_1(a)), \dots, \text{Re}\,(\sigma_s(a)), \text{Im}\,(\sigma_1(a)), \dots, \text{Im}\,(\sigma_s(a))).$$

Further we define

$$< a, b >= \sum_{i=1}^{r} \rho_i(a)\rho_i(b) + 2\sum_{i=1}^{s} \text{Re }(\sigma_i(a))\text{Re }(\sigma_i(b)) + 2\sum_{i=1}^{s} \text{Im }(\sigma_i(a))\text{Im }(\sigma_i(b)).$$

We may extend this to an inner product on $\mathbb{R}^{r+2s}$. For the following we consider the metric defined by this inner product.

Theorem 6.4.17 *If $I \neq 0$ is an ideal in $\mathcal{O}_K$ then $\Gamma = f(I)$ is a complete lattice in $\mathbb{R}^{r+2s}$ with*

$$vol(\Gamma) = \sqrt{|d_K|}[\mathcal{O}_K : I],$$

where d_K is the discriminant of K

Proof Let $\alpha_1, \dots, \alpha_n$ be an integral basis for I such that

$$\Gamma = \mathbb{Z}f(\alpha_1) + \cdots + \mathbb{Z}f(\alpha_n).$$

We number the embeddings $\tau : K \to \mathbb{C}$ via $\tau_1, \dots, \tau_n$ and consider the matrix $A = (\tau_l(\alpha_i))$. Then

$$d(I) = (\det(A))^2 = [\mathcal{O}_K : I]^2 d_K$$

and

$$vol(\Gamma) = |\det(< f(\alpha_i), f(\alpha_j) >)|^{\frac{1}{2}} = |\det(A)|.$$

□

In the Minkowski theory we consider in $\mathbb{R}^n$ the parallelopipeds

$$X = \{x_1, \dots, x_r, u_1, \dots, u_s, v_1, \dots, v_s; |x_i| \leq c_i, i = 1, \dots, r$$

$$u_j^2 + v_j^2 \leq d_i, j = 1, \dots, s\}$$

with $c_i, d_j > 0$.

Using Minkowski's theorem on the existence of lattice points in this type of subsets of $\mathbb{R}^n$ (see [Co]) and an analytic evaluation with respect to the above metric we get the following.

Theorem 6.4.18 *If d_K is the discriminant of $\mathcal{O}_K$ then*

$$\sqrt{|d_k|} \geq \frac{n^n}{n!}(\frac{\pi}{4})^{\frac{n}{2}}$$

As a direct consequence we have the result of Minkowski.

Theorem 6.4.19 *(Minkowski) If $K \neq \mathbb{Q}$ then $|d_K| \neq 1$.*

A refinement of the analytic evaluation leads to a result of Hermite.

Theorem 6.4.20 *If $D > 0$ is constant then there are only finitely many algebraic number fields with $|d_K| \leq D$.*

For detailed proofs of Theorems 6.4.18, 6.4.19 and 6.4.20 see [Ne].

6.4.6 Dirichlet's Unit Theorem

We mentioned when discussing real quadratic fields that each unit is up to ±1 a power of a fundamental unit. This is a special case of the theorem below called the Dirichlet unit theorem. We state it in general and then give a proof for the quadratic case.

Theorem 6.4.21 *(Dirichlet unit theorem) The group of units $U(\mathcal{O}_K)$ of $\mathcal{O}_K$ is the direct product of the finite cyclic group $U(K)$ of roots of unity which are contained in K and a free abelian group of rank $r + s - 1$, whereas in the last section r is the number of real embeddings $K \rightarrow \mathbb{R}$ and s is the number of pairs of complex non-real embeddings $K \rightarrow \mathbb{C}$.*

Equivalently there exist units $\epsilon_1, \ldots, \epsilon_t$ in $U(\mathcal{O}_K)$ with $t = r + s - 1$ called **fundamental units** *such that each unit $u \in U(\mathcal{O}_K)$ is a product*

$$u = \zeta \epsilon_1^{\nu_1} \cdots \epsilon_t^{\nu_t}$$

with $\nu_i \in \mathbb{Z}$ and ζ is a root of unity contained in K.

We prove only the case for quadratic fields $K = \mathbb{Q}(\sqrt{d})$ with d squarefree. For a proof in the general case see [Ne]. We have already considered the units in quadratic imaginary number fields (Theorem 6.4.8) The structure of the unit groups (see [Co]) can be given by

(1) If $d = -1$ then $U(\mathcal{O}_K) = \{\pm 1, \pm i\}$. This is cyclic of order 4.

(2) If $d = -3$ then $U(\mathcal{O}_K) = \{\pm 1, \pm \omega, \pm \overline{\omega}\}$. This is cyclic of order 6 (see exercises).

(3) If $d \neq -1, -3$ and $d \ < 0$ squarefree then $U(\mathcal{O}_K) = \{-1, 1\}$ which is cyclic of order 2.

For the remainder of this section we assume that d is a positive squarefree integer. As explained in the proof of Theorem 6.4.9, for real quadratic fields we must consider solutions of Pell's equation $x^2 - dy^2 = 1$. We will show that there are infinitely many solutions. First we need some technical results.

Lemma 6.4.8 *If ζ is an irrational real number then there are infinitely many rational numbers $\frac{x}{y}$ with $(x, y) = 1$ and $|\frac{x}{y} - \zeta| \ < \frac{1}{y^2}$.*

Proof Consider the partition of the half-open interval $[0, 1)$ by

$$[0, 1) = [0, \frac{1}{n}) \cup [\frac{1}{n}, \frac{2}{n}) \cup \cdots \cup [\frac{n-1}{n}, 1).$$

If $\alpha \in \mathbb{R}$ then the **fractional part** of α is $\alpha - [\alpha]$, whereas usual [ ] is the greatest integer function. The fractional part of any irrational number lies in a unique member of the above partition.

Consider the fractional parts of $0, \zeta, 2\zeta, \ldots, n\zeta$. At least two of these must lie in the same subinterval. Hence there must exist j, k with $j > k$, $0 < j \leq n$ such that

$$|j\zeta - [j\zeta] - (k\zeta - [k\zeta])| < \frac{1}{n}.$$

Put $y = j - k$, $x = [k\zeta] - [j\zeta]$ so that $|x - y\zeta| < \frac{1}{n}$. We may assume that $(x, y) = 1$ for dividing by (x, y) only strengthens the inequality. Further $0 < y < n$ implies that

$$|\frac{x}{y} - \zeta| < \frac{1}{ny} < \frac{1}{y^2}.$$

To obtain infinitely many solutions note that $|\frac{x}{y} - \zeta| \neq 0$ and then choose any integer $m > \frac{1}{|\frac{x}{y} - \zeta|}$. The above procedure then gives the existence of integers x_1, y_1 such that

$$|\frac{x_1}{y_1} - \zeta| < \frac{1}{my_1} < |\frac{x}{y} - \zeta|$$

and $0 < y < m$. Continuing like this then leads to an infinite number of solutions. □

Lemma 6.4.9 *There is a constant $M = M(d)$ such that $|x^2 - dy^2| < M$ has infinitely many integral solutions.*

Proof Write $x^2 - dy^2 = (x + \sqrt{d}y)(x - \sqrt{d}y)$. From Lemma 6.4.8 there exist infinitely many pairs of relatively prime integers (x, y), $y > 0$, satisfying $|x - \sqrt{d}y| < \frac{1}{y}$. It follows that

$$|x + \sqrt{d}y| \leq |x - \sqrt{d}y| + 2\sqrt{d}y < \frac{1}{y} + 2\sqrt{d}y.$$

Then

$$|x^2 - dy^2| < |\frac{1}{y} + 2\sqrt{d}y|\frac{1}{y} \leq 2\sqrt{d} + 1.$$

□

Theorem 6.4.22 *(Pell's equation) $x^2 - dy^2 = 1$ has infinitely many integral solutions. Further there is a particular solution (x_1, y_1) such that every solution has the form $\pm(x_n, y_n)$, where $x_n + y_n\sqrt{d} = (x_1 + y_1\sqrt{d})^n$ for $n \in \mathbb{Z}$.*

Proof From Lemma 6.4.9 there is a positive integer m such that $x^2 - dy^2 = m$ for infinitely many integral pairs (x, y) with $x > 0, y > 0$. We may assume that the x components are distinct. Further since there are only finitely many residue classes modulo $|m|$ one can find pairs $(x_1, y_1), (x_2, y_2)$ such that $x_1 \neq x_2$ and $x_1 \equiv x_2 \mod |m|$ and $y_1 \equiv y_2 \mod |m|$.

Let $\alpha = x_1 - y_1\sqrt{d}$, $\beta = x_2 - y_2\sqrt{d}$. If $\gamma = x - y\sqrt{d}$ let $\overline{\gamma} = x - y\sqrt{d}$ the conjugate of γ and $N(\gamma) = x^2 - dy^2$ the norm of γ.

Then $\alpha\overline{\beta} = A + B\sqrt{d}$ with $m|A$ and $m|B$. Thus $\alpha\overline{\beta} = m(u + v\sqrt{d})$ for some integers u, v. Taking norms on both sides yields

$$m^2 = m^2(u^2 - v^2 d) \implies u^2 - v^2 d = 1.$$

It remains to show that $v \neq 0$.

If $v = 0$ then $u = \pm 1$ and then $\alpha\overline{\beta} = \pm m$. Multiplying by β gives $\alpha m = \pm m\beta$ or $\alpha = \pm\beta$. But this implies $x_1 = x_2$ a contradiction. Therefore there is a solution to Pell's equation with $xy \neq 0$.

We now prove the second assertion. We say that a solution (x, y) is greater than a solution (u, v) if $x + y\sqrt{d} > u + v\sqrt{d}$. Now consider the smallest solution $\alpha = x + y\sqrt{d}$ with $x > 0, y > 0$. Such a solution clearly exists and is unique. It is called a **fundamental solution**. Consider any solution $\beta = u + v\sqrt{d}$ with $u > 0, v > 0$. We show that there is a positive integer n such that $\beta = \alpha^n$.

Suppose not. Then choose $n > 0$ such that $\alpha^n < \beta < \alpha^{n+1}$. Then $1 < (\overline{\alpha})^n\beta < \alpha$ since $\overline{\alpha} = \alpha^{-1}$. However if $(\overline{\alpha})^n\beta = A + B\sqrt{d}$ then (A, B) is a solution to Pell's equation and $1 < A + B\sqrt{d} < \alpha$.

Now $A + B\sqrt{d} > 0$ so $A - B\sqrt{d} = (A + B\sqrt{d})^{-1} > 0$. Hence $A > 0$. Also $A - B\sqrt{d} = (A + B\sqrt{d})^{-1} < 1$ and hence $B\sqrt{d} > A - 1 \geq 0$. Thus $B > 0$. This contradicts the minimality of α. If $\beta = a + b\sqrt{d}$ is a solution with $a > 0, b < 0$ then $\beta^{-1} = a - b\sqrt{d} = \alpha^n$ by the above argument so $\beta = \alpha^{-n}$.

The cases $a < 0, b > 0$ and $a < 0, b < 0$ lead to $-\alpha^n$ for $n \in \mathbb{Z}$. This proves the theorem.

□

We can now prove Dirichlet's Unit Theorem for real quadratic fields.

Theorem 6.4.23 *Let $K = \mathbb{Q}(\sqrt{d})$ with $d > 0$ and squarefree be a real quadratic field. Then there exists a unit $\epsilon_0 \in \mathcal{O}_K$ such that every unit in $\mathcal{O}_K$ is of the form $\pm\epsilon_0^n$ for $n \in \mathbb{Z}$. It follows that $U(\mathcal{O}_K) = \mathbb{Z}_2 \times \mathbb{Z}$ the direct product of $\mathbb{Z}$ and $\mathbb{Z}_2$.*

Proof From Theorem 6.4.22 there exist positive nonzero integers x, y such that $x^2 - dy^2 = 1$. Thus $\epsilon = x + y\sqrt{d}$ is a unit in $\mathcal{O}_K$ with $\epsilon > 1$. Let M be a fixed real number greater than ϵ. There are at most finitely many $\alpha \in \mathcal{O}_K$, $\alpha = p + q\sqrt{d}$, $p, q, \in \mathbb{Q}$ with $|\alpha| < M$ and also $|\overline{\alpha}| < M$. This is clear since there are only finitely many integers k with $|k| < M$.

Let β be a unit with $1 < \beta < M$. Such a β exists since $M > \epsilon$. Then $N(\beta)N(\overline{\beta}) = \pm 1$. If $\overline{\beta} = -\frac{1}{\beta}$ then $-M < -\frac{1}{\beta} < M$ and if $\overline{\beta} = \frac{1}{\beta}$ then also $-M < \frac{1}{\beta} < M$. Thus there are only finitely many units β with $1 < \beta < M$ and of course there is at least one ϵ.

Let ϵ_0 be the smallest positive unit greater than 1. If β is any positive unit then there is a unique integer s with $\epsilon^s \leq \beta < \epsilon^{s+1}$. Then $1 \leq \beta\epsilon_0^{-s} < \epsilon_0$. Since $\beta\epsilon_o^{-s}$ is also a unit we must have $\beta\epsilon^{-s} = 1$. If $\beta < 0$ then $-\beta$ is positive and $-\beta = \epsilon_0^s$ for some $s \in \mathbb{Z}$, completing the proof.

□

If $d = 2$ the fundamental unit is $\epsilon_0 = 1 + \sqrt{2}$ and for $d = 5$ a fundamental unit is $\frac{1}{2}(1 + \sqrt{5})$ (see exercises). However even for small discriminants, computation of the fundamental unit can be quite difficult. For example the fundamental unit for $d = 34$ is $35 + 6\sqrt{34}$.

6.5 The Theory of Ideals

In analyzing the proofs of unique factorization, the uniqueness part, whether in $\mathbb{Z}$, a general Euclidean domain or a principal ideal domain, hinged on the respective analog of Euclid's Lemma. That is, if p is a prime and $p|ab$ then $p|a$ or $p|b$. In these cases this lemma depended on the fact that the principal ideal $< p >$ generated by a prime p was both a prime ideal and a maximal ideal. For the algebraic number rings $\mathcal{O}_K$ we have seen that there are always prime factorizations (Theorem 6.4.1) but these are not always unique. Hence Euclid's Lemma cannot hold in general. The problem is that the principal ideal generated by a prime $\pi \in \mathcal{O}_K$ need not be a prime ideal. Kummer addressed this problem by adjoining to $\mathcal{O}_K$ **ideal numbers** which generated prime ideals. He could recover unique factorization but the components of the factorization did not always lie in the ring $\mathcal{O}_K$. Dedekind took a different approach. Rather than work with factorizations of the elements of $\mathcal{O}_K$ he worked with ideals in $\mathcal{O}_K$. He was then able to show that for all $\mathcal{O}_K$ there is unique factorization of ideals into prime ideals. Further as consequences of this factorization many results in elementary number theory such as Fermat's theorem and the Chinese Remainder Theorem can be recovered, albeit in terms of ideals.

Since each algebraic number ring $\mathcal{O}_K$ is an integral domain we can apply the material on ideals introduced in Section 6.2. Recall that an ideal I in $\mathcal{O}_K$ is a subring of $\mathcal{O}_K$ such that $\lambda I \subset I$ for all $\lambda \in \mathcal{O}_K$. Equivalently $I \subset \mathcal{O}_K$ is an ideal if $\lambda\alpha + \tau\beta \in I$ whenever $\alpha, \beta \in I$ and $\lambda \cdot \tau \in \mathcal{O}_K$. If $\alpha_1, \ldots, \alpha_k \in \mathcal{O}_K$ then the set

$$< \alpha_1, \ldots, \alpha_k > = \{\lambda_1\alpha_1 + \cdots + \lambda_k\alpha_k; \lambda_i \in \mathcal{O}_K\}$$

forms an ideal called the ideal generated by $\alpha_1, \ldots, \alpha_k$. If k is finite then $< \alpha_1, \ldots, \alpha_k >$ is **finitely generated**. The ideal $< \alpha >$ is the **principal ideal** generated by α. An ideal I is a **prime ideal** if whenever $\alpha\beta \in I$ then either $\alpha \in I$ or $\beta \in I$. I is a **maximal ideal** if whenever $\alpha \notin I$ then $< \alpha, I > = \mathcal{O}_K$.

First we show that every ideal $I \subset \mathcal{O}_K$ has an integral basis and hence is finitely generated. This fact follows directly from the fact that $\mathcal{O}_K$ is a finitely generated free $\mathbb{Z}$-module and results on submodules of such modules or more simply from the basis theorem for finitely generated abelian groups (see Chapter 2 or [CFR]). However we give a direct proof mimicking the existence of an integral basis for all of $\mathcal{O}_K$.

Theorem 6.5.1 *If K has degree n over $\mathbb{Q}$ then each ideal $I \subset \mathcal{O}_K$ has an integral basis of rank n. That is there exists $\omega_1, \ldots, \omega_n \in I$ such that any $\alpha \in I$ can be expressed uniquely as*

$$\alpha = m_1\omega_1 + \cdots + m_n\omega_n$$

with $m_i \in \mathbb{Z}$. In particular any ideal in I is finitely generated of rank $\leq n$.

Proof Suppose $A \subset \mathcal{O}_K \subset K$ is a nonzero ideal and suppose $|K : \mathbb{Q}| = n$. If A has an integral basis $\omega_1, \ldots, \omega_k$ then these are linearly independent (as elements of K) over $\mathbb{Q}$. Since the dimension of K over $\mathbb{Q}$ is n it follows that $k \leq n$. Suppose then that $\beta_1, \ldots, \beta_n$ are integers in $\mathcal{O}_K$ which form a basis for K over $\mathbb{Q}$. In the proof of Theorem 6.4.4 it was shown that K has such a basis. If $\alpha \in A$ with $\alpha \neq 0$ then $\alpha\beta_1, \ldots, \alpha\beta_n$ are all in A since A is an ideal and are linearly independent. However since they are in A they can be linearly expressed in terms of $\omega_1, \ldots, \omega_k$ which is impossible if $k < n$. Therefore if A has an integral basis then it must have n elements in it.

The proof that A does indeed have an integral basis is almost identical to the proof of Theorem 6.4.4. Consider all sets $\omega_1, \ldots, \omega_n$ in A which are linearly independent over $\mathbb{Q}$. The set $\alpha\beta_1, \ldots, \alpha\beta_n$ is an example. For each such set the discriminant $\Delta(\omega_1, \ldots, \omega_n)$ is then a nonzero rational integer. Therefore we can choose a set $\omega_1, \ldots, \omega_n$ for which the discriminant is minimal. This is an integral basis for A the detail identical to those in Theorem 6.4.4 (see exercises).

□

The fact that each ideal in $\mathcal{O}_K$ has bounded rank implies immediately that each $\mathcal{O}_K$ is **Noetherian**. That is each ring of algebraic integers satisfies the ascending chain condition on ideals. Hence each ascending chain of ideals in any $\mathcal{O}_K$ eventually becomes stationary (see Section 6.2.3).

Clearly two ideals $A = <\alpha_1, \ldots, \alpha_m>$, $B = <\beta_1, \ldots, \beta_k>$ are the same if each α_i is an integral linear combination of the β_j and each β_i is an integral linear combination of the α_j. From this we obtain.

Lemma 6.5.1 *If $\alpha, \beta \neq 0$ then $<\alpha> = <\beta>$ if and only if α and β are associates.*

Crucial to unique factorization in $\mathbb{Z}$ and in Euclidean domains in general was that each prime ideal is maximal. This is true in all $\mathcal{O}_K$.

Theorem 6.5.2 *An ideal $I \subset \mathcal{O}_K$ with $I \neq <0>$ is a prime ideal if and only if it is a maximal ideal.*

Proof Suppose $P = <\omega_1, \dots, \omega_s>$ is a maximal ideal in $\mathcal{O}_K$. We show that P is also a prime ideal. Suppose $\alpha\beta \in P$ and suppose that $\alpha \notin P$ we must show that $\beta \in P$. Let $P' = <\omega_1, \dots, \omega_s, \alpha>$. Since $\{\omega_1, \dots, \omega_s\} \subset P'$ it follows that $P \subset P'$. Since P is maximal either $P' = P$ or $P' = \mathcal{O}_K$. If $P = P'$ then $\alpha \in P' = P$ contradicting the assumption that $\alpha \notin P$. Therefore $P' = \mathcal{O}_K$ and hence $1 \in P'$. It follows that

$$1 = \alpha_1\omega_1 + \cdots + \alpha_s\omega_s + \alpha_{s+1}\alpha$$

with $\alpha_1, \dots, \alpha_s, \alpha_{s+1} \in \mathcal{O}_K$. Multiplying through by β yields

$$\beta = (\beta\alpha_1)\omega_1 + \cdots + (\beta\alpha_s)\omega_s + \alpha_{s+1}\alpha\beta.$$

Since $\omega_1, \dots, \omega_s \in P$ and $\alpha\beta \in P$ and P is an ideal, it follows that $\beta \in P$. Therefore P is a prime ideal.

Conversely suppose P is a prime ideal. We show that it is maximal. Recall that if R is a commutative ring and I is an ideal then I is maximal if and only if R/I is a field (see Section 6.2). If $\alpha \neq 0$ is an element of P then its norm $N(\alpha)$ is also in P. Since the norm is a rational integer it follows that $P \cap \mathbb{Z} \neq <0>$. Since P is a prime ideal then $P \cap \mathbb{Z}$ is a nonzero prime ideal in $\mathbb{Z}$. Hence $P \cap \mathbb{Z} = p\mathbb{Z}$ for some rational prime p. Then $\mathbb{Z}/p\mathbb{Z} = \mathbb{Z}_p$ a finite field. Now the quotient ring $\mathcal{O}_K/P$ is formed by adjoining algebraic elements to the finite field $k = \mathbb{Z}/pZ$. However adjoining algebraic elements to a field forms a field. Therefore the quotient ring $\mathcal{O}_K/P$ forms a field and therefore P is a maximal ideal.

□

6.5.1 Unique Factorization of Ideals

We now introduce a product on the set of ideals of $\mathcal{O}_K$. Relative to this product we will show that there is unique factorization in terms of prime ideals.

