

Under the Banyan Tree

Decoding Numbers



Sumit Gupta

Under the Banyan Tree: Decoding Numbers

“Numbers, like humans, have relationships and families.”

Sumit Gupta

Under the Banyan Tree: Decoding Numbers

by Sumit Gupta

Copyright © 2025 Sumit Gupta. All rights reserved.

First Edition: October 2025

While the author has used good faith effort to ensure that the information and instructions contained in this work are accurate, the author disclaim all responsibility for errors or omissions, or for damages resulting from the use of or reliance on the information contained herein.

No part of the material protected by this copyright notice may be reproduced or utilized in any form or by any means, electronic or mechanical, including photocopying, recording, or by any information storage and retrieval system without the written permission of the copyright owner.

This book is dedicated to my elder brother who taught me mathematics.

Contents

List of Notations	xiii
Preface	xxi
Why This Book?	xxi
Why The Title?	xxii
A Bit About Indian Statistical Institute	xxii
How to Use This Book	xxv
Acknowledgements	xxvi
About the Author	xxix
I Theoretical Foundation	1
1 Set Theory	5
1.1 Fundamental Concepts of Sets	6
1.1.1 Basic Definitions and Notation	6
1.2 Set Relations and Properties	7
1.2.1 Subset Relations	7
1.2.2 Power Sets	10
1.3 Set Operations and Their Representations	11
1.3.1 Basic Set Operations and Venn Diagrams	11
1.3.2 Algebraic Properties of Set Operations	14

1.4	Counting Techniques in Set Theory	16
1.4.1	Fundamentals of Set Cardinality	16
1.4.2	The Inclusion-Exclusion Principle	16
1.5	Practice Exercises	19
2	Number System	23
2.1	Real Numbers	24
2.1.1	Historical Development of Number Systems	24
2.1.2	Definition and Properties of Real Numbers	25
2.1.3	Properties of Real Numbers	26
2.1.4	Rational Numbers	29
2.1.5	Irrational Numbers	30
2.2	Integer Functions	33
2.2.1	Greatest Integer Function	33
2.2.2	Smallest Integer Function	36
2.2.3	The Fractional Part Function	38
2.2.4	The Factorial Function	40
2.3	Complex Numbers	43
2.3.1	Basic Operations with Complex Numbers	44
2.3.2	Complex Conjugate	44
2.3.3	Modulus of Complex Numbers	46
2.3.4	Polar Form	47
2.3.5	De Moivre's Theorem	48
2.3.6	Roots of Complex Numbers	49
2.3.7	Geometric Interpretation of n -th Roots	50
2.4	Practice Exercises	54
3	Foundational Mathematics	57
3.1	The Pigeonhole Principle	58

3.2 Mathematical Induction	59
3.2.1 Strong Induction	61
3.3 The Arithmetic Mean-Geometric Mean Inequality	64
3.4 Mathematical Functions	71
3.4.1 Introduction to Functions	71
3.4.2 Types of Functions Based on Mapping	72
3.4.3 Monotonic Functions	77
3.4.4 Convex Functions	81
3.5 Polynomials and Their Roots	85
3.5.1 Polynomial Division	85
3.5.2 Polynomial Roots	87
3.5.3 The Fundamental Theorem of Algebra	92
3.5.4 The Complete Factorization Theorem	93
3.5.5 Vieta's Formulas	95
3.6 Binomial Theorem	99
3.6.1 Binomial Coefficients	99
3.6.2 The Binomial Expansion	100
3.6.3 Important Related Results	101
3.7 Practice Exercises	107
3.8 Practice Exercises	108
 II Introduction to Number Theory	 111
 4 Divisibility	 115
4.1 Divisibility	116
4.1.1 Basic Properties of Divisibility	116
4.1.2 Division Algorithm	117
4.2 Greatest Common Divisor	119

4.2.1	Greatest Common Divisor (GCD)	119
4.2.2	Bézout's Identity	122
4.3	Least Common Multiple	127
4.4	Linear Diophantine Equations	129
4.4.1	Existence of Solutions	130
4.4.2	Finding All Solutions	130
4.4.3	Finding a Particular Solution	132
4.4.4	Diophantine Equations with Constraints	133
4.5	Practice Exercises	136
5	Prime Numbers	139
5.1	Introduction to Prime Numbers	140
5.2	Fundamental Properties of Prime Numbers	142
5.2.1	Prime Factorization	142
5.2.2	Prime Factorization and its Relation to GCD and LCM	144
5.3	Divisor Function	146
5.3.1	Sum of Divisors	149
5.3.2	Mersenne Primes	153
5.4	Distribution of Prime Numbers	156
5.5	Practice Exercises	158
6	Modular Arithmetic	161
6.1	Introduction to Modular Arithmetic	162
6.2	Multiplicative Inverses in Modular Arithmetic	163
6.2.1	Definition and Existence	164
6.3	Euler's Totient Function	166
6.3.1	Definition	166
6.3.2	Properties of Euler's Totient Function	166
6.3.3	Sum of Euler's Totient Function Over Divisors	169

