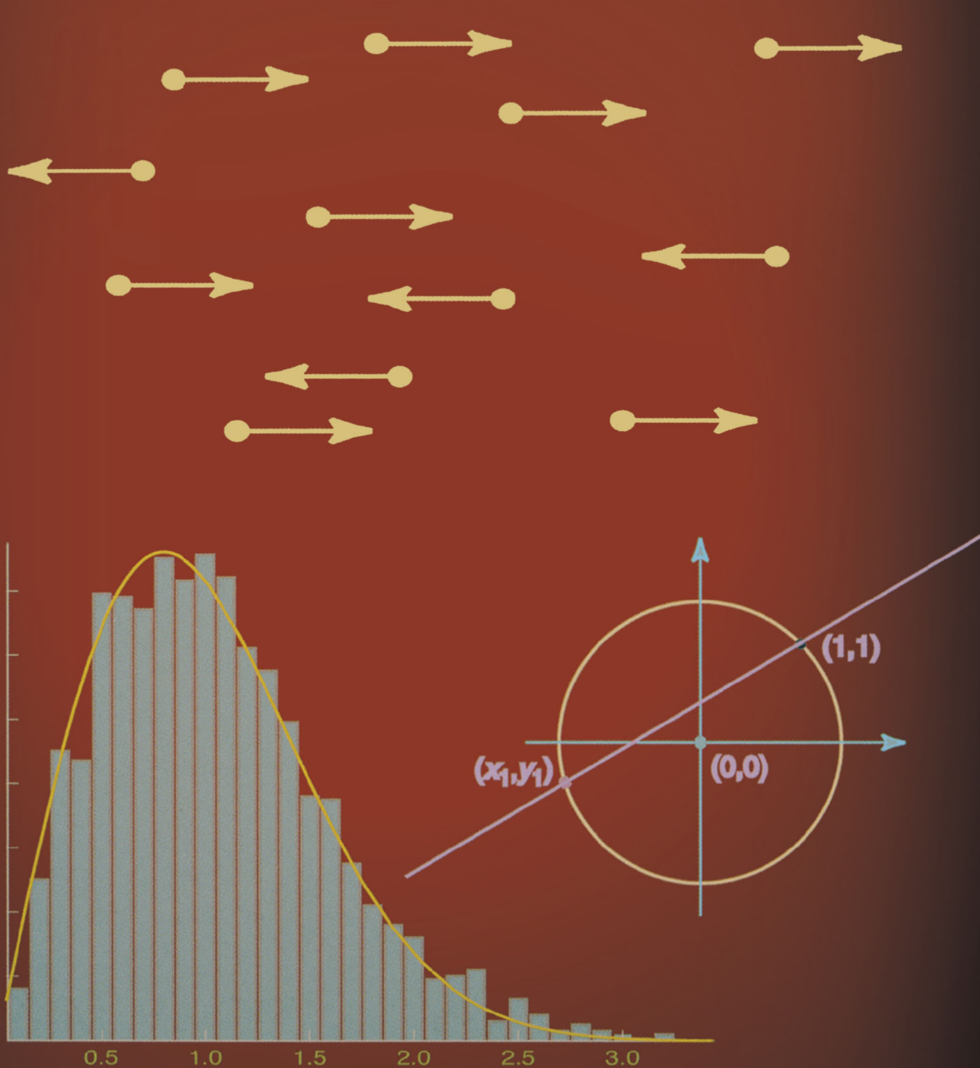
An Invitation to Modern Number Theory



Steven J. Miller & Ramin Takloo-Bighash

An Invitation to Modern Number Theory

An Invitation to Modern Number Theory

Steven J. Miller and
Ramin Takloo-Bighash

PRINCETON UNIVERSITY PRESS
PRINCETON AND OXFORD

Published by Princeton University Press, 41 William Street, Princeton, New Jersey 08540

In the United Kingdom: Princeton University Press, 3 Market Place, Woodstock, Oxfordshire OX20 1SY

Library of Congress Cataloging-in-Publication Data

Miller, Steven J.
An invitation to modern number theory / Steven J. Miller and Ramin Takloo-Bighash.
p. cm.
Includes bibliographical references and index.
ISBN-13: 978-0-691-12060-7 (cloth : alk. paper)
ISBN-10: 0-691-12060-9 (cloth : alk. paper)
1. Number theory. I. Takloo-Bighash, Ramin. II. Title.

QA241.M5344 2006
512.7—dc22 2005052165

British Library Cataloging-in-Publication Data is available

This book has been composed in Times Roman in LaTeX

The publisher would like to acknowledge the authors of this volume for providing the camera-ready copy from which this book was printed.

Printed on acid-free paper. ∞

pup.princeton.edu

Printed in the United States of America

10 9 8 7 6 5 4 3

ISBN-13: 978-0-691-12060-7 (cloth)

To our wives, for their patience and constant encouragement, and to our students and colleagues, whose insights and exuberance made this project both possible and enjoyable.

Contents

Foreword

The beginning student of physics, chemistry, or computer science learns early on that in order to gain a proper understanding of the subject, one has to understand seemingly different topics and their relation to one another. While this is equally true in mathematics, this feature is not brought to the fore in most modern texts of pure mathematics. Different fields are usually presented as complete and isolated topics, and for the most part this is how it should be. However, modern mathematics, abstract as it might appear to the beginner, is driven by concrete and basic problems. In fact many of the different areas were developed in attempts, sometimes successful, to resolve such fundamental questions. Hence it should not be surprising that often the solution to a concrete long-standing problem involves combining different areas. This is especially true of modern number theory. The formulation of the problems is mostly elementary, and the expected truths were many times discovered by numerical experimentation.

In this lovely book the authors introduce number theory in terms of its connections to other fields of mathematics and its applications. While adhering to this theme, they also emphasize concrete problems (solved and unsolved). They develop the requisite mathematical background along the way to ensure proper and clear treatment of each of the many topics discussed. This allows the beginner to get an immediate taste of modern mathematics as well as of mathematical research. Naturally the treatments of various theories and theorems cannot be as complete as in books which are devoted to a single topic; however, through the indicated further reading, the many excellent exercises and the proposed research projects, the reader will get an excellent first understanding of the material.

Parts of this book have been used very successfully in undergraduate courses at the junior level at Princeton, New York University, Ohio State, and Brown. It will no doubt find similar success much more broadly, and it should appeal to both more- and less-advanced students. Covering most of the material in this book is a challenging task for both the student (or reader) and the instructor. My experience in co-teaching (with one of the authors) a version of part of the material is that this effort results in rich rewards for both the student and the instructor.

Peter Sarnak
Princeton, NJ
December 2005

Preface

This book on modern number theory grew out of undergraduate research seminars taught at Princeton University (2001–2003), and similar courses taught at New York University, Ohio State, Brown University and a summer Research Experience for Undergraduates at the American Institute of Mathematics. The purpose of these classes was to expose undergraduates to current research topics in mathematics. To supplement the standard lecture-homework classes, we wanted a course where students could work on outstanding conjectures and open problems and experience firsthand the kinds of problems mathematicians study. In the sciences and engineering, undergraduates are often exposed to state of the art problems in experimental laboratories. We want to bring a similar experience to students interested in mathematics. This book is the outcome of that effort, providing the novice with hints as to what we feel is a good path through the immense landscape of number theory, as well as the needed background material. We have tried to give students and their teachers a model which can be used to develop their own research program; to this end, throughout the book are detailed descriptions of accessible open problems and references to the literature. Though we encourage students and teachers to attempt some of the open problems, the book stands alone and may be used for a standard lecture course (especially for new subjects such as Random Matrix Theory where there are not many introductory works accessible to undergraduates). Our goal is to supplement the classic texts in the field by showing the connections between seemingly diverse topics, as well as making some of the subjects more accessible to beginning students and whetting their appetite for continuing in mathematics.

The book has five parts, though several themes run throughout the book.

- Part I deals with basic number theory (cryptography and basic group theory), elementary L-functions (including the connections between zeros of $\zeta(s)$ and primes), and solutions to Diophantine equations. The material in this part is fairly standard, and could serve as an introduction to number theory. In some sections a little group theory and first semester complex analysis is assumed for some advanced topics. Our purpose in the first chapter is not to write a treatise on cryptography, but to review some of the background necessary from basic number theory for later chapters. It is possible to *motivate* this material in the context of cryptography; though these applications are very important, this connection is meant only to interest the reader, as this is not a exposition on cryptography. Similarly, elliptic curves are a terrific example for some of the material in Chapter 4 (and later in the book); as such, we introduce just enough for these purposes. As there are numerous excellent books on both of these subjects, we have kept our treatments short and refer the interested reader to these for more details. One theme in these chapters is the search for efficient algorithms, which appears frequently in later parts as well.

- Part II has two connected themes: approximating numbers with rationals, and continued fractions. In the first, the basic properties of algebraic and transcendental numbers are discussed, and a proof of Roth's Theorem (on how well algebraic numbers can be approximated by rationals) is given in full detail. This is one of the great achievement of 20th century number theory. Roth's Theorem has now been greatly generalized, and there are a few different ways to prove it. Our formulation and proof follow Roth's original proof. The proof we present here, though long and technical, requires only knowledge of elementary calculus and linear algebra. The second part is an introduction to continued fractions (a subject of interest in its own right, but also of use in approximation theory) and culminates in several open problems; this chapter is independent of Roth's Theorem and may serve as a survey to the subject. Also, time and again (especially in Part III when we study digit bias and spacings between terms in certain sequences), we see that answers to many number theoretic questions depend on properties of the numbers in the problem; often the continued fraction expansion highlights these properties. There are references to open problems in continued fractions, many of which concern the distribution of digits (see Part III).

- Part III encompasses three themes. The first is the distribution of the first digit of several interesting sequences (for example, the Fibonacci numbers and iterates of the $3x + 1$ map). We use this problem as a motivation for hypothesis testing (whether or not numerical data supports or contradicts conjectured behavior). Hypothesis testing is an extremely important subject, especially as computers are used more and more frequently in mathematics. The second theme centers around the Gauss-Kuzmin Theorem for the distribution of digits of continued fractions. We then develop enough Fourier Analysis to prove various basic results, including a sketch of the proof of the Central Limit Theorem and Poisson Summation (one of the most used tools in number theory). We use these results to investigate the behavior of $n^k\alpha \bmod 1$ for fixed k, α (specifically, the spacings between these numbers in $[0, 1]$; for many k and α these spacings appear to be the same as the spacings between adjacent primes); we study other spacing problems in Part V; in fact, our results on the Fourier transform are needed in Chapter 18 when we investigate zeros of L-functions. Numerous open problems and references to the current literature are provided.

- Part IV is a brief introduction to the Circle Method, a powerful theory to study questions in additive number theory (such as writing a number as a sum of a fixed number of k^{th} powers or primes). After developing the basics of the theory, we discuss in some detail why, using these methods, we cannot (yet?) show that any even number is the sum of two primes but we can show any large odd number is the sum of three primes. We use the Circle Method to predict how many Germain primes (p and $\frac{p-1}{2}$ both prime) are less than x. This example illustrates many of the key techniques of the theory, as well as the problems that arise in applications. Further, the density of these primes has recently been connected to fast primality testing algorithms. As usual we conclude with some open problems.

- Part V is an introduction to Random Matrix Theory and its interplay with number theory. What began as a model in the 1950s for physicists to study the energy levels of heavy nuclei has become a powerful tool after a chance encounter one day at tea in the 1970s (see [Ha] for an entertaining account of the meeting) for predicting the behavior of zeros of $\zeta(s)$ and other L-functions; knowledge of these zeros is intimately connected to properties of primes. The general result is that there is a striking similarity between the spacings between energy levels of heavy nuclei, eigenvalues of sets of matrices and zeros of L-functions. We take a classical approach to the subject. Results from linear algebra and occasionally first semester complex analysis are used (especially in the final chapter); a review of enough of the background material is provided for students to follow the key ideas in the proofs. There are numerous open problems requiring only elementary probability theory and linear algebra (at the level covered in this book); many have already been successfully investigated by our students.

There are several chapters throughout the book covering background material in basic number theory, algebra, Fourier analysis and probability theory, as well as two appendices on needed calculus, analysis and linear algebra results. Clearly our book is not meant to replace standard textbooks in these fields. We have two reasons for including these background chapters (in addition to the material being interesting in its own right). First, waiting for students to assemble such a background takes time, and the main purpose of our book is to show students in the early stages of their education what mathematicians do, and the interplay between the various parts of number theory and mathematics. Second, often very little of the background subjects is needed to understand the basic formulation and set-up of current work. Therefore a student who has not seen such material in a previous course can get a feel for these subjects by reading the review and background chapters, and then move on to the current research chapters. We have, however, written the chapters in such a way that there are often additional remarks or sections for students with stronger backgrounds. We have also included references throughout the book showing how the same methods and techniques are used for many different problems.

We have strived to keep the pre-requisites to a minimum: what is required is more a willingness to explore than a familiarity with the landscape. Several times we use results from later in the book in earlier investigations; our hope is that after seeing how these theorems are used and needed the reader will be motivated and interested enough to study the proofs. For most of the book one-variable calculus is the only requirement. We have also tried to emphasize common techniques in proofs (the reader is strongly encouraged to study the *techniques* entry in the index).