Definition 6.5.1 *If* $A = <\alpha_1, \dots, \alpha_m>$, $B = <\beta_1, \dots, \beta_k>$ *are ideals in* $\mathcal{O}_K$ *then their* **product**

$$AB = <\alpha_1\beta_1, \alpha_1\beta_2, \dots, \alpha_i\beta_j, \dots, \alpha_m\beta_k>$$

is the ideal generated by all products of the generating elements.

It is a simple exercise to show that this definition is independent of the generating systems chosen.

Now we say that A **divides** B denoted $A|B$ if there exists an ideal C such that $B = AC$. A is then called a **factor** of B. A is a **divisor** of B if $B \subset A$. Finally A is an **irreducible** ideal if the only factors of A are A and $<1> = \mathcal{O}_K$.

The concepts of factor and divisor will turn out to be equivalent but we will prove the main theorem before proving this. We would like to use the irreducible ideals in the role of **primes**. However for the time being we will not call them prime ideals reserving that term for the previous definition. However we will eventually prove that an ideal $I \subset \mathcal{O}_K$ is irreducible if and only if it is a prime ideal. Therefore as in the case of rational integers, for ideals, the terms prime and irreducible will be interchangeable.

First we show that a factor is a divisor.

Lemma 6.5.2 *If $A|B$ then $B \subset A$, that is, a factor is a divisor.*

Proof Suppose $B = AC$ so that $A|B$. Let

$$A = <\alpha_1, \ldots, \alpha_s>, B = <\beta_1, \ldots, \beta_t>, \; C = <\gamma_1, \ldots, \gamma_u> .$$

Then

$$<\beta_1, \ldots, \beta_t> = <\alpha_1\gamma_1, \ldots, \alpha_i\gamma_j, \ldots \alpha_s\gamma_u> .$$

Therefore for each $k = 1, \ldots .t$,

$$\beta_k = \sum_{i,j} \theta_{ij}\alpha_i\gamma_j \text{ with } \theta_{ij} \in \mathcal{O}_K.$$

This implies that

$$\beta_k = \sum_i (\sum_j \theta_{i,j}\gamma_j)\alpha_i.$$

Hence each β_k is an integral (from $\mathcal{O}_K$) linear combination of the α_i and thus $\beta_k \in A$. Therefore $B \subset A$.

□

To arrive at the prime factorization we need certain finiteness conditions.

Lemma 6.5.3 *A rational integer $m \neq 0$ belongs to at most finitely many ideals in $\mathcal{O}_K$.*

Proof Suppose m is a rational integer and $m \in A$, where A is an ideal in $\mathcal{O}_K$. Since both $\pm m \in A$ we may assume that $m > 0$. Let $\omega_1, \ldots, \omega_n$ be an integral basis for K. If $A = <\alpha_1, \ldots, \alpha_s>$ then each α_i may be written as

$$\alpha_i = \sum_{i=1}^{n} c_{ij}\omega_j,$$

where the c_{ij} are rational integers. Then for each $j = 1, \ldots, n$

$$c_{ij} = q_{ij}m + r_{ij}, 0 \leq r_{ij} \; < m.$$

Then

$$\alpha_i = \sum (q_{ij}m + r_{ij})\omega_j = m \sum q_{ij}\omega_j + \sum r_{ij}\omega_j = m\gamma_i + \beta_i,$$

where γ_i and β_i are integers and β_i can take on only finitely many values since $r_{ij} \; < m$. Now since $m \in A$ we have

$$A = < \alpha_1, \ldots, \alpha_s > = < \alpha_1, \ldots, \alpha_s, m > = < m\gamma_1 + \beta_1, \ldots, m\gamma_s + \beta_s > .$$

However since $m \in A$ it follows that $m\gamma_i \in A$ for all i and thus

$$A = < \beta_1, \ldots, \beta_s > .$$

Since there are only finitely many choices for each β_i there are only finitely many choices for A.

□

Lemma 6.5.4 *An ideal $A \neq < 0 >$ has only a finite number of divisors and hence only a finite number of factors.*

Proof Let A be an ideal with $A \neq < 0 >$. If $\alpha \in A$ with $\alpha \neq 0$ then the norm $N(\alpha) \in A$. Since α is an algebraic integer $N(\alpha) \in \mathbb{Z}$. It follows that $A \cap \mathbb{Z} \neq \{0\}$. But then $N(\alpha)$ can belong to only finitely many ideals and A can have only finitely many divisors. Since each factor is a divisor, A has only finitely many factors.

□

We now state the main result.

Theorem 6.5.3 *(Unique Factorization of Ideals) Every ideal $I \subset \mathcal{O}_K$ with $I \neq < 0 >$ and $I \neq < 1 >$ can be factored into a product of prime ideals. This factorization is unique except for the ordering of the factors.*

The proof is broken into several steps. First we introduce some further general ideas from algebra.

Definition 6.5.2 *If R is a commutative ring with an identity then a* **module** *over R or an* **R-module** *is an abelian group M which allows scalar multiplication from R satisfying*

1. $rv \in M$ *if* $r \in R, v \in M$.
2. $r(u + v) = ru + rv$ *for* $r \in R, u, v \in M$.
3. $(r + s)v = rv + sv$ *for* $r, s \in R, v \in M$.
4. $(rs)v = r(sv)$ *for* $r, s \in R, v \in M$.
5. $1v = v$ *for* $v \in M$.

Therefore we can think of a module as a vector space where the set of scalars is just a commutative ring rather than a field. Clearly any abelian group is a $\mathbb{Z}$-module.

A subset $\{m_i\}$ of elements of M generates M if every element of M is a finite R-linear combination of finitely many elements from $\{m_i\}$. If a set of generators is finite then M is a finitely generated module over R. If M is a module then an **R-basis** for M is a generating set which is linearly independent over R. Not every R-module has an R-basis. An R-module which has an R-basis is called a **free R-module**. A **submodule** N is a subgroup of M which is also a module. The following is important for our further work. For a proof see [CFR].

Theorem 6.5.4 *Let R be a principal ideal domain and M a free R-module. If $m_1, \dots, m_s$ is a finite R-basis and N is a nonzero submodule of M then N is also free and has a finite basis with $\leq s$ elements.*

Since each abelian group is a $\mathbb{Z}$-module and $\mathbb{Z}$ is a principal ideal domain if we apply this theorem to abelian groups we get the basis theorem for finitely generated abelian groups.

Now we return to the proof of the main theorem. To obtain the existence of a unique factorization we extend the definition of an ideal.

Definition 6.5.3 *A* **fractional ideal** *in K is a nonzero finitely generated $\mathcal{O}_K$-submodule of K. That is $I \subset K$ is a fractional ideal if I is an additive subgroup of K closed under multiplication from $\mathcal{O}_K$.*

An ordinary ideal $A \subset \mathcal{O}_K$ is then also a fractional ideal. In this context we call an ordinary ideal an **integral ideal**.

Notice that fractional ideals can be multiplied in the same manner as ordinary ideals to obtain other fractional ideals. We next define an addition of fractional ideals.

Definition 6.5.4 *If A and B are fractional ideals then the sum is defined by*

$$A + B = \{\alpha + \beta; \alpha \in A, \beta \in B\}$$

The sum of fractional ideals is again a fractional ideal (see exercises).

Lemma 6.5.5 *Every integral ideal contains a product of prime ideals.*

Proof Let S consists of the set of integral ideals for which this statement is false. If S is nonempty, since $\mathcal{O}_K$ satisfies the ACC on ideals (is Noetherian), it follows that S must have a maximal element A. Therefore A is an integral ideal which is not prime and for which any ideal properly containing A must contain a product of prime ideals. Since A is not a prime ideal there must exist elements α, β both not in A but with $\alpha\beta \in A$. Then $A_1 = < A, \alpha >$ and $B_1 = < A, \beta >$ both properly contain A and hence both contain a product of primes ideals. Then $A_1 B_1$ also contains a product of prime ideals. But

$$A_1 B_1 \subset AA + \alpha A + \beta A + < \alpha\beta > \subset A$$

since $\alpha\beta \in A$. But then A contains a product of prime ideals which is a contradiction. Therefore the set S must be empty and hence every integral ideal contains a product of prime ideals.

□

We also need the following which gives an inverse under this multiplication for ordinary ideals.

Definition 6.5.5 *For an integral ideal $A \subset \mathcal{O}_K$ we define*

$$A^{-1} = \{\alpha \in K; \alpha A \in \mathcal{O}_K\}$$

Lemma 6.5.6 *For $A \subset \mathcal{O}_K$, an integral ideal, the set A^{-1} is a fractional ideal and $\mathcal{O}_K \subset A^{-1}$. Further if A is a proper ideal then A^{-1} properly contains $\mathcal{O}_K$.*

Proof We leave the proof that A^{-1} is again a fractional ideal to the exercises and prove that if A is a proper ideal then A^{-1} properly contains $\mathcal{O}_K$. We must show that there is an element of A^{-1} which is not an algebraic integer. Choose an $\alpha \in A$ with $\alpha \neq 0$. From Lemma 6.5.4 there is a set of prime ideals $P_1, \dots, P_s$ satisfying

$$P_1 \cdots P_s \subset < \alpha > \subset A.$$

Choose such a set of prime ideals with minimal possible s. Since $A \neq \mathcal{O}_K$ by the Noetherian property it follows that A must be contained in some maximal (and hence prime) ideal P. Therefore we have

$$P_1 \cdots P_s \subset P.$$

If $P \neq P_i$ for all $i = 1, \dots, s$ then there is an $\alpha_i \in P_i$ with $\alpha_i \notin P$ and with $\alpha_1 \cdots \alpha_s \in P$. This contradicts the fact that P is a prime ideal. Therefore $P = P_i$ for some i. Without loss of generality assume $P = P_1$. We now have

$$P P_2 \cdots P_s \subset < \alpha > \subset A \subset P.$$

Since s was minimal $P_2 \cdots P_s$ is not contained in $< \alpha >$. Therefore there is a $\beta \in P_2 \cdots P_s$ with $\beta \notin < \alpha >$. Let $\gamma = \alpha^{-1}\beta$. Then γ is not an algebraic integer. However

$$\gamma A = \alpha^{-1}\beta A \subset \alpha^{-1}\beta P \subset \alpha^{-1} P P_2 \cdots P_s \subset \mathcal{O}_K.$$

Hence by definition $\gamma \in A^{-1}$. □

Lemma 6.5.7 *If A is an integral ideal then $A^{-1}A = \mathcal{O}_K$.*

Proof Let $B = A^{-1}A$. Then $B \subset \mathcal{O}_K$ so BB is an integral ideal. Then

$$AA^{-1}B = BB^{-1} \subset \mathcal{O}_K \implies A^{-1}B^{-1} \subset A.$$

It follows that for any $\alpha \in B^{-1}$ we must have $A^{-1}\alpha \subset A^{-1}$ and therefore $A^{-1}\alpha^n \subset A^{-1}$ for all natural numbers n. But then $A^{-1} < \alpha >$ is an $\mathcal{O}_K$-submodule of A^{-1} and is therefore finitely generated (see Theorem 6.5.1). However $\mathcal{O}_K[\alpha]$ being a submodule of $A^{-1} < \alpha >$ is also finitely generated. Since $\mathcal{O}_K$ is integrally closed in K it follows that $\alpha \in \mathcal{O}_K$. Therefore $B^{-1} \subset \mathcal{O}_K$ and hence $B^{-1} = \mathcal{O}_K$. It follows that $B = \mathcal{O}_K$ for otherwise by Lemma 6.5.5 $\mathcal{O}_K$ would be proper in B^{-1}. □

Lemma 6.5.8 *Every integral ideal is a product of prime ideals.*

Proof From Lemma 6.5.4 we know that any integral ideal contains a product of prime ideals. If an integral ideal contains a single prime ideal it must coincide with that ideal since prime ideals are maximal. We now do induction on the length of a product of prime ideals contained in an integral ideal and assume that any integral ideal containing a product of fewer than n prime ideals is a product of prime ideals. Now suppose A is an integral ideal and A contains a product of n prime ideals;

$$P_1 P_2 \cdots P_n \subset A.$$

As in the proof of Lemma 6.5.6 choose a maximal ideal P containing A so that we have

$$P_1 P_2 \cdots P_n \subset A \subset P.$$

Again as in the proof of Lemma 6.5.6 P must coincide with one of the P_i say P_1 so that we have

$$P P_2 \cdots P_n \subset A \subset P \implies P^{-1} P P_2 \cdots P_n \subset P^{-1} A \subset \mathcal{O}_K.$$

The integral ideal $P^{-1}A$ now contains a product of fewer than n prime ideals so by our inductive hypothesis we have

$$P^{-1} A = Q_1 \cdots Q_s,$$

where each Q_i is a prime ideal. But then

$$A = P P^{-1} A = P Q_1 \cdots Q_s$$

is a product of prime ideals.

□

Now that we have established that each integral ideal is a product of prime ideals we must show that this product is unique up to ordering.

Lemma 6.5.9 *Let $P_1 \cdots P_s \subset Q_1 \cdots Q_t$, where the P_i and Q_j are all prime ideals. Then $s = t$ and the set of Q_j is just a rearrangement of the set of P_i.*

Proof The proof mimics the proof of the uniqueness of factorization of the rational integers. Since $Q_1 \cdots Q_t \subset Q_1$ we have

$$P_1 \cdots P_s \subset Q_1 \cdots Q_t \subset Q_1.$$

Since Q_1 is prime and hence maximal as in the proofs of the previous lemmas Q_1 must coincide with some P_i. Without loss of generality we may assume then that $Q_1 = P_1$. We then have

$$P_1^{-1} P_1 P_2 P_3 \cdots P_s \subset P_1^{-1} P_1 Q_2 \cdots Q_t \implies P_2 \cdots P_s \subset Q_1 \cdots Q_t.$$

Continuing in this manner we get the result.

□

As an immediate consequence of this lemma we get the following corollary which is the required unique factorization.

Corollary 6.5.1 *Suppose $A = P_1 \cdots P_s = Q_1 \cdots Q_t$ are two expressions for the integral ideal A as a product of prime ideals. Then $s = t$ and the set of Q_j are just a rearrangement of the set of P_i.*

This series of lemmas completes the proof of the unique factorization theorem. If A is a nonzero proper integral ideal then from Lemma 6.5.6 it can be expressed as a product of prime ideals. Then from Corollary 6.5.1 this expression is unique.

Finally we show that a divisor is a factor. Hence by the uniqueness theorem if A is a prime ideal it is also an irreducible ideal. Therefore for ideals the terms prime and irreducible become interchangeable.

Lemma 6.5.10 *Let A and B be integral ideals. Then A is a divisor of B if and only if A is a factor of B.*

Proof We have already seen that if A is a factor of B then A is a divisor, that is, if $A|B$ then $B \subset A$. We must show then that if A is a divisor of B, that is, $B \subset A$, then A is a factor of B. Hence we must show that if $B \subset A$ then there is an ideal C with $B = AC$. Now from unique factorization we have

$$A = P_1^{e_1} \cdots P_r^{e_r}$$

for some prime ideals $P_1, \ldots, P_r$. Here we have combined identical prime ideals to an exponent as in the standard form of a rational integer. Since $B \subset A$ it is an easy consequence of the unique factorization theorem that the factorization of B will contain all the prime ideals in the factorization of A and to a higher exponent. Hence

$$B = P_1^{f_1} \cdots P_r^{f_r} Q_1 \cdots Q_s$$

with each $f_i \geq e_i$ and $Q_1, \ldots, Q_s$ prime ideals. Then

$$C = P_1^{f_1 - e_1} \cdots P_r^{f_r - e_r} Q_1 \cdots Q_s$$

is an integral ideal and $B = AC$.

□

6.5.2 An Application of Unique Factorization

As we saw in Chapter 2 many results are direct consequences of the Fundamental Theorem of Arithmetic. In a similar manner, as a consequence of the unique factorization theorem for ideals many of these results have lovely analogs for ideals in algebraic number rings. In this section we will look at one of these, the Chinese Remainder Theorem. In the final section, after we discuss the ideal class group, an analog of Fermat's theorem will also be presented.

Recall that for the rational integers the following is the Chinese Remainder Theorem.

Theorem 6.5.5 *(Chinese Remainder Theorem) Suppose that $m_1, m_2, \dots, m_k$ are k positive integers that are relatively prime in pairs. If $a_1, \dots, a_k$ are any integers then the simultaneous congruences*

$$x \equiv a_i \; mod \; m_i, i = 1, \dots, k$$

have a common solution which is unique modulo $m_1 m_2 \cdots m_k$.

To extend this result we need to give the analogs of greatest common divisors (gcds) and least common multiples (lcm's) for ideals. Since these concepts are defined in terms of divisibility the definitions are identical.

Definition 6.5.6 *If A and B are integral ideals in $\mathcal{O}_K$ then*

1.

$$gcd(A, B) = D,$$

where D is an integral ideal such that $D|A, D|B$ and if D_1 is another integral ideal such that $D_1|A$ and $D_1|B$ then $D_1|D$.

2.

$$lcm(A, B) = L,$$

where L is an integral ideal such that $A|L$, $B|L$ and if $A|L_1$, $B|L_1$ for some integral ideal L_1 then $L|L_1$.

From the unique factorization theorem it easily follows, in exactly the same manner as for the integers, that if

$$A = P_1^{e_1} \cdots P_r^{e_r} \text{ and } B = P_1^{f_1} \cdots P_r^{f_r}$$

with $P_1, \dots, P_r$ distinct prime ideals and $e_i, f_i \geq 0$ and $P_i^0 = \mathcal{O}_K$ then

$$gcd(A, B) = P_1^{min(e_1, f_1)} \dots P_r^{min(e_r, f_r)}$$

and

$$lcm(A, B) = P_1^{max(e_1, f_1)} \cdots P_r^{max(e_r, f_r)}.$$

Further since an ideal is a factor if and only if it is a divisor, that is $D|A$ if and only if $A \subset D$ it follows that $gcd(A, B)$ is the smallest ideal containing both A and B while $lcm(A, B)$ is the largest ideal contained in both A and B. Now the sum $A + B$ is the smallest ideal containing both A and B and the intersection $A \cap B$ is the largest ideal contained in both A and B. Hence

$$gcd(A, B) = A + B,$$

$$lcm(A, B) = A \cap B.$$

Further, exactly as for the rational integers

$$AB = gcd(A, B) \cdot lcm(A, B) = (A + B)(A \cap B).$$

We summarize all these observations in the next theorem.

Theorem 6.5.6 *Let A, B be integral ideals in $\mathcal{O}_K$ and suppose*

$$A = P_1^{e_1} \cdots P_r^{e^r} \text{ and } B = P_1^{f_1} \cdots P_r^{f_r}$$

with $P_1, \ldots, P_r$ distinct prime ideals and e_i, $f_i \geq 0$ and $P_i^0 = \mathcal{O}_K$ then

(1) $gcd(A, B) = A + B = P_1^{min(e_1, f_1)} \cdots P_r^{min(e_r, f_r)}$.

(2) $lcm(A, B) = A \cap B = P_1^{max(e_1, f_1)} \cdots P_r^{max(e_r, f_r)}$.

(3) $AB = (A + B)(A \cap B)$.

Now to get the Chinese Remainder Theorem we need to extend the concept of **relatively prime** or **coprime**. Since $P_i^0 = \mathcal{O}_K$ we have:

Definition 6.5.7 *The integral ideals A, B are* **relatively prime** *or* **coprime** *if they have no common prime factor. Equivalently they are coprime if $A + B = \mathcal{O}_K$.*

We now get:

Theorem 6.5.7 *(Chinese Remainder Theorem for Ideals) Let $\{A_1, \ldots, A_n\}$ be a set of integral ideals in $\mathcal{O}_K$ which are pairwise relatively prime, that is $A_i + A_j = \mathcal{O}_K$ if $i \neq j$, and let $\{\alpha_1, \ldots, \alpha_n\}$ be an arbitrary set of algebraic integers in $\mathcal{O}_K$. Then there exists an element $\alpha \in \mathcal{O}_K$ such that*

$$\alpha \equiv \alpha_i \text{ mod } A_i \text{ for } 1 \leq i \leq n$$

and further α is unique modulo $A_1 A_2 \cdots A_n$.

Proof The proof mimics the proof for the rational integers, that is we actually construct the element α (see Chapter 2).

Since $A_1, \ldots, A_n$ are pairwise relatively prime it follows that A_i is relatively prime to $\prod_{j \neq i} A_j$. Hence for $1 \leq i \leq n$ there exist elements β_i, β_i' with $\beta_i \in A_i$ and $\beta_i' \in \prod_{j \neq i} A_j$ such that $\beta_i + \beta_i' = 1$. Now let

$$\alpha = \alpha_1 \beta_1' + \alpha_2 \beta_2' + \cdots + \alpha_n \beta_n'.$$

Since $\beta_i + \beta_i' = 1$ and $\beta_i \in A_i$ it follows that $\beta_i' \equiv 1 \bmod A_i$. Further $\beta_i' \in A_j$ if $i \neq j$, so $\beta_i' \equiv 0 \bmod A_j$. Therefore

$$\alpha \equiv \alpha_i \bmod A_i \text{ for } i = 1, \ldots, n.$$

Suppose α' is another simultaneous solution to the given congruences. Then

$$\alpha - \alpha' \in A_1 \cap A_2 \cap \cdots \cap A_n.$$

Since they are pairwise relatively prime

$$A_1 \cap A_2 \cdots \cap A_n = A_1 A_2 \cdots A_n.$$

and hence $\alpha \equiv \alpha' \bmod A_1 \cdots A_n$. □

6.5.3 The Ideal Class Group

Out of the set of fractional ideals in $\mathcal{O}_K$ we will now form a group, called the **ideal class group**, which in a sense will measure how close $\mathcal{O}_K$ is to being a principal ideal domain and hence a unique factorization domain. In particular this group will be trivial if and only if $\mathcal{O}_K$ is a principal ideal domain.

First of all, note that fractional ideals can be multiplied exactly as the ordinary integral ideals of $\mathcal{O}_K$. That is if A, B are fractional ideals with

$$A = <\alpha_1, \ldots, \alpha_m>, B = <\beta_1, \ldots, \beta_k>$$

then their **product**

$$AB = <\alpha_1\beta_1, \alpha_1\beta_2, \ldots, \alpha_i\beta_j, \ldots, \alpha_m\beta_k>$$

is the ideal generated by all products of the generating elements.