6.4 Euler's Theorem	172
6.4.1 Statement and Motivation	172
6.4.2 Intuitive Understanding of Euler's Theorem	173
6.4.3 Fermat's Little Theorem as a Special Case	177
6.5 Wilson's Theorem	178
6.6 Wilson's Theorem for Composite Number	182
6.7 Polynomial Congruences	185
6.7.1 Solutions to Polynomial Congruences Modulo a Prime	186
6.8 Practice Exercises	189
7 The Möbius Function	193
7.1 Dirichlet Convolution	194
7.1.1 Properties of Dirichlet Convolution	195
7.1.2 Geometric Visualization	200
7.2 The Möbius Function	201
7.2.1 Geometric Visualization	203
7.2.2 Properties of the Möbius Function	204
7.3 Möbius Inversion Formula	208
7.4 Applications to Arithmetic Functions	213
7.4.1 Euler's Totient Function	213
7.5 Practice Exercises	217
8 Primitive Root	221
8.1 Introduction to Primitive Roots	222
8.2 Primitive Roots	222
8.2.1 Exploring Order of an Integer in Modular Arithmetic	223
8.2.2 The Polynomial Connection: Roots of $x^k \equiv 1 \pmod{p}$	225
8.2.3 Counting Elements by Order	228
8.2.4 The Existence of Primitive Roots for Primes	229

8.3 Primitive Roots for Composite Numbers	230
8.3.1 Primitive Roots for Prime Powers	231
8.4 Non-Existence of Primitive Roots	237
8.5 Practice Exercises	241
9 Quadratic Residues and Reciprocity	245
9.1 Quadratic Residues	246
9.2 The Legendre Symbol	248
9.2.1 Properties of the Legendre Symbol	249
9.3 Quadratic Reciprocity	252
9.4 Computing Legendre Symbols Efficiently	259
9.5 The Jacobi Symbol	260
9.6 Quadratic Forms and the Legendre Symbol	261
9.6.1 Representation of Primes	261
9.6.2 Extension to Composite Numbers	264
9.7 Practice Exercises	266
III Past ISI Exam Questions	269
10 Questions from Past UGA Papers	273
10.1 2025	273
10.2 2024	275
10.3 2023	277
10.4 2022	279
10.5 2021	281
10.6 2020	282
10.7 2019	283
10.8 2018	285
10.9 2017	286

11 Questions from Past UGB Papers	289
11.1 2025	289
11.2 2024	291
11.3 2023	292
11.4 2022	293
11.5 2021	294
11.6 2020	295
11.7 2019	296
11.8 2018	297
11.9 2017	298
IV Mock Test	299
Afterword	309

List of Notations

This book uses the following notations throughout. Familiarizing yourself with these symbols will enhance your reading experience.

Number Sets

Symbol	Description
\mathbb{N}	The set of natural numbers: $\{1, 2, 3, \dots\}$
\mathbb{N}_0	The set of natural numbers including zero: $\{0, 1, 2, 3, \dots\}$
\mathbb{Z}	The set of integers: $\{\dots, -2, -1, 0, 1, 2, \dots\}$
\mathbb{Z}^+	The set of positive integers: $\{1, 2, 3, \dots\}$
\mathbb{Z}^-	The set of negative integers: $\{\dots, -3, -2, -1\}$
\mathbb{Q}	The set of rational numbers: $\{a/b \mid a \in \mathbb{Z}, b \in \mathbb{Z}, b \neq 0\}$
\mathbb{R}	The set of real numbers
\mathbb{C}	The set of complex numbers
\mathbb{P}	The set of prime numbers: $\{2, 3, 5, 7, 11, \dots\}$

Set Theory

Symbol	Description
$A \cup B$	Union of sets A and B
$A \cap B$	Intersection of sets A and B
$A \setminus B$	Set difference: elements in A but not in B
$A \Delta B$	Symmetric difference: elements in either A or B but not both
$A \times B$	Cartesian product of sets A and B

Continued on next page

Set Theory (continued)

Symbol	Description
$a \in A$	a is an element of set A
$a \notin A$	a is not an element of set A
\emptyset	The empty set
$ A $	Cardinality (size) of set A
\subseteq	Subset relation
\subset	Proper subset relation
$\mathcal{P}(A), 2^A$	Power set of A