The book breaks naturally into five parts. Depending on the background of the students, and whether or not a class is going to explore open problems further, a typical semester class would cover material from one part of the book (as well as whatever background material is needed), though we recommend everyone at least skim Chapter 1 to ensure familiarity with the language and some of the motivating influences and themes of number theory. Many topics (such as applications to cryptography, algebraic structure of numbers and spacings between events) occur in various forms throughout the book. In a two semester course, one can cover two of the advanced parts and see these connections. We have also tried to give students the opportunity to discover the theory by themselves by giving many exercises. Mathematics is not meant to be a passive pursuit. Some of the problems are mere warm-ups; others are real problems that require time and effort. The reader should not be discouraged at being unable to work out all the problems. The value of an exercise is often in the time and energy spent on it, rather than the final solution. Many of the more difficult problems are standard theorems and can be seen proved in other textbooks. In this regard our manuscript is in the spirit of [Mu2]. **In Appendix C we have provided hints and further remarks to certain exercises; these problems are marked with either an (h) or (hr) in the text.**

We have assembled an extensive bibliography to aid the reader in further study. In addition to the excellent texts [AZ, Apo, BS, Da1, Da2, EE, Est2, Fe, HW, IR, IK, Kh, Kn, La2, Meh2, Na, NZM, ST, vdP6] on continued fractions, number theory and random matrix theory, we recommend the recent work of Narkiewicz [Nar] (where the reader will find proofs of many number theory results, as well as over 1800 references) as well as [Guy] (where there are extensive bibliographies for open problems). We conclude in Appendix D with some remarks on common themes running through this book and number theory.

The students in our courses used computers to assemble large amounts of data for some of the problems mentioned in the text, which then led us to appropriate conjectures and in some cases even gave us ideas on how to prove them. For links to previous student reports as well as some of the research papers mentioned in the bibliography, please visit

http://www.math.princeton.edu/mathlab/book/index.html

These include student programs (mostly in C++, Maple, Mathematica, MATLAB, or PARI) and detailed references for those interested in continuing these studies. Students should also consult MathSciNet [AMS], the arXiv [Cor1] and Project Euclid [Cor2] to find and download additional references.

It is a pleasure to thank the professors and teaching assistants who have helped run the class over the years (Alex Barnett, Vitaly Bergelson, João Boavida, Alexander Bufetov, Salman Butt, Brian Conrey, David Farmer, Harald Helfgott, Chris Hughes, James Mailhot, Atul Pokharel, Michael Rubinstein, Peter Sarnak, Lior Silberman, Yakov Sinai, Warren Sinnott, Florin Spinu and Andrew Wiles), as well as the students.

We would also like to thank several of our colleagues. In particular, we thank Eduardo Dueñez, Rob Gross and Amir Jafari for reviewing an early draft and providing numerous helpful suggestions to improve the presentation, and Timothy Abbot, Mike Buchanan, Scott Craver, Kevin Dayaratna, Dean Eiger, Manfred Einsiedler, Dan File, Chris Hammond, Ted Hill, Alex Kontorovich, Josh Krantz, Matt Michelini, Jeff Miller, Liz Miller, Paria Mirmonsef, C. J. Mozzochi, Anna Pierrehumbert, Amitabha Roy, Zeév Rudnick, Eitan Sayag, Aaron Silberstein, Dan Stone, Howard Straubing, Yuri Tschinkel, Akshay Venkatesh and Bill Zaboski for discussions and comments on various chapters. The first author gave several lectures on the material to a summer research group and the Ross Program at Ohio State (summer 2004), and is indebted to the students for their comments and suggestions. We are grateful to Nicole, Michelle and Leo Beaupre, Andrew and David Norris, Joe Silverman and the staff at Princeton University Press for help with the illustrations, and to Stephen Kudla for mutually productive LaTeX discussions.

We are extremely grateful to Princeton University Press, especially to our editor Vickie Kearn, our production editor Lucy Day W. Hobor and our copyeditor Jennifer Slater, for all their help and aid, to Bob Gunning, for initiating contact between us and PUP and encouraging us to write the book, and to the National Science Foundation's VIGRE program, which helped fund many of the classes at Princeton, NYU and Ohio State.

The first author was partially supported by VIGRE post-doctoral fellowships at Princeton, New York University, The Ohio State University and Brown University, and enjoyed the hospitality of Boston University during the final stages of the project. The second author enjoyed the hospitality of the University of Maryland at College Park, Johns Hopkins University and The Ohio State University at various stages of working on the project. His work was partially supported by a Young Investigator's Award from the National Security Agency.

Steven J. Miller
Providence, RI
December 2005

Ramin Takloo-Bighash
Princeton, NJ
December 2005

Notation

$\mathbb{W}$: the set of whole numbers: $\{1, 2, 3, 4, \ldots\}$.

$\mathbb{N}$: the set of natural numbers: $\{0, 1, 2, 3, \ldots\}$.

$\mathbb{Z}$: the set of integers: $\{\ldots, -2, -1, 0, 1, 2, \ldots\}$.

$\mathbb{Q}$: the set of rational numbers: $\{x : x = \frac{p}{q}, p, q \in \mathbb{Z}, q \neq 0\}$.

$\mathbb{R}$: the set of real numbers.

$\mathbb{C}$: the set of complex numbers: $\{z : z = x + iy,\ x, y \in \mathbb{R}\}$.

$\Re z, \Im z$: the real and imaginary parts of $z \in \mathbb{C}$; if $z = x + iy$, $\Re z = x$ and $\Im z = y$.

$\mathbb{Z}/n\mathbb{Z}$: the additive group of integers mod n: $\{0, 1, \ldots, n-1\}$.

$(\mathbb{Z}/n\mathbb{Z})^*$: the multiplicative group of invertible elements mod n.

$\mathbb{F}_p$: the finite field with p elements: $\{0, 1, \ldots, p-1\}$.

$a|b$: a divides b.

$p^k||b$: p^k divides b and p^{k+1} does not divide b.

(a, b) : greatest common divisor (gcd) of a and b, also written $\gcd(a, b)$.

prime, composite : a positive integer a is prime if $a > 1$ and the only divisors of a are 1 and a; if $a > 1$ is not prime, we say a is composite.

coprime (relatively prime) : a and b are coprime (or relatively prime) if their greatest common divisor is 1.

$x \equiv y \mod n$: there exists an integer a such that $x = y + an$.

$\forall$: for all.

$\exists$: there exists.

Big-Oh notation : $A(x) = O(B(x))$, read "$A(x)$ is of order (or big-Oh) $B(x)$", means $\exists C > 0$ and an x_0 such that $\forall x \geq x_0$, $|A(x)| \leq C\,B(x)$. This is also written $A(x) \ll B(x)$ or $B(x) \gg A(x)$.

Little-Oh notation : $A(x) = o(B(x))$, read "$A(x)$ is little-Oh of $B(x)$", means $\lim_{x\to\infty} A(x)/B(x) = 0$.

$|S|$ or $\#S$: number of elements in the set S.

p : usually a prime number.

i, j, k, m, n : usually an integer.

$[x]$ or $\lfloor x \rfloor$: the greatest integer less than or equal to x, read "the floor of x".

$\{x\}$: the fractional part of x; note $x = [x] + \{x\}$.

supremum : given a sequence $\{x_n\}_{n=1}^{\infty}$, the supremum of the set, denoted $\sup_n x_n$, is the smallest number c (if one exists) such that $x_n \leq c$ for all n, and for any $\epsilon > 0$ there is some n_0 such that $x_{n_0} > c - \epsilon$. If the sequence has finitely many terms, the supremum is the same as the maximum value.

infimum : notation as above, the infimum of a set, denoted $\inf_n x_n$, is the largest number c (if one exists) such that $x_n \geq c$ for all n, and for any $\epsilon > 0$ there is some n_0 such that $x_{n_0} < c + \epsilon$. If the sequence has finitely many terms, the infimum is the same as the minimum value.

$\Box$: indicates the end of a proof.

PART 1
Basic Number Theory

Chapter One

Mod p Arithmetic, Group Theory and Cryptography

In this chapter we review the basic number theory and group theory which we use throughout the book, culminating with a proof of quadratic reciprocity. Good introductions to group theory are [J, La3]; see [Da1, IR] for excellent expositions on congruences and quadratic reciprocity, and [Sil2] for a friendly introduction to much of the material below. We use cryptographic applications to motivate some basic background material in number theory; see [Ga] for a more detailed exposition on cryptography and [Lidl, vdP2] for connections with continued fractions. The guiding principle behind much of this chapter (indeed, much of this book and number theory) is the search for efficient algorithms. Just being able to write down an expression does not mean we can evaluate it in a reasonable amount of time. Thus, while it is often easy to prove a solution exists, doing the computations as written is sometimes impractical; see Chapter 6 of [BB] and [Wilf] for more on efficient algorithms.

1.1 CRYPTOGRAPHY

Cryptography is the science of encoding information so that only certain specified people can decode it. We describe some common systems. To prove many of the properties of these crypto-systems will lead us to some of the basic concepts and theorems of algebra and group theory.

Consider the following two password systems. In the first we choose two large distinct primes p and q; for example, let us say p and q have about 200 digits each. Let $N = pq$ and display the 400 digit number N for everyone to see. The password is any divisor of N greater than 1 and less than N. One very important property of the integers is unique factorization: any integer can be written uniquely as a product of prime powers. This implies that the only factorizations of N are $1 \cdot N$, $N \cdot 1$, $p \cdot q$ and $q \cdot p$. Thus there are two passwords, p and q. For the second system, we choose a 5000 digit number. We keep this number secret; to gain access the user must input this number.

Which method is more secure? While it is harder to correctly guess 5000 digits then 200, there is a danger in the second system: the computer needs to store the password. As there is no structure to the problem, the computer can only determine if you have entered the correct number by comparing your 5000 digit number to the one it was told is the password. Thus there is a code-book of sorts, and code-books can be stolen. In the first system there is no code-book to steal. The computer does not need to know p or q: it only needs to know N and how to divide, and it will know the password when it sees it!

There are so many primes that it is not practical to try all 200 digit prime numbers. The Prime Number Theorem (Theorem 2.3.7) states that there are approximately $\frac{x}{\log x}$ primes smaller than x; for $x = 10^{200}$, this leads to an impractically large number of numbers to check. What we have is a process which is easy in one direction (multiplying p and q), but hard in the reverse (knowing N, right now there is no "fast" algorithm to find p and q).

It is trivial to write an algorithm which is guaranteed to factor N: simply test N by all numbers (or all primes) at most $\sqrt{N}$. While this will surely work, this algorithm is so inefficient that it is useless for such large numbers. This is the first of many instances where we have an algorithm which will give a solution, but the algorithm is so slow as to be impractical for applications. Later in this chapter we shall encounter other situations where we have an initial algorithm that is too slow but where we can derive faster algorithms.

Exercise 1.1.1. *There are approximately 10^{80} elementary objects in the universe (photons, quarks, et cetera). Assume each such object is a powerful supercomputer capable of checking 10^{20} numbers a second. How many years would it take to check all numbers (or all primes) less than $\sqrt{10^{400}}$? What if each object in the universe was a universe in itself, with 10^{80} supercomputers: how many years would it take now?*

Exercise 1.1.2. *Why do we want p and q to be distinct primes in the first system?*

One of the most famous cryptography methods is RSA (see [RSA]). Two people, usually named Alice and Bob, want to communicate in secret. Instead of sending words they send numbers that represent words. Let us represent the letter a by 01, b by 02, all the way to representing z by 26 (and we can have numbers represent capital letters, spaces, punctuation marks, and so on). For example, we write 030120 for the word "cat." Thus it suffices to find a secure way for Alice to transmit numbers to Bob. Let us say a message is a number M of a fixed number of digits.

Bob chooses two large primes p and q and then two numbers d and e such that $(p-1)(q-1)$ divides $ed-1$; we explain these choices in §1.5. Bob then makes publicly available the following information: $N = pq$ and e, but keeps secret p, q and d. It turns out that this allows Alice to send messages to Bob that only Bob can easily decipher. If Alice wants to send the message $M < N$ to Bob, Alice first calculates M^e, and then sends Bob the remainder after dividing by N; call this number X. Bob then calculates X^d, whose remainder upon dividing by N is the original message M! The proof of this uses modulo (or clock) arithmetic and basic group theory, which we describe below. Afterwards, we return and prove the claim.

Exercise 1.1.3. *Let $p = 101$, $q = 97$. Let $d = 2807$ and $e = 23$. Show that this method successfully sends "hi" (0809) to Bob. Note that $(0809)^{23}$ is a sixty-six digit number! See Remark 9.5.6 for one way to handle such large numbers.*

Exercise$^{\textbf{(hr)}}$ 1.1.4. *Use a quadratic polynomial $ax^2 + bx + c$ to design a security system satisfying the following constraints:*

1. *the password is the triple (a, b, c);*

2. *each of* 10 *people is given some information such that any three of them can provide* (a, b, c)*, but no two of them can.*

Generalize the construction: consider a polynomial of degree N *such that some people "know more" than others (for example, one person can figure out the password with anyone else, another person just needs two people, and so on).*

Remark 1.1.5. We shall see another important application of unique factorization in §3.1.1 when we introduce the Riemann zeta function. Originally defined as an infinite sum over the integers, by unique factorization we shall be able to express it as a product over primes; this interplay yields numerous results, among them a proof of the Prime Number Theorem.