Theorem 6.5.8 *The fractional ideals of K form an abelian group under the above multiplication called the* **ideal group** $\mathcal{I}_K$ *of K. The unit element is $<1> = \mathcal{O}_K$ and the inverse element for a fractional ideal A is*

$$A^{-1} = \{x \in K; xA \subset \mathcal{O}_K\}.$$

Proof Associativity and commutativity are clear. Further for any fractional ideal A we have $A\mathcal{O}_K = A$ so $\mathcal{O}_K$ is a unit element. Hence we must show inverses.

If A is an integral ideal then from Lemma 6.5.6 we have $A^{-1}A = \mathcal{O}_K$ with A^{-1} as defined in the theorem. Hence A^{-1} is an inverse for integral ideals. Now let B be a fractional ideal. Then there exists an $\alpha \in \mathcal{O}_K$ with $\alpha \neq 0$ such that $\alpha B \subset \mathcal{O}_K$. Then $(\alpha B)^{-1} = \alpha^{-1}B^{-1}$ as defined above and hence $BB^{-1} = \mathcal{O}_K$.

□

Corollary 6.5.2 *Each fractional ideal A has, up to order, a unique product decomposition*

$$A = \prod_P P^{e_p}$$

with $e_p \in \mathbb{Z}$, at most finitely many $e_p \neq 0$ (recall $P^0 = \mathcal{O}_K$) and $\{P\}$ is the set of prime ideals in $\mathcal{O}_K$.

Proof This mimics the proof that any rational number is a product of rational primes. Each fractional ideal V can be written as a quotient $V = \frac{A}{B} = AB^{-1}$ of two integral ideals A, B. Since each of A, B has a unique expression as a product of prime ideals the result follows.

□

The above corollary can also be phrased as:

Corollary 6.5.3 *The ideal group $\mathcal{I}_K$ is a free abelian group generated by the prime ideals $P \neq < 0 >$ in $\mathcal{O}_K$.*

If $a \in K^\star = K\backslash\{0\}$ then $a\mathcal{O}_K$ forms a fractional ideal. Any fractional ideal of this form is called a **fractional principal ideal**

Theorem 6.5.9 *The set of fractional principal ideals $\{a\mathcal{O}_K\}$ with $a \in K^\star$ forms a normal subgroup of the ideal group $\mathcal{I}_K$. We denote this subgroup by $\mathcal{P}_\mathcal{K}$.*

Proof Now $(a\mathcal{O}_K)(b\mathcal{O}_K) = ab\mathcal{O}_K$ and $(a\mathcal{O}_K)^{-1} = a^{-1}\mathcal{O}_K$ so the set of fractional principal ideals is closed under product and inverse. Therefore $\mathcal{P}_\mathcal{K}$ forms a subgroup. Since the ideal group is abelian any subgroup is normal and hence $\mathcal{P}_\mathcal{K}$ is a normal subgroup.

□

Since $\mathcal{P}_\mathcal{K}$ is a normal subgroup we can form the factor group.

Definition 6.5.8 *The factor group*

$$\mathcal{C}l_K = \mathcal{I}_K/\mathcal{P}_K$$

is called the **ideal class group** *or the* **class group** *of K.*

Let $\mathcal{O}_K^\star$ be the group of units of $\mathcal{O}_K$. Then there is an exact sequence

$$1 \to \mathcal{O}_K^\star \to K^\star \to \mathcal{I}_K \to \mathcal{C}l_K \to 1.$$

The following is immediate.

Theorem 6.5.10 *$\mathcal{O}_K$ is a principal ideal domain if and only if $\mathcal{C}l_K = \{1\}$.*

In general the problem of determining the class group $\mathcal{C}l_K$ is quite complicated.

6.5.4 Norms of Ideals

We define a norm for an ideal which is related to the norm of an element. Further we show that this norm is multiplicative.

Definition 6.5.9 *If A is an ideal in $\mathcal{O}_K$ then we define the* **norm** *of A by*

$$\mathcal{N}(A) = [\mathcal{O}_K : A].$$

First of all notice that the norm of an ideal is always finite since

$$d(A) = [\mathcal{O}_K : A]^2 d_K,$$

where $d(A)$ is the discriminant of the ideal and d_K is the discriminant of the field.

The following result shows how the norm of an ideal is related to the norm of an element.

Theorem 6.5.11 *If $< a >$ is a principal ideal in $\mathcal{O}_K$ then*

$$\mathcal{N}(< a >) = |N_K(a)|.$$

Proof Suppose $\omega_1, \dots, \omega_n$ is a $\mathbb{Z}$-basis for $\mathcal{O}_K$. Then $a\omega_1, \dots, a\omega_n$ is a $\mathbb{Z}$-basis for $a\mathcal{O}_K$. If $a\omega_i = \sum_{j=1}^n a_{ij}\omega_j$ and $A = (a_{ij})$ then

$$|\det(A)| = [\mathcal{O}_K : a\mathcal{O}_K]$$

on one side while $\det(A) = N_K(a)$ by definition.

□

Further this norm is multiplicative on the set of ideals.

Theorem 6.5.12 *Let A be a nonzero integral ideal in $\mathcal{O}_K$. If*

$$A = P_1 P_2 \cdots P_r$$

is the prime ideal decomposition of A then

$$\mathcal{N}(A) = \mathcal{N}(P_1)\mathcal{N}(P_2) \cdots \mathcal{N}(P_r).$$

In particular

$$\mathcal{N}(AB) = \mathcal{N}(A)\mathcal{N}(B)$$

for nonzero integral ideals A, B.

Proof Suppose A is a nonzero integral ideal and $A \neq \mathcal{O}_K$. Then A has a canonical prime ideal decomposition

$$A = P_1^{e_1} \cdots P_s^{e_s}, s \geq 1, e_i \geq 1$$

with pairwise different P_i. We must show that

$$\mathcal{N}(A) = \prod_{i=1}^{s} \mathcal{N}(P_i)^{e_i}.$$

By the Chinese Remainder Theorem we have

$$\mathcal{O}_K / A = \bigoplus_{i=1}^{s} \mathcal{O}_K / P_i^{e_i}$$

which gives

$$\mathcal{N}(A) = \prod_{i=1}^{s} \mathcal{N}(P_i^{e_i}).$$

It remains to show that for each prime ideal P and each natural number n we have $[P^n : P^{n+1}] = \mathcal{N}(P)$. For this we choose a $t \in P^n / P^{n+1}$ and consider the homomorphism of abelian groups given by $x \to tx + P^{n+1}$ from $\mathcal{O}_K$ into the factor group P^n / P^{n+1}.

The kernel of this map is an ideal in $\mathcal{O}_K$. The kernel does not contain all of $\mathcal{O}_K$ since $t \notin P^{n+1}$ but it does contain P since $tP \subset P^{n+1}$. Therefore since P is maximal this kernel must be P. The image of this homomorphism is the factor group T / P^{n+1}, where $T = t\mathcal{O}_K + P^{n+1}$ is an ideal in $\mathcal{O}_K$ contained in P^n but not contained in P^{n+1}. Therefore we must have precisely $T = P^n$. The isomorphism theorem for abelian groups then gives

$$\mathcal{O}_K / P \cong P^n / P^{n+1}.$$

Hence in particular

$$[\mathcal{O}_K : P] = \mathcal{N}(P) = [P^n : P^{n+1}]$$

completing the proof.

□

Suppose P is a nonzero prime ideal in $\mathcal{O}_K$. Then it is a maximal ideal and hence the factor ring $\mathcal{O}_K/P$ is a field and hence a finite field since $[\mathcal{O}_K : P]$ is finite. If its characteristic is p then $P \cap \mathbb{Z} = p\mathbb{Z}$, where p is a rational prime. Now $\mathcal{N}(P)$ is the number of elements in $\mathcal{O}_K/P$ and therefore $\mathcal{N}(P) = p^f$ for some $f \in \mathbb{N}$. This exponent is called the **residue class degree** of the prime ideal P. It is the degree of the field $\mathcal{O}_K/P$ over its prime field $\mathbb{Z}_p$. The multiplicative group $(\mathcal{O}_K/P)^\star$ is cyclic being the finite multiplicative group of a field (see Chapter 2 and the exercises). From this we obtain the analog of Fermat's theorem for ideals in $\mathcal{O}_K$.

Theorem 6.5.13 *(Fermat) If $P \neq < 0 >$ is a prime ideal in $\mathcal{O}_K$ then*

$$\alpha^{\mathcal{N}(P)} \equiv \alpha \; mod \; P$$

for all $\alpha \in \mathcal{O}_K$.

We saw in Section 6.4.3 that rational primes in quadratic integer rings may be decomposed in $\mathcal{O}_K$. Further we can classify all possible situations. We generalize this.

Theorem 6.5.14 *(Decomposition of a rational prime). Let p be a rational prime. The exponent $e(p) = \nu_P(p\mathcal{O}_K)$ of a prime ideal P with $P|p\mathcal{O}_K$ in the prime ideal decomposition is called the* **ramification index** *of p in K over $\mathbb{Q}$. Then*

$$\sum_{P|p\mathcal{O}_K} e(p)f(p) = [K : \mathbb{Q}],$$

where $f(p)$ is the residue class degree of p.

Proof Let $n = [K : \mathbb{Q}]$ be the degree of K over $\mathbb{Q}$ and let p be a rational prime. Then

$$\mathcal{N}(p\mathcal{O}_K) = |N(p)| = p^n.$$

On the other hand by the Chinese Remainder Theorem $\mathcal{O}_K/p\mathcal{O}_K$ is isomorphic to the direct sum of the factor rings $\mathcal{O}_K/(P^{e(p)})$, where $P|p\mathcal{O}_K$. Hence

$$p^n = |\mathcal{O}_K/p\mathcal{O}_K| = \prod_{P|p\mathcal{O}_K} \mathcal{N}(P)^{e(p)} = \prod_{P|p\mathcal{O}_K} P^{f(p)e(p)}.$$

□

Finally we show that there are only finitely many elements α in $\mathcal{O}_K$ of a given norm.

Theorem 6.5.15 *Up to units there are only finitely many elements $\alpha \in \mathcal{O}_K$ with a given norm $N_K(\alpha) = a$.*

Proof Let a be a rational integer with $a > 1$. We first claim that in each of the finitely many residue classes of $\mathcal{O}_K/a\mathcal{O}_K$ there are, up to units, at most one element α with $|N_K(\alpha)| = a$.

To see this suppose $\beta = \alpha + a\gamma$ with $\gamma \in \mathcal{O}_K$ is another element with $|N_K(\beta)| = a$. Then

$$\frac{\alpha}{\beta} = 1 \pm \frac{N(\beta)}{\beta}\gamma \in \mathcal{O}_K$$

since $\frac{N(\beta}{\beta} \in \mathcal{O}_K$. Analogously

$$\frac{\beta}{\alpha} = 1 \pm \frac{N(\alpha)}{\alpha}\gamma \in \mathcal{O}_K.$$

This implies that α, β are associates, that is $\alpha = \epsilon\beta$ with ϵ a unit.

It follows that up to units there are at most $[\mathcal{O}_K : a\mathcal{O}_K]$ elements in $\mathcal{O}_K$ with the norm $\pm a$.

□

6.5.5 Class Number

In this final section we show that the ideal class group must be finite giving another finite integer invariant for each number field.

The Minkowski Theory (see Section 6.4.5) leads to the following which we state without proof (see [Ne]).

Theorem 6.5.16 *Each ideal $A \neq <0>$ in $\mathcal{O}_K$ contains an element $a \in A$ with*

$$|N_K(a)| \le (\frac{2}{\pi})^s \sqrt{|d_K|}\mathcal{N}(A),$$

whereas before s denotes the number of pairs of complex, non-real embeddings of K into $\mathbb{C}$

Using this result we obtain:

Theorem 6.5.17 *For each algebraic number field K the ideal class group*

$$\mathcal{C}l_K = \mathcal{I}_K/\mathcal{P}_K$$

is finite. Its order $h_K = [\mathcal{I}_K : \mathcal{P}_K]$ is called the **class number** *of K.*

Proof Let $P \neq (0)$ be a prime ideal in $\mathcal{O}_K$ and suppose $P \cap \mathbb{Z} = p\mathbb{Z}$ with p a rational prime. Then $\mathcal{O}_K/P$ is a finite extension of its prime field $F_p = \mathbb{Z}/\mathbb{Z}_p$ of degree $f \geq 1$. Hence $\mathcal{N}(P) = p^f$.

For a fixed rational prime p there are only finitely many prime ideals P with $P \cap \mathbb{Z} = p\mathbb{Z}$ since then $P|p\mathbb{Z}$. Therefore there are only finitely many prime ideals P with bounded absolute norm. Now each nonzero integral ideal A has a prime ideal decomposition

$$A = P_1^{e_1} \cdots P_r^{e_r} \text{ with } e_r \geq 1$$

and then we have

$$\mathcal{N}(A) = (\mathcal{N}(P_1))^{e_1} \cdots (\mathcal{N}(P_r))^{e_r}.$$

Putting this altogether we have that there are only finitely many ideals $A \neq (0)$ in $\mathcal{O}_K$ with bounded absolute norm $\mathcal{N}(A) \leq M$.

Hence it is enough to show that each class $[A] \in \mathcal{C}l_K$ contains an integral ideal A_1 with

$$\mathcal{N}(A_1) \leq M = (\frac{2}{\pi})^s \sqrt{d_K},$$

where s is as in Theorem 6.5.16.

To show this, choose an arbitrary representative $A \neq (0)$ in this class and a nonzero $\gamma \in \mathcal{O}_K$ with $B = \gamma A^{-1} \subset \mathcal{O}_K$. By Theorem 6.5.5.1 there exists an $\alpha \in B$ with $\alpha \neq 0$ such that

$$|N_K(\alpha)|(\mathcal{N}(B))^{-1} = \mathcal{N}((\alpha\mathcal{O}_K)B^{-1}) = \mathcal{N}(\alpha B^{-1}) \leq M.$$

The ideal $A_1 = \alpha B^{-1} = \alpha\gamma^{-1}A \in [A]$ has the desired property.

□

We remarked before that an algebraic number ring $\mathcal{O}_K$ is a principal ideal domain if and only if its ideal class group is trivial. Hence in the present language we can say that $\mathcal{O}_K$ is a principal ideal domain if and only if the class number of K is 1.

For quadratic imaginary number fields $\mathbb{Q}(\sqrt{-d})$ Heegner, Stark, and Baker proved the following.

Theorem 6.5.18 *Let $K = \mathbb{Q}(\sqrt{-d})$, where d is a squarefree positive integer. Then K has class number 1, that is $h_K = 1$, if and only if*

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

For more on this see [Ri 3]. We end with the following conjecture.

Conjecture 6.5.19 *There are infinitely many algebraic number fields with class number one.*

6.6 Exercises

6.1 Show that in any ring R with identity 1 (commutative or not) that if $uv = 1$ and $wu = 1$ then $v = w$. Hence if an element has both a left and right inverse it is a unit.

6.2 Let T be an $n \times n$ matrix over a field F. Suppose $TU = I$ for some matrix U. Show that $UT = I$ also.

(Hint: Consider T as a linear transformation. If $TU = I$ it must have rank n. Hence there exists a matrix V such that $VT = I$. Apply Problem 6.1)

6.3 Show that the set of units in a commutative ring R with identity forms an abelian group under multiplication.

6.4 Show that if $a \in \mathbb{Z}_n$ then a is a unit if and only if $(a, n) = 1$.

6.5 Show that in any UFD there are infinitely many primes. (Hint: Use Euclid's Proof)

6.6 Prove Lemma 6.2.1. Let F be a field and let $P(x) \neq 0$, $Q(x) \neq 0$ be nonzero polynomials in $F[x]$. Then:

1. $\deg P(x)Q(x) = \deg P(x) + \deg Q(x)$.
2. $\deg (P(x) \pm Q(x)) \leq \max(\deg P(x), \ \deg Q(x))$ if $P(x) \pm Q(x) \neq 0$.

6.7 Let F be a field and $F[x]$ the set of polynomials over F. Verify the ring properties for $F[x]$.

6.8 Fill in the details for a proof of the division algorithm in $F[x]$. (Hint: Consider the degrees of the polynomials.)

6.9 Let S be a subring of the field F (such as $\mathbb{Z}$ in $\mathbb{R}$). Let $S[x]$ consist of the polynomials in $F[x]$ with coefficients from S. Show that $S[x]$ is a subring of $F[x]$. Recall that to show a subset is a subring we must only show that it is nonempty and closed under addition, subtraction, and multiplication.

6.10 Use the division algorithm to find the quotient and remainder for the following pairs of polynomials in the indicated polynomial rings.

(a) $f(x) = x^3 + 5x^2 + 6x + 1$, $g(x) = x - 1$ in $\mathbb{R}[x]$.

(b) $f(x) = x^3 + 5x^2 + 6x + 1$, $g(x) = x - 1$ in $\mathbb{Z}_5[x]$.

(c) $f(x) = x^3 + 5x^2 + 6x + 1$, $g(x) = x - 1$ in $\mathbb{Z}_{13}[x]$.

6.11 Use the **Euclidean algorithm** to find the gcd of the following pairs of polynomials in $\mathbb{Q}[x]$.
(a) $f(x) = 2x^3 - 4x^2 + x - 2$, $g(x) = x^3 - x^2 - x - 2$.
(b) $f(x) = x^4 + x^3 + x^2 + x + 1$, $g(x) = x^3 - 1$.

6.12 Show that if $f(x) \in \mathbb{R}[x]$ and $\alpha \in \mathbb{C}$ is a root then $\overline{\alpha}$, its complex conjugate, is also a root.

6.13 Use the Fundamental Theorem of Algebra coupled with Problem 6.12 to show that if $p(x) \in \mathbb{R}[x]$ is irreducible then $p(x)$ is of degree 1 or of degree 2.

6.14 Prove Lemma 6.2.8: Let R be a Euclidean domain and let $r_1, r_2 \in R$. Then any two gcds of $r_1, r_2 \in R$ are associates. Further an associate of a gcd of r_1, r_2 is also a gcd.

6.15 Prove Lemma 6.2.9: Suppose that R is a Euclidean domain and $r_1, r_2 \in R$ with $r_2 \neq 0$. Then a gcd d for r_1, r_2 exists and is expressible as a linear combination with minimal norm. That is there exists $x, y \in R$ with

$$d = r_1 x + r_2 y$$

and $N(d) \leq N(d_1)$ for any other linear combination of r_1, r_2.

Further if $r_1 \neq 0, r_2 \neq 0$ then a gcd can be found by the Euclidean algorithm exactly as in $\mathbb{Z}$ and $F[x]$. (Hint: Mimic the proof in the ordinary integers $\mathbb{Z}$.)

6.16 Suppose D is a Euclidean domain and assume $r \in D$ has two prime factorizations

$$r = r_1 \cdots r_k = s_1 \cdots s_t$$

with $r_1, \dots, r_k, s_1, \dots, s_t$ all primes in D. Show that each r_i is an associate of some s_j and $k = t$. (Hint: Use Euclid's Lemma repeatedly.)

6.17 Prove Lemma 6.2.11: If $\alpha, \beta \in \mathbb{Z}[i]$ then:

1. $N(\alpha)$ is an integer for all $\alpha \in \mathbb{Z}[i]$.
2. $N(\alpha) \geq 0$ for all $\alpha \in \mathbb{Z}[i]$.
3. $N(\alpha) = 0$ if and only if $\alpha = 0$.
4. $N(\alpha) \geq 1$ for all $\alpha \neq 0$.
5. $N(\alpha\beta) = N(\alpha)N(\beta)$ that is the norm is multiplicative.

6.18 (a) Find the gcd and lcm of the Gaussian integers $5 + 3i$ and $6 - 4i$.
(b) Determine if $1 + 4i$ and $13i$ are primes or not in $\mathbb{Z}[i]$.
(c) Determine the prime decomposition in $\mathbb{Z}[i]$ of $3 + 5i$.

6.19 Solve the congruence in $\mathbb{Z}[i]$.

$$(2+3i)x \equiv 1 \bmod 1+3i$$

6.20 Suppose that $p(x) = a_k x^k + \cdots + a_0 \in \mathbb{Z}[x]$ and $p(r) = 0$ with $r = \frac{m}{n} \in \mathbb{Q}$. Show that $m|a_0, n|a_n$. (This is called the rational root theorem).

6.21 Use the rational root theorem coupled with polynomial factorization to show that

$$p(x) = x^3 - x + 5$$

is irreducible over $\mathbb{Q}$.

6.22 Use the multiplicativity of the norm to show that in $\mathbb{Z}[\sqrt{-5}]$ the numbers $3, 7, 1+2i\sqrt{5}, 1-2i\sqrt{5}$ are all primes and not associates of each other. Recall that $N(a+bi\sqrt{5}) = a^2 + 5b^2$.

Since $21 = 3 \cdot 7 = (1+2i\sqrt{5})(1-2i\sqrt{5})$ this shows that prime factorization is not unique in $\mathbb{Z}[\sqrt{-5}]$.

6.23 Prove that any Euclidean domain is a principal ideal domain. (Hint: Let $I \subset D$, with $I \neq \{0\}$, be an ideal with D a Euclidean domain. Let $r \in I$ with minimal norm. Mimic the proof in $\mathbb{Z}$ to show that $I = (r)$.

6.24 Show that the following properties hold in a PID.

(i) $a|b$ if and only if $< b > \subset < a >$.

(ii) $< b > = < c >$ if and only if b and c are associates.

(iii) $< a > = R$ if and only if a is a unit.

6.25 Prove that if R is a UFD then the polynomial ring $R[x]$ is also a UFD.

6.26 Let F be a field and I the set of polynomials in $F[x, y]$ with constant term 0. Show that this forms an ideal which is not principal.

6.27 Let R be an integral domain and $I \subset R$ an ideal. Show that $r_1 \sim r_2$ if $r_1 - r_2 \in I$ defines an equivalence relation on R. (Since the equivalence classes are the cosets of I this shows that the cosets partition R.)