Divisibility and Congruence

Symbol	Description
$a b$	a divides b (i.e., b is divisible by a)
$a \nmid b$	a does not divide b
$\gcd(a, b)$	Greatest common divisor of a and b
$\text{lcm}(a, b)$	Least common multiple of a and b
$a \equiv b \pmod{m}$	a is congruent to b modulo m (i.e., $m (a - b)$)
$a \bmod m$	The remainder when a is divided by m
\mathbb{Z}_m	The set of residue classes modulo m : $\{0, 1, 2, \dots, m - 1\}$
a^{-1}	Multiplicative inverse of a modulo m
$\text{ord}_n(a)$	The multiplicative order of a modulo n

Number-Theoretic Functions

Symbol	Description
$\lfloor x \rfloor$	Floor function: the greatest integer not exceeding x
$\lceil x \rceil$	Ceiling function: the least integer not less than x
$\{x\}$	Fractional part of x : $\{x\} = x - \lfloor x \rfloor$
$n!$	Factorial of n : $n! = n \times (n - 1) \times \dots \times 2 \times 1$
$\tau(n)$	Number of positive divisors of n
$\sigma(n)$	Sum of all positive divisors of n

Continued on next page

Number-Theoretic Functions (continued)

Symbol	Description
$\sigma_k(n)$	Sum of the k -th powers of all positive divisors of n
$\phi(n), \varphi(n)$	Euler's totient function: number of integers k in range $1 \leq k \leq n$ coprime to n
$\mu(n)$	Möbius function
$f * g$	Dirichlet convolution of arithmetic functions f and g
f^{-1}	Dirichlet inverse of arithmetic function f
$\omega(n)$	Number of distinct prime factors of n
$\Lambda(n)$	von Mangoldt function

Primes and Factorization

Symbol	Description
p_n	The n -th prime number ($p_1 = 2, p_2 = 3, \dots$)
$\pi(x)$	Prime counting function: the number of primes not exceeding x
M_p	Mersenne number: $M_p = 2^p - 1$

Complex Numbers

Symbol	Description
i	Imaginary unit: $i^2 = -1$
\bar{z}	Complex conjugate of z
$ z $	Modulus (absolute value) of complex number z
$\operatorname{Re}(z)$	Real part of complex number z
$\operatorname{Im}(z)$	Imaginary part of complex number z
$\arg(z)$	Argument of complex number z
$e^{i\theta}$	Euler's formula: $\cos \theta + i \sin \theta$

Binomial Coefficients

Symbol	Description
$\binom{n}{k}$, $C(n, k)$	Binomial coefficient: $\binom{n}{k} = \frac{n!}{k!(n-k)!}$

Quadratic Residues and Legendre Symbol

Symbol	Description
$\left(\frac{a}{p}\right)$	Legendre symbol: quadratic residue symbol
$\left(\frac{a}{n}\right)$	Jacobi symbol: generalization of Legendre symbol

Intervals and Real Analysis

Symbol	Description
$[a, b]$	Closed interval from a to b (includes endpoints)
(a, b)	Open interval from a to b (excludes endpoints)
$[a, b)$	Half-open interval (includes a , excludes b)
$(a, b]$	Half-open interval (excludes a , includes b)
$[a, \infty)$	Ray from a to infinity (includes a)
$(-\infty, b]$	Ray from negative infinity to b (includes b)

Logical Symbols

Symbol	Description
\forall	Universal quantifier: “for all”
\exists	Existential quantifier: “there exists”
\nexists	“There does not exist”
$\exists!$	“Such that”
\Rightarrow	Implication: “implies”
\Leftarrow	Reverse implication: “is implied by”
\Leftrightarrow	Logical equivalence: “if and only if”
\wedge	Logical conjunction: “and”
\vee	Logical disjunction: “or”

Continued on next page

Logical Symbols (continued)

Symbol	Description
\neg	Logical negation: “not”
\therefore	“Therefore”
\because	“Because”

Summation and Product Notation

Symbol	Description
\sum	Summation symbol
\prod	Product symbol
$\sum_{i=1}^n a_i$	Sum of a_i from $i = 1$ to n
$\prod_{i=1}^n a_i$	Product of a_i from $i = 1$ to n
$\sum_{d n} f(d)$	Sum over all divisors d of n

Functions and Relations

Symbol	Description
$f : A \rightarrow B$	Function f from set A to set B
f^{-1}	Inverse function of f
$f \circ g$	Composition of functions f and g
id_A	Identity function on set A

Other Mathematical Symbols

Symbol	Description
\sqrt{a}	Square root of a
$\sqrt[n]{a}$	n -th root of a
a^b	a raised to the power b
\log	Logarithm (typically base 10 or natural)
\ln	Natural logarithm (base e)