1.2 EFFICIENT ALGORITHMS

For computational purposes, often having an algorithm to compute a quantity is not enough; we need an algorithm which will compute it *quickly*. We have seen an example of this when we tried to factor numbers; while we can factor any number, current algorithms are so slow that crypto-systems based on "large" primes are secure. For another example, recall Exercise 1.1.3 where we needed to compute a sixty-six digit number! Below we study three standard problems and show how to either rearrange the operations more efficiently or give a more efficient algorithm than the obvious candidate. See Chapter 6 of [BB] and [Wilf] for more on efficient algorithms.

1.2.1 Exponentiation

Consider x^n. The obvious way to calculate it involves $n-1$ multiplications. By writing n in base two we can evaluate x^n in at most $2\log_2 n$ steps, an enormous savings. One immediate application is to reduce the number of multiplications in cryptography (see Exercise 1.1.3). Another is in §1.2.33, where we derive a primality test based on exponentiation.

We are used to writing numbers in base 10, say

$$x \;=\; a_m 10^m + a_{m-1}10^{m-1} + \cdots + a_1 10^1 + a_0, \quad a_i \in \{1,2,3,4,5,6,7,8,9\}. \tag{1.1}$$

Base two is similar, except each digit is now either 0 or 1. Let k be the largest integer such that $2^k \le x$. Then

$$x \;=\; b_k 2^k + b_{k-1}2^{k-1} + \cdots + b_1 2 + b_0, \quad b_i \in \{0,1\}. \tag{1.2}$$

It costs k multiplications to evaluate x^{2^i} for all $i \le k$. How? Consider $y_0 = x^{2^0}$, $y_1 = y_0 \cdot y_0 = x^{2^0} \cdot x^{2^0} = x^{2^1}$, $y_2 = y_1 \cdot y_1 = x^{2^2}, \dots, y_k = y_{k-1} \cdot y_{k-1} = x^{2^k}$.

To evaluate x^n, note

$$\begin{aligned}
x^n &= x^{b_k 2^k + b_{k-1} 2^{k-1} + \cdots + b_1 2 + b_0} \\
&= x^{b_k 2^k} \cdot x^{b_{k-1} 2^{k-1}} \cdots x^{b_1 2} \cdot x^{b_0} \\
&= \left(x^{2^k}\right)^{b_k} \cdot \left(x^{2^{k-1}}\right)^{b_{k-1}} \cdots \left(x^2\right)^{b_1} \cdot \left(x^1\right)^{b_0} \\
&= y_k^{b_k} \cdot y_{k-1}^{b_{k-1}} \cdots y_1^{b_1} \cdot y_0^{b_0}.
\end{aligned} \tag{1.3}$$

As each $b_i \in \{0, 1\}$, we have at most $k+1$ multiplications above (if $b_i = 1$ we have the term y_i in the product, if $b_i = 0$ we do not). It costs k multiplications to evaluate the x^{2^i} $(i \leq k)$, and at most another k multiplications to finish calculating x^n. As $k \leq \log_2 n$, we see that x^n can be determined in at most $2 \log_2 n$ steps. Note, however, that we do need more storage space for this method, as we need to store the values $y_i = x^{2^i}$, $i \leq \log_2 n$. For n large, $2 \log_2 n$ is much smaller than $n - 1$, meaning there is enormous savings in determining x^n this way. See also Exercise B.1.13.

Exercise 1.2.1. *Show that it is possible to calculate x^n storing only two numbers at any given time (and knowing the base two expansion of n).*

Exercise 1.2.2. *Instead of expanding n in base two, expand n in base three. How many calculations are needed to evaluate x^n this way? Why is it preferable to expand in base two rather than any other base?*

Exercise 1.2.3. *A better measure of computational complexity is not to treat all multiplications and additions equally, but rather to count the number of digit operations. For example, in 271×31 there are six multiplications. We then must add two three-digit numbers, which involves at most four additions (if we need to carry). How many digit operations are required to compute x^n?*

1.2.2 Polynomial Evaluation (Horner's Algorithm)

Let $f(x) = a_n x^n + a_{n-1} x^{n-1} + \cdots + a_1 x + a_0$. The obvious way to evaluate $f(x)$ is to calculate x^n and multiply by a_n (n multiplications), calculate x^{n-1} and multiply by a_{n-1} ($n-1$ multiplications) and add, et cetera. There are n additions and $\sum_{k=0}^{n} k$ multiplications, for a total of $n + \frac{n(n+1)}{2}$ operations. Thus the standard method leads to about $\frac{n^2}{2}$ computations.

Exercise 1.2.4. *Prove by induction (see Appendix A.1) that $\sum_{k=0}^{n} k = \frac{n(n+1)}{2}$. In general, $\sum_{k=0}^{n} k^d = p_{d+1}(n)$, where $p_{d+1}(n)$ is a polynomial of degree $d+1$ with leading term $\frac{n^{d+1}}{d+1}$; one can find the coefficients by evaluating the sums for $n = 0, 1, \ldots, d$ because specifying the values of a polynomial of degree d at $d+1$ points uniquely determines the polynomial (see also Exercise 1.1.4). See [Mil4] for an alternate proof which does not use induction.*

Exercise 1.2.5. *Notation as in Exercise 1.2.4, use the integral test from calculus to show the leading term of $p_{d+1}(n)$ is $\frac{n^{d+1}}{d+1}$ and bound the size of the error.*

Exercise 1.2.6. *How many operations are required if we use our results on exponentiation?*

Consider the following grouping to evaluate $f(x)$, known as **Horner's algorithm**:

$$(\cdots((a_n x + a_{n-1})x + a_{n-2})\,x + \cdots + a_1)\,x + a_0. \tag{1.4}$$

For example,

$$7x^4 + 4x^3 - 3x^2 - 11x + 2 = (((7x+4)x-3)\,x - 11)\,x + 2. \tag{1.5}$$

Evaluating term by term takes 14 steps; Horner's Algorithm takes 8 steps. One common application is in fractal geometry, where one needs to iterate polynomials (see also §1.2.4 and the references there). Another application is in determining decimal expansions of numbers (see §7.1).

Exercise 1.2.7. *Prove Horner's Algorithm takes at most $2n$ steps to evaluate $a_n x^n + \cdots + a_0$.*

1.2.3 Euclidean Algorithm

Definition 1.2.8 (Greatest Common Divisor)**.** *Let $x, y \in \mathbb{N}$. The greatest common divisor of x and y, denoted by $\gcd(x,y)$ or (x,y), is the largest integer which divides both x and y.*

Definition 1.2.9 (Relatively Prime, Coprime)**.** *If for integers x and y, $\gcd(x,y) = 1$, we say x and y are relatively prime (or coprime).*

The **Euclidean algorithm** is an efficient way to determine the greatest common divisor of x and y. Without loss of generality, assume $1 < x < y$. The obvious way to determine $\gcd(x,y)$ is to divide x and y by all positive integers up to x. This takes at most $2x$ steps; we show a more efficient way, taking at most about $2\log_2 x$ steps.

Let $[z]$ denote the **greatest integer** less than or equal to z. We write

$$y \;=\; \left[\frac{y}{x}\right]\cdot x + r_1, \quad 0 \le r_1 < x. \tag{1.6}$$

Exercise 1.2.10. *Prove that $r_1 \in \{0, 1, \dots, x-1\}$.*

Exercise 1.2.11. *Prove $\gcd(x,y) = \gcd(r_1, x)$.*

We proceed in this manner until r_k equals zero or one. As each execution results in $r_i < r_{i-1}$, we proceed at most x times (although later we prove we need to apply these steps at most about $2\log_2 x$ times).

$$\begin{aligned}
x &= \left[\frac{x}{r_1}\right] \cdot r_1 + r_2, \;\; 0 \le r_2 < r_1 \\
r_1 &= \left[\frac{r_1}{r_2}\right] \cdot r_2 + r_3, \;\; 0 \le r_3 < r_2 \\
r_2 &= \left[\frac{r_2}{r_3}\right] \cdot r_3 + r_4, \;\; 0 \le r_4 < r_3 \\
&\;\;\vdots \\
r_{k-2} &= \left[\frac{r_{k-2}}{r_{k-1}}\right] \cdot r_{k-1} + r_k, \;\; 0 \le r_k < r_{k-1}.
\end{aligned} \tag{1.7}$$

Exercise 1.2.12. *Prove that if* $r_k = 0$ *then* $\gcd(x, y) = r_{k-1}$*, while if* $r_k = 1$*, then* $\gcd(x, y) = 1$.

We now analyze how large k can be. The key observation is the following:

Lemma 1.2.13. *Consider three adjacent remainders in the expansion:* r_{i-1}*,* r_i *and* r_{i+1} *(where* $y = r_{-1}$ *and* $x = r_0$*). Then* $\gcd(r_i, r_{i-1}) = \gcd(r_{i+1}, r_i)$*, and* $r_{i+1} < \frac{r_{i-1}}{2}$.

Proof. We have the following relation:

$$r_{i-1} \;=\; \left[\frac{r_{i-1}}{r_i}\right] \cdot r_i + r_{i+1}, \; 0 \le r_{i+1} < r_i. \tag{1.8}$$

If $r_i \le \frac{r_{i-1}}{2}$ then as $r_{i+1} < r_i$ we immediately conclude that $r_{i+1} < \frac{r_{i-1}}{2}$. If $r_i > \frac{r_{i-1}}{2}$, then we note that

$$r_{i+1} \;=\; r_{i-1} - \left[\frac{r_{i-1}}{r_i}\right] \cdot r_i. \tag{1.9}$$

Our assumptions on r_{i-1} and r_i imply that $\left[\frac{r_{i-1}}{r_i}\right] = 1$. Thus $r_{i+1} < \frac{r_{i-1}}{2}$. □

We count how often we apply these steps. Going from $(x, y) = (r_0, r_{-1})$ to (r_1, r_0) costs one application. Every two applications gives three pairs, say (r_{i-1}, r_{i-2}), (r_i, r_{i-1}) and (r_{i+1}, r_i), with r_{i+1} at most half of r_{i-1}. Thus if k is the largest integer such that $2^k \le x$, we see have at most $1 + 2k \le 1 + 2\log_2 x$ pairs. Each pair requires one integer division, where the remainder is the input for the next step. We have proven

Lemma 1.2.14. *Euclid's algorithm requires at most* $1 + 2\log_2 x$ *divisions to find the greatest common divisor of* x *and* y.

Euclid's algorithm provides more information than just the $\gcd(x, y)$. Let us assume that $r_i = \gcd(x, y)$. The last equation before Euclid's algorithm terminated was

$$r_{i-2} \;=\; \left[\frac{r_{i-2}}{r_{i-1}}\right] \cdot r_{i-1} + r_i, \; 0 \le r_i < r_{i-1}. \tag{1.10}$$

Therefore we can find integers a_{i-1} and b_{i-2} such that

$$r_i \;=\; a_{i-1}r_{i-1} + b_{i-2}r_{i-2}. \tag{1.11}$$

We have written r_i as a linear combination of r_{i-2} and r_{i-1}. Looking at the second to last application of Euclid's algorithm, we find that there are integers a'_{i-2} and b'_{i-3} such that

$$r_{i-1} \;=\; a'_{i-2}r_{i-2} + b'_{i-3}r_{i-3}. \tag{1.12}$$

Substituting for r_{i-1} in the expansion of r_i yields that there are integers a_{i-2} and b_{i-3} such that

$$r_i \;=\; a_{i-2}r_{i-2} + b_{i-3}r_{i-3}. \tag{1.13}$$

Continuing by induction and recalling $r_i = \gcd(x, y)$ yields

Lemma 1.2.15. *There exist integers a and b such that* $\gcd(x, y) = ax + by$. *Moreover, Euclid's algorithm gives a* constructive *procedure to find a and b.*

Thus, not only does Euclid's algorithm show that a and b exist, it gives an efficient way to find them.

Exercise 1.2.16. *Find a and b such that* $a \cdot 244 + b \cdot 313 = \gcd(244, 313)$.

Exercise 1.2.17. *Add the details to complete an alternate, non-constructive proof of the existence of a and b with $ax + by = \gcd(x, y)$:*

1. *Let d be the smallest positive value attained by $ax + by$ as we vary $a, b \in \mathbb{Z}$. Such a d exists. Say $d = \alpha x + \beta y$.*
2. *Show* $\gcd(x, y)|d$.
3. *Let $e = Ax + By > 0$. Then $d|e$. Therefore for any choice of $A, B \in \mathbb{Z}$, $d|(Ax + By)$.*
4. *Consider $(a, b) = (1, 0)$ or $(0, 1)$, yielding $d|x$ and $d|y$. Therefore $d \leq \gcd(x, y)$. As we have shown $\gcd(x, y)|d$, this completes the proof.*

Note this is a non-constructive proof. By minimizing $ax + by$ we obtain $\gcd(x, y)$, but we have no idea how many steps are required. Prove that a solution will be found either among pairs (a, b) with $a \in \{1, \dots, y-1\}$ and $-b \in \{1, \dots, x-1\}$, or $-a \in \{1, \dots, y-1\}$ and $b \in \{1, \dots, x-1\}$. Choosing an object that is minimal in some sense (here the minimality comes from being the smallest integer attained as we vary a and b in $ax + by$) is a common technique; often this number has the desired properties. See the proof of Lemma 6.4.3 for an additional example of this method.