6.28 Suppose F is a field and $p(x) \in F[x]$ is irreducible. Then show that if $\langle x \rangle = x + < p(x) >$ in the factor ring

$$F' = F[x]/ < p(x) >$$

then $p(< x >) = < p(x) >$. (Consider the operations in F'.)

6.29 Prove Lemma 6.3.1: If $F \subset F' \subset F''$ are fields with F'' a finite extension of F, then $|F' : F|$ and $|F'' : F'|$ are also finite, and

$$|F'' : F| = |F'' : F'||F' : F|.$$

6.30 Show that if $F \subset F'$ are fields and $\alpha \in F'$ then the intersection of all subfields of F' containing both α and F is again a subfield.

6.31 Let K be an algebraic number field of degree n. On the set of n embeddings $K \to \mathbb{C}$ fixing $\mathbb{Q}$ define the relation $\sigma \sim \tau$ if $\sigma(\alpha) = \tau(\alpha)$ for $\alpha \in K$. Show that this is an equivalence relation.

6.32 Let $\alpha \in \mathbb{R}$ be algebraic over $\mathbb{Q}$ and β be transcendental. Show that $\alpha \pm \beta$, $\alpha\beta$, $\frac{\alpha}{\beta}$ are all transcendental.

6.33 Let F be a field and $x_0, x_1, \dots, x_n$ are $n+1$ distinct elements of F. Prove that the Vandermonde determinant has the value

$$V(x_0, \dots, x_n) = \begin{vmatrix} 1 & x_0 & \dots & x_0^n \\ 1 & x_1 & \dots & x_1^n \\ \dots \\ 1 & x_n & \dots & x_n^n \end{vmatrix} = \prod_{i<j}(x_j - x_i).$$

(Hint: Use the following steps)

(i) Show that it is true for $n = 2$.

(ii) Let $V_n(x) = V(x_0, \dots, x_{n-1}, x)$ with x as a variable. Show that $V_n(x)$ is a polynomial of degree n with roots $x_0, \dots, x_{n-1}$.

(iii) Use part (ii) to show that

$$V_n(x) = V(x_0, \dots, x_{n-1})(x - x_0) \cdots (x - x_n).$$

(iv) Substitute x_n to complete the induction and the proof.

6.34 Let $K = \mathbb{Q}(\theta)$ be an algebraic number field of degree n. For $\alpha \in K$ define the mapping $T_\alpha : K \to K$ by

$$T_\alpha(x) = \alpha x.$$

Show that this is a linear transformation of the n-dimensional $\mathbb{Q}$-vector space K.

6.35 A **primitive integral polynomial** is a polynomial $p(x) \in \mathbb{Z}[x]$ such that the gcd of all its coefficients is 1. Prove the following:

(a) If $f(x)$ and $g(x)$ are primitive then so is $f(x)g(x)$.

(b) If $f(x)$ is monic then it is primitive.

(c) If $f(x) \in \mathbb{Q}[x]$ then there exists a rational number c such that $f(x) = cf_1(x)$ with $f_1(x)$ primitive.

6.36 Let $K = \mathbb{Q}(\sqrt{-d})$ with d squarefree. Let $\omega = \sqrt{d}$ if $d \equiv 2 \bmod 4$ or $d \equiv 3 \bmod 4$ and let $\omega = \frac{1+\sqrt{d}}{2}$ if $d \equiv 1 \bmod 4$. Show that every integer in $\mathcal{O}_K$ is uniquely of the form $m + n\omega$, $m, n \in \mathbb{Z}$ and $\{1, \omega\}$ is an integral basis.

6.37 Let $d = 3$, $K = Q(\sqrt{-d})$ and $\omega = \frac{-1+i\sqrt{3}}{2}$. Show that $\pm\omega, \pm\overline{\omega}$ are units in $\mathcal{O}_K$. (Note that $\omega^3 = 1$.)

6.38 Complete the proof of Theorem 6.5.1, that is that A does indeed have an integral basis. (Hint: Mimic the proof of Theorem 6.4.4.)

6.39 Show that the product of two ideal is independent of generating system, that is if $A =< \alpha_1, \ldots, \alpha_m >$, $B =< \beta_1, \ldots, \beta_k >$ are ideals in $\mathcal{O}_K$ and also $A =< \alpha_1', \ldots, \alpha_m' >$, $B =< \beta_1', \ldots, \beta_k' >$ then

$$< \alpha_1\beta_1, \alpha_1\beta_2, \ldots, \alpha_i\beta_j, \ldots, \alpha_m\beta_k >=< \alpha_1'\beta_1', \alpha_1'\beta_2', \ldots, \alpha_i'\beta_j', \ldots, \alpha_m'\beta_k' > .$$

6.40 Prove that the sum of fractional ideals is again a fractional ideal.

6.41 Express the symmetric polynomial $f(x_1, x_2, x_3) = x_1^3 + x_2^3 + x_3^3$ as a polynomial in the elementary symmetric polynomials s_1, s_2, s_3.

6.42 Find the minimal polynomial of $\sqrt{2} + \sqrt{3}$ over $\mathbb{Q}$. (How do you know its algebraic?) (Hint: $\mathbb{Q}(\sqrt{2}, \sqrt{3})$ has degree 4 over $\mathbb{Q}$ and hence $\sqrt{2} + \sqrt{3}$ has degree 2 or degree 4 over $\mathbb{Q}$. Show that it cannot have degree 2).

6.43 Let p be a prime and θ a rational number not a pth power. Let $K = \mathbb{Q}(\theta^{\frac{1}{p}})$. Show that if K_1 is a field with $\mathbb{Q} \subset K_1 \subset K$ then either $K_1 = \mathbb{Q}$ or $K_1 = K$.

6.44 Let $\alpha_1, \ldots, \alpha_n$ be algebraic integers in K. Show that if $\alpha_1, \ldots, \alpha_n$ is a basis for K over $\mathbb{Q}$ and $\Delta(\alpha_1, \ldots, \alpha_n)$ is squarefree then $\alpha_1, \ldots, \alpha_n$ is an integral basis.

6.45 Let α, β be algebraic integers in K and $< \alpha >$, $< \beta >$ the principal ideals they generate. Show that if $< \alpha > \mid < \beta >$ then $\alpha | \beta$.

6.46 Classify the algebraic number fields K with discriminant

$$-100 \leq d_K \leq 100.$$

Chapter 7
The Fields $\mathbb{Q}_p$ of p-Adic Numbers: Hensel's Lemma

7.1 The p-Adic Fields and p-Adic Expansions

In the previous chapter, we described algebraic extensions of the rational numbers. We then saw that the arithmetic of the integers within these algebraic number fields was similar to that of the ordinary integers and further that many algebraic number fields allowed unique factorization while all these fields allowed unique factorization in terms of ideals.

In this chapter, we look at a separate type of extension of the rational field motivated by both analysis and algebra. For each prime p, we will get a new field called the **field of p-adic numbers** denoted by $\mathbb{Q}_p$. These fields will be constructed in a manner analogous to the way the real number system $\mathbb{R}$ is constructed from $\mathbb{Q}$. The p-adic numbers can be used to consider and study congruences modulo p and modulo p^n and have many applications in classical number theory. In particular they were used in the proof of Fermat's last theorem by A. Wiles (see [W]).

The p-adic numbers were first developed by Kurt Hensel in 1897 and for each prime p they can be considered as a completion of the rational numbers. To understand this, let us recall some facts about the real number system. We will go deeply into these in the next section. The real numbers have the property that every Cauchy sequence (see Section 7.2) of real numbers has a limit. This is not true for the rational numbers. Because of this we say that the real numbers are **complete**. Further each real number is actually the limit of a sequence of rationals. We say that $\mathbb{Q}$ is **dense** in $\mathbb{R}$ and that $\mathbb{R}$ is the **completion** of $\mathbb{Q}$.

Convergence of sequences in $\mathbb{R}$ and $\mathbb{Q}$ depends upon measuring distance. For the standard approach we measure distance in terms of absolute value, that is if $r, s \in \mathbb{R}$ then $d(r, s) = |r - s|$. We say that absolute value is a **norm** on the field $\mathbb{R}$ and the reals are a **normed field**. What Hensel and others noticed is that the completion of $\mathbb{Q}$ can be carried out for any normed field. A norm can be placed on $\mathbb{Q}$ depending on a given prime p and the resulting normed field can be completed as was $\mathbb{R}$. The resulting field is the field of p-adic numbers. The actual details will be given in Section 7.2.

B. Fine and G. Rosenberger, *Number Theory*,
DOI 10.1007/978-3-319-43875-7_7

To do p-adic arithmetic we must recall the p-ary expansion of real numbers. Any real number r can be expressed as a decimal expansion

$$r = \sum_{i=-\infty}^{n} a_i 10^i$$

where $a_i \in \{0, 1, 2, 3, 4, 5, 6, 7, 8, 9\}$ and there are finitely many decimal places to the left of the decimal point and possibly infinitely many to the right.

Although in common practice we use a decimal expansion, that is base 10, in reality any base $m \in \mathbb{N}$ can be used. Historically, we use base 10 because we have 10 fingers or digits. We have the theorem:

Theorem 7.1.1 *Let $r \in \mathbb{R}$ and $m \in \mathbb{N}$ with $m \geq 2$. Then, r can be expressed as*

$$r = \sum_{i=-\infty}^{n} a_i m^i$$

where $a_i \in \{0, 1, \ldots, m-1\}$.

The expansion in Theorem 7.1.1 is called the m-**ary expansion**. We give an example for an integer in base 5.

EXAMPLE 7.1.1 Determine the 5-ary expansion of 371.

The method uses the division algorithm and is related to the Euclidean algorithm. We first consider the highest power of 5 that is less than 371. This is $5^3 = 125$. We then use the division algorithm to obtain

$$371 = (2)(125) + 121 = 2 \cdot 5^3 + 121.$$

We now repeat the process with 121 to obtain

$$371 = 2 \cdot 5^3 + 4 \cdot 5^2 + 21 = 2 \cdot 5^3 + 4 \cdot 5^2 + 4 \cdot 5 + 1.$$

This gives the 5-ary expansion. We write this as $(2441)_5$. Writing 371 without the base indicates the standard base 10.

Arithmetic can be done exactly as in standard decimal expansions but carries must be done modulo m.

EXAMPLE 7.1.2 Add the numbers $(2441)_5$ and $(3244)_5$

Here, we write

$$\begin{array}{rrrrr} & 2 & 4 & 4 & 1 \\ & 3 & 2 & 4 & 4 \\ \\ 1 & 1 & 2 & 4 & 0 \end{array}$$

In base 10 this is $1 \cdot 5^4 + 1 \cdot 5^3 + 2 \cdot 5^2 + 4 \cdot 5 + 0 = 820$. In base 10 $(2441)_5 = 371$ and $(3244)_5 = 449$. We have $371 + 449 = 820$ and, as they must, the additions agree.

Base 2 expansions are called **binary expansions**. Because these only use two digits, 0, 1, binary expansions become extremely important in representing numbers on a computer. The digit 0 can be expressed as an open electrical gate while the digit 1 by a closed gate. Thus, any integer can be expressed as a sequence of open and closed circuits.

7.2 The Construction of the Real Numbers

The construction of the p-adic fields is entirely analogous to the construction of the real numbers from the rational numbers. What differs is the way distance is measured. We first describe the construction of $\mathbb{R}$.

7.2.1 The Completeness of Real Numbers

There are several different constructions of the real number system $\mathbb{R}$ starting with the rational numbers $\mathbb{Q}$. The two best known are the **Dedekind cut construction** and the **Cauchy completion procedure**. For our purposes in studying the p-adic fields the second is the most important. We recall first some basic facts about sequences and completeness in $\mathbb{R}$ and then in the next subsection show how $\mathbb{R}$ can be constructed starting from $\mathbb{Q}$.

The analytic properties of the real numbers depend upon **distance** which in turn depends upon **absolute value**. Recall that if $x \in \mathbb{R}$ then its **absolute value** is defined by

$$|x| = \begin{cases} x & \text{if } x \geq 0 \\ -x & \text{if } x < 0. \end{cases}$$

Lemma 7.2.1 *We have the following properties for absolute value:*

1. $|x| \geq 0$ *and* $|x| = 0$ *iff* $x = 0$
2. $|xy| = |x||y|$
3. $|x + y| \leq |x| + |y|$ *(triangle inequality)*

Absolute value forms a norm on $\mathbb{R}$ and we say that $\mathbb{R}$ is a **normed field**.

Absolute value allows us to define distance on $\mathbb{R}$. In particular, if $x, y \in \mathbb{R}$ then $d(x, y) = |x - y|$. This then satisfies the common properties of a **metric**

1. $d(x, y) \geq 0$ and $d(x, y) = 0$ if and only if $x = y$
2. $d(x, y) = d(y, x)$
3. $d(x, y) \leq d(x, z) + d(z, y)$(triangle inequality).

The completion of $\mathbb{Q}$ depends upon the convergence of sequences.

Definition 7.2.1 *A sequence (x_n) in $\mathbb{R}$* **converges** *or has a* **limit** *$x \in \mathbb{R}$ denoted $x_n \to x$ if for all $\epsilon > 0$ there exists an $N = N(\epsilon)$ such that $|x_n - x| < \epsilon$ for all $n \geq N$. If a sequence has no limit we say it* **diverges**.

The following is clear but important.

Lemma 7.2.2 *The limit of a sequence is unique; that is if* $\lim x_n = x$ *and* $\lim x_n = y$ *then $x = y$.*

Crucial to our construction are **Cauchy sequences**.

Definition 7.2.2 *A sequence (x_n) is a* **Cauchy sequence** *if for all $\epsilon > 0$ there exists an $N = N(\epsilon)$ such that $|x_n - x_m| < \epsilon$ for all $n, m \geq N$. This means the terms of the sequence cluster close to each other after a certain point.*

Roughly in a convergent sequence all the terms of the sequence after a certain point are close to the limit. In a Cauchy sequence, all the terms after a certain point are close to each other. Clearly a convergent sequence must be a Cauchy sequence. However within the rationals there are Cauchy sequences that do not converge within $\mathbb{Q}$ as the next example shows.

EXAMPLE 7.2.1.1 Consider $x = \sqrt{2}$. This is an irrational number so it has a non-repeating infinite decimal expansion

$$x = \sqrt{2} = 1.414 \cdots .$$

Let $x_1 = 1, x_2 = 1.4$ and in general x_n the $(n - 1)$-st decimal approximation of $\sqrt{2}$. Now (x_n) is a sequence of rational numbers. Within $\mathbb{R}$ we have $\lim x_n = x$ so within $\mathbb{R}$ the sequence (x_n) is a convergent sequence and hence a Cauchy sequence. Since distance is measured the same in $\mathbb{Q}$ as in $\mathbb{R}$ this is also a Cauchy sequence in $\mathbb{Q}$. However there is no limit within $\mathbb{Q}$ since limits of sequences are unique and $x \notin \mathbb{Q}$.

Within the real number system though the following theorem is true.

Theorem 7.2.1 *A sequence in $\mathbb{R}$ converges if and only if it is a Cauchy sequence.*

As we will see in the next section, this theorem is a direct consequence of the construction of $\mathbb{R}$ starting with $\mathbb{Q}$. In most analysis courses a proof of this theorem depends on the least upper bound property, which we introduce below.

Because of the above theorem, we say that $\mathbb{R}$ is **complete**. Geometrically the completeness of $\mathbb{R}$ is essentially equivalent to the fact that $\mathbb{R}$ is in 1-1 correspondence with the points on a line. Further, convergence and completeness allow us to define and study all the **analytic properties** of functions—continuity, differentiability, and integrability.

The real numbers $\mathbb{R}$ are an **ordered field**. That is, on $\mathbb{R}$ there is an ordering such that if $a, b \in \mathbb{R}$ then either $a < b$ or $a = b$ or $a > b$. The common properties of $\mathbb{R}$ can be defined for any ordered field.

Lemma 7.2.3 *In any ordered field, F, squares must be positive, that is, $x^2 > 0$ for all $x \neq 0$. In particular in $\mathbb{R}$ the equation $x^2 + 1 = 0$ has no solution.*

Proof Let $x \in \mathbb{R}$ with $x \neq 0$. The either $x > 0$ or $-x > 0$. If $x > 0$ then $x^2 = xx > 0$ since the positive elements are closed under multiplication. If $x < 0$ then $-x > 0$ and $x^2 = xx = (-x)(-x)$ by the laws of signs. But then $(-x)(-x) > 0$ and hence $x^2 > 0$. □

Definition 7.2.3 *Let F be any ordered field and $S \subset F$.*

(1) Then, S is **bounded** *if there exist $a, b \in F$ with $a \leq s \leq b$ for all $s \in S$. The element b is called an* **upper bound** *for S and the element a is called a* **lower bound** *for S. An element $b \in F$ is a* **least upper bound** *or LUB for S if b is an upper bound for S and if b_1 is another upper bound for S then $b \leq b_1$.*

(2) Suppose $a, b \in F$. Then the closed interval with endpoints a, b is the set

$$[a, b] = \{x \in F; a \leq x \leq b\}.$$

Note that by reversing all the inequalities in the above definition, we could also define the greatest lower bound for S or GLB. The GLB is not necessary for our discussions.

Definition 7.2.4 *(1) An ordered field F satisfies the* ***least upper bound property*** *(the LUB property) if every nonempty subset $S \subset F$ which has an upper bound in F also has a least upper bound in F.*

(2) An ordered field F satisfies the ***nested intervals property*** *if whenever ($I_n = [a_n, b_n] \subset F$ where $a_n \leq b_n$ for all n) is a sequence of nested closed intervals ($I_{n+1} \subset I_n$) whose lengths go to zero then there exists a unique point in F common to all the intervals, that is, $\bigcap_n I_n \in F$.*

The key result on the completeness of $\mathbb{R}$ is that these properties are equivalent and further equivalent to the fact that Cauchy sequences converge.

Theorem 7.2.2 *Let F be an ordered field. Then the following are equivalent*

(1) F satisfies the LUB property.

(2) F satisfies the nested intervals property.

(3) Every Cauchy sequence in F actually converges.

Definition 7.2.5 *An ordered field is* ***complete*** *if it satisfies any (and hence all) of the properties in the last theorem.*

We then have:

Theorem 7.2.3 *The real number field $\mathbb{R}$ is a complete ordered field.*

7.2.2 The Construction of $\mathbb{R}$

As we mentioned there are several constructions that arrive at the reals $\mathbb{R}$ beginning with the rationals $\mathbb{Q}$. In this section, and most relevant to the construction of the p-adic numbers, we describe a construction known as **Cauchy completion**. We present the proofs for $\mathbb{R}$ starting with $\mathbb{Q}$ but these proofs are entirely general for any ordered field and will refer back to them when we construct the p-adic fields.

Cauchy completion is a general procedure to embed an incomplete metric space M as a dense subset of a complete metric space $\overline{M}$. The complete metric space $\overline{M}$ is called the **Cauchy completion** of M. We explain these terms which are in essence generalizations of properties of the reals.

Definition 7.2.6 *A* **metric space** *is a set M with a* **distance function** *on it, that is, a function $d : M \times M \to \mathbb{R}$ satisfying*

(1) $d(x, y) \geq 0$ and $d(x, y) = 0$ iff $x = y$;

(2) $d(x, y) = d(y, x)$;

(3) $d(x, y) \leq d(x, z) + d(z, y)$ (triangle inequality).

The rational numbers $\mathbb{Q}$ and the real numbers $\mathbb{R}$ are metric spaces where $d(x, y) = |x - y|$.

In any metric space we can define sequences, convergence and Cauchy sequences exactly as in the real numbers. In general we say that a metric space M is **complete** if every Cauchy sequence in M converges to an element of M.

A subset S in a metric space M is **dense** in M if given any $x \in M$ and real number $\epsilon > 0$ there is a $s \in S$ with $d(x, s) < \epsilon$. This means that any point in M is arbitrarily close to a point in S. This is equivalent to the fact that given $x \in S$ there exists a sequence $(x_n) \subset M$ whose limit is x. For example the rationals are dense in the reals (see the exercises).

Notice that the equivalence of Cauchy sequence completeness to the least upper bound property, that holds in an ordered field, does not necessarily hold in a general metric space. In a metric space we may not have any order.

Starting with the rationals $\mathbb{Q}$, we want to construct an ordered field F which is the completion of the rationals with respect to absolute value distance. That is, we want to construct a field $\mathbb{R}$ such that $\mathbb{Q} \subset \mathbb{R}$ and $\mathbb{R}$ is complete as a metric space. Further $\mathbb{Q}$ is a dense subset of $\mathbb{R}$.

The **Cauchy completion** of $\mathbb{Q}$ proceeds in the following manner:

Step (1): Consider the set $\overline{\mathbb{Q}}$ of all Cauchy sequences of rationals. That is, an element of $\overline{\mathbb{Q}}$ is a Cauchy sequence $(q_1, q_2, \dots)$ of rational numbers. Define on $\overline{\mathbb{Q}}$ the relation

$$(q_1, q_2, \dots) \sim (s_1, s_2, \dots) \text{ iff } \lim(q_i - s_i) = 0.$$

That is, after some index i, the two sequences get arbitrarily close.

Lemma 7.2.4 *This defines an equivalence relation on $\overline{\mathbb{Q}}$.*

We leave the proof of this lemma to the exercises.

Step (2): Let $\mathbb{R}$ be the set of all equivalence classes of Cauchy sequences of rationals under the equivalence relation above. We now want to show five things:

(1) $\mathbb{R}$ is an ordered field.
(2) $\mathbb{Q} \subset \mathbb{R}$.
(3) $\mathbb{R}$ is a metric space.
(4) $\mathbb{Q}$ is dense in $\mathbb{R}$.
(5) $\mathbb{R}$ is complete.

Step (3): We have the following theorem.

Theorem 7.2.4 $\mathbb{R}$ *is an ordered field.*

Proof To show that $\mathbb{R}$ is a field we have to show that we can define addition, additive inverses, multiplication and multiplicative inverses to satisfy the field axioms. To prove this we need to know that Cauchy sequences are bounded so we prove this first.