Continued on next page

Other Mathematical Symbols (continued)

Symbol	Description
\approx	Approximately equal to
\sim	Asymptotically equivalent to
\square	End of proof (Q.E.D.)
∞	Infinity
\pm	Plus or minus

Abbreviations

Abbreviation	Meaning
WLOG	Without loss of generality
iff	If and only if
s.t.	Such that
w.r.t.	With respect to
GCD	Greatest Common Divisor
LCM	Least Common Multiple
AM-GM	Arithmetic Mean-Geometric Mean

Throughout this book, we use standard mathematical notation whenever possible. Any deviations or special notations are explained when they are introduced.

Preface

Why This Book?

This is a book that I had wanted to write for a very long time. An initial plan was to write this during the first year of undergraduate studies at the Indian Statistical Institute. However, other interests took precedence over this book and what started as mild procrastination gradually transformed into a decades-long delay.

The primary motivation for this book stemmed from the challenges I faced while preparing for the B.Stat (Hons) entrance examination. Apart from one generic Number Theory text, no book existed that was specifically tailored for this entrance examination. While past examination questions were available, there was a glaring absence of resources to guide students through solving these problems. Such preparation becomes particularly burdensome for students from underprivileged backgrounds who lack proper guidance.

During my high school years, I had no access to private tutoring and frequently found myself stuck trying to decipher solutions independently. I was fortunate to have my elder brother's assistance, but not every aspiring student enjoys such privilege. This book aims to bridge that gap and provide the support I wished I had during my own preparation journey.

What finally triggered me to write this book was not only reflecting on my past struggles, but also seeing the same challenges mirrored in today's students. Over the last couple of months, numerous aspirants have reached out to me through social media, seeking information about the institute and the entrance examination, including recommendations for relevant study materials.

Their questions reminded me of my own difficulties and rekindled my desire to create the resource I once needed. What had been a dormant idea was suddenly brought back to life by these interactions.

I realized that the information on the Indian Statistical Institute continues to remain sparse among high school students. Even though I appeared for the entrance examination long ago, I find it surprising that a book aimed at the entrance examination still does not exist. This lack of awareness particularly affects talented students from smaller towns who might never discover this opportunity. Hence, the motivation for writing the book remains just as relevant today as it was two decades ago.

Life, much like numbers, is full of unexpected discoveries and patterns. I am grateful to have finally completed this long-planned project despite my current responsibilities. I hope it provides the guidance and support that I once wished for during my own preparation journey.

Why The Title?

This book is the first in a series that I plan to write covering various topics for the entrance examination. Initially, I had planned to write one comprehensive book covering all topics. However, I quickly realized that this approach was impractical, and I would not be able to do justice to the material in a single volume. Hence, I decided to organize the content by topic and devoted this first book to my favorite subject, Number Theory.

When I first encountered Number Theory, I fell in love with the subject (and it remains my favorite to this day). There is something magical about the most fundamental building blocks of mathematics: numbers.

Even though the concepts introduced in the book might seem complex at times, you will be astonished to realize that all the concepts lead to greatest common divisor (GCD), a concept introduced to us as highest common factor in primary school.

Instead of “Decoding Numbers,” a more appropriate title for the book might have been “Decoding GCD,” as that signifies the core learning of the book. I preferred the first title since it offers broader scope to incorporate additional related concepts that extend beyond just GCD.

This book series is titled “Under the Banyan Tree” because I connect with the emblem of ISI, which features *Ficus benghalensis*, symbolizing resilience, unity, and the interconnectedness of life. The way a banyan grows, dropping roots from branches that become new trunks creating a network of connected supports, reminds me of how mathematical ideas develop and connect to each other.

Mathematics, like nature, reveals its beauty to those who spend time with it. Just as one might contemplate the world’s mysteries under the shade of a banyan tree, this book invites you to explore the fascinating patterns and structures of number theory.

A Bit About Indian Statistical Institute

History and Academic Programs

The Indian Statistical Institute (ISI) was founded by Professor P.C. Mahalanobis in 1931 at Presidency College in Kolkata. In the 1940s, it shifted to its current location in Baranagar when Mahalanobis purchased a plot there.

ISI officially operated from a building called “Amrapali,” named by the renowned poet and Nobel

laureate Rabindranath Thakur. The name was chosen because the plot was mostly occupied with mango trees. Even today, a cluster of mango trees surrounds the building, reminiscent of its original landscape.

In 1959, ISI gained the status of an Institution of National Importance by an act of Parliament. Today, it is headquartered in Kolkata with additional centers in Delhi, Bangalore, Chennai, and Tezpur.