Exercise 1.2.18. *How many steps are required to find the greatest common divisor of $x_1, \dots, x_N$?*

Remark 1.2.19. In bounding the number of computations in the Euclidean algorithm, we looked at three adjacent remainders and showed that a desirable relation held. This is a common technique, where it can often be shown that at least one of several consecutive terms in a sequence has some good property; see also Theorem 7.9.4 for an application to continued fractions and approximating numbers.

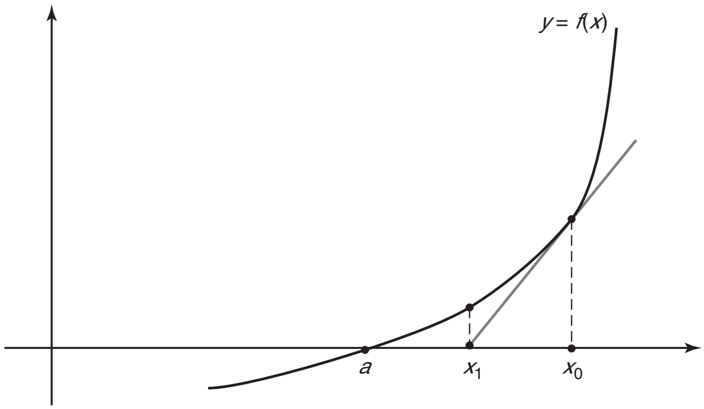


Figure 1.1 Newton's Method

1.2.4 Newton's Method and Combinatorics

We give some examples and exercises on efficient algorithms and efficient ways to arrange computations. The first assumes some familiarity with calculus, the second with basic combinatorics.

Newton's Method: Newton's Method is an algorithm to approximate solutions to $f(x) = 0$ for f a differentiable function on $\mathbb{R}$. It is much faster than the method of **Divide and Conquer** (see §A.2.1), which finds zeros by looking at sign changes of f, though this method is of enormous utility (see Remark 3.2.24 where Divide and Conquer is used to find zeros of the Riemann zeta function).

Start with x_0 such that $f(x_0)$ is small; we call x_0 the initial guess. Draw the tangent line to the graph of f at x_0, which is given by the equation

$$y - f(x_0) \;=\; f'(x_0) \cdot (x - x_0). \tag{1.14}$$

Let x_1 be the x-intercept of the tangent line; x_1 is the next guess for the root α. See Figure 1.1. Simple algebra gives

$$x_1 \;=\; x_0 - \frac{f(x_0)}{f'(x_0)}. \tag{1.15}$$

We now iterate and apply the above procedure to x_1, obtaining

$$x_2 \;=\; x_1 - \frac{f(x_1)}{f'(x_1)}. \tag{1.16}$$

If we let $g(x) = x - \frac{f(x)}{f'(x)}$, we notice we have the sequence

$$x_0, \;\; g(x_0), \;\; g(g(x_0)), \;\; \dots \tag{1.17}$$

We hope this sequence will converge to the root, at least for x_0 close to the root and for f sufficiently nice. How close x_0 has to be is a delicate matter. If there are several roots to f, which root the sequence converges to depends crucially on the initial value x_0 and the function f. In fact its behavior is what is known technically as **chaotic**. Informally, we say that we have chaos when tiny changes in the initial value give us very palpable changes in the output. One common example is in iterates of polynomials, namely the limiting behavior of $f(x_0)$, $f(f(x_0))$, $f(f(f(x_0)))$ and so on; see [Dev, Edg, Fal, Man].

Exercise 1.2.20. *Let $f(x) = x^2 - a$ for some $a > 0$. Show Newton's Method converges to $\sqrt{a}$, and discuss the rate of convergence; i.e., if x_n is accurate to m digits, approximately how accurate is x_{n+1}? For example, look at $a = 3$ and $x_0 = 2$. Similarly, investigate $\sqrt[n]{a}$. Compare this with Divide and Conquer, where each iteration basically halves the error (so roughly every ten iterations yields three new decimal digits, because $\frac{1}{2^{10}} \approx \frac{1}{10^3}$).*

Remark 1.2.21. One big difference between Newton's Method and Divide and Conquer is that while both require us to evaluate the function, Newton's Method requires us to evaluate the derivative as well. Hence Newton's Method is not applicable to as wide of a class of functions as Divide and Conquer, but as it uses more information about f it is not surprising that it gives better results (i.e., converges faster to the answer).

Exercise 1.2.22. *Modify Newton's Method to find maxima and minima of functions. What must you assume about these functions to use Newton's method?*

Exercise 1.2.23. *Let $f(x)$ be a degree n polynomial with complex coefficients. By the Fundamental Theorem of Algebra, there are n (not necessarily distinct) roots. Assume there are m distinct roots. Assign m colors, one to each root. Given a point $x \in \mathbb{C}$, we color x with the color of the root that x approaches under Newton's Method (if it converges to a root). Write a computer program to color such sets for some simple polynomials, for example for $x^n - 1 = 0$ for $n = 2, 3$ or 4.*

Exercise 1.2.24. *Determine conditions on f, the root a and the starting guess x_0 such that Newton's Method will converge to the root. See page 212 of [BB] or page 118 of [Rud] for more details.*

Exercise$^{\text{(h)}}$ 1.2.25 (Fixed Points). *We say x_0 is a fixed point of a function h if $h(x_0) = x_0$. Let f be a continuously differentiable function. If we set $g(x) = x - \frac{f(x)}{f'(x)}$, show a fixed point of g corresponds to a solution to $f(x) = 0$.*

Assume that $f : [a, b] \to [a, b]$ and there is a $C < 1$ such that $|f'(x)| < C$ for $x \in [a, b]$. Prove f has a fixed point in $[a, b]$. Is the result still true if we just assume $|f'(x)| < 1$? Fixed points have numerous applications, among them showing optimal strategies exist in n-player games. See [Fr] for more details.

Combinatorics: Below we describe a combinatorial problem which contains many common features of the subject. Assume we have 10 identical cookies and 5 distinct people. How many different ways can we divide the cookies among the people, such that all 10 cookies are distributed? Since the cookies are identical, we cannot tell which cookies a person receives; we can only tell how many. We could enumerate all possibilities: there are 5 ways to have one person receive 10 cookies, 20 ways to have one person receive 9 and another receive 1, and so on. While in principle we can solve the problem, in practice this computation becomes intractable, especially as the numbers of cookies and people increase.

We introduce common combinatorial functions. The first is the **factorial function**: for a positive integer n, set $n! = n \cdot (n-1) \cdots 2 \cdot 1$. The number of ways to choose r objects from n when order matters is $n \cdot (n-1) \cdots (n-(r-1)) = \frac{n!}{(n-r)!}$ (there are n ways to choose the first element, then $n-1$ ways to choose the second element, and so on). The **binomial coefficient** $\binom{n}{r} = \frac{n!}{r!(n-r)!}$ is the number of ways to choose r objects from n objects when order does not matter. The reason is that once we have chosen r objects there are $r!$ ways to order them. For convenience, we define $0! = 1$; thus $\binom{n}{0} = 1$, which may be interpreted as saying there is one way to choose zero elements from a set of n objects. For more on binomial coefficients, see §A.1.3.

We show the number of ways to divide 10 cookies among 5 people is $\binom{10+5-1}{5-1}$. In general, if there are C cookies and P people,

Lemma 1.2.26. *The number of distinct ways to divide C identical cookies among P different people is $\binom{C+P-1}{P-1}$.*

Proof. Consider $C+P-1$ cookies in a line, and number them 1 to $C+P-1$. Choose $P-1$ cookies. There are $\binom{C+P-1}{P-1}$ ways to do this. This divides the cookies into P sets: all the cookies up to the first chosen (which gives the number of cookies the first person receives), all the cookies between the first chosen and the second chosen (which gives the number of cookies the second person receives), and so on. This divides C cookies among P people. Note different sets of $P-1$ cookies correspond to different partitions of C cookies among P people, and every such partition can be associated to choosing $P-1$ cookies as above. $\square$

Remark 1.2.27. In the above problem we do not care *which* cookies a person receives. We introduced the numbers for convenience: now cookies 1 through i_1 (say) are given to person 1, cookies $i_1 + 1$ through i_2 (say) are given to person 2, and so on.

For example, if we have 10 cookies and 5 people, say we choose cookies $3, 4, 7$ and 13 of the $10+5-1$ cookies:

$$\odot \quad \odot \quad \otimes \quad \otimes \quad \odot \quad \odot \quad \otimes \quad \odot \quad \odot \quad \odot \quad \odot \quad \odot \quad \otimes \quad \odot$$

This corresponds to person 1 receiving two cookies, person 2 receiving zero, person 3 receiving two, person 4 receiving five and person 5 receiving one cookie.

The above is an example of a partition problem: we are solving $x_1 + x_2 + x_3 + x_4 + x_5 = 10$, where x_i is the number of cookies person i receives. We may interpret Lemma 1.2.26 as the number of ways to divide an integer N into k non-negative integers is $\binom{N+k-1}{k-1}$.

Exercise 1.2.28. *Prove that*

$$\sum_{n=0}^{N} \binom{n+k-1}{k-1} = \binom{N+1+k-1}{k-1}. \tag{1.18}$$

We may interpret the above as dividing N cookies among k people, where we do not assume all cookies are distributed.

Exercise^(h) 1.2.29. *Let $\mathcal{M}$ be a set with $m > 0$ elements, $\mathcal{N}$ a set with $n > 0$ elements and $\mathcal{O}$ a set with $m+n$ elements. For $\ell \in \{0, \dots, m+n\}$, prove*

$$\sum_{k=\max(0,\ell-n)}^{\min(m,\ell)} \binom{m}{k}\binom{n}{\ell-k} \ = \ \binom{m+n}{\ell}. \tag{1.19}$$

This may be interpreted as partitioning $\mathcal{O}$ into two sets, one of size ℓ.

In Chapter 13 we describe other partition problems, such as representing a number as a sum of primes or integer powers. For example, the famous Goldbach problem says any even number greater than 2 is the sum of two primes (known to be true for integers up to $6 \cdot 10^{16}$ [Ol]). While to date this problem has resisted solution, we have good heuristics which predict that, not only does a solution exist, but how many solutions there are. Computer searches have verified these predictions for large N of size 10^{10}.

Exercise 1.2.30 (Crude Prediction)**.** *By the Prime Number Theorem, there are $\frac{N}{\log N}$ primes less than N. If we assume all numbers $n \le N$ are prime with the same likelihood (a crude assumption), predict how many ways there are to write N as a sum of two primes.*

Exercise 1.2.31. *In partition problems, often there are requirements such as that everyone receives at least one cookie. How many ways are there to write N as a sum of k non-negative integers? How many solutions of $x_1 + x_2 + x_3 = 1701$ are there if each x_i is an integer and $x_1 \ge 2$, $x_2 \ge 4$, and $x_3 \ge 601$?*

Exercise 1.2.32. *In solving equations in integers, often slight changes in the coefficients can lead to wildly different behavior and very different sets of solutions. Determine the number of non-negative integer solutions to $x_1 + x_2 = 1996$, $2x_1 + 2x_2 = 1996$, $2x_1 + 2x_2 = 1997$, $2x_1 + 3x_2 = 1996$, $2x_1 + 2x_2 + 2x_3 + 2x_4 = 1996$ and $2x_1 + 2x_2 + 3x_3 + 3x_4 = 1996$. See Chapter 4 for more on finding integer solutions.*

Exercise^(h) 1.2.33. *Let f be a homogenous polynomial of degree d in n variables. This means*

$$f(x_1, \dots, x_n) \ = \ \sum_{\substack{0 \le k_1, \dots, k_n \le d \\ k_1 + \cdots + k_n = d}} a_{k_1,\dots,k_n} x_1^{k_1} \cdots x_n^{k_n}, \ \ a_{k_1,\dots,k_n} x_1^{k_1} \in \mathbb{C}. \tag{1.20}$$

Prove for any $\lambda \in \mathbb{C}$ that

$$f(\lambda x_1, \dots, \lambda x_n) \ = \ \lambda^d f(x_1, \dots, x_n). \tag{1.21}$$

As a function of n and d, how many possible terms are there in f (each term is of the form $x_1^{k_1} \cdots x_n^{k_n}$)?

The above problems are a small set of interesting results in combinatorics; see also [Mil4] for other techniques to prove combinatorial identities. We give some additional problems which illustrate the subject; the Binomial Theorem (Theorem A.1.8) is useful for these and other investigations.

Exercise(h) 1.2.34. *Let k be a positive integer and consider the sequence $1^k, 2^k, 3^k$, ... (so $x_n = n^k$). Consider the new sequence obtained by subtracting adjacent terms: $2^k - 1^k$, $3^k - 2^k$, ... and so on. Continue forming new sequences by subtracting adjacent terms of the previous terms. Prove that each term of the k^{th} sequence is $k!$.*

Exercise(hr) 1.2.35. *Let k and d be positive integers. Prove*

$$k^d = \sum_{m=0}^{d-1} \sum_{\ell=0}^{k-1} \binom{d}{m} \ell^m. \tag{1.22}$$

1.3 CLOCK ARITHMETIC: ARITHMETIC MODULO n

Let $\mathbb{Z}$ denote the set of integers and for $n \in \mathbb{N}$ define $\mathbb{Z}/n\mathbb{Z} = \{0, 1, 2, \ldots, n-1\}$. We often read $\mathbb{Z}/n\mathbb{Z}$ as the **integers modulo** n.