Lemma 7.2.5 *If* $x = x_1, x_2, \dots$ *is a Cauchy sequence then* (x_n) *is bounded, that is, there is a* $B > 0$ *with* $|r_n| \leq B$ *for all* n.

Proof Let $\epsilon = 1$. Then since (x_n) is a Cauchy sequence it follows that there exists an N such that $|x_n - x_m| < 1$ for all $n, m \geq N$. In particular if $n > N$ we have $|x_n - x_N| < 1$. Then if $n \geq N$ we have

$$|x_n| = |x_n - x_N + x_N| \leq |x_n - x_N| + |x_N| < |x_N| + 1.$$

Now let $B = \max\{|x_1|, \dots, |x_N|, |x_N| + 1\}$. Then from the above $|x_n| \leq B$ for all n. □

Now let $r, s \in \mathbb{R}$, i.e., r and s are equivalence classes of Cauchy sequences. So let $r = [(q_1, \dots, q_n, \dots)]$ and $s = [(t_1, \dots, t_n, \dots)]$ be the equivalence classes of Cauchy sequences of rationals (q_n) and (t_n), respectively. We define

$$r \pm s = [(q_1 \pm t_1, \dots, q_n \pm t_n, \dots)], \tag{7.1}$$

$$r \cdot s = [(q_1 t_1, \dots, q_n t_n, \dots)]. \tag{7.2}$$

For this to make sense, we have to show that $(q_n \pm t_n)$ and $(q_n t_n)$ are again Cauchy sequences and that addition and multiplication of these equivalences classes are independent of the equivalence class representative chosen. That is, $+$ and $\times$ defined this way are well-defined. Here we will show that multiplication of equivalence classes is well-defined and leave all the other verifications to the exercises. For this purpose, suppose that $(q_n) \sim (q_n^{'})$ and $(t_n) \sim (t_n^{'})$. Then we have $\lim(q_n - q_n^{'}) = \lim(t_n - t_n^{'}) = 0$. We must show that

$$\lim(q_n t_n - q_n^{'} t_n^{'}) = 0.$$

But all sequences here are Cauchy, hence they are all bounded. In particular, there exists M_1 and there exists M_2 such that $|t_n| \leq M_1$ and $\left|q_n^{'}\right| \leq M_2$ for all $n \in \mathbb{N}$.

Now let $\epsilon > 0$. Since $q_n - q_n^{'} \to 0$ and $t_n - t_n^{'} \to 0$, there exists N_1 and N_2 such that $\left|q_n - q_n^{'}\right| < \frac{\epsilon}{2M_1}$ for all $n \geq N_1$ and $\left|t_n - t_n^{'}\right| < \frac{\epsilon}{2M_2}$ for all $n \geq N_2$. Taking $N = \max\{N_1, N_2\}$. We have, using properties of absolute values, that for all $n \geq N$

$$\left|q_n t_n - q_n^{'} t_n^{'}\right| = \left|(q_n t_n - q_n^{'} t_n) + (q_n^{'} t_n - q_n^{'} t_n^{'})\right| \tag{7.3}$$

$$\leq \left|q_n - q_n^{'}\right| |t_n| + \left|t_n - t_n^{'}\right| \left|q_n^{'}\right| \tag{7.4}$$

$$< \frac{\epsilon}{2M_1} M_1 + \frac{\epsilon}{2M_2} M_2 = \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon. \tag{7.5}$$

This shows that $(q_n t_n - q_n^{'} t_n^{'}) \to 0$. (The other verifications are done in a similar manner.)

Clearly $[(0, 0, \ldots)]$ and $[(1, 1, \ldots)]$ are additive and multiplicative identities, respectively. For this to make sense, $(0, 0, \ldots)$ and $(1, 1, \ldots)$ must be Cauchy sequences of rationals. The properties of commutativity, associativity, and distributivity follow from the fact that these are true in $\mathbb{Q}$. The additive inverse of an equivalence class $r \in \mathbb{R}$, is if $r = [(q_n)]$, defined as $-r = [(-q_n)]$. It is clear that if (q_n) is a Cauchy sequence of rationals, then so is $(-q_n)$. It follows that $[(-q_n)]$ makes sense an element of $\mathbb{R}$. Thus far we have shown that $\mathbb{R}$ is a commutative ring with unity.

It remains to show that every nonzero element of $\mathbb{R}$ has a multiplicative inverse. If $r = [(q_n)]$ is an equivalence class of a Cauchy sequence and $r \neq 0$ then $\lim\ q_n \neq 0$. We leave it as an exercise to show then that there exits a N such that for all $n \geq N$ we have that $q_n \neq 0$. Therefore it makes sense to define $\frac{1}{r} = [(0, 0, \ldots, 0, \frac{1}{q_n}, \frac{1}{q_{n+1}}, \ldots)]$. We need to show that the sequence $(0, 0, \ldots, 0, \frac{1}{q_n}, \frac{1}{q_{n+1}}, \ldots)$ is a Cauchy sequence. This is also left as an exercise. Also note that $r \cdot \frac{1}{r} = [(0, \ldots, 0, 1, 1, \ldots)] = [(1, 1, \ldots, 1, 1, \ldots)]$. Thus we have shown that the set $\mathbb{R}$ with these operations is indeed a field.

Given $r = [(q_1, \ldots, q_n, \ldots)] \in \mathbb{R}$, then we define $r > 0$ or r is positive to mean $r \neq 0$, that is, it is not equal to the equivalence class $[(0, 0, \ldots, 0, \ldots)]$) and $r = [(q_n)]$ for some Cauchy sequence of rationals such that $\lim\ q_n \neq 0$ and there exists an N such that $q_n > 0$ for all $n \geq N$. Again since $r > 0$ was defined on equivalence classes (r is an equivalence class), we must show this is well-defined. We leave this verification to the exercises. If $r, s \in \mathbb{R}$, define $r > s$ to mean $r - s > 0$. This defines an order on $\mathbb{R}$ and hence $\mathbb{R}$ is an ordered field. Again we leave the details to the exercises.

Step (4): We now show that $\mathbb{Q} \subset \mathbb{R}$.

Theorem 7.2.5 $\mathbb{Q} \subset \mathbb{R}$. *More precisely, we can embed* $\mathbb{Q}$ *as a subring of* $\mathbb{R}$.

Proof To each $q \in \mathbb{Q}$ associate the sequence $(q, q, q, q, \dots)$. This is clearly a Cauchy sequence of rationals. Hence, $[(q, q, q, q, \dots)] \in \mathbb{R}$. Note that $(q, q, q, q, \dots) \sim (\dots, q, q, q, \dots)$. So that $[(q, q, q, q, \dots.)] = [(\dots, q, q, q, \dots)]$. Consider the map $q \mapsto (q, q, q, q, \dots.)$. This mapping embeds $\mathbb{Q}$ into $\mathbb{R}$ and hence we can consider $\mathbb{Q}$ as a subset of $\mathbb{R}$. □

Note that this theorem implies that when we talk about a *rational* number $\overline{q}$ in $\mathbb{R}$ what is meant is $\overline{q} = [(\dots, q, q, q, \dots)]$ where q is a rational number.

Step (5): We must show that $\mathbb{R}$ is a metric space.

Theorem 7.2.6 *$\mathbb{R}$ is a metric space.*

Proof To make $\mathbb{R}$ a metric space we define an absolute value on $\mathbb{R}$ and then use this to define distance by $d(r, s) = |r - s|$. If $r = [(q_1, q_2, \dots)] \in \mathbb{R}$ then we define $|r| = [(|q_1|, |q_2|, \dots)]$. We must first show that $(|q_1|, |q_2|, \dots)$ is again a Cauchy sequence of rationals. To see this, consider any $\epsilon > 0$ then there exists an $N = N(\epsilon)$ such that for all $n, m \geq N$ we have $|q_n - q_m| < \epsilon$. This must be true since $(q_1, q_2, \dots)$ is a Cauchy sequence. But now

$$|q_n| = |(q_n - q_m) + q_m| \leq |q_n - q_m| + |q_m| \Longrightarrow$$

$$|q_n| - |q_m| \leq |q_n - q_m|.$$

Similarly, we can get $|q_m| - |q_n| \leq |q_m - q_n| = |q_n - q_m|$. But this implies $|q_n| - |q_m| \geq -|q_n - q_m|$. Combining this with the display above, gives

$$-|q_n - q_m| \leq |q_n| - |q_m| \leq |q_n - q_m| \Longrightarrow ||q_n| - |q_m|| \leq |q_n - q_m|.$$

But this implies from above that $||q_n| - |q_m|| < \epsilon$ for all $n, m \geq N$. Thus $(|q_1|, |q_2|, \dots)$ is a Cauchy sequence of rationals and hence $|r| = [(|q_1|, |q_2|, \dots)]$ is in $\mathbb{R}$. We also need to show that this definition of absolute value is well-defined because it was defined on equivalence classes. This is not hard using the inequality proved above and so is left for the exercises. It is also not hard to show that $|r|$ satisfies the usual absolute value properties (see the exercises). Therefore $d(x, y) = |x - y|$ defines a metric on $\mathbb{R}$. □

Lemma 7.2.6 *If $x, y \in \mathbb{R}$ and $\overline{\omega}$ is any rational number in $\mathbb{R}$, then $|x - y| < \overline{\omega}$ means that if $x = [(x_n])$, $y = [(y_n)]$, and $\overline{\omega} = [(\omega, \omega, \dots)]$, there exists $N = N(\omega)$ such that $|x_n - y_n| < \omega$ for all $n \geq N$.*

Proof By the above definitions, $|x - y| = [(|x_n - y_n|)] < \overline{\omega}$ means that $\overline{\omega}- |x - y| > 0$. Thus we must have that $\overline{\omega}- |x - y| = [(a_n)]$ where (a_n) is a Cauchy sequence of rationals such that there exists N with $a_n > 0$ for all $n \geq N$. But $\overline{\omega}- |x - y| = |(\omega - |x_n - y_n|)|$ Since (x_n) and (y_n) are Cauchy so is $\overline{\omega}- |x - y|$. Given any $\epsilon > 0$ there exists N_1 and there exists N_2 such that $|x_n - x_m| < \epsilon/2$ and $|y_n - y_m| < \epsilon/2$ for all $n, m \geq N_1$ and $n.m \geq N_2$. So that for if $n, m \geq \max\{N_1, N_2\}$, then

$|\omega - |x_n - y_n| - (\omega - |x_m - y_m|| = ||x_m - y_m| - |x_n - y_n||$ by the inequality established in the proof of the theorem above gives that this is $\leq \big|(x_m - x_n) + (y_n - y_{m)}\big|$ which by the triangle inequality is $\leq \epsilon/2 + \epsilon/2 = \epsilon$. Thus $(\omega - |x_n - y_n|)$ is a Cauchy sequence of rationals and so it does make sense to consider the real number given by its equivalence class. But our given inequality would imply that $|(\omega - |x_n - y_n|)| > 0$, which in turn by definition means there exists $N = N(\omega)$ such that $|x_n - y_n| < \omega$ for all $n \geq N$. □

Step (6): We must show that $\mathbb{Q}$ is dense in $\mathbb{R}$.

Theorem 7.2.7 *$\mathbb{Q}$ is a dense subset of $\mathbb{R}$.*

Proof To prove this we have to show that any real number is arbitrarily close to a rational number. Here, we must be careful about what we mean by a *rational* number and what we mean by *arbitrarily* close. A *rational* number is an equivalence class of a sequence where the elements of the sequence are all just one and the same rational number (or at least are eventually that). When we say this rational is *arbitrarily* close to a real number, we mean that we can make the distance between these two real numbers as defined above less than any preassigned positive rational. (Also as defined above.) So let $r = [(q_1, q_2, q_3, \dots)] \in \mathbb{R}$ and suppose that $\omega > 0$ is any (small) rational number—not an equivalence class yet! We need to show that there exists a rational number (here an equivalence class) call it $\overline{q}$ such that $|r - \overline{q}| < \overline{\omega}$ where $\overline{\omega} = [(\dots, \omega, \omega, \omega, \dots)]$. (Note that $\overline{\omega} > 0$ by our definition of the ordering on $\mathbb{R}$.) Since (q_n) is a Cauchy sequence there exists an $N = N(\omega)$ such that $|q_n - q_m| < \omega$ for all $n, m \geq N$. Choose a particular k with $k \geq N$ and set $\overline{q} = [(\dots, q_k, q_k, q_k, \dots)]$. This is an equivalence class of a Cauchy sequence of rationals and so $\overline{q} \in \mathbb{R}$. By the embedding above, we associate $\overline{q}$ with the rational number, q_k. Now $|r - \overline{q}| = [(\dots, |q_n - q_k|, |q_{n+1} - q_k|, \dots)]$. Now using the above Lemma, since for $n \geq N$ $|q_n - q_k| < \omega$, we have that $|r - \overline{q}| < \overline{\omega}$. □

Step (7): Finally, we must show that $\mathbb{R}$ is complete. Here we show completeness by Cauchy sequences but since we have shown that $\mathbb{R}$ is an ordered field this is equivalent to the LUB property and the nested intervals property.

Theorem 7.2.8 *$\mathbb{R}$ is complete.*

Proof To prove this we have to show that any Cauchy sequence of real numbers converges to a real number. Let $r_1, r_2, \dots, r_n, \dots$ be a Cauchy sequence of reals. We show that it has a limit which is a real number. Realize that each r_i is itself an equivalence class of Cauchy sequences of rationals. For each n choose a rational $\overline{q_n} = [(\dots, q_n, q_n, \dots)]$, that is, rational in the sense of our embedding, such that $|r_n - \overline{q_n}| < \overline{\frac{1}{n}}$ where $\overline{\frac{1}{n}} = [(\dots, \frac{1}{n}, \frac{1}{n}, \dots)]$. This can be done since $\mathbb{Q}$ is dense in $\mathbb{R}$. Consider the sequence of rationals $(q_1, q_2, q_3, \dots, q_n, \dots)$. We claim this sequence is a Cauchy sequence. Fix $\epsilon > 0$ a (small) positive rational not an equivalence class. Choose $N \in \mathbb{N}$ such that $1/N < \epsilon/3$. Since (r_n) is Cauchy there exists $\dot{M} > 0$ such that $n, m \geq M \implies |r_n - r_m| < \overline{\epsilon/3}$ (here $\overline{\epsilon/3}$ =$[(\dots, \epsilon/3, \epsilon/3, \dots)]$) and $n \geq N$

$$|\overline{q_n} - \overline{q_m}| \leq |\overline{q_n} - r_n| + |r_n - r_m| + |r_m - \overline{q_m}|$$

by the triangle inequality. Thus, we have for $n, m \geq \max\{M, N\}$, $|\overline{q_n} - \overline{q_m}| < \overline{\epsilon}$. (Here $\overline{\epsilon} = [(\dots, \epsilon, \epsilon, \dots)]$.) This means that $|q_n - q_m| < \epsilon$ for all $n, m \geq \max\{M, N\}$. Thus (q_n) is a Cauchy sequence of rationals. So it makes sense to consider the real number $r = [(q_n)] \in \mathbb{R}$. Further, we claim that $r_n \to r$. By construction, $\overline{q_n} \to r_n$. But the fact that $\overline{q_n} \to r = [(q_n)]$ just follows from (q_n) being a Cauchy sequence. For we need, $|r - \overline{q_m}| = |[(q_n)] - \overline{q_m}|$ to be made small for n, m sufficiently large. But Lemma 7.2.6 says this is true if $|q_n - q_m|$ can be made small for n, m sufficiently large. This is precisely what it means for (q_n) to be a Cauchy sequence. Thus we have proven $\lim(\overline{q_n} - r_n) = 0$ and $\lim(r - \overline{q_n}) = 0$. So that $\lim(r - r_n) = \lim((r - \overline{q_n}) + (\overline{q_n} - r_n)) = 0$. Thus the Cauchy sequence $r_1, r_2, \dots, r_n, \dots$ has the limit r showing that $\mathbb{R}$ is complete. □

Having completed the **Cauchy completion** of $\mathbb{Q}$, we no longer consider $\mathbb{R}$ to be a set of equivalence classes of Cauchy sequences of rationals.

7.2.3 *The Characterization of* $\mathbb{R}$

Before moving on to the p-adic numbers, we provide a complete algebraic characterization of the reals. We need one additional property besides completeness. First note that an ordered field must have characteristic zero and hence contains a subring isomorphic to the integers.

Definition 7.2.7 *An ordered field F is* **archimedean** *if for any pair $f_1, f_2 \in F$ with $f_2 > f_1 > 0$ there exists an $n \in \mathbb{N}$ such that $nf_1 > f_2$.*

The complete characterization of $\mathbb{R}$ is then given by completeness together with the archimedean property.

Theorem 7.2.9 *$\mathbb{R}$ is a complete archimedean ordered field. Further, any other complete archimedean ordered field is isomorphic to $\mathbb{R}$.*

7.3 Normed Fields and Cauchy Completions

The real numbers $\mathbb{R}$ are a completion of the rationals $\mathbb{Q}$ and are characterized as the unique (up to isomorphism) complete archimedean ordered field. The question arises as to whether there are other completions of the rationals. The answer is yes but they must be, by necessity, non-archimedean, and further are of a very special type. Notice that the construction of $\mathbb{R}$ from $\mathbb{Q}$ used the absolute value prominently and Cauchy sequences and denseness were in terms of this distance. For the additional completions of $\mathbb{R}$ we must define different distance functions on $\mathbb{Q}$. We do this in general.

Definition 7.3.1 *A* **norm** *on a field F is a function $|\ |: F \to \mathbb{R}$ satisfying*

(1) $|x| \geq 0$ for all $x \in F$,

(2) $|x| = 0$ if and only if $x = 0$,

(3) $|xy| = |x||y|$ for all $x, y \in F$,

(4) $|x + y| \leq |x| + |y|$ for all $x, y \in F$ (triangle inequality).

A **normed field** *is a field F with a norm.*

For example, $\mathbb{Q}$ and $\mathbb{R}$ are normed fields with the usual absolute value. Any normed field F is a metric space under $d(x, y) = |x - y|$. Since a normed field F is a metric space the concepts of convergence, Cauchy sequence, completeness and denseness of subsets are all defined on F. As before we say that a normed field is **complete** if every Cauchy sequence within the field converges to an element in the field, that is within the field F the concepts of Cauchy sequence and convergent sequence coincide.

The basic result is that given any normed field F it can be embedded as a dense subset of a complete ordered field $\overline{F}$. The complete ordered field obtained in this manner is called the **Cauchy completion** of F.

Theorem 7.3.1 *Given an ordered field F then there exists a complete ordered $\overline{F}$ for which F is a dense subfield. The field $\overline{F}$ is called the* **Cauchy completion** *of F.*

The proof of Theorem 7.3.1 is identical to the proof that $\mathbb{R}$ can be constructed from $\mathbb{Q}$. That proof used only the absolute value properties which are the general norm properties. To construct $\overline{F}$ from F we follow exactly the same steps as in Section 7.2.3. We let $\overline{F}$ be the set of Cauchy sequences from F under the equivalence relation that two Cauchy sequences are equivalent if their differences go to zero. We then show that $\overline{F}$ is a complete ordered field and that F is a dense subset of $\overline{F}$. We leave the details to the exercises.

7.4 The p-Adic Fields

Considering $\mathbb{R}$ as the completion of $\mathbb{Q}$ depended upon absolute value as the norm on $\mathbb{Q}$. The question arose as to whether $\mathbb{Q}$ could be completed in any other way. The answer is yes but it requires a completely different norm on the rationals. As we saw in Theorem 7.2.9 the reals are characterized as a complete archimedean ordered field. Hence, if we are to complete $\mathbb{Q}$ relative to a different norm this norm must not be non-archimedean. Before describing this new norm (actually infinitely many new norms) on $\mathbb{Q}$ we discuss some properties of norms in general.

Since the completion of a normed field depends on Cauchy sequences we consider two norms to be equivalent if they give rise to exactly the same Cauchy sequences.

Definition 7.4.1 *Two norms on a normed field F are* **equivalent** *if their induced metrics are equivalent. That is $|\ |_1$ is equivalent to $|\ |_2$ if a sequence is Cauchy with respect to one metric if and only if it is Cauchy with respect to the other.*

The next result gives a condition for equivalence of norms.

Theorem 7.4.1 *Two norms $|\ |_1$ and $|\ |_2$ on a normed field F are equivalent if and only if there exists an $\alpha > 0$ such that*

$$|x|_2 = |x|_1^\alpha$$

for all $x \in F$

Proof Suppose that $|x|_2 = |x|_1^\alpha$ for all $x \in F$ and suppose that (x_n) is a Cauchy sequence relative to the first norm. Given $\epsilon > 0$ and N be found for $\epsilon^{1/\alpha}$. Then, for $m, n > N$ we have $|x_n - x_m|_1 < \epsilon^{1/\alpha}$ so that $|x_n - x_m|_2 < \epsilon$. Therefore (x_n) is a Cauchy sequence relative to the second norm and the two norms are equivalent.

Conversely, suppose the two norms are equivalent. Choose an $a \in F$ with $|a|_1 < 1$. This is possible since we have a nontrivial norm. Then, let

$$\alpha = \frac{\log(|a|_2)}{\log(|a|_1)}.$$

It follows that $|a|_2 = (|a|_1)^\alpha$. We show this is true for all $x \in F$. We show this for $|x|_1 < 1$. The other cases follow the same argument.

Consider the set $S = \{r = \frac{m}{n}, m, n \in \mathbb{N}; (|x|_1)^r < |a|_1\}$. Then for any $r \in S$ we have $(|x|_1)^m < (|a|_1)^n$ so that $|\frac{x^m}{a^n}|_1 < 1$. But then $|\frac{x^m}{a^n}|_2 < 1$ and so $(|x|_2)^m < (|a|_2)^n$ and therefore $(|x|_2)^r < |a|_2$. The same argument with the $|\ |_2$ replacing $|\ |_1$ shows that for the same S we have

$$S = \{r = \frac{m}{n}, m, n \in \mathbb{N}; (|x|_2)^r < |a|_2\}.$$

By taking logarithms, we then must have

$$r > \frac{\log |a|_1}{\log |x|_1} \text{ and } r > \frac{\log |a|_2}{\log |x|_2}.$$

Since the logarithms involved are all negative we then must have

$$\frac{\log |a|_1}{\log |x|_1} = \frac{\log |a|_2}{\log |x|_2}$$

because otherwise there would be a rational number between these two values. However this equality implies

$$\alpha = \frac{\log |a|_2}{\log |a|_1} = \frac{\log |x|_2}{\log |x|_1}$$

and we have the result. □

On the rational numbers the absolute value is a norm. The next lemma describes norms equivalent to absolute value on $\mathbb{Q}$.