ISI offers undergraduate, postgraduate, and doctoral programs in statistics, mathematics, computer science, and quantitative economics. The institute offers the B.Stat (Hons) program at its Kolkata campus, while the B.Math (Hons) program is offered at the Bengaluru campus.

The institute does not charge any academic fees for the two undergraduate programs. Rather, it provides a monthly stipend, along with a yearly contingency grant for academic purchases. Undergraduate students from ISI automatically progress to their respective Masters programs, after which they typically either pursue doctoral studies or begin careers in the corporate sector.

Academic Experience and Placements

The education at ISI, Kolkata is of exceptional quality and has a very high focus on rigor. The faculty to student ratio is really attractive, which makes professors accessible to students and one could easily work with them on independent research projects. I had worked on many such research projects throughout my studies. One such research project was my Masters' thesis, which is a requirement for every M.Stat student to submit at the end of the final year. I had done my thesis under the tutelage of Dr. Sourabh Bhattacharya and Dr. Guruprasad Kar, who supported me immensely and I will forever remain grateful to them. I am generally proud of my thesis, which was on Bayesian Nonparametric and had created a novel class of time-series models fusing Dynamic Linear Models (DLMs) with Dirichlet Process (DPs).

Top researchers from all over the world regularly visit ISI, Kolkata. One can attend their talk, which is generally open to students, and can get introduced to them. The library at ISI, Kolkata is huge and is a collection of books, journals, and manuscripts from all over the world. In short, ISI provides one with ample resources to succeed in their studies and engage in their own independent research.

I have often been enquired about the placements post the Masters degree at ISI. I think the placements, be it academic or industry, are comparable to the other top institutions in the country. Post graduation, quite a number of my classmates pursued doctoral programs at the world's top universities. Like many others, I joined a quantitative modeling team at one of the investment banks and worked on designing trading strategies in the high-frequency space.

ISI has a distinguished alumni network present in many prestigious academic institutions worldwide, making significant contributions to their respective areas of research. Quite a number of the alumni have also established themselves in leadership positions of the world's leading firms. I hear that with the growth of India and the increased demand for quantitative talent,

the industry placements have improved significantly, and many graduates now secure positions at prestigious investment management firms and hedge funds.

Campus Life and Personal Experiences

Being a B.Stat (Hons) student, I spent most of my time at the Kolkata campus. Hence, my reflections are biased towards it.

As a student in the mid 2000s, I witnessed the institute undergo major renovations. More or less a new version of every single building was created. The undergraduate stipend was revised from rupees 500 to 800, and the hostel and main campus were connected via an underground passage. A brand new football ground, which also doubled as a cricket ground, was also created.

In my first year, a brand new Boys' hostel was launched. Each student had a separate room with a private lane connecting two rooms and hosting an attached washroom to be shared between two students. It was really luxurious by Indian university standards back then!

A discussion about the Kolkata campus is incomplete without a mention of the day canteen. The canteen was open for everyone and would serve three rosogollas for a rupee. A chapati would cost 15 paise and it always proved to be more troublesome to get the correct change than the food. By graduation, I had secured a good collection of 5 paise coins, which might have become antique pieces by now if I had decided to retain them.

I also had another canteen at the Boys' hostel, which was reserved for the residents. The food quality there was lacking though. Fortunately, the Kolkata campus is centrally located in the city, with vibrant markets and food establishments just outside its gates. This convenient location offered plenty of affordable and delicious options available throughout the day.

These impressive renovations and developments at ISI during my time were not coincidental. They coincided with the chairmanship of the late Mr. Pranab Mukherjee, who was a frequent visitor to the campus. I recall one particular visit when he came solely to inaugurate our football ground, his security detail outnumbered the students present on campus. On another memorable occasion, he arrived accompanied by several prominent leaders of the CPI(M) party, including the late Shri Buddhadeb Bhattacharya, former Chief Minister of West Bengal.

Under the late Mr. Mukherjee's Chairmanship, ISI received substantial funding. He seemed invested in the growth of the institute and I think his position as a senior Cabinet Minister likely helped secure these funds.

An interesting tidbit about the institute is that there is a dinosaur named after ISI. The Geology Building hosts a museum on its ground floor and it has a complete fossil skeleton of a 47 feet long sauropod from the early Jurassic period (about 160 million years back).

I could muster to enter the museum only two times. The first time was because I had taken Geology as an optional subject and students were given a tour. The structure is just marvelous and breathtaking and is supported by an internal metal framework that holds the bones in

their proper anatomical position. A human could not even reach upto its knees. It was a fun experience, which could have easily turned into a disaster if it suddenly started walking.