Definition 1.3.1 (Congruence)**.** *$x \equiv y \bmod n$ means $x - y$ is an integer multiple of n. Equivalently, x and y have the same remainder when divided by n.*

When there is no danger of confusion, we often drop the suffix mod n, writing instead $x \equiv y$.

Lemma 1.3.2 (Basic Properties of Congruences)**.** *For a fixed $n \in \mathbb{N}$ and a, a', b, b' integers we have*

1. *$a \equiv b \bmod n$ if and only if $b \equiv a \bmod n$.*
2. *$a \equiv b \bmod n$ and $b \equiv c \bmod n$ implies $a \equiv c \bmod n$.*
3. *$a \equiv a' \bmod n$ and $b \equiv b' \bmod n$, then $ab \equiv a'b' \bmod n$. In particular $a \equiv a' \bmod n$ implies $ab \equiv a'b \bmod n$ for all b.*

Exercise 1.3.3. *Prove the above relations. If $ab \equiv cb \bmod m$, must $a \equiv c \bmod m$?*

For $x, y \in \mathbb{Z}/n\mathbb{Z}$, we define $x + y$ to be the unique number $z \in \mathbb{Z}/n\mathbb{Z}$ such that $n|(x + y - z)$. In other words, z is the unique number in $\mathbb{Z}/n\mathbb{Z}$ such that $x + y \equiv z \bmod n$. One can show that $\mathbb{Z}/n\mathbb{Z}$ is a finite group under addition; in fact, it is a finite ring. (See §1.4.1 for the definition of a group).

Exercise(h) 1.3.4 (Arithmetic Modulo n)**.** *Define multiplication of $x, y \in \mathbb{Z}/n\mathbb{Z}$ by $x \cdot y$ is the unique $z \in \mathbb{Z}/n\mathbb{Z}$ such that $xy \equiv z \bmod n$. We often write xy for $x \cdot y$. Prove that this multiplication is well defined, and that an element x has a multiplicative inverse if and only if $(x, n) = 1$. Conclude that if every non-zero element of $\mathbb{Z}/n\mathbb{Z}$ has a multiplicative inverse, then n must be prime.*

Arithmetic modulo n is also called clock arithmetic. If $n = 12$ we have $\mathbb{Z}/12\mathbb{Z}$. If it is 10 o'clock now, in 5 hours it is 3 o'clock because $10 + 5 = 15 \equiv 3 \bmod 12$. See [Bob] for an analysis of the "randomness" of the inverse map in clock arithmetic.

Definition 1.3.5 (Least Common Multiple)**.** *Let $m, n \in N$. The least common multiple of m and n, denoted by $\text{lcm}(m, n)$, is the smallest positive integer divisible by both m and n.*

Exercise 1.3.6. *If $a \equiv b \bmod n$ and $a \equiv b \bmod m$, then $a \equiv b \bmod \text{lcm}(m, n)$.*

Exercise 1.3.7. *Prove for all positive integers m, n that $\text{lcm}(m, n) \cdot \text{gcd}(m, n) = mn$.*

Are there integer solutions to the equation $2x + 1 = 2y$? The left hand side is always odd, the right hand side is always even. Thus there are no integer solutions. What we did is really arithmetic modulo 2 or arithmetic in $\mathbb{Z}/2\mathbb{Z}$, and indicates the power of congruence arguments.

Consider now $x^2 + y^2 + z^2 = 8n + 7$. This never has integer solutions. Let us study this equation modulo 8. The right hand side is 7 modulo 8. What are the squares modulo 8? They are $1^2 \equiv 1, 2^2 \equiv 4, 3^2 \equiv 1, 4^2 \equiv 0$, and then the pattern repeats (as modulo 8, k and $(8-k)$ have the same square). We see there is no way to add three squares and get 7. Thus there are no solutions to $x^2+y^2+z^2 = 8n+7$.

Remark 1.3.8 (Hasse Principle)**.** In general, when searching for integer solutions one often tries to solve the equation modulo different primes. If there is no solution for some prime, then there are no integer solutions. Unfortunately, the converse is not true. For example, Selmer showed $3x^3 + 4y^3 + 5z^3 = 0$ is solvable modulo p for all p, but there are no rational solutions. We discuss this in more detail in Chapter 4.

Exercise 1.3.9 (Divisibility Rules)**.** *Prove a number is divisible by 3 (or 9) if and only if the sum of its digits are divisible by 3 (or 9). Prove a number is divisible by 11 if and only if the alternating sum of its digits is divisible by 11 (for example, 341 yields 3-4+1). Find a rule for divisibility by 7.*

Exercise 1.3.10 (Chinese Remainder Theorem)**.** *Let m_1, m_2 be relatively prime positive integers. Prove that for any $a_1, a_2 \in \mathbb{Z}$ there exists a unique $x \bmod m_1 m_2$ such that $x \equiv a_1 \bmod m_1$ and $x \equiv a_2 \bmod m_2$. Is this still true if m_1 and m_2 are not relatively prime? Generalize to $m_1, \dots, m_k$ and $a_1, \dots, a_k$.*

1.4 GROUP THEORY

We introduce enough group theory to prove our assertions about RSA. For more details, see [Art, J, La3].

1.4.1 Definition

Definition 1.4.1 (Group)**.** *A set G equipped with a map $G \times G \to G$ (denoted by $(x, y) \mapsto xy$) is a group if*

1. *(Identity) $\exists e \in G$ such that $\forall x \in G$, $ex = xe = x$.*

2. *(Associativity)* $\forall x, y, z \in G$, $(xy)z = x(yz)$.

3. *(Inverse)* $\forall x \in G, \exists y \in G$ *such that* $xy = yx = e$.

4. *(Closure)* $\forall x, y \in G$, $xy \in G$.

We have written the group multiplicatively, $(x, y) \mapsto xy$; if we wrote $(x, y) \mapsto x + y$, we say the group is written additively. We call G a finite group if the set G is finite. If $\forall x, y \in G, xy = yx$, we say the group is **abelian** or **commutative**.

Exercise 1.4.2. *Show that under addition* $\mathbb{Z}/n\mathbb{Z}$ *is an abelian group.*

Exercise 1.4.3. *Consider the set of* $N \times N$ *matrices with real entries and non-zero determinant. Prove this is a group under matrix multiplication, and show this group is not commutative if* $N > 1$. *Is it a group under matrix addition?*

Exercise 1.4.4. *Let* $(\mathbb{Z}/p\mathbb{Z})^* = \{1, 2, \dots, p-1\}$ *where* $a \cdot b$ *is defined to be* $ab \bmod p$. *Prove this is a multiplicative group if* p *is prime. More generally, let* $(\mathbb{Z}/m\mathbb{Z})^*$ *be the subset of* $\mathbb{Z}/m\mathbb{Z}$ *of numbers relatively prime to* m. *Show* $(\mathbb{Z}/m\mathbb{Z})^*$ *is a multiplicative group.*

Exercise 1.4.5 (Euler's ϕ-function (or totient function))**.** *Let* $\phi(n)$ *denote the number of elements in* $(\mathbb{Z}/n\mathbb{Z})^*$. *Prove that for* p *prime,* $\phi(p) = p - 1$ *and* $\phi(p^k) = p^k - p^{k-1}$. *If* p *and* q *are distinct primes, prove* $\phi(p^j q^k) = \phi(p^j)\phi(q^k)$. *If* n *and* m *are relatively prime, prove that* $\phi(nm) = \phi(n)\phi(m)$. *Note* $\phi(n)$ *is the size of the group* $(\mathbb{Z}/n\mathbb{Z})^*$.

Definition 1.4.6 (Subgroup)**.** *A subset* H *of* G *is a subgroup if* H *is also a group.*

Our definitions imply any group G has at least two subgroups, itself and the empty set.

Exercise 1.4.7. *Prove the following equivalent definition: A subset* H *of a group* G *is a subgroup if for all* $x, y \in H$, $xy^{-1} \in H$.

Exercise 1.4.8. *Let* G *be an additive subgroup of* $\mathbb{Z}$. *Prove that there exists an* $n \in \mathbb{N}$ *such that every element of* G *is an integral multiple of* n.

Exercise 1.4.9. *Let* $GL_n(\mathbb{R})$ *be the multiplicative group of* $n \times n$ *invertible matrices with real entries. Let* $SL_n(\mathbb{Z})$ *be the subset with integer entries and determinant 1. Prove* $SL_n(\mathbb{Z})$ *is a subgroup. This is a very important subgroup in number theory; when* $n = 2$ *it is called the* ***modular group***. *See §7.7 for an application to continued fractions.*

1.4.2 Lagrange's Theorem

We prove some basic properties of **finite groups** (groups with finitely many elements).

Definition 1.4.10 (Order)**.** *If* G *is a finite group, the number of elements of* G *is the order of* G *and is denoted by* $|G|$. *If* $x \in G$, *the order of* x *in* G, $\mathrm{ord}(x)$, *is the least positive power* m *such that* $x^m = e$, *where* $e \in G$ *is the identity of the group.*

Exercise(h) 1.4.11. *Prove all elements in a finite group have finite order.*

Theorem 1.4.12 (Lagrange)**.** *Let H be a subgroup of a finite group G. Then $|H|$ divides $|G|$. In particular, taking H to be the subgroup generated by $x \in G$,* $\mathrm{ord}(x)|\mathrm{ord}(G)$.

We first prove two useful lemmas.

Lemma 1.4.13. *Let H be a subgroup of G, and let $h \in H$. Then $hH = H$.*

Proof. It suffices to show $hH \subset H$ and $H \subset hH$. By closure, $hH \subset H$. For the other direction, let $h' \in H$. Then $hh^{-1}h' = h'$; as $h^{-1}h' \in H$, every $h' \in H$ is also in hH. □

Lemma 1.4.14. *Let H be a subgroup of a group G. Then for all $g_i, g_j \in G$ either $g_iH = g_jH$ or the two sets are disjoint.*

Proof. Assume $g_iH \cap g_jH$ is non-empty; we must show they are equal. Let $x = g_ih_1 = g_jh_2$ be in the intersection. Multiplying on the right by $h_1^{-1} \in H$ (which exists because H is a subgroup) gives $g_i = g_jh_2h_1^{-1}$. So $g_iH = g_jh_2h_1^{-1}H$. As $h_2h_1^{-1}H = H$, we obtain $g_iH = g_jH$. □

Definition 1.4.15 (Coset)**.** *We call a subset gH of G a* coset *(actually, a left coset) of H. In general the set of all gH for a fixed H is not a subgroup.*

Exercise(h) 1.4.16. *Show not every set of cosets is a subgroup.*

We now prove Lagrange's Theorem.

Proof of Lagrange's theorem. We claim

$$G = \bigcup_{g \in G} gH. \tag{1.23}$$

Why is there equality? As $g \in G$ and $H \subset G$, each $gH \subset G$, hence their union is contained in G. Further, as $e \in H$, given $g \in G$, $g \in gH$. Thus, G is a subset of the right side, proving equality.

By Lemma 1.4.13, two cosets are either identical or disjoint. By choosing a subset of the cosets, we show the union in (1.23) equals a union of disjoint cosets. There are only finitely many elements in G. As we go through all g in G, if the coset gH equals one of the cosets already chosen, we do not include it; if it is new, we do. Continuing this process, we obtain

$$G = \bigcup_{i=1}^{k} g_iH \tag{1.24}$$

for some finite k, and the k cosets are disjoint. If $H = \{e\}$, k is the number of elements of G; in general, however, k will be smaller. Each set g_iH has $|H|$ elements, and no two cosets share an element. Thus $|G| = k|H|$, proving $|H|$ divides $|G|$. □

Exercise 1.4.17. *Let $G = (\mathbb{Z}/15\mathbb{Z})^*$. Find all subgroups of G and write G as the union of cosets for some proper subgroup H (H is a* **proper subgroup** *of G if H is neither $\{1\}$ nor G).*

Exercise 1.4.18. *Let $G = (\mathbb{Z}/p_1p_2\mathbb{Z})^*$ for two distinct primes p_1 and p_2. What are the possible orders of subgroups of G? Prove that there is either a subgroup of order p_1 or a subgroup of order p_2 (in fact, there are subgroups of both orders).*

1.4.3 Fermat's Little Theorem

We deduce some consequences of Lagrange's Theorem which will be useful in our cryptography investigations.