Lemma 7.4.1 *On the rational numbers $\mathbb{Q}$ with absolute value $|\ |$, the function $|x|_\alpha = |x|^\alpha$ is a norm on $\mathbb{Q}$ if and only if $\alpha \le 1$. In this case it is equivalent to absolute value $|\ |$.*

Proof Let $|x|_\alpha = |x|^\alpha$ with $\alpha \le 1$. We show that this is a norm on $\mathbb{Q}$. The first two properties of a norm are direct so we must only show the triangle inequality.

Consider $|(x+y)|_\alpha = |x+y|^\alpha$. Assume that $|y| \le |x|$. Then

$$|(x+y)|_\alpha = |x+y|^\alpha \le (|x|+|y|)^\alpha = |x|^\alpha(1+\frac{|y|}{|x|})^\alpha \le |x|^\alpha(1+\frac{|y|}{|x|})$$

$$\le |x|^\alpha(1+\frac{|y|^\alpha}{|x|^\alpha}) = |x|^\alpha + |y|^\alpha = |x|_\alpha + |y|_\alpha.$$

Conversely if $\alpha > 1$ then the triangle inequality is not satisfied. For example

$$|1+1|^\alpha = 2^\alpha > 1^\alpha + 1^\alpha.$$

□

The archimedean property and its negation are crucial for our additional completions of $\mathbb{Q}$ so we make the definitions formal.

Definition 7.4.2 *A norm $|\ |$ on a field F is* **archimedean** *if given $x, y \in F$ with $x \ne 0$ there exists an integer n with $|nx| > |y|$. If a norm is not archimedean it is called* **non-archimedean**.

Non-archimedean norms satisfy a very special version of the triangle inequality.

Lemma 7.4.2 *A norm $|\ |$ on F is* **non-archimedean** *if and only if it satisfies*

$$|x+y| \le max(|x|, |y|).$$

The inequality above is called the **strong triangle inequality**. *The induced metric is called an* **ultra-metric** *and satisfies*

$$d(x,z) \le max(d(x,y), d(x,z)).$$

We leave the proof to the exercises. However, recall that an ordered field must have characteristic 0 and hence contains a copy of the rational integers $\mathbb{Z}$. Non-archimedean norms on a field F are also characterized by the norms of the integers.

Theorem 7.4.2 *(1) The norm $| \ |$ is non-archimedean if and only if $|n| \leq 1$ for all integers n.*

(1) The norm $| \ |$ is archimedean if and only if

$$sup(\{|n|; n \in \mathbb{Z}\}) = \infty.$$

Proof Suppose that $| \ |$ is non-archimedean. For any norm we have $|1| = 1$. Now we do induction on the natural numbers which we may assume to be in F. Assume $|k| \leq 1$ and consider $|k+1|$. Then $|k+1| \leq max\{|k|, 1\} \leq 1$ so the assertion is true for all natural numbers by induction. We have the equality the $|-x| = |x|$ so the assertion is true for all integers.

Conversely, suppose that $|x| \leq 1$ for all integers x. We show that $|x+y| \leq max\{|x|, |y|\}$. Now we have

$$|x+y|^n = |(x+y)^n| = |\sum_{k=0}^{n} \binom{n}{k} x^k y^{n-k}| \leq \sum_{k=0}^{n} |\binom{n}{k}| |x|^k |y|^{n-k}.$$

But $\binom{n}{k}$ is an integer so

$$|x+y|^n \leq \sum_{k=0}^{n} |x|^k |y|^{n-k} \leq (n+1) max\{|x|, |y|\}^n.$$

Hence

$$|x+y| \leq (n+1)^{1/n} max\{|x|, |y|\} \text{for all } n.$$

Taking the limit as $n \to \infty$ gives us the non-archimedean inequality. This completes part (1).

For part (2) it is clear that if $| \ |$ is archimedean there must be integers with arbitrarily large norms. □

7.4.1 The p-Adic Norm

For each prime p, we now introduce a non-archimedean norm on the rational numbers. Completion of $\mathbb{Q}$ with respect to this norm will give us the field of p-adic numbers. Since it is non-archimedean this p-adic norm is not equivalent to absolute value and hence as a normed field none of the p-adic fields are isomorphic to $\mathbb{R}$. Further we will show that for different primes p_1 and p_2 the corresponding p_1-adic norm is not equivalent to the p_2-adic norm.

Let $x = \frac{m}{n}$ be a rational number where $(m, n) = 1$. As we remarked in Chapter 2 the fundamental theorem of arithmetic implies that x also has a unique prime decomposition

$$x = p_1^{e_1} \cdots p_k^{e_k}$$

where here the exponents e_i are allowed to be negative. Now let p be a fixed prime and $x \in \mathbb{Q}$. Then it follows from the prime decomposition that

$$x = p^{\alpha}(\frac{a}{b})$$

with integers a, b such that $(a, b) = 1$, $p \nmid ab$ and $\alpha \in \mathbb{Z}$. We now define the **p-adic norm** of the rational number x by

$$|x|_p = p^{-\alpha} \text{ if } x \neq 0 \text{ and } 0 \text{ if } x = 0.$$

The map $ord : \mathbb{Q} \to \mathbb{Z}$ by $ord(x) = \alpha$ is called the **p-adic valuation**.

Lemma 7.4.3 *For any prime p, the p-adic norm is a non-archimedean norm on $\mathbb{Q}$. Further $| \ |_p$ can take on only a discrete set of values.*

Proof The basic norm properties are straightforward computations and we leave them to the exercises. From the definition the p-adic norm for any rational is p^{-m} for some integer m. Therefore, the p-adic norm can take on only a discrete set of values. Finally, for any integer n it is clear that the p-adic norm is 1 or less. Therefore, this norm must be non-archimedean. □

Since for any prime p the p-adic norm is a norm hence it defines a p-adic distance function on $\mathbb{Q}$ given by

$$d(x, y) = |x - y|_p.$$

Further since the norm is non-archimedean it follows that the p-adic distance function is an ultra-metric and satisfies

$$d(x, z) \leq \max(d(x, y), d(x, z)).$$

The p-adic norm for any natural number n is less than or equal to one. On the other hand if $n > 1$ we have for the ordinary absolute value $|n| > 1$. It follows that $|n|_p \neq |n|^{\alpha}$ for any real number α and hence for no prime is the p-adic norm equivalent to the standard absolute value.

Lemma 7.4.4 *For each prime p the corresponding p-adic norm on $\mathbb{Q}$ is not equivalent to the standard absolute value on $\mathbb{Q}$.*

Next, we show that for distinct primes p_1 and p_2 the corresponding norms are inequivalent.

Lemma 7.4.5 *If p_1, p_2 are distinct primes then the corresponding p-adic norms are inequivalent.*

Proof Suppose that $p_1 \neq p_2$. Let $x_n = (\frac{p_1}{p_2})^n$. In the p_1-adic norm this goes to zero and hence (x_n) converges and is therefore a Cauchy sequence. However in the p_2-adic norm the sequence (x_n) goes to infinity and hence diverges and is therefore not a Cauchy sequence. It follows that the two norms are not equivalent. □

Finally, we show that being close in the p-adic norm is equivalent to being congruent modulo p^n. That is:

Lemma 7.4.6 *If $a, b \in \mathbb{N}$ then $a \equiv b$ mod p^n if and only if $|a-b|_p \leq p^{-n}$.*

Proof Suppose that $a, b \in \mathbb{N}$ and $a \equiv b \mod p^n$. Then $p^n|(a-b)$. It follows that $\text{ord}(a-b) \geq n$ and hence $|a-b|_p \leq p^{-n}$. Conversely if $|a-b|_p \leq p^{-n}$ then $p^n|(a-b)$ and hence $a \equiv b \mod p^n$. □

7.5 The Construction of $\mathbb{Q}_p$

For each prime p, the rational numbers equipped with the p-adic norm provides a non-archimedean ordered field. Using the Cauchy completion procedure we can construct a complete ordered field that has the rationals as a dense subset, with respect to the induced p-adic distance. For a given prime p this is the **field of p-adic numbers** that we will denote by $\mathbb{Q}_p$. Each of these fields is non-archimedean and hence non-isomorphic to the real numbers $\mathbb{R}$. Further from Lemma 7.4.5, for differing primes p_1, p_2 the corresponding norms are inequivalent and therefore the corresponding fields are distinct as ordered fields. We therefore have the following theorem.

Theorem 7.5.1 *For each prime p, the field $\mathbb{Q}_p$ of p-adic numbers is a complete non-Archimedean ordered field which contains the rational numbers $\mathbb{Q}$ as a dense subset. Further each of these fields is distinct from the real numbers $\mathbb{R}$ and for different primes p_1, p_2 the fields are distinct.*

In Section 7.7, we will prove a type of converse to this result (Ostrowski's theorem) and show that $\mathbb{R}$ and the $\mathbb{Q}_p$ are the only complete ordered extensions of $\mathbb{Q}$ that have $\mathbb{Q}$ as a dense subfield. In Section 7.8, we use a property of the p-adic fields to prove that $\mathbb{R}$ is not isomorphic (as fields) to any $\mathbb{Q}_p$ and if p_1, p_2 are distinct primes then Q_{p_1} and $\mathbb{Q}_{p_2}$ are non-isomorphic.

7.5.1 p-Adic Arithmetic and p-Adic Expansions

As we remarked at the beginning of this chapter, the common way to handle real number arithmetic is via decimal expansions. As we pointed out though any base can be utilized and for computer hardware purposes usually some form of binary expansion is used. In a similar manner, given a fixed prime p each p-adic number has a unique p-adic expansion which allows arithmetic to be carried out. This expansion uses p-adic digits, that is the numbers 0, 1, ..., $p-1$ and arithmetic on the digits must be done modulo p. In real arithmetic decimal expansions, there is always the ambiguity with 9 and 0, that is, for example, $.399999\cdots$ and $.40000\cdots$ define the same number. Because of the uniqueness of the p-adic expansions this ambiguity does not occur and often p-adic representations are preferable for computer arithmetic.

Theorem 7.5.2 *Let p be a fixed prime and $\mathbb{Q}_p$ the field of p-adic numbers. Then each p-adic number $x \in \mathbb{Q}_p$ has a canonical p-adic expansion*

$$x = \sum_{n=-m}^{\infty} d_n p^n$$

with $d_i \in \{0, 1, \ldots, p-1\}$. This expansion is unique.

Proof Let p be a fixed prime and $\mathbb{Q}_p$ the corresponding p-adic field. To start, consider rational numbers x with $|x|_p \leq 1$. We show that for any $i \in \mathbb{Z}$ there exists a rational integer α with $|\alpha - x|_p \leq p^{-i}$ and further we can take $\alpha \in \{0, 1, 2, \ldots, p-1\}$. In this range the rational integer α is unique.

To see this consider $x = \frac{a}{b}$ with a, b integers such that $(a, b) = 1$ and let $i \in \mathbb{Z}$. Since $|x|_p \leq 1$ we must have that both a, b are relatively prime to p and hence also relatively prime to p^i. Hence there exists $m, n \in \mathbb{Z}$ with $mb + np^i = 1$. Now let $\alpha = am$. It follows that

$$|\alpha - x|_p = |am - \frac{a}{b}|_p = |\frac{a}{b}|_p |mb - 1|_p \leq |mb - 1|_p = |np^i|_p \leq p^{-i}.$$

Recall that the strong triangle inequality holds for the p-adic norm. That is, $|a + b|_p \leq \max\{|a|_p, |b|_p\}$. Hence in the inequality given above, we can add a multiple of p^i to α to get an integer α^* in the range $\{0, 1, \ldots, p-1\}$ for which

$$|\alpha^* - x|_p \leq p^{-i}.$$

There is only one such integer in this range congruent modulo $p^i - 1$ giving the uniqueness.

Next recall that if $a \in \mathbb{Q}_p$ then a is given by an equivalence class of Cauchy sequences (under the p-adic norm) of rationals. We claim that if $|a|_p \leq 1$ then there is a unique such Cauchy sequence $(a_1, a_2, \ldots)$ representing a with

$$a_i \in \mathbb{Z} \text{ and } a_i \equiv a_{i+1} \bmod p^i.$$

Let (b_i) be a Cauchy sequence of rationals representing a. We show that there is an equivalent Cauchy sequence (a_i) satisfying the conditions above and which is unique. Since $|b_i|_p \to |a|_p \le 1$ as $i \to \infty$ we may assume, after throwing away some initial terms if necessary, that $|b_i|_p \le 1$ for all i.

Now for each $j = 1, 2, ...$ let N_j be a positive integer such that $|b_i - b_i'|_p \le p^{-j}$ for all $i, i' \ge N_j$. We may assume that the sequence N_j is increasing so that $N_j \ge j$ for all j. From the first part of the proof there are then integers a_j with $0 \le a_j < p^j$ with

$$|a_j - b_{N_j}| \le \frac{1}{p^j}.$$

Now consider the sequence (a_j). For $j \in \mathbb{N}$ then for $i \ge N_j$ we have

$$|a_i - b_i|_p = |a_i - a_j + a_j - b_{N_j} + b_{N_j} - b_i|_p.$$

From the strong triangle inequality we obtain

$$|a_i - b_i|_p \le \max\{|a_i - a_j|_p, |a_j - b_{N_j}|_p, |b_i - b_{N_j}|_p\} \le \max\{\frac{1}{p^j}, \frac{1}{p^j}, \frac{1}{p^j}\} = \frac{1}{p^j}.$$

It follows that

$$|a_i - b_i|_p \to 0$$

and hence (a_j) is also a Cauchy sequence in the p-adic norm and also represents a.

Further

$$|a_{j+1} - a_j|_p = |a_{j+1} - b_{N_{j+1}} + b_{N_{j+1}} - b_{N_j} + b_{N_j} - a_j|_p.$$

Again using the strong triangle inequality

$$|a_{j+1} - a_j|_p \le \max\{|a_{j+1} - b_{N_{j+1}}|_p, |b_{N_{j+1}} - b_{N_j}|_p, |a_j - b_{N_j}|_p\}$$

$$\le \max\{\frac{1}{p^j}, \frac{1}{p^j}, \frac{1}{p^j}\} = \frac{1}{p^j}.$$

Therefore it follows that $a_j \equiv a_{j+1} \bmod p^j$.

We show that the sequence (a_j) is unique. Suppose that (a_i') is another Cauchy sequence representing a and satisfying $a_i' \equiv a_{i+1}' \bmod p^i$. Suppose that for some j we have $a_j \ne a_j'$. Then for any $i > j$ we must have

$$a_i \equiv a_j \not\equiv a_j' \equiv a_i' \bmod p^j.$$

This then implies that

$$|a_i - a_i|' > \frac{1}{p^j} \text{ for all } i \ge j,$$

contradicting the fact that (a_i) and (a_i') are equivalent Cauchy sequences both representing a.

It is from this unique sequence (a_i) that we construct the canonical p-adic expansion. Assume first as before that $|a|_p \leq 1$. Each $a_i \in \mathbb{Z}$ with $0 \leq a < p^i$ and hence each a_i has a p-ary expansion

$$a_i = d_0 + d_1 p + \cdots + d_{i-1} p^{i-1}$$

with each $d_j \in \{0, 1, \ldots, p-1\}$. Since $a_i \equiv a_{i+1} \bmod p^i$. It follows that

$$a_{i+1} = d_0 + d_1 p + \cdots + d_{i-1} p^{i-1} + d_i p^i$$

with $d_i \in \{0, 1, \ldots, p-1\}$. From this it follows that a is represented by an infinite series

$$a = \sum_{i=0}^{\infty} d_i p^i$$

which converges in the p-adic norm to the sum a. Thus, a can be considered as sequence of p-adic digits which extends infinitely far to the left

$$a = \cdots d_n d_{n-1} \cdots d_0.$$

This sequence uniquely represents a and is called the **canonical p-adic expansion**.

Up to now, we have considered p-adic norms less than or equal to 1. Now suppose that $|a|_p > 1$. If $|a|_p = p^m$ then $a = p^m a'$ with $|a'|_p \leq 1$. It follows that a is a convergent series of p-adic digits $d_i \in \{0, 1, \ldots, p-1\}$ of the form

$$a = \sum_{i=-m}^{\infty} d_i p^i$$

with $d_m \neq 0$. Thus we can represent a as a sequence of p-adic digits with a point and infinitely many digits to the left of the point and finitely many digits to the right. That is,

$$a = \cdots d_n d_{n-1} \cdots d_0.d_{-1} d_{-2} \cdots d_{-m}.$$

For any $a \in \mathbb{Q}_p$ there is a unique such expansion and this is called the **canonical p-adic expansion**. □

We will call the dot after d_0, because of familiarity, the **decimal point** in the expansion, although this of course has nothing to do with the standard decimal point. Notice further that in order to do real number arithmetic there is also ambiguity in the expansion no matter what base is used. In decimal arithmetic for example, that is base 10, we have 4.000000... and 3.999999... representing the same number. In

p-adic arithmetic there is no such ambiguity. This can often be used advantageously in doing rational arithmetic on a computer.

In order to do arithmetic in $\mathbb{Q}_p$ we first need to discuss how to find the p-adic expansions, especially for rational numbers. We start with representations of the integers. Notice first that multiplying by p^n with $n > 0$ moves the decimal point n places to the right while multiplying by p^{-n} moves it n places to the left. So for example in $\mathbb{Q}_5$ we have

$$23434134. \times 5^2 = 2343413400.$$

while

$$23434134. \times 5^{-2} = 234341.34$$

Now, let $n \in \mathbb{N}$. Then as we showed in Section 7.1, n has a p-ary expansion

$$n = a_0 + a_1 p + \cdots + a_{k-1} p^{k-1} + a_k p^k$$

with $a_i \in \{0, 1, \ldots, p-1\}$. The p-adic expansion of n is then this p-ary expansion in the reverse order

$$n = \cdots a_k a_{k-1} \cdots a_0.$$

Lemma 7.5.1 *Consider $n \in \mathbb{N}$. Then it has the p-ary expansion*

$$n = a_0 + \cdots + a_k p^k$$

Then the p-adic expansion is $n = a_k \cdots a_0.$, the p-ary expansion with the digits in the reverse order.

EXAMPLE 7.5.1 Find the 5-adic expansion of 17.

Here we first find $17 = 2 + 3 \cdot 5$. Hence the 5-adic expansion of 17 is 32.

p-adic arithmetic is then done much as standard decimal arithmetic but "carries" must be done modulo p and taken to the left. We will discuss this more later but we show how this is done in $\mathbb{Q}_5$.

EXAMPLE 7.5.2 Show that $4 \times 17 = 68$ using the 5-adic expansion.

We have the 5-adic expansions $4_5 = 4.$ and $17_5 = 32.$ Then 4×17 is given by

$$4. \times 32. = 233.$$

To see how we obtain this notice that $4 \times 2 = 8$ and 8 has the 5-adic expansion 13. Therefore the first digit is 3 and we carry the 1 to the left. Then $4 \times 3 + 1 = 13$ which has the 5-adic expansion 23. Therefore the 5-adic expansion 233. is the final result. A quick computation shows that this 5-adic expansion has the value 68 as expected.

We next consider the representations of negative numbers.

Lemma 7.5.2 *If*

$$\alpha = \sum_{i=n}^{\infty} a_i p^i$$

then

$$-\alpha = \sum_{i=n}^{\infty} b_i p^i$$

where $b_n = p - a_n$ *and* $b_i = (p-1) - a_i$ *if* $i > n$.

We defer the proof to the exercises but here instead present an example which exhibits the method.

EXAMPLE 7.5.3 Find the 5-adic expansion of -3. Suppose that $-3 = \cdots a_n a_{n-1} \cdots a_0$. Then $\cdots a_n a_{n-1} \cdots a_0 + 3. = 0$. Working from the left we then have that $a_0 + 3 = 0 \bmod 5$ and hence $a_0 = 2$. It follows that $-3 = \cdots a_n \cdots a_1 \cdot 2$. Adding 3. to this we get that $3 + 2 = 5$ which has the 5-adic expansion 10, so we carry the 1 to get $a_1 + 1 = 0$. Hence $a_1 = 4$ since arithmetic on the digits is done modulo 5. Continuing in this manner we obtain

$$-3 = \cdots 44442.$$

Now, we consider the p-adic representation of rational numbers. We use the following lemma which is essentially p-adic division by an integer and then give an example.

Lemma 7.5.3 *The fractions* $\frac{a}{b}$ *have a periodic* p*-adic expansion. Suppose that*

$$\alpha = p^n\left(\frac{c_1}{d_1}\right)$$

with c_1, d_1 *integers such that* $(c_1, d_1) = 1$ *and* $p \nmid c_1 d_1$, *and suppose that* $\alpha = a_n p^n + a_{n+1} p^{n+1} + \ldots + \ldots$. *Then*

$$a_n = c_1 d_1^{-1} \bmod p.$$

Consider then

$$\frac{c_1}{d_1} - a_n = p\left(\frac{c_2}{d_2}\right).$$

Then $a_{n+1} = c_2 d_2^{-1}$ *mod* p *and so on.*

EXAMPLE 7.5.4 We find the 7-adic expansion of $\frac{3}{4}$.