I was smitten with the giant and whenever I would go past the building, I would try to take a peek at the structure through the transparent mirror in the door. Another chance came when once by a sheer stroke of luck the door to the museum was open with very few folks inside. I took swift steps inside, saw it hurriedly and ran out quickly.

I have many more memories to share, but I must restrain myself or the preface would rival the book itself in length.

How to Use This Book

This book is primarily meant for students preparing for the B.Stat (Hons) and B.Math (Hons) entrance examinations. However, given that a lot of concepts overlap with other similar exams, one might also find it useful to prepare for the Mathematics Olympiad, undergraduate programs at Chennai Mathematical Institute (CMI) and other such competitive examinations.

This book is divided into five parts:

Part I focuses on the foundational concepts, which are not directly related to Number Theory but are needed to get a comprehensive understanding of the subject.

Part II of the book starts with the basic definitions of the number system and then quickly ramps up to defining more complicated theorems, ending in primitive roots. Each chapter comes with multiple examples spread throughout and exercises at the end of the chapter. This is the core of the book and deserves the most attention.

Part III provides Number Theoretic questions that had appeared in the past examination papers. Part IV complements them by providing two mock tests replicating the feel for the ISI entrance examination.

Part V provides solutions to the end of the chapter exercises, past exam questions, and mock tests. Use this section as a reference after attempting the problems. To limit the size of the book, I have moved this section online. You can access the solutions by visiting vatvriksha.com.

Mathematical theorems are nothing more than the outcome of properties of various mathematical structures. Once you have a conceptual understanding of mathematical structures, then you can understand and derive the properties of mathematical structures and the related theorems intuitively. Hence, the primary focus of the book is on developing a visualization of the mathematical structures. I do this by breaking down concepts into smaller components and working on examples.

Having developed an intuitive understanding, each chapter then adds a formal proof to each theorem. This helps an aspirant learn how to correctly write and prove mathematical results, which is essentially tested in an ISI entrance examination.

Each chapter ends with exercises for one to cement their conceptual understanding. Solutions to these questions are also provided for one to cross-check and learn. My suggestion is to make multiple attempts at solving the exercises before looking at the solutions. This process of grappling with problems through multiple attempts is fundamental to developing mathematical maturity and analytical thinking. Over time, the number of attempts needed to solve an exercise will decrease, reflecting the mathematical maturity that you have developed.

I also provide past exam questions with their solutions. The past examinations help an aspirant to get used to the pattern of questions asked in the entrance examination and prepare accordingly. I also provide two mock tests for one to simulate an actual exam setting.

Hope this book helps you achieve your goals.

Acknowledgements

I am deeply grateful to everyone who contributed to bringing this book to life.

My heartfelt thanks go to all those who read the manuscript and helped improve the initial draft. Special recognition goes to Manish Kumar Choudhary (FinanceOps), who was the first to read the book in its entirety. His encouraging words about the content sustained my motivation throughout the post-writing stage.

I am indebted to Soumendra Dhanee (Invideo AI) for providing an innovative alternative to the usual inclusion-exclusion solution to the co-prime counting problem. His creative approach, discussed in Example 6.3.2, focuses on finding the count of all numbers co-prime to any of the primes among 3, 5, and 7. This contribution enriched the mathematical content of this book.

I acknowledge the assistance of Claude AI in formatting and typesetting this manuscript, which allowed me to focus on the mathematical substance rather than technical formatting challenges.

I would also like to express my gratitude to the Indian Statistical Institute, which not only shaped my academic foundation but continues to inspire students like those this book aims to serve. The institute's commitment to excellence in mathematical education remains the driving force behind this work.

Finally, after decades of delay, completing this long-planned project feels deeply fulfilling. Most importantly, I thank you, the reader, for embarking on this journey through number theory with me. I hope you discover the same wonder and beauty in numbers that has captivated mathematicians for centuries, and that this book serves as a stepping stone toward your academic aspirations.

Sumit Gupta
October 2025

About the Author

The author earned his B.Stat.(Hons.) and M.Stat. degrees from the Indian Statistical Institute, Kolkata in 2008, and later completed his masters studies at the University of California, Berkeley. During high school, he qualified for the Regional Mathematical Olympiad and was selected for Chennai Mathematical Institute's B.Sc.(Hons.) program in Mathematics and Computer Science.

His professional journey spans multiple domains: four years in quantitative trading at Morgan Stanley, two years in quantitative investing at Goldman Sachs, and a leadership role as Head of Artificial Intelligence at a fintech unicorn. He founded pandainuniv.com, an educational platform that guides students through PhD decision-making.