Corollary 1.4.19 (Fermat's Little Theorem)**.** *For any prime p, if $\gcd(a, p) = 1$ then $a^{p-1} \equiv 1 \bmod p$.*

Proof. As $|(\mathbb{Z}/p\mathbb{Z})^*| = p - 1$, the result follows from Lagrange's Theorem. □

Exercise(h) 1.4.20. *One can reformulate Fermat's Little Theorem as the statement that if p is prime, for all a we have $p|a^p - a$. Give a proof for this formulation* without *using group theory. Does $n|a^n - a$ for all n?*

Exercise 1.4.21. *Prove that if for some a, $a^{n-1} \not\equiv 1 \bmod n$ then n is composite.*

Thus Fermat's Little Theorem is a fast way to show certain numbers are composite (remember exponentiation is fast: see §1.2.1); we shall also encounter Fermat's Little Theorem in §4.4.3 when we count the number of integer solutions to certain equations. Unfortunately, it is not the case that $a^{n-1} \equiv 1 \bmod n$ implies n is prime. There are composite n such that for all positive integers a, $a^{n-1} \equiv 1 \bmod n$. Such composite numbers are called Carmichael numbers (the first few are 561, 1105 and 1729). More generally, one has

Theorem 1.4.22 (Euler)**.** *If $\gcd(a, n) = 1$, then $a^{\phi(n)} \equiv 1 \bmod n$.*

Proof. Let $(a, n) = 1$. By definition, $\phi(n) = |(\mathbb{Z}/n\mathbb{Z})^*|$. By Lagrange's Theorem the order of $a \in (\mathbb{Z}/n\mathbb{Z})^*$ divides $\phi(n)$, or $a^{\phi(n)} \equiv 1 \bmod n$. □

Remark 1.4.23. For our applications to RSA, we only need the case when n is the product of two primes. In this case, consider the set $\{1, \ldots, pq\}$. There are pq numbers, q numbers are multiples of p, p numbers are multiples of q, and one is a multiple of both p and q. Thus, the number of numbers in $\{1, \ldots, pq\}$ relatively prime to pq is $pq - p - q + 1$ (why?). Note this equals $\phi(p)\phi(q) = (p-1)(q-1)$. This type of argument is known as **Inclusion - Exclusion**. See also Exercise 2.3.18.

Exercise 1.4.24. *Korselt [Kor] proved that a composite number n is a Carmichael number if and only if n is square-free and if a prime $p|n$, then $(p-1)|(n-1)$. Prove that if these two conditions are met then n is a Carmichael number.*

Research Project 1.4.25 (Carmichael Numbers)**.** It is known (see [AGP]) that there are infinitely many Carmichael numbers. One can investigate the spacings between adjacent Carmichael numbers. For example, choose a large X and look at all Carmichael numbers in $[X, 2X]$, say $c_1, \dots, c_{n+1}$. The average spacing between these numbers is about $\frac{2X-X}{n}$ (they are spread out over an interval of size X, and there are n differences: $c_2 - c_1, \dots, c_{n+1} - c_n$. How are these differences distributed? Often, it is more natural to rescale differences and spacings so that the average spacing is 1. The advantage of such a renormalization is the results are often scale invariant (i.e., unitless quantities). For more on investigating such spacings, see Chapter 12.

Exercise[(h)] 1.4.26. *Prove an integer is divisible by* 3 *(resp.,* 9*) if and only if the sum of its digits is divisible by* 3 *(resp.,* 9*).*

Exercise[(h)] 1.4.27. *Show an integer is divisible by* 11 *if and only if the alternating sum of its digits is divisible by* 11*; for example,* 924 *is divisible by* 11 *because* $11|(9-2+4)$*. Use Fermat's Little Theorem to find a rule for divisibility by* 7 *(or more generally, for any prime).*

Exercise[(h)] 1.4.28. *Show that if x is a positive integer then there exists a positive integer y such that the product xy has only zeros and ones for digits.*

1.4.4 Structure of $(\mathbb{Z}/p\mathbb{Z})^*$

The multiplicative group $(\mathbb{Z}/p\mathbb{Z})^*$ for p prime has a rich structure which will simplify many investigations later.

Theorem 1.4.29. *For p prime, $(\mathbb{Z}/p\mathbb{Z})^*$ is cyclic of order $p-1$. This means there is an element $g \in (\mathbb{Z}/p\mathbb{Z})^*$ such that*

$$(\mathbb{Z}/p\mathbb{Z})^* \ = \ \{1, 2, \dots, p-2, p-1\} \ = \ \{g^1, g^2, \dots, g^{p-2}, g^{p-1}\}. \tag{1.25}$$

We say g is a **generator** of the group. For each x there is a unique integer $k \in \{1, \dots, p-1\}$ such that $x \equiv g^k \bmod p$. We say k is the **index** of x relative to g. For each $x \in (\mathbb{Z}/p\mathbb{Z})^*$, the **order** of x is the smallest positive integer n such that $x^n \equiv 1 \bmod p$. For example, if $p = 7$ we have

$$\{1, 2, 3, 4, 5, 6\} \ = \ \{3^6, 3^2, 3^1, 3^4, 3^5, 3^3\}, \tag{1.26}$$

which implies 3 is a generator (and the index of 4 relative to 3 is 4, because $4 \equiv 3^4 \bmod 7$). Note 5 is also a generator of this group, so the generator need not be unique.

Sketch of the proof. We will use the fact that $(\mathbb{Z}/p\mathbb{Z})^*$ is a commutative group: $xy = yx$. Let $x, y \in (\mathbb{Z}/p\mathbb{Z})^*$ with orders m and n for the exercises below. The proof comes from the following:

Exercise 1.4.30. *Assume $m = m_1 m_2$, with m_1, m_2 relatively prime. Show x^{m_1} has order m_2.*

Exercise[(h)] 1.4.31. *Let ℓ be the least common multiple of m and n (the smallest number divisible by both m and n). Prove that there is an element z of order ℓ.*

Exercise 1.4.32. *By Lagrange's Theorem, the order of any x divides $p-1$ (the size of the group). From this fact and the previous exercises, show there is some d such that the order of every element divides $d \leq p-1$,* and *there is an element of order d and no elements of larger order.*

The proof is completed by showing $d = p - 1$. The previous exercises imply that every element satisfies the equation $x^d - 1 \equiv 0 \bmod p$. As every element in the group satisfies this, and there are $p-1$ elements in the group, we have a degree d polynomial with $p - 1$ roots. We claim this can only occur if $d = p - 1$.

Exercise[(h)] 1.4.33. *Prove the above claim.*

Therefore $d = p-1$ and there is some element g of order $p-1$; thus, g's powers generate the group. □

Exercise 1.4.34. *For $p > 2$, $k > 1$, what is the structure of $(\mathbb{Z}/p^k\mathbb{Z})^*$? If all the prime divisors of m are greater than 2, what is the structure of $(\mathbb{Z}/m\mathbb{Z})^*$? For more on the structure of these groups, see any undergraduate algebra textbook (for example, [Art, J, La3]).*

1.5 RSA REVISITED

We have developed sufficient machinery to prove why RSA works. Remember Bob chose two primes p and q, and numbers d (for decrypt) and e (for encrypt) such that $de \equiv 1 \bmod \phi(pq)$. He made public $N = pq$ and e and kept secret the two primes and d. Alice wants to send Bob a number M (smaller than N). She encrypts the message by sending $X \equiv M^e \bmod N$. Bob then decrypts the message by calculating $X^d \bmod N$, which we claimed equals M.

As $X \equiv M^e \bmod N$, there is an integer n such that $X = M^e + nN$. Thus $X^d = (M^e + nN)^d$, and the last term is clearly of the form $(M^e)^d + n'N$ for some n'. We need only show $(M^e)^d \equiv M \bmod N$. As $ed \equiv 1 \bmod \phi(N)$, there is an m such that $ed = 1 + m\phi(N)$. Therefore

$$(M^e)^d = M^{ed} = M^{1+m\phi(N)} = M \cdot M^{m\phi(N)} = M \cdot (M^{\phi(N)})^m. \quad (1.27)$$

If M is relatively prime to N then By Euler's Theorem (Theorem 1.4.22), $M^{\phi(N)} \equiv 1 \bmod N$, which completes the proof. Thus we can only send messages relatively prime to N. In practice this is not a problem, as it is very unlikely to stumble upon a message that shares a factor with N; of course, if we did find such a message we could quickly find the factors of N. If our initial message has a factor in common with N, we need only tweak our message (add another letter or spell a word incorrectly).

Why is RSA secure? Assume a third person (say Charlie) intercepts the encrypted message X. He knows X, N and e, and wants to recover M. Knowing d such that $de \equiv 1 \bmod \phi(N)$ makes decrypting the message trivial: one need only compute $X^d \bmod N$. Thus Charlie is trying to solve the equation $ed \equiv 1 \bmod \phi(N)$; fortunately for Alice and Bob this equation has two unknowns, d and $\phi(N)$! Right now, there is no known fast way to determine $\phi(N)$ from N. Charlie can of course factor N; once he has the factors, he knows $\phi(N)$ and can find d; however, the fastest factorization algorithms make 400 digit numbers unaccessible for now.

This should be compared to primality testing, which was only recently shown to be fast ([AgKaSa]). Previous deterministic algorithms to test whether or not a number is prime were known to be fast only if certain well believed conjectures are true. It was an immense achievement showing that there is a deterministic, efficient algorithm. The paper is very accessible, and worth the read.

Remark 1.5.1. Our simple example involved computing a sixty-six digit number, and this was for a small N ($N = 9797$). Using binary expansions to exponentiate, as we need only transmit our message modulo N, we never need to compute anything larger than the product of four digit numbers.

Remark 1.5.2. See [Bon] for a summary of attempts to break RSA. Certain products of two primes are denoted RSA challenge numbers, and the public is invited to factor them. With the advent of parallel processing, many numbers have succumbed to factorization. See http://www.rsasecurity.com/rsalabs/node.asp?id=2092 for more details.

Exercise 1.5.3. *If $M < N$ is not relatively prime to N, show how to quickly find the prime factorization of N.*

Exercise 1.5.4 (Security Concerns)**.** *In the system described, there is no way for Bob to verify that the message came from Alice! Design a system where Alice makes some information public (and keeps some secret) so that Bob can verify that Alice sent the message.*

Exercise 1.5.5. *Determining $\phi(N)$ is equivalent to factoring N; there is no computational shortcut to factoring. Clearly, if one knows the factors of $N = pq$, one knows $\phi(N)$. If one knows $\phi(N)$ and N, one can recover the primes p and q. Show that if $K = N + 1 - \phi(N)$, then the two prime factors of N are $(K \pm \sqrt{K^2 - 4N})/2$, and these numbers are in fact integers.*

Exercise$^{\text{(hr)}}$ 1.5.6 (Important)**.** *If e and $(p-1)(q-1)$ are given, show how one may efficiently find a d such that $ed - 1$ divides $(p-1)(q-1)$.*

1.6 EISENSTEIN'S PROOF OF QUADRATIC RECIPROCITY

We conclude this introduction to basic number theory and group theory by giving a proof of quadratic reciprocity (we follow the beautiful exposition in [LP] of Eisenstein's proof; see the excellent treatments in [IR, NZM] for alternate proofs). In §1.2.4, we described Newton's Method to find square roots of real numbers. Now we turn our attention to a finite group analogue: for a prime p and an $a \not\equiv 0 \bmod p$, when is $x^2 \equiv a \bmod p$ solvable? For example, if $p = 5$ then $(\mathbb{Z}/p\mathbb{Z})^* = \{1, 2, 3, 4\}$. Squaring these numbers gives $\{1, 4, 4, 1\} = \{1, 4\}$. Thus there are two solutions if $a \in \{1, 4\}$ and no solutions if $a \in \{2, 3\}$. The problem of whether or not a given number is a square is solvable: we can simply enumerate the group $(\mathbb{Z}/p\mathbb{Z})^*$, square each element, and see if a is a square. This takes about p steps; quadratic reciprocity will take about $\log p$ steps. For applications, see §4.4.

1.6.1 Legendre Symbol

We introduce notation. From now on, p and q will always be distinct odd primes.

Definition 1.6.1 (Legendre Symbol $\left(\frac{\cdot}{p}\right)$)**.** *The Legendre Symbol $\left(\frac{a}{p}\right)$ is*

$$\left(\frac{a}{p}\right) = \begin{cases} 1 & \text{if } a \text{ is a non-zero square modulo } p \\ 0 & \text{if } a \equiv 0 \text{ modulo } p \\ -1 & \text{if } a \text{ is a not a square modulo } p. \end{cases} \tag{1.28}$$

The Legendre symbol is a function on $\mathbb{F}_p = \mathbb{Z}/p\mathbb{Z}$. We extend the Legendre symbol to all integers by $\left(\frac{a}{p}\right) = \left(\frac{a \bmod p}{p}\right)$.

Note a is a square modulo p if there exists an $x \in \{0, 1, \ldots, p-1\}$ such that $a \equiv x^2 \bmod p$.

Definition 1.6.2 (Quadratic Residue, Non-Residue)**.** *For $a \not\equiv 0 \bmod p$, if $x^2 \equiv a \bmod p$ is solvable (resp., not solvable) we say a is a quadratic residue (resp., non-residue) modulo p. When p is clear from context, we just say residue and non-residue.*

Exercise 1.6.3. *Show the Legendre symbol is multiplicative: $\left(\frac{ab}{p}\right) = \left(\frac{a}{p}\right)\left(\frac{b}{p}\right)$.*

Exercise$^{(h)}$ 1.6.4 (Euler's Criterion)**.** *For odd p, $\left(\frac{a}{p}\right) \equiv a^{\frac{p-1}{2}} \bmod p$.*

Exercise 1.6.5. *Show $\left(\frac{-1}{p}\right) = (-1)^{\frac{p-1}{2}}$ and $\left(\frac{2}{p}\right) = (-1)^{\frac{p^2-1}{8}}$.*

Lemma 1.6.6. *For p an odd prime, half of the non-zero numbers in $(\mathbb{Z}/p\mathbb{Z})^*$ are quadratic residues and half are quadratic non-residues.*

Proof. As p is odd, $\frac{p-1}{2} \in \mathbb{N}$. Consider the numbers $1^2, 2^2, \ldots, (\frac{p-1}{2})^2$. Assume two numbers a and b are equivalent mod p. Then $a^2 \equiv b^2 \bmod p$, so $(a-b)(a+b) \equiv 0 \bmod p$. Thus either $a \equiv b \bmod p$ or $a \equiv -b \bmod p$; in other words, $a \equiv p - b$. For $1 \leq a, b \leq \frac{p-1}{2}$ we cannot have $a \equiv p - b \bmod p$, implying the $\frac{p-1}{2}$ values above are distinct. As $(p-r)^2 \equiv r^2 \bmod p$, the above list is all of the non-zero squares modulo p. Thus half the non-zero numbers are non-zero squares, half are non-squares. □

Remark 1.6.7. By Theorem 1.4.29, $(\mathbb{Z}/p\mathbb{Z})^*$ is a cyclic group with generator g. Using the group structure we can prove the above lemma directly: once we show there is at least one non-residue, the g^{2k} are the quadratic residues and the g^{2k+1} are the non-residues.