The easiest way to proceed is to find the 7-adic expansion of $\frac{1}{4}$ and then multiply (using 7-adic multiplication) by 3. Suppose that

$$x = \frac{1}{4} = \cdots b_n b_{n-1} \cdots b_1 b_0.$$

Then $4x = 1$ so $4b_0 \equiv 1 \bmod 7$. It follows that $b_0 = 2$ so that

$$x = \cdots b_n b_{n-1} \cdots b_1 \cdot 2.$$

Multiply this by 4 to obtain

$$4x = \cdots (4b_1 + 1) \cdot 1.$$

To see this $4 \cdot 2 = 8 = 7 + 1$ so we *carry* the 1 to get in the second digit $4b_1 + 1$. Since the result is $1 = \cdots 0001.$ we then have

$$4b_1 + 1 = 0 \bmod 7 \implies b_1 = 5.$$

Therefore

$$x = \cdots b_2 \cdot 5 \cdot 2.$$

Continuing in this manner we get

$$\frac{1}{4} = \cdots 15152.$$

Now, we multiply this by 3 to get the 7-adic expansion of $\frac{3}{4}$:

$$\frac{3}{4} = \cdots 15151516.$$

To see this we start the multiplication, $3 \times \cdots 151512.$ at the far right to first get 6. There is nothing to carry. Then $3 \times 5 = 15 = 2 \times 7 + 1$. Hence we write down the 1 and carry the 2. Then we have $(3)(1) + 2 = 5$ and continue to the left.

From the method and from the example, it is clear that if this is done for any rational number the resulting p-adic expansion must eventually be periodic. The proof is essentially the same as showing the decimal expansion for any rational must eventually be periodic. We leave the proof to the exercises.

Corollary 7.5.1 *Let p be a fixed prime and $x \in \mathbb{Q}_p$. Then, the p-adic expansion for x is periodic if and only if $x \in \mathbb{Q}$.*

For a fixed p the arithmetic in $\mathbb{Q}_p$ can be done as in decimal arithmetic but the carries must be done mod p and to the left.

EXAMPLE 7.5.5 Let $x = \cdots 45213.$ and $y = \cdots 61115.$ in $\mathbb{Q}_7$, Find $x + y$. Using carrying mod 7 we get

$$x + y = \cdots 36331.$$

To see this, we start at the far right and add $3 + 5 = 8$ which has the 7-adic representation 11. Hence we write the 1 and carry the 1 to the left. Then, we have $1 + 1 = 2$ plus the carry 1 to get 3. We then continue to the left.

7.6 The *p*-Adic Integers

If p is a fixed prime, then the p-adic norm is non-archimedean and hence the norm of any rational integer $x \in \mathbb{Z}$ is less than or equal to one. It follows that the p-adic expansion of any rational integer extends only to the left of the decimal point. We extend this to form the ring of p-adic integers which is a subring of the field of p-adic numbers and contains the rational integers. It is a unique factorization domain like $\mathbb{Z}$ but has many properties quite different than the ordinary rational integers.

Definition 7.6.1 *A p-adic number $\alpha \in \mathbb{Q}_p$ is a* **p-adic integer** *if its p-adic norm is less than or equal to* 1, $|\alpha|_p \leq 1$. *We denote the set of p-adic integers by $\mathbb{Z}_p$ and hence*

$$\mathbb{Z}_p = \{\alpha \in \mathbb{Q}_p; |\alpha|_p \leq 1\}.$$

Note that $\mathbb{Z}_p$ also denotes the modular ring $\mathbb{Z}/p\mathbb{Z}$. For the remainder of this chapter $\mathbb{Z}_p$ will denote the ring of p-adic integers and we will use $\mathbb{Z}/p\mathbb{Z}$ for the modular ring mod p.

Since the p-adic norm of a p-adic integer is less than or equal to 1 it follows that in the p-adic expansion of a p-adic integer the digits (possibly infinitely many) are always to the left of the decimal point. This can be taken as an alternative definition of a p-adic integer.

Lemma 7.6.1 *A p-adic number $\alpha \in \mathbb{Q}_p$ is a p-adic integer if and only if its canonical expansion has only positive powers of p. That is*

$$\mathbb{Z}_p = \{\alpha \in \mathbb{Q}_p; \alpha = \sum_{i=0}^{\infty} a_i p^i\}.$$

The p-adic integers form a subring of $\mathbb{Q}_p$ which contains $\mathbb{Z}$.

Theorem 7.6.1 *The set $\mathbb{Z}_p$ of p-adic integers forms a subring of $\mathbb{Q}_p$ which contains the rational integers $\mathbb{Z}$.*

Proof It is clear that $\mathbb{Z} \subset \mathbb{Z}_p$. To show that $\mathbb{Z}_p$ is a subring we must show that it is closed under addition, multiplication and additive inverses. Since the p-adic norm is non-archimedean it satisfies the strong triangle inequality and hence these closure properties are straightforward. We leave the details to the exercises. □

Later, we will see that this ring is actually a unique factorization domain.

Recall that a **unit** in a ring R with identity is an element which has a multiplicative inverse. In the rational integers $\mathbb{Z}$ the only units are ± 1. The situation is quite different in $\mathbb{Z}_p$ where there are many units and in fact every rational integer m relatively prime to p is invertible.

Theorem 7.6.2 *A p-adic integer $\alpha \in \mathbb{Z}_p$ is a unit if and only if $\alpha = ...a_2a_1a_0$ with $a_0 \neq 0$. Hence the group of units*

$$U(\mathbb{Z}_p) = \{\sum_{i=0}^{\infty} a_i p^i; a_0 \neq 0\}.$$

Proof Let $\alpha \in \mathbb{Z}_p$ and suppose that

$$\cdots a_n \cdots a_1 a_0.$$

is the p-adic expansion for α. Consider $\beta \in \mathbb{Z}_p$ with

$$\beta = \cdots b_n \cdots b_1 b_0.$$

Now consider the equation

$$\beta\alpha = 1 = \cdots 001.$$

Since $a_0 \neq 0 \bmod p$ we can solve for the expansion of β in the above equation. First we would have $a_0 b_0 = 1$ where a_0 and b_0 are p-adic digits. Since a_0 is not equal to 0 it has an inverse mod p and thus a solution for b_0. Thus, we have found the first digit of β. Now we multiply again (see Section 7.3) to get

$$a_0 b_1 + a_1 b_0 + \text{ carry } \equiv 0 \bmod p.$$

This is now solvable for b_1 and we obtain the second digit of β. Continuing in this manner we can solve $\beta\alpha = 1$ and hence α is a unit. □

The next result shows that any element of $\mathbb{Q}_p$ is a product of an invertible p-adic integer and a power of p.

Lemma 7.6.2 *Let $x \in \mathbb{Q}_p$ with $|x|_p = p^{-n}$. Then $x = p^n u$ with $u \in U(\mathbb{Z}_p)$.*

Proof Let $x \in \mathbb{Q}_p$ with $|x|_p = p^{-n}$. Then $p^{-n}x = u$ has norm 1 since $|p^{-n}x|_p = |p^{-n}|_p |x|_p = p^n p^{-n} = 1$. Since $|u|_p = 1$ it follows that u is a p-adic integer and further since its norm is exactly 1 from the previous theorem it is invertible and hence a p-adic unit. □

Recall that any integral domain can be embedded into its field of fractions, which is the smallest field containing it. The field of fractions for $\mathbb{Z}$ is of course $\mathbb{Q}$. This last lemma shows that $\mathbb{Q}_p$ is actually the field of fractions for $\mathbb{Z}_p$.

Theorem 7.6.3 *The field of p-adic numbers is the field of fractions for the ring of p-adic integers.*

This last theorem provides an alternative approach to the construction of the p-adic numbers. Start with p-ary expansions of integers and complete them to form the ring of p-adic integers $\mathbb{Z}_p$. Then take the field of fractions of $\mathbb{Z}_p$ to find the field of p-adic numbers and show that this is complete. This was the approach followed originally by Hensel. We refer to [H] for details.

7.6.1 Principal Ideals and Unique Factorization

Although the p-adic integers differ radically from the rational integers in the structure of their unit groups here we show that the p-adic integers $\mathbb{Z}_p$, like the rational integers $\mathbb{Z}$, form a unique factorization domain.

Recall from Chapter 6 (see Section 6.2) that a **unique factorization domain** or **UFD** is an integral domain R such that for each $r \in R$ either $r = 0$, r is a unit or r has a factorization into primes which is unique up to ordering and unit factors.

In this more general algebraic language, the Fundamental Theorem of Arithmetic states that the rational integers $\mathbb{Z}$ form a UFD. Gauss proved that the complex integers were also a UFD as well as the ring of polynomials over any field F (see Chapter 6).

In Chapter 6, we also examined **principal ideal domains** abbreviated as PID, which are integral domains where every ideal is a principal ideal. We showed that any principal ideal domain is a UFD. Using the p-adic norm, we show that any ideal in the p-adic integers $\mathbb{Z}_p$ is either (0) or $p^k\mathbb{Z}_p$ for some $k \in \mathbb{N}$. It follows that $\mathbb{Z}_p$ is a principal ideal domain and therefore a unique factorization domain. Further, $\mathbb{Z}_p$ has a unique maximal ideal.

Theorem 7.6.4 *The ring of p-adic integers $\mathbb{Z}_p$ is a principal ideal domain. The ideals are the principal ideal* (0) *and $p^k\mathbb{Z}_p$ for all $k \in \mathbb{N} \cup \{0\}$. The ideal $p\mathbb{Z}_p = \mathbb{Z}_p \backslash U(\mathbb{Z}_p)$ is the unique maximal ideal.*

Proof Let $a \in \mathbb{Z}_p$ with $a = \sum_{i=0}^{\infty} a_i p^i$. Consider the evaluation map from the p-adic integers to the integers modulo p, $f : \mathbb{Z}_p \to \mathbb{Z}/p\mathbb{Z}$ given by $f(a) = a_0$. For a prime p the modular integers $\mathbb{Z}/p\mathbb{Z}$ form a field. The evaluation map is then a homomorphism onto a finite field with kernel $p\mathbb{Z}_p$ and hence $p\mathbb{Z}_p$ is a maximal ideal. We show that it is unique.

Let J be another proper maximal ideal we show $J = p\mathbb{Z}_p$. It is clear that $p\mathbb{Z}_p$ contains all the p-adic integers with norm strictly less than 1. Suppose that $\alpha \in J$. If $|\alpha|_p = 1$ then α is a unit in $\mathbb{Z}_p$ and $J = \mathbb{Z}_p$. Therefore, if J is a proper ideal it follows that $|\alpha|_p < 1$ and hence $\alpha \in p\mathbb{Z}_p$. Therefore $J \subset p\mathbb{Z}_p$ and by maximality $J = p\mathbb{Z}_p$.

From the proof above it follows that if I is any proper ideal in $\mathbb{Z}_p$ we must have $I \subset p\mathbb{Z}_p$. The ideals in $p\mathbb{Z}_p$ are precisely the principal ideals $p^n\mathbb{Z}_p$ for some natural number n. Therefore, $\mathbb{Z}_p$ is a principal ideal domain with unique maximal ideal $p\mathbb{Z}_p$. □

Since a PID must be a unique factorization domain, we have the following corollary.

Corollary 7.6.1 *The ring of p-adic integers $\mathbb{Z}_p$ is a unique factorization domain.*

We mention that the development of the p-adic integers as a PID can be generalized to what are termed **discrete valuation rings**. A **discrete valuation** is an integer valuation on a field K, that is a function

$$\rho : K \to \mathbb{Z} \cup \{\infty\}$$

satisfying the conditions

(1) $\rho(xy) = \rho(x) + \rho(y)$,
(2) $\rho(x + y) \geq min(\rho(x), \rho(y))$,
(3) $\rho(x) = \infty$ iff $x = 0$.

A field with a non-trivial discrete valuation is called a **discrete valuation field**. A **discrete valuation ring** is an integral domain whose field of fractions is a discrete valuation field. The p-adic norm defines a discrete valuation and hence $\mathbb{Z}_p$ is a discrete valuation ring.

It can be proved that for a discrete valuation ring, the discrete valuation makes it a principal ideal domain and any irreducible elements generate its unique maximal ideal.

7.6.2 The Completeness of $\mathbb{Z}_p$

Consider a convergent sequence of p-adic integers (x_n) with $\lim x_n = x$. Here the limit is with respect to the p-adic norm. Since each $x_n \in \mathbb{Z}_p$ we have $|x_n|_p \leq 1$ and therefore $|\lim x_n|_p \leq 1$ also. It follows that x must also be a p-adic integer and hence the limit of any convergent sequence of p-adic integers is a p-adic integer. It follows that as a subset of the metric space $\mathbb{Q}_p$ the set $\mathbb{Z}_p$ is closed. It is known that a closed subset of a complete metric space is also complete and therefore the p-adic integers are complete. We have thus proved.

Theorem 7.6.5 *The p-adic integers $\mathbb{Z}_p$ are complete as a metric subspace of the field of p-adic numbers $\mathbb{Q}_p$.*

7.7 Ostrowski's Theorem

We have seen that the field of real numbers is up to isomorphism the only archimedean completion of $\mathbb{Q}$. That is, if F is any other complete archimedean ordered field that contains $\mathbb{Q}$ as a dense subset then F is isomorphic to $\mathbb{R}$. Ostrowki's theorem, that we present in this section says that besides the reals, the only completions of $\mathbb{Q}$ are the fields of p-adic numbers.

Theorem 7.7.1 *(Ostrowski) Every nontrivial norm $| \ |$ on $\mathbb{Q}$ is equivalent to either absolute value $| \ |$ or a p-adic norm $| \ |_p$ for some prime p. Therefore, the only complete fields containing $\mathbb{Q}$ are the reals $\mathbb{R}$ and the p-adic fields $\mathbb{Q}_p$.*

Proof Let $| \ |$ be a norm on $\mathbb{Q}$. Assume first that its archimedean. Then there exists an integer n with $|n| > 1$. Let n_0 be the least such integer and suppose that $|n_0| = n_0^{\alpha}$. We show that $|n| = n^{\alpha}$ for all positive integers n.

Write n in its n_0-expansion so that $n = a_0 + a_1 n_0 + \cdots + a_s n_o^s$. By our assumption on n_0 we have $|a_i| \leq 1$ for all i. Therefore $|n| \leq Cn^{\alpha}$. Using n^N gives us $|n| \leq C^{1/N} n^{\alpha}$. Letting $N \to \infty$ we get that $n \leq n^{\alpha}$.

Use the expansion again to get $n \geq n^{\alpha}$ so therefore $n = n^{\alpha}$. This then implies that if $q \in \mathbb{Q}$ with $q > 0$ then $|q| = q^{\alpha}$ and hence the norm is equivalent to absolute value. Therefore if the norm is archimedean it is equivalent to absolute value.

Now suppose the norm is non-archimedean. Then $|n| \leq 1$ for all integers n. Let n_0 be the least integer for which $|n_0| < 1$. Claim first that n_0 is a prime. If not $n_0 = n_1 n_2$ with $n_1 < n, n_2 < n$. From this $|n_1| = |n_2| = 1$ and hence $|n_0| = 1$ a contradiction. Therefore, $n_0 = p$ a prime and we claim the norm is equivalent to the p-adic norm.

If p does not divide n then $n = rp + s$ and $|s| = 1$. But then $|rp| < 1$ and so $|n - s| < |s|$ and so $|n| = |s| = 1$. Thus if p does not divide n we have $|n| = 1$. Given $n \in \mathbb{N}$ we have $n = p^k m$ with $(m, p) = 1$. Then $|n| = |p^k||m| = |p|^k$. If $|p| < 1$ then $|p| = p^{-\alpha} = (\frac{1}{p})^{\alpha}$ for some α and hence this norm is equivalent to the p-adic norm. □

7.8 Hensel's Lemma and Applications

For fixed primes p the p-adic numbers have many applications to ordinary number theory especially to solving congruences modulo p. Important in this regard is **Hensel's Lemma**. First, we define congruence in $\mathbb{Q}_p$.

Definition 7.8.1 *$a \equiv b$ mod p^n in $\mathbb{Q}_p$ if $|a - b|_p \leq p^{-n}$.*

Now, we present Hensel's Lemma that is a result in modular arithmetic. The lemma says that if a polynomial equation has a simple root modulo a prime number p, then this root corresponds to a unique root of the same equation modulo any higher power of p. This root can be found by iteratively lifting the solution modulo successive powers of p and is an analog of Newton's method.

Theorem 7.8.1 *(Hensel's Lemma) Let $f(x) = c_0 + c_1x + \cdots + c_nx^n$ be a polynomial in $\mathbb{Z}_p[x]$ (coefficients are p-adic integers). Let $f'(x)$ be the formal derivative of $f(x)$. Suppose $\overline{a}_0 \in \mathbb{Z}_p$ with $f(\overline{a}_0) \equiv 0$ mod p and $f'(\overline{a}_0) \not\equiv 0$ mod p. Then, there exists a unique p-adic integer a such that $f(a) = 0$ and $a \equiv \overline{a}_0$ mod p.*

As preparation for the proof of Hensel's lemma we recall **Newton's method** for solving a non-linear equation $f(x) = 0$ over the reals where $f(x)$ is a differentiable real-valued function. We start with an initial guess x_0. This initial guess must be sufficiently close to a solution for this method to work but we will ignore this here and refer to [A] for the technical requirements. Given x_0 we form the tangent line to the curve $y = f(x)$ at the point $(x_0, f(x_0))$. This has the equation

$$y - f(x_0) = f'(x_0)(x - x_0).$$

Let x_1 be where the tangent line crosses the x-axis, that is where $y = 0$. We then have

$$-f(x_0) = f'(x_0)(x_1 - x_0) \implies x_1 = x_0 - \frac{f(x_0)}{f'(x_0)}$$

assuming that $f'(x_0) \neq 0$. This provides the initial step in an iteration scheme. Consider the tangent line at $(x_1, f(x_1))$ and obtain

$$x_2 = x_1 - \frac{f(x_1)}{f'(x_1)} \text{ assuming } f'(x_1) \neq 0$$

and in general

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)} \text{ assuming } f'(x_n) \neq 0.$$

Under appropriate conditions (see [A]) this iteration scheme will converge to a solution of $f(x) = 0$. How close the initial guess must be to a solution for the method to converge depends on the function $f(x)$ (see [A]).

This method can be applied to polynomial equations $P(x) = 0$ over the reals. The proof of Hensel's lemma in the p-adic field $\mathbb{Q}_p$ utilizes a p-adic version of Newton's technique.

Proof Let $f(x)$ be an p-adic integral polynomial, that is, $f(x)$ has p-adic coefficients, and let $\overline{a}_0$ be as in the statement of Hensel's lemma. We will prove the existence of a solution a by inductively constructing its canonical p-adic expansion

$$a = d_0 + d_1p + \cdots + d_kp^k + \cdots$$

where d_i are p-adic digits to be determined. Let a_k be the k-th convergent for a,

$$a_k = d_0 + d_1 + \cdots + d_kp^k.$$

We will use an induction and a p-adic version of Newton's method to show that we can find p-adic digits so that $f(a_k) \equiv 0 \bmod p^{k+1}$ and $a_k \equiv \overline{a}_0 \bmod p$. Then as $a_k \to a$ we have a as the desired solution.

Let $\overline{a}_0$ have the canonical p-adic expansion

$$\overline{a}_0 = b_0 + b_1 p + \cdots + b_k p^k + \cdots$$

Take $a_0 = d_0 = b_0$. Then $a_0 \equiv \overline{a}_0 \bmod p$ and $f(a_0) \equiv 0 \bmod p$. This establishes the lowest level of an induction.

Now, suppose we have a_{k-1} satisfying $f(a_{k-1}) \equiv 0 \bmod p^k$ and $a_{k-1} \equiv \overline{a}_0 \bmod p$. Now let

$$a_k = a_{k-1} + d_k p^k$$

where d_k is a p-adic digit to be determined. Then

$$f(a_k) = f(a_{k-1} + d_k p^k) = \sum_{i=0}^{n} c_i (a_{k-1} + d_k p^k)^i.$$

Then

$$f(a_k) = c_0 + \sum_{i=1}^{n} c_i (a_{k-1}^i + i(a_{k_1}^{i-1} d_k p^k + \text{ terms in powers higher than } p^{k+1})).$$

This implies that

$$f(a_k) = f(a_{k-1}) + d_k p^k f'(a_{k-1}).$$

By the inductive hypothesis we have $f(a_{k-1}) \equiv 0 \bmod p^k$ and hence there is a p-adic digit e_k with

$$f(a_k) = e_k p^k + d_k p^k f'(a_{k-1}).$$

To obtain the appropriate digit d_k we must then have

$$e_k + d_k f'(a_{k-1}) \equiv 0 \bmod p.$$

Since $a_{k-1} \equiv \overline{a}_0 \bmod p$ we have $f'(a_{k-1}) \equiv f'(\overline{a}_0) \not\equiv 0 \bmod p$. Therefore, the digit d_k can be found by

$$d_k = -\frac{e_k}{f'(a_{k-1})} \bmod p$$

and hence $f(a_k) \equiv 0 \bmod p$. Notice that approximating the p-adic digits uses essentially the same iteration scheme as Newton's method over the reals.

Now consider

$$a = d_0 + d_1 p + \cdots + d_k p^k + \cdots$$

Since $f(a) \equiv f(a_k) \bmod p^{k+1}$ for all k we must have $f(a) = 0$.

Now assume that a_{k-1} has the desired properties and consider a_k. Let d_k be a p-adic digit to be determined and consider

$$a_k = a_{k-1} + d_k p^k.$$

The uniqueness of a follows from the uniqueness of the sequence of convergents a_k. □

The proof of Hensel's lemma provides an algorithm for constructing the solution to an equation $f(x) = 0$ with $f(x) \in \mathbb{Z}_p[x]$. This algorithm is analogous to Newton's Method for solving real polynomial equations.

Suppose $\overline{a}_0$ is a solution to $f(x) \equiv 0 \bmod p$. Then follow the procedure outlined in the proof. Take d_0 the first p-adic digit of $\overline{a}_0$ and let $a_0 = \overline{a}_0$. Let $a_k = a_{k-1} + d_k p^k$ and iteratively find the digits d_k by $d_k = \frac{-a_{k-1}}{f'(a_{k-1})}$ for $k \geq 1$.

Theorem 7.8.2 *A polynomial with rational integer coefficients (in $\mathbb{Z}[x]$) has a root in $\mathbb{Z}_p$ if and only if it has an integer root modulo p^k for any $k \geq 1$.*

Proof Suppose that $f(x) \in \mathbb{Z}[x]$ and suppose that $f(a) = 0$ where $a \in \mathbb{Z}_p$. Then from the proof of Hensel's lemma there exists a sequence of integers (a_k) with $a_k \equiv a \bmod p^k$. Since $f(a_k) \equiv f(a) \bmod p^k$ and $f(a) = 0$ we must have an integer solution mod p^k for each k.