This cross-disciplinary background in quantitative finance and artificial intelligence informs his approach to demonstrating how mathematical theory connects to practical applications. His own struggles with entrance examinations inspired him to develop the preparatory resources he once needed. His enthusiasm for Number Theory began in high school, when he first encountered the field's elegant structures.

Contact: sumiitguptame@gmail.com.

Part I

Theoretical Foundation

Chapter 1

Set Theory

Bernard Bolzano (1781-1848)



Bernard Bolzano was a Bohemian mathematician, logician, philosopher, and theologian who anticipated many key concepts of set theory decades before Georg Cantor formalized the field. Working in relative isolation in Prague, Bolzano developed surprisingly modern ideas about infinite collections and their properties that were far ahead of his time. In his work “Paradoxes of the Infinite” (published posthumously in 1851), Bolzano distinguished between different types of infinity and explored ideas related to the “size” of infinite sets, approaching the concept of cardinality that Cantor would later formalize. Additionally, Bolzano worked on rigorous foundations for calculus, developing early formulations of concepts like limits and continuity using set-theoretic ideas.

Bolzano was forced to leave his university position because his progressive theological and political views conflicted with the Austrian authorities. He spent much of his later life under police surveillance and was forbidden from publishing or teaching. Despite these restrictions, he continued his mathematical work in isolation, leaving behind thousands of unpublished pages of manuscripts that weren’t fully appreciated until the 20th century, long after his ideas had been independently rediscovered by others.

Welcome to the beginning of your journey toward the ISI entrance examination!

In this chapter, we will explore set theory, which is the fundamental mathematical language that provides the framework to formalize mathematical concepts, making it essential for rigorous definitions and proofs. We introduce the essential elements of set theory: definitions, notations, operations, and properties. We will explore how sets can be combined, compared, and counted, developing a toolkit of techniques that will prove invaluable throughout the ISI examination.

1.1 Fundamental Concepts of Sets

Set theory forms the foundation of modern mathematics, and hence a thorough understanding of sets and their operations is essential for success in mathematical entrance examinations.

1.1.1 Basic Definitions and Notation

Definition: Set

A set is a well-defined collection of distinct objects. These objects are called the elements or members of the set.

Notation:

- If x is an element of set A , we write $x \in A$.
- If x is not an element of set A , we write $x \notin A$.

We represent standard number sets using established mathematical notation: \mathbb{Z} denotes integers, \mathbb{N} denotes natural numbers (positive integers starting from 1), \mathbb{Q} denotes rational numbers, and \mathbb{R} denotes real numbers. Sets are enclosed within braces, allowing us to represent them in two principal ways.

It is important to acknowledge that the definition of natural numbers varies across different mathematical contexts and regions. Some mathematicians and textbooks define $\mathbb{N} = \{0, 1, 2, 3, \dots\}$, including zero as a natural number. In this book, we follow the convention where $\mathbb{N} = \{1, 2, 3, 4, \dots\}$, excluding zero. When zero is included, the set is denoted as \mathbb{N}_0 .

Sets are enclosed within braces, allowing us to represent them in two principal ways. For explicit enumeration, we list all elements directly, as in $A = \{1, 2, 3, 4, 5\}$ or $C = \{-2, -1, 0, 1, 2\}$. For sets defined by properties, we use set-builder notation with a conditional statement following a vertical bar, such as $B = \{x \in \mathbb{N} \mid x < 6\}$, which represents all natural numbers less than 6, or $D = \{x \in \mathbb{Z} \mid -3 \leq x \leq 3\}$, which represents integers between -3 and 3, inclusive.

There are several types of sets with special characteristics. The **empty set** contains no elements and is denoted by \emptyset or $\{\}$. The **universal set**, often denoted by U , contains all elements under

consideration in a given context.

For instance, the set of real numbers \mathbb{R} might serve as our universal set U when working with various number systems. Alternatively, if our problem statement revolves around integers and their subsets, then \mathbb{Z} would be an appropriate choice for the universal set. The selection of U depends on the specific context of the mathematical problem we're addressing.

A **finite set** has a countable number of elements that eventually ends, while an **infinite set** contains an endless number of elements that cannot be completely enumerated. A **singleton set** is a specific type of finite set that contains exactly one element.

It's important to note that both \mathbb{R} and \mathbb{Z} are examples of infinite sets, as they contain endlessly many elements. In contrast, the sets $A = \{1, 2, 3, 4, 5\}$, $B = \{x \in \mathbb{N} \mid x < 6\}$, $C = \{-2, -1, 0, 1, 2\}$, and $D = \{x \in \mathbb{Z} \mid -3 \leq x \leq 3\}$ that we introduced earlier are all finite sets.