Exercise 1.6.8. *Show for any $a \not\equiv 0 \bmod p$ that*

$$\sum_{t=0}^{p-1} \left(\frac{t}{p}\right) = \sum_{t=0}^{p-1} \left(\frac{at+b}{p}\right) = 0. \tag{1.29}$$

Exercise 1.6.9. *For $x \in \{0, \dots, p-1\}$, let $F_p(x) = \sum_{a \le x} \left(\frac{n}{p}\right)$; note $F_p(0) = F_p(p-1) = 0$. If $\left(\frac{-1}{p}\right) = 1$, show $F_p\left(\frac{p-1}{2}\right) = 0$. Do you think $F(x)$ is more likely to be positive or negative? Investigate its values for various x and p.*

Initially the Legendre symbol is defined only when the bottom is prime. We now extend the definition. Let $n = p_1 \cdot p_2 \cdots p_t$ be the product of t distinct odd primes. Then $\left(\frac{a}{n}\right) = \left(\frac{a}{p_1}\right)\left(\frac{a}{p_2}\right)\cdots\left(\frac{a}{p_t}\right)$; this is the **Jacobi symbol**, and has many of the same properties as the Legendre symbol. We will study only the Legendre symbol (see [IR] for more on the Jacobi symbol). Note the Jacobi symbol does *not* say that if a is a square (a quadratic residue) mod n, then a is a square mod p_i for each prime divisor.

The main result (which allows us to calculate the Legendre symbol quickly and efficiently) is the celebrated

Theorem 1.6.10 (The Generalized Law of Quadratic Reciprocity)**.** *For m, n odd and relatively prime,*

$$\left(\frac{m}{n}\right)\left(\frac{n}{m}\right) = (-1)^{\frac{m-1}{2}\frac{n-1}{2}}. \tag{1.30}$$

Gauss gave eight proofs of this deep result when m and n are prime. If either p or q are equivalent to $1 \bmod 4$ then we have $\left(\frac{q}{p}\right) = \left(\frac{p}{q}\right)$, i.e., p has a square root modulo q if and only if q has a square root modulo p. We content ourselves with proving the case with m, n prime.

Exercise 1.6.11. *Using the Generalized Law of Quadratic Reciprocity, Exercise 1.6.5 and the Euclidean algorithm, show one can determine if $a < m$ is a square modulo m in logarithmic time (i.e., the number of steps is at most a fixed constant multiple of $\log m$). This incredible efficiency is just one of many important applications of the Legendre and Jacobi symbols.*

1.6.2 The Proof of Quadratic Reciprocity

Our goal is to prove

Theorem 1.6.12 (Quadratic Reciprocity)**.** *Let p and q be distinct odd primes. Then*

$$\left(\frac{q}{p}\right)\left(\frac{p}{q}\right) = (-1)^{\frac{p-1}{2}\frac{q-1}{2}}. \tag{1.31}$$

As p and q are distinct, odd primes, both $\left(\frac{q}{p}\right)$ and $\left(\frac{p}{q}\right)$ are ± 1. The difficulty is figuring out which signs are correct, and how the two signs are related. We use Euler's Criterion (Exercise 1.6.4).

The idea behind Eisenstein's proof is as follows: $\left(\frac{q}{p}\right)\left(\frac{p}{q}\right)$ is -1 to a power. Further, we only need to determine the power modulo 2. Eisenstein shows many expressions are equivalent modulo 2 to this power, and eventually we arrive at an expression which is trivial to calculate modulo 2. We repeatedly use the fact that as p and q are distinct primes, the Euclidean algorithm implies that q is invertible modulo p and p is invertible modulo q.

We choose to present this proof as it showcases many common techniques in mathematics. In addition to using the Euclidean algorithm and modular arithmetic, the proof shows that quadratic reciprocity is equivalent to a theorem about the number of integer solutions of some inequalities, specifically the number of pairs of integers strictly inside a rectangle. This is just one of many applications of counting solutions; we discuss this topic in greater detail in Chapter 4.

1.6.3 Preliminaries

Consider all multiples of q by an even $a \le p-1$: $\{2q, 4q, 6q, \dots, (p-1)q\}$. Denote a generic multiple by aq. Recall $[x]$ is the greatest integer less than or equal to x. By the Euclidean algorithm,

$$aq \;=\; \left[\frac{aq}{p}\right] p + r_a, \quad 0 < r_a < p-1. \tag{1.32}$$

Thus r_a is the least non-negative number equivalent to $aq \bmod p$. The numbers $(-1)^{r_a} r_a$ are equivalent to even numbers in $\{0, \dots, p-1\}$. If r_a is even this is clear; if r_a is odd, then $(-1)^{r_a} r_a \equiv p - r_a \bmod p$, and as p and r_a are odd, this is even. Finally note $r_a \neq 0$; if $r_a = 0$ then $p|aq$. As p and q are relatively prime, this implies $p|a$; however, p is prime and $a \le p-1$. Therefore p cannot divide a and thus $r_a \neq 0$.

Lemma 1.6.13. *If* $(-1)^{r_a} r_a \equiv (-1)^{r_b} r_b$ *then* $a = b$.

Proof. We quickly get $\pm r_a \equiv r_b \bmod p$. If the plus sign holds, then $r_a \equiv r_b \bmod p$ implies $aq \equiv bq \bmod p$. As q is invertible modulo p, we get $a \equiv b \bmod p$, which yields $a = b$ (as a and b are even integers between 2 and $p-1$).

If the minus sign holds, then $r_a + r_b \equiv 0 \bmod p$, or $aq + bq \equiv 0 \bmod p$. Multiplying by $q^{-1} \bmod p$ now gives $a + b \equiv 0 \bmod p$. As a and b are even integers between 2 and $p-1$, $4 < a + b \le 2p - 2$. The only integer strictly between 4 and $2p-2$ which is equivalent to $0 \bmod p$ is p; however, p is odd and $a + b$ is even. Thus the minus sign cannot hold, and the elements are all distinct. □

Remark 1.6.14. The previous argument is very common in mathematics. We will see a useful variant in Chapter 5, where we show certain numbers are irrational by proving that if they were not then there would have to be an integer strictly between 0 and 1.

Lemma 1.6.15. *We have*

$$\left(\frac{q}{p}\right) \;=\; (-1)^{\sum_{a \text{ even}, a \neq 0} r_a}, \tag{1.33}$$

where a even, $a \neq 0$ *means* $a \in \{2, 4, \dots, p-3, p-1\}$.

Proof. For each even $a \in \{2, \dots, p-1\}$, $aq \equiv r_a \bmod p$. Thus modulo p

$$\begin{aligned} \prod_{\substack{a \text{ even} \\ a \neq 0}} aq &\equiv \prod_{\substack{a \text{ even} \\ a \neq 0}} r_a \\ q^{\frac{p-1}{2}} \prod_{\substack{a \text{ even} \\ a \neq 0}} a &\equiv \prod_{\substack{a \text{ even} \\ a \neq 0}} r_a \\ \left(\frac{q}{p}\right) \prod_{\substack{a \text{ even} \\ a \neq 0}} a &\equiv \prod_{\substack{a \text{ even} \\ a \neq 0}} r_a, \end{aligned} \tag{1.34}$$

where the above follows from the fact that we have $\frac{p-1}{2}$ choices for an even a (giving the factor $q^{\frac{p-1}{2}}$) and Euler's Criterion (Exercise 1.6.4). As a ranges over all even numbers from 2 to $p-1$, so too do the distinct numbers $(-1)^{r_a} r_a \bmod p$. Note how important it was that we showed $r_a \neq 0$ in (1.32), as otherwise we would just have $0 = 0$ in (1.34). Thus modulo p,

$$\begin{aligned} \prod_{\substack{a \text{ even} \\ a \neq 0}} a &\equiv \prod_{\substack{a \text{ even} \\ a \neq 0}} (-1)^{r_a} r_a \\ \prod_{\substack{a \text{ even} \\ a \neq 0}} a &\equiv (-1)^{\sum_{a \text{ even}, a \neq 0} r_a} \prod_{\substack{a \text{ even} \\ a \neq 0}} r_a. \end{aligned} \tag{1.35}$$

Combining gives

$$\left(\frac{q}{p}\right) (-1)^{\sum_{a \text{ even}, a \neq 0} r_a} \prod_{\substack{a \text{ even} \\ a \neq 0}} r_a \equiv \prod_{\substack{a \text{ even} \\ a \neq 0}} r_a \bmod p. \tag{1.36}$$

As each r_a is invertible modulo p, so is the product. Thus

$$\left(\frac{q}{p}\right) (-1)^{\sum_{a \text{ even}, a \neq 0} r_a} \equiv 1 \bmod p. \tag{1.37}$$

As $\left(\frac{q}{p}\right) = \pm 1$, the lemma follows by multiplying both sides by $\left(\frac{q}{p}\right)$. $\square$

Therefore it suffices to determine $\sum_{a \text{ even}, a \neq 0} r_a \bmod 2$. We make one last simplification. By the first step in the Euclidean algorithm (1.32), we have $aq = \left[\frac{aq}{p}\right] p + r_a$ for some $r_a \in \{2, \dots, p-1\}$. Hence

$$\sum_{\substack{a \text{ even} \\ a \neq 0}} aq = \sum_{\substack{a \text{ even} \\ a \neq 0}} \left(\left[\frac{aq}{p}\right] p + r_a \right) = \sum_{\substack{a \text{ even} \\ a \neq 0}} \left[\frac{aq}{p}\right] p + \sum_{\substack{a \text{ even} \\ a \neq 0}} r_a. \tag{1.38}$$

As we are summing over even a, the left hand side above is even. Thus the right hand side is even, so

$$\begin{aligned} \sum_{\substack{a \text{ even} \\ a \neq 0}} \left[\frac{aq}{p}\right] p &\equiv \sum_{\substack{a \text{ even} \\ a \neq 0}} r_a \mod 2 \\ \sum_{\substack{a \text{ even} \\ a \neq 0}} \left[\frac{aq}{p}\right] &\equiv \sum_{\substack{a \text{ even} \\ a \neq 0}} r_a \mod 2, \end{aligned} \tag{1.39}$$

where the last line follows from the fact that p is odd, so modulo 2 dropping the factor of p from the left hand side does not change the parity. We have reduced the proof of quadratic reciprocity to calculating $\sum_{a \text{ even}, a\neq 0} \left[\frac{aq}{p}\right]$. We summarize our results below.

Lemma 1.6.16. *Define*

$$\mu = \sum_{\substack{a \text{ even} \\ a\neq 0}} \left[\frac{aq}{p}\right]$$
$$\nu = \sum_{\substack{a \text{ even} \\ a\neq 0}} \left[\frac{ap}{q}\right]. \tag{1.40}$$

Then

$$\left(\frac{q}{p}\right) = (-1)^{\mu}$$
$$\left(\frac{p}{q}\right) = (-1)^{\nu}. \tag{1.41}$$

Proof. By (1.37) we have

$$\left(\frac{q}{p}\right) = (-1)^{\sum_{a \text{ even}, a\neq 0} r_a}. \tag{1.42}$$

By (1.39) we have

$$\sum_{\substack{a \text{ even} \\ a\neq 0}} \left[\frac{aq}{p}\right] \equiv \sum_{\substack{a \text{ even} \\ a\neq 0}} r_a \bmod 2, \tag{1.43}$$

and the proof for $\left(\frac{q}{p}\right)$ is completed by recalling the definition of μ; the proof for the case $\left(\frac{p}{q}\right)$ proceeds similarly. □

1.6.4 Counting Lattice Points

As our sums are not over all even $a \in \{0, 2, \dots, p-1\}$ but rather just over even $a \in \{2, \dots, p-1\}$, this slightly complicates our notation and forces us to be careful with our book-keeping. We urge the reader not to be too concerned about this slight complication and instead focus on the fact that we are able to show quadratic reciprocity is equivalent to counting the number of pairs of integers satisfying certain relations.

Consider the rectangle with vertices at $A = (0,0)$, $B = (p,0)$, $C = (p,q)$ and $D = (0,q)$. The upward sloping diagonal is given by the equation $y = \frac{q}{p}x$. As p and q are distinct odd primes, there are no pairs of integers (x,y) on the line AC. See Figure 1.2.