Conversely, suppose that for each k there is an integer a_k with $f(a_k) \equiv 0 \bmod p^k$. We have seen that the p-adic integers are complete so the sequence a_k has a convergent subsequence $(a_{k'})$. Suppose that the limit of this subsequence is a. A polynomial is a continuous function on any normed field (see exercises) and hence

$$f(a) = \lim f(a_k).$$

However $f(a_k) \equiv 0 \bmod p^k$ for all k and therefore $f(a) \equiv 0 \bmod p^k$ for all k and hence $f(a) = 0$. □

Corollary 7.8.1 *If a polynomial $F(x)$ with integer coefficients has no roots modulo p then it has no roots.*

Hensel's lemma can be used to describe the roots of unity in $\mathbb{Q}_p$.

Theorem 7.8.3 *For any prime p and $(m, p) = 1$ there exists a primitive m-th root of unity in $\mathbb{Q}_p$ if and only if $m|(p-1)$. In this case every m-th root of unity is also a $(p-1)$-th root of unity. The set of $(p-1)$-th roots of unity forms a cyclic subgroup of $U(\mathbb{Z}_p)$ of order $p-1$.*

Proof If $m|(p-1)$ then $p-1=km$ and hence every m-th root of unity in $\mathbb{Q}_p$ is also a $(p-1)$-th root of unity. Consider the polynomial $f(x)=x^{p-1}-1$. Then its formal derivative is $f'(x)=(p-1)x^{p-2}$. Now let a be a rational integer with $1 \leq a \leq p-1$. Then from Fermat's theorem, we have $f(a)=0$ and further $f'(a) \neq 0$ since $|f'(a)|_p=1$. Therefore Hensel's lemma implies that there are exactly $p-1$ solutions to $f(x)=0$ and they are all $(p-1)$-th roots of unity.

Conversely suppose that $a \in \mathbb{Q}_p$ with $a^m=1$ then $|a|_p=1$ and a is a p-adic integer. Let $\cdots a_1a_0.=a$ then $a \equiv a_0 \bmod p$ and hence $a_0^m=1$. Since a_0 is a rational integer this implies that $m|(p-1)$.

The set of $(p-1)$-th roots of unity in $\mathbb{Q}_p$ is then a finite subgroup of a field and as we saw in Theorem 2.4.13 this must be cyclic. □

As we have seen in this book, quadratic residues modulo a prime are important in several different areas of number theory. In fact determining quadratic residues was crucial in the Rabin encryption system. The final result of this section ties quadratic residues modulo a prime p to square roots in the p-adic integers.

Lemma 7.8.1 *A rational integer a not divisible by p has a square root in $\mathbb{Z}_p$ ($p \neq 2$) if and only if a is a quadratic residue modulo p.*

Proof Let $a \in \mathbb{Z}$ with $(a,p)=1$. Consider the polynomial $P(x)=x^2-a$ in $\mathbb{Z}_p[x]$. Suppose that a is a quadratic residue mod p. Then there exists $\overline{a}_0$ with $\overline{a}_0 \in \{1,2,\ldots,p-1\}$ and $a^2 \equiv \overline{a}_0^2 \bmod p$. Further $P'(x)=2x$ and $P'(\overline{a}_0)=2\overline{a}_0 \neq 0 \bmod p$ since $(a,p)=1$. Therefore by Hensel's lemma $P(x)$ has a solution in $\mathbb{Z}_p$.

Conversely suppose that a is not a quadratic residue. Then $P(x) \not\equiv 0 \bmod p$ and hence $P(x) \not\equiv 0 \bmod p^k$ for any k. It follows that $P(x)$ can have no solution in $\mathbb{Z}_p$.

7.8.1 The Non-isomorphism of the *p*-Adic Fields

Since each p-adic field is non-archimedean we have seen from the characterization of $\mathbb{R}$ that for any prime p the p-adic field $\mathbb{Q}_p$ is not isomorphic to $\mathbb{R}$. In the next theorem we use the results on square roots in $\mathbb{Q}_p$ to provide another proof of this and to show that p-adic fields for different primes are non-isomorphic.

Theorem 7.8.4 *The p-adic field $\mathbb{Q}_p$ is not isomorphic to $\mathbb{R}$ for any prime p. Further, if p_1 and p_2 are distinct primes then the corresponding p-adic fields are non-isomorphic.*

Proof Let p be a prime and suppose that $f:\mathbb{R} \to \mathbb{Q}_p$ is an isomorphism. Then p has a square root in $\mathbb{R}$ and hence by the isomorphism $f(p)$ has a square root in $\mathbb{Q}_p$. However, p is not a quadratic residue mod p and therefore p has no square root in $\mathbb{Q}_p$ providing a contradiction.

If $p_1 \neq p_2$ then there are $\frac{p_1-1}{2}$ quadratic residues mod p_1 and $\frac{p_2-1}{2}$ quadratic residues mod p_2. It follows that if $p_2 > p_1$ there must exist an integer a which is a quadratic residue mod p_2 but not mod p_1. Use this integer a and then follow the same proof as above. We leave the details to the exercises. □

As a final application of both Hensel's lemma and the utility of the p-adic fields in general we mention without proof the **local-global principle** of Hasse. The rational numbers $\mathbb{Q}$ are called a **global field** while its Completions, the real numbers $\mathbb{R}$ and the p-adic fields $\mathbb{Q}_p$ are called **local fields**. Any relationship among a set of rational numbers which is true globally, that is in $\mathbb{Q}$ is also true locally, that is in $\mathbb{R}$ and all the p-adic fields $\mathbb{Q}_p$.

Hasse's **Global-Local Principle** provides a partial converse for equations involving quadratic forms with integer coefficients:

$$\sum_{i,j} a_{ij}x_ix_j + \sum_i b_ix_i + c = 0.$$

If such an equation has solutions in $\mathbb{R}$ and in $\mathbb{Q}_p$ for every prime p, then it has a rational solution in $\mathbb{Q}$. In other words, a quadratic equation with integer coefficients has a global solution, that is in $\mathbb{Q}$ if and only if it has solutions in all the local fields, that is in $\mathbb{R}$ and in $\mathbb{Q}_p$ for all p.

7.9 Exercises

7.1 Find the p-adic norm and p-adic expansion in $\mathbb{Q}_7$ of:

(a) 15

(b) -1

(c) -3

(d) $\frac{1}{3}$

7.2 Describe in detail, analogously as for $\mathbb{R}$, the Cauchy completion of the rational numbers $\mathbb{Q}$ equipped with the p-adic norm for a prime p.

7.3 Fill in the details of the proof of Theorem 7.8.4, that is if $p_1 \neq p_2$ then the p-adic fields Q_{p_1} and $\mathbb{Q}_{p_2}$ are not isomorphic.

7.4 Let p be a prime number and $\mathbb{Z}_p$ the p-adic integers. Show that $\mathbb{Z}_p/p^n\mathbb{Z}_p$ is isomorphic to $Z/p^n\mathbb{Z}$ for any $n > 0$.

7.5 Let p be a prime number and $\mathbb{Z}_p$ the p-adic integers. Show that the additive group of $\mathbb{Z}_p$ is torsion-free.

7.6 Use the algorithm in the proof of Hensel's Lemma to find a solution (if there exists one) of the polynomial equations:

(a) $x^3 - 3x^2 + 2x + 1 = 0$ in $\mathbb{Q}_7$

(b) $x^4 - 6$ in $\mathbb{Q}_{11}$

7.7 Complete the proof that a p-adic expansion for x is periodic if and only if x is rational.

7.8 Show that if $x \in \mathbb{Q}_p$ and $x \equiv 0 \bmod p^k$ for all $k \geq 1$ then $x = 0$.

7.9 Let $f(x) \in \mathbb{Q}_p[x]$ that is a polynomial with p-adic coefficients. Show that $f(x)$ is a continuous function of $\mathbb{Q}_p$.

7.10 Complete the proof of Theorem 7.8.4 and show that if p_1, p_2 are distinct primes then the corresponding p-adic fields are non-isomorphic.

7.11 Prove that the rationals $\mathbb{Q}$ are dense in $\mathbb{Q}_p$.

7.12 Prove that the p-adic integers $\mathbb{Z}_p$ are compact as a metric space using the p-adic norm.

7.13 Show that for any prime p and any positive integer m not divisible by p, there exists a primitive m-th root of unity in $\mathbb{Q}_p$ if and only if m divides $p - 1$.

7.14 Show that the set of roots of unity in $\mathbb{Q}_p$ is a subgroup of the group of p-adic units.

7.15 Prove that a rational number $x \in \mathbb{Q}$ is a square if and only if it is a square in every p-adic field $\mathbb{Q}_p$ and in the real numbers $\mathbb{R}$.

7.16 Let $\mathbb{Z}_2$ be the 2-adic integers, Show that if $b \in \mathbb{Z}_2$ and $b \equiv 1 \bmod 8$ then b is a square in $\mathbb{Z}_2$.

7.17 Show that the equation $(x^2 - 2)(x^2 - 17)(x^2 - 34) = 0$ has a solution in the real numbers $\mathbb{R}$ and in all the p-adic field $\mathbb{Q}_p$ with p prime, but has no solution in the rational numbers $\mathbb{Q}$.

Bibliography

[AKS] M. Agrawal, N. Kayal, N. Saxena, PRIMES is in P. Ann. Math. **160**(2), 2781–2793 (2004)

[AGR] W.R. Alford, A. Granville, C. Pomerance, There are infinitely many carmichael numbers. Ann. Math. **139**, 703–722 (1994)

[AG] I. Anshel, M. Anshel, D. Goldfeld, An algebraic method for public key cryptography. Math. Res. Lett. **6**, 287–291 (1999)

[A] T.M. Apostol, *Introduction to Analytic Number Theory* (Springer, New York, 1976)

[Apo] T.M. Apostol, *The Most Surprising Result in Mathematics*. The Mathematical Intelligencer (2001)

[Ah] L. Ahlfors, *Introduction to Complex Analysis* (Springer, New York, 1976)

[Ba] A. Baker, *Transcendental Number Theory* (Cambridge University Press, Cambridge, 1975)

[B] C.X. Barnes, The infinitude of primes: a proof using continued fractions. L'Enseig. Mathematique **22**, 313–316 (1976)

[BH96] R.C. Baker, G. Harmon, The Brun-Titchmarsh theorem on average. Proceedings of a Conference in Honor of Heini Halberstam **1**, 39–103 (1996)

[BFKR] G. Baumslag, B. Fine, M. Kreuzer, G. Rosenberge, *A Course in Mathematical Cryptography* (DeGruyter, Berlin, 2015)

[BFX] G. Baumslag, B. Fine, X. Xu, Cryptosystems using linear groups, in *Proceedings of the International Conference on Algebraic Cryptography* (2005)

[Be] D. Bernstein, Proving primality after Agrawal, Kayena and Saxena (to appear)

[BK] M. Berry, J.P. Keating, The Riemann zeros and eigenvalue asymptotics. SIAM Rev. **41**(2), 236–266 (1999)

[Bo] F. Bornemann, PRIMES is in P: a breakthrough for everyman. Not. AMS **50**(5), 545–552 (2003)

[BFK] J.I. Burgos Gil, J. Fresan, U. Kühn, Classical and motivic multiple zeta values, in *Clay Mathematical Proceedings*, in preparation

[Br] J.W. Bruce, A really trivial proof of the Lucas Lehmer test. Am. Math. Monthly **100**, 370–371 (1993)

[C] E. Cohen, Legendre's identity. Am. Math. Monthly **76**, 611–616 (1969)

[Co] H. Cohn, *A Classical Invitation to Algebraic Numbers and Class Fields* (Springer, New York, 1978)

[CP] R. Crandall, C. Pomerance, *Prime Numbers; A Computational Perspective* (Springer, New York, 2001)

[CR] C. Curtis, I. Reiner, *Representation Theory of Finite Groups* (Wiley Interscience, New York, 1966)

[Da] H. Davenport, *Multiplicative Number Theory* (Springer, New York, 1980)

B. Fine and G. Rosenberger, *Number Theory*,
DOI 10.1007/978-3-319-43875-7

[DP] C.J. del la Vallee Poussin, Recherches analytiques sur la theorie des nombres: Premier partie: La fonction (s) de Riemann et les nombres premiers en general, Annales de la Soc. scientifique de Bruxelles **20**, 183–256 (1896)

[Di] H.G. Diamond, Elementary methods in the study of the distribution of prime numbers. Bull. Am. Math. Soc. **7**, 553–589 (1982)

[D] L.E. Dickson, *History of the Theory of Numbers* (Chelsea, New York, 1950)

[DH] W. Diffie, M. Hellman, New directions in cryptography. IEEE Trans. Inf. Theory **22**, 644–654 (1976)

[E] P.T.D. Elliott, *Probabilistic Number Theory I: Mean Value Theorems* (Springer, New York, 1979)

[E 1] P.T.D. Elliott, *Probabilistic Number Theory II: Central Limit Theorems* (Springer, New York, 1980)

[Er] P. Erdos, On a new method in elementary number theory which leads to an elementary proof of the prime number theorem. Proc. Nat. Acad. Sci. USA **35**, 374–384 (1949)

[F] B. Fine, *The Algebraic Theory of the Bianchi Groups* (Marcel Dekker, New York, 1989)

[F1] B. Fine, Sums of squares rings. Can. J. Math. **29**, 155–160 (1977)

[F2] B. Fine, A note on the two-square theorem. Can. Math. Bull. **20**, 93–94 (1977)

[F3] B. Fine, Cyclotomic equations and square properties in rings. Int. J. Math. Math. Sci. **9**, 89–95 (1986)

[FGR 1] B. Fine, A. Gaglione, G. Rosenberger, *Abstract Algebra* (Johns Hopkins Press, Baltimore, 2015)

[FR 1] B. Fine, G. Rosenberger, *Algebraic Generalizations of Discrete Groups* (Marcel Dekker, New York, 2001)

[FR 2] B. Fine, G. Rosenberger, *The Fundamental Theorem of Algebra* (Springer, New York, 1999)

[Fou85] E. Fouvry, Theoreme de Brun-Titchmarsh: application au theoreme de Fermat. Invent. Math. **79**, 383–407 (1985)

[Fr] J. Fraleigh, *A First Course in Abstract Algebra*, 7th edn. (Addison-Wesley, Reading, 2003)

[Fu] H. Furstenberg, On the infinitude of primes. Am. Math. Monthy **62**, 353 (1955)

[G] C.F. Gauss, *Disquisitiones Arithmeticae*, English edn. (Yale University Press, New Haven, 1966)

[Go] S.W. Golomb, A connected topology for integers. Am. Math. Monthly **24**, 663–665 (1966)

[GT] B. Green, T. Tao, The primes contain arbitarily long arithmetic progressions. Ann. Math. **167**, 481–547 (2008)

[Gr] M.D. Greenberg, *Advanced Engineering Mathematics* (Prentice Hall, Englewood Cliffs, 1988)

[Ha] J. Hadamard, Sur la distribution des zeros de la fonction $\zeta(s)$ et ses consequences arithmetiques. Bull. de la Soc. Math, de France **24**, 199–220 (1896)

[HR] H. Halberstam, H.E. Richert, *Sieve Methods* (Academic Press, London, 1974)

[HL] G.H. Hardy, J.E. Littlewood, A new solution of Waring's problem. Q. J. Math. **48**, 272–293 (1919)

[HW] G.H. Hardy, E.M. Wright, *An Introduction to the Theory of Numbers*, 5th edn. (Clarendon Press, Oxford, 1979)

[He] H.A. Helfgott, The ternary Goldbach conjecture is true, nal (2013). arXiv:1312.7748

[Ho] P. Hoffman, *Archimedes Revenge* (Fawcett Crest, New York, 1988)

[J] D. Johnson, *Presentations of Groups* (Cambridge University Press, Cambridge, 1990)

[K] S. Katok, *Fuchsian Groups* (University of Chicago Press, Chicago, 1992)

[KR 1] G. Kern-Isberner, G. Rosenberger, A note on numbers of the form $x^2 + Ny^2$. Arch. Math. **43**, 148–155 (1984)

[KR 2] G. Kern-Isberner, G. Rosenberger, Normalteiler vom Geschlecht eins in freien Produkten endlicher zyklischer Gruppen. Results Math. **11**, 272–288 (1987)

[Ko] N. Koblitz, *A Course in Number Theory and Cryptography* (Springer, New York, 1984)

[L] E. Landau, *Elementary Number Theory* (Chelsea, New York, 1958)

[Le] N. Levinson, More than one third of the zeros of Riemann's zeta function are on $\sigma = 1/2$. Adv. Math. **13**, 383–436 (1974)

[Lin] Y.V. Linnik, An elementary solution of Waring's problem by Shnirelman's method. Math. Sbornik **12**, 225–230 (1943)

[Li] J.E. Littlewood, Sur la distribution des nombres premiers. Comptes Rendus Acad. Sci. Paris **158**, 1869–1872 (1914)

[MSU] A.G. Myasnikov, V. Shpilrain, A. Ushakov, *Group-Based Cryptography, Advanced Courses in Mathematics - CRM Barcelona* (Birkhäuser, Basel, 2008)

[Ma] J. Maynard, Small gaps between primes. Ann. Math. **181**, 383–413 (2015)

[Mc] N. McCoy, *Elementary Number Theory* (Chelsea, New York, 1958)

[N] M. Nathanson, *Elementary Methods in Number Theory* (Springer, New York, 2000)

[Na] W. Narkiewicz, *The Development of Prime Number Theory* (Springer, New York, 2000)

[Neu] J. Neukirch, *Algebraic Number Theory* (Springer, New York, 1999)

[Ne] D.J. Newman, Simple analytic proof of the prime number theorem. Am. Math. Monthly **87**, 693–696 (1980)

[New 1] M. Newman, *Integral Matrices* (Academic Press, New York, 1972)

[New 2] M. Newman, Matrix representations of groups, National Bureau of Standards (1968)

[NP] I. Niven, B. Powell, Primes in certain arithmetic progressions. Am. Math. Monthly **83**, 467–475 (1976)

[NZ] I. Niven, H.S. Zuckerman, *The Theory of Numbers*, 4th edn. (Wiley, New York, 1980)

[NZM] I. Niven, H.S. Zuckerman, H.L. Montgomery, *The Theory of Numbers*, 5th edn. (Wiley, New York, 1991)

[O] O. Ore, *Number Theory and Its History* (McGraw-Hill, New York, 1949)

[P] O. Perron, *Die Lehre von den Kettenbrücken* (Chelsea, New York, 1957)

[PP] Prime pages. http://primes.utm

[PD] H. Pollard, H. Diamond, *The Theory of Algebraic Numbers*, vol. 9 (Carus Mathematical Monographs (Mathematical Association of America, Washington, 1975)

[Ri] P. Ribenboim, *The Book of Prime Number Records* (Springer, New York, 1989)

[Ri 2] P. Ribenboim, *The Little Book of Bigger Primes* (Springer, New York, 2004)

[Ri 3] P. Ribenboim, *Die Welt der Primzahlen* (Springer, New York, 2011)

[Re] H.J.J. Te Riele, On the sign of the difference $\pi(x) - Li(x)$. Math. Comput. **48**, 323–328 (1986)

[Rie] B. Riemann, Ueber die Anzahl der Primzahlen unter einer gegebener Groesse Monatsber. Kgl. Preuss, Akad. Wiss, Berlin, 671–680 (1860)

[RSA] R. Rivest, A. Shamir, L. Adelman, A method for obtaining digital signatures and public-key cryptosystems. Commun. ACM **21**, 120–126 (1978)

[Ro] J. Rotman, *The Theory of Groups* (W.C. Brown, Dubuque, 1984)

[Sc1] R. Schoof, Elliptic curves over finite fields and the computation of square roots mod p. Math. Comput. **44**, 483–494 (1985)

[Sc2] R. Schoof, Counting points on elliptic curves over finite fields. J. Theor. Nombres Bordeaux **7**, 219–254 (1995)

[S] B. Segal, Generalization du theoreme de Brun. Dokl. Akad. Nauk SSSR, 501–507 (1930)

[Se] A. Selberg, An elementary proof of the prime number theorem. Ann. Math. **50**, 305–313 (1949)

[Ser] J.P. Serre, *A Course in Arithmetic* (Springer, New York, 1973)

[Sil] J.H. Silverman, *The Arithmetic of Elliptic Curves* (Springer, Berlin, 1986)

[St] D.R. Stinson, *Cryptography: Theory and Practice* (Chapman and Hall/CRC, Boca Raton, 2002)

[TM] G. Tenenbaum, M. Mendes-France, *The Prime Numbers and Their Distribution*, vol. 6 Student Mathemtical Library (American Mathematical Society, Providence, 2000)

[Tu] J. Tupper, Lucas-Lehmer primality test. http://www.jt-actuary.com/lucas-le.htm

[VC] J.G. Van der Corput, Ueber Summen von Primzahlen und Primzahlquadraten. Math. Ann. **116**, 1–50 (1939)

[V] I.M. Vinogradov, *On Waring's Theorem Iz* (Akad, Nauk SSSR (English translation in selected works) (Springer, New York, 1985)

[W] A. Weil, *Number Theory; An Approach Through History* (Birkhauser, Boston, 1984)

[Wa] M. Waldschmidt, Lectures on multiple zeta values, IMSC 2011 (2012). http://www.math.jussieu.fr/~miw/articles/pdf/MZV2011IMSc.pdf

[Za] D. Zagier, *Die Ersten 50 Millionen Primzahlen* (Birkhauser, Boston, 1977)

[Zag] D. Zagier, Newman's short proof of the prime number theorem. Am. Math. Monthly **104**, 705–708 (1997)

[Zh] Y. Zhang, Bounded gaps between primes. Ann. Math. **179**, 1121–1174 (2014)

[Zud] V.V. Zudilin, Algebraic relations for multiple zeta values. Uspekhi Mat. Nauk **58**, 3–32 (2003)

Index

B. Fine and G. Rosenberger, *Number Theory*,
DOI 10.1007/978-3-319-43875-7