We add some non-integer points: $E = (\frac{p}{2}, 0)$, $F = (\frac{p}{2}, \frac{q}{2})$, $G = (0, \frac{q}{2})$ and $H = (\frac{p}{2}, q)$. Let $\#ABC_{\text{even}}$ denote the number of integer pairs **strictly inside** the triangle ABC with even x-coordinate, and $\#AEF$ denote the number of integer

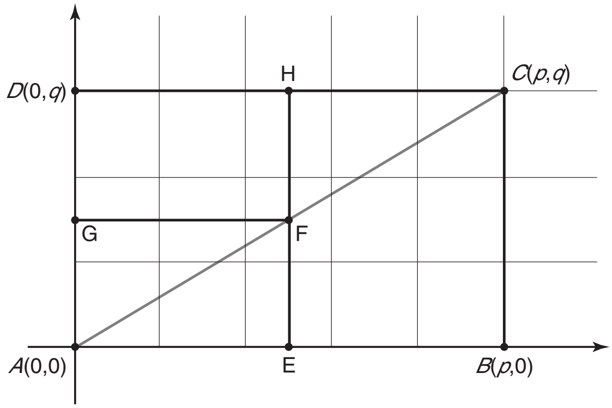


Figure 1.2 Lattice for the proof of Quadratic Reciprocity. Points $E(\frac{p}{2},0)$, $F(\frac{p}{2},\frac{q}{2})$, $G(0,\frac{q}{2})$, $H(\frac{p}{2},q)$

pairs **strictly inside** the triangle AEF; thus, we do not count any integer pairs on the lines AB, BC, CD or DA.

We now interpret $\sum_{a \text{ even}, a\neq 0}\left[\frac{aq}{p}\right]$. Consider the vertical line with x-coordinate a. Then $\left[\frac{aq}{p}\right]$ gives the number of pairs (x,y) with x-coordinate equal to a and y-coordinate a positive integer at most $\left[\frac{aq}{p}\right]$. To see this, consider the line AC (which is given by the equation $y = \frac{q}{p}x$). For definiteness, let us take $p = 5$, $q = 7$ and $a = 4$. Then $\left[\frac{aq}{p}\right] = \left[\frac{28}{5}\right] = 5$, and there are exactly five integer pairs with x-coordinate equal to 4 and positive y-coordinate at most $\left[\frac{28}{5}\right]$: $(4,1)$, $(4,2)$, $(4,3)$, $(4,4)$ and $(4,5)$. The general proof proceeds similarly.

Thus $\sum_{a \text{ even}, a\neq 0}\left[\frac{aq}{p}\right]$ is the number of integer pairs **strictly inside** the rectangle $ABCD$ with even x-coordinate that are below the line AC, which we denote $\#ABC_{\text{even}}$. We prove

Lemma 1.6.17. *The number of integer pairs under the line AC strictly inside the rectangle with even x-coordinate is congruent modulo 2 to the number of integer pairs under the line AF strictly inside the rectangle. Thus $\#ABC_{\text{even}} = \#AEF$.*

Proof. First observe that if $0 < a < \frac{p}{2}$ is even then the points under AC with x-coordinate equal to a are exactly those under the line AF with x-coordinate equal to a. We are reduced to showing that the number of points under FC strictly inside the rectangle with even x-coordinate is congruent modulo 2 to the number of points under the line AF strictly inside the rectangle with odd x-coordinate. Therefore let us consider an even a with $\frac{p}{2} < a < p-1$.

The integer pairs on the line $x = a$ strictly inside the rectangle are $(a,1)$, $(a,2), \ldots, (a,q-1)$. There are $q-1$ pairs. As q is odd, there are an even number of integer pairs on the line $x = a$ strictly inside the rectangle. As there are no integer pairs on the line AC, for a fixed $a > \frac{p}{2}$, modulo 2 there are the same number of integer pairs *above* AC as there are *below* AC. The number of integer pairs *above* AC on the line $x = a$ is equivalent modulo 2 to the number of integer pairs below AF on the line $x = p - a$. To see this, consider the map which takes (x, y) to $(p - x, q - y)$. As $a > \frac{p}{2}$ and is even, $p - a < \frac{p}{2}$ and is odd. Further, every odd $a < \frac{p}{2}$ is hit (given $a_{\text{odd}} < \frac{p}{2}$, start with the even number $p - a_{\text{odd}} > \frac{p}{2}$). A similar proof holds for $a < \frac{p}{2}$. □

Exercise 1.6.18. *Why are there no integer pairs on the line AC?*

We have thus shown that

$$\sum_{\substack{a \text{ even} \\ a \neq 0}} \left[\frac{aq}{p}\right] \equiv \#AEF \bmod 2; \tag{1.44}$$

remember that $\#AEF$ is the number of integer pairs strictly inside the triangle AEF. From Lemma 1.6.16 we know the left hand side is μ and $\left(\frac{q}{p}\right) = (-1)^{\mu}$. Therefore

$$\left(\frac{q}{p}\right) = (-1)^{\mu} = (-1)^{\#AEF}. \tag{1.45}$$

Reversing the rolls of p and q, we see that

$$\left(\frac{p}{q}\right) = (-1)^{\nu} = (-1)^{\#AGF}, \tag{1.46}$$

where $\nu \equiv \#AGF \bmod 2$, with $\#AGF$ equal to the number of integer pairs strictly inside the triangle AGF.

Exercise 1.6.19. *Prove 1.46.*

Combining our expressions for μ and ν yields

$$\mu + \nu = \#AEF + \#AGF \bmod 2, \tag{1.47}$$

which is the number of integer pairs strictly inside the rectangle $AEFG$. There are $\frac{p-1}{2}$ choices for x ($x \in \{1, 2, \ldots, \frac{p-1}{2}\}$) and $\frac{q-1}{2}$ choices for $y \in \{1, 2, \ldots, \frac{q-1}{2}\}$), giving $\frac{p-1}{2}\frac{q-1}{2}$ pairs of integers strictly inside the rectangle $AEFG$. Thus,

$$\begin{aligned} \left(\frac{q}{p}\right)\left(\frac{p}{q}\right) &= (-1)^{\mu+\nu} \\ &= (-1)^{\#AEF+\#AGF} \\ &= (-1)^{\frac{p-1}{2}\frac{q-1}{2}}, \end{aligned} \tag{1.48}$$

which completes the proof of Quadratic Reciprocity.

Exercise 1.6.20 (Advanced)**.** *Let p be an odd prime. Are there infinitely many primes q such that q is a square mod p? The reader should return to this problem after Dirichlet's Theorem (Theorem 2.3.4).*

Chapter Two

Arithmetic Functions

We introduce many of the common functions of number theory, and the techniques to analyze them. We determine how rapidly many of these functions grow; for example, how many numbers at most x are prime. As seen in §1.1, this has immediate applications to cryptography (ensuring that there are sufficiently many primes to make certain codes reasonably secure), though such applications are not the driving force of our investigations. Primes are the building blocks of numbers; it is natural to study how many primes there are, and how they are distributed. Many of the other functions introduced here will reappear throughout the book (in Chapter 3 when we investigate primes in arithmetic progression, and Chapters 13 and 14 when we investigate writing integers as the sum of primes); for more on basic number theory see [Apo, HW, IR, NZM] as well as the first chapter of [IK]. In studying the number of primes we introduce numerous standard techniques of number theory, such as partial summation, telescoping series and estimating integrals.

2.1 ARITHMETIC FUNCTIONS

Definition 2.1.1. *An arithmetic function is a function from the natural numbers* $\mathbb{N} = \{0, 1, 2, \dots\}$ *to the complex numbers* $\mathbb{C}$.

Some of the most important examples of arithmetic functions are the following:

1. Set $e(n)$ equal to one if $n = 1$ and zero otherwise.

2. For s a complex number, set $\sigma_s(0) = 0$ and for positive integers n,

$$\sigma_s(n) \ = \ \sum_{d|n} d^s. \tag{2.1}$$

   $\sigma_0(n)$ is the number of divisors of n. The usual notation for the **divisor function** σ_0 is $d(n)$ or $\tau(n)$.

3. Set $\phi(n)$ to be the number of natural numbers $a \in \{1, \dots, n\}$ with the property that $(a, n) = 1$ (see Exercise 1.4.5); ϕ is called the **totient function**.

4. The **Möbius function** μ is defined by $\mu(1) = 1$ and

$$\mu(n) \ = \ \begin{cases} (-1)^s & \text{if } n = p_1 p_2 \cdots p_s,\ p_i \text{ distinct primes} \\ 0 & \text{otherwise.} \end{cases} \tag{2.2}$$

5. The **von Magnoldt function** Λ is defined by

$$\Lambda(n) \;=\; \begin{cases} \log p & \text{if } n = p^k, p \text{ prime} \\ 0 & \text{otherwise.} \end{cases} \tag{2.3}$$

6. Set $\pi(n) = \#\{p : p \text{ is a prime}, p \leq n\}$, the number of primes at most x.

Note the Möbius function is zero if n is not square-free, and the von Magnoldt function is zero if n is not a power of a prime.

Definition 2.1.2 (Multiplicative Function)**.** *An arithmetic function f is said to be **multiplicative** if for all $m, n \in \mathbb{N}$ with $(m, n) = 1$, we have*

$$f(mn) \;=\; f(m)f(n). \tag{2.4}$$

It is called completely multiplicative *if $f(mn) = f(m)f(n)$ for all $m, n \in \mathbb{N}$.*

It is not hard to see that $e(n)$, $\mu(n)$ and $\sigma_s(n)$ are multiplicative. The function $\phi(n)$ is also multiplicative, though it is slightly harder to prove. The von Magnoldt function is not multiplicative.

Exercise(h) 2.1.3. *Prove the above statements.*

Exercise(hr) 2.1.4. *In this exercise we collect some basic properties of $\phi(n)$. Verify the following statements:*

1. *$\phi(p^\alpha) = p^\alpha - p^{\alpha-1}$, p prime.*
2. *$\phi(n) = n\prod_{p|n}\left(1 - \frac{1}{p}\right)$.*
3. *$\sum_{d|n}\phi(d) \;=\; n$.*
4. *If $(a, n) = 1$, we have $a^{\phi(n)} \equiv 1 \bmod n$.*

Exercise 2.1.5. *Here we list some problems related to the divisor function $d = \sigma_0$.*

1. *Let $n = p_1^{\alpha_1}\cdots p_k^{\alpha_k}$. Prove that*

$$d(n) \;=\; \prod_{i=1}^{k}(\alpha_i + 1). \tag{2.5}$$

2. *For any fixed $A, B > 0$ one can find an n such that $d(n) \geq A\log^B n$; however, given any $\epsilon > 0$ there exists a C_ϵ such that for all n, $d(n) \leq C_\epsilon n^\epsilon$.*
3. *Prove that a number n is a perfect square if an only if $d(n)$ is odd.*
4. *Suppose we have 10000 doors numbered $1, 2, \ldots, 10000$. Suppose at first all of them are locked. First we go through all the doors and unlock all of them. Then we pick all even numbered doors, and change the status of their lock. Then we do the same with all the doors whose number is a multiple of three, then with four and so on, all the way to 10000. How many doors will be left unlocked?*

If f, g are two arithmetic functions, we define their **convolution** $f * g$ by

$$f * g(n) \;=\; \sum_{d|n} f(d) g\left(\frac{n}{d}\right). \tag{2.6}$$

The convolution is also know as the **Dirichlet product**.

Lemma 2.1.6. *The convolution $*$ has the following properties: for all arithmetic functions f, g, h we have*

1. $f * e = f$,
2. $f * g = g * f$,
3. $f * (g * h) = (f * g) * h$,
4. $f * (g + h) = f * g + f * h$.

Exercise 2.1.7. *Prove Lemma 2.1.6.*

Remark 2.1.8 (For Those Knowing Abstract Algebra)**.** Lemma 2.1.6 shows that the collection of all arithmetic functions under pointwise addition and $*$ multiplication forms a commutative ring. Note an arbitrary arithmetic function need not have a $*$ inverse.

Lemma 2.1.9. *If f, g are multiplicative functions, then so is $f * g$.*

Exercise 2.1.10. *Prove Lemma 2.1.9.*

We often let 1 denote the constant function that takes the value 1 at all natural numbers: $1(n) = 1$. We have

Lemma 2.1.11. $\mu * 1 = e$.

Proof. We need to show that

$$\sum_{d|n} \mu(d) \;=\; 0 \tag{2.7}$$

whenever $n \neq 1$; the case $n = 1$ is straightforward. Let $n = p_1^{\alpha_1} \cdots p_r^{\alpha_r}$. Any divisor d of n can be written as $d = p_1^{\beta_1} \cdots p_r^{\beta_r}$, with $\beta_i \leq \alpha_i$. By definition $\mu(d) = 0$ unless $\beta_i \in \{0, 1\}$. The number of divisors d with exactly k of the $\beta_i = 1$ is equal to $\binom{r}{k}$. For such divisors we have $\mu(d) = (-1)^r$. Consequently

$$\sum_{d|n} \mu(d) \;=\; \sum_{k=0}^{r} (-1)^k \binom{n}{k}, \tag{2.8}$$

which by the Binomial Theorem (Theorem A.1.8) equals $(1-1)^k = 0$. □

Exercise 2.1.12. *Prove*

$$\sum_{d^k|n} \mu(d) \;=\; \begin{cases} 1 & \text{if } n \text{ is } k^{\text{th}}\text{-power-free,} \\ 0 & \text{otherwise.} \end{cases} \tag{2.9}$$