1.2 Set Relations and Properties

Understanding the relationships between sets is fundamental to mastering set theory. This section explores relationships between sets via subset relationships and power sets.

1.2.1 Subset Relations

Definition: Subset

A set A is a subset of set B , denoted by $A \subseteq B$, if every element of A is also an element of B . If $A \subseteq B$ and $A \neq B$, then A is a proper subset of B , denoted by $A \subset B$.

Earlier we discussed the empty set, which contains no elements. At the opposite end of the spectrum is the set itself. Every set is a subset of itself, since all of its elements are, by definition, contained within it. This might seem obvious, but it represents a critical distinction when we consider proper subsets, which specifically exclude this case of self-containment.

Mathematically, $A \subset B$ (proper subset) if $\exists y \in B$ such that $y \notin A$ (where the notation \exists means “there exists”). A proper subset must exclude at least one element of the original set, making the two sets distinctly different.

The relationship between \subseteq and \subset mirrors that between \leq and $<$ in inequalities. Just as $a \leq b$ allows for $a = b$, while $a < b$ requires a to be strictly less than b , the notation $A \subseteq B$ permits $A = B$, whereas $A \subset B$ demands that A be strictly contained within B .

Subsets have several important properties that form the foundation of set relationships.

Theorem: Subset Properties

For any sets A , B , and C :

1. $\emptyset \subseteq A$ for any set A
2. $A \subseteq A$ (reflexivity)
3. If $A \subseteq B$ and $B \subseteq C$, then $A \subseteq C$ (transitivity)
4. $A = B$ if and only if $A \subseteq B$ and $B \subseteq A$

Proof

For property 1, we need to prove that $\emptyset \subseteq A$ for any set A .

By definition, $\emptyset \subseteq A$ means that every element of \emptyset is also an element of A . Since the empty set contains no elements, there is no element in \emptyset that is not in A . Therefore, the statement “every element of \emptyset is also an element of A ” is vacuously true, and we can conclude that $\emptyset \subseteq A$.

For property 2, we need to prove that $A \subseteq A$ (reflexivity).

By definition, $A \subseteq A$ means that every element of A is also an element of A . This is clearly true since any element belongs to itself. Therefore, $A \subseteq A$ holds for any set A .

For property 3, we need to prove that if $A \subseteq B$ and $B \subseteq C$, then $A \subseteq C$ (transitivity).

Assume that $A \subseteq B$ and $B \subseteq C$. Let x be any element of A . Since $A \subseteq B$, we know that $x \in B$. Now, since $B \subseteq C$, we know that $x \in C$. Therefore, every element of A is also an element of C , which means $A \subseteq C$.

□

The fourth property establishes a key biconditional relationship: “ $A = B$ if and only if $A \subseteq B$ and $B \subseteq A$.” This statement involves the phrase “if and only if,” which represents a special type of logical relationship that appears frequently in mathematical theorems and proofs. To prove such statements rigorously, we need to understand the logical structure behind them.

Therefore, before proceeding with the proof of this important set equality, let us develop the necessary background in logical reasoning that will serve as a foundation for rigorous mathematical proofs throughout this book.

Logical Equivalence and Biconditional Statements

In mathematical logic, the phrase “if and only if” (often abbreviated as “iff”) establishes a biconditional relationship between two statements. The biconditional relationship “if and only

if" is symbolized by " \iff " in mathematical notation.

When we write $P \iff Q$, we are asserting that the statements P and Q have identical truth values in all cases. In other words, $P \iff Q$ implies that statements P and Q are logically equivalent.

Definition: Logical Equivalence

Two statements P and Q are logically equivalent if they have the same truth value in all possible cases. We denote this equivalence by $P \iff Q$.

To understand logical equivalence, we need to define two fundamental logical operators:

- $\neg P$ (negation): Represents "not P " or the denial of statement P . If P is true, then $\neg P$ is false, and vice versa.
- $P \wedge Q$ (conjunction): Represents " P and Q ." This statement is true only when both P and Q are true; otherwise, it is false.

Consider the statement P : "7 is a prime number," which is true. Its negation $\neg P$: "7 is not a prime number" is false. In set theory, if $x \in A$ means " x is an element of set A ," then $x \notin A$ (or $\neg(x \in A)$) means " x is not an element of set A ."

The biconditional $P \iff Q$ can be decomposed into two implications:

$$P \iff Q \equiv (P \implies Q) \wedge (Q \implies P) \quad (1.2.1)$$

This means that $P \iff Q$ is true precisely when both $P \implies Q$ and $Q \implies P$ are true. The conjunction symbol \wedge indicates that both implications must be satisfied simultaneously.

An important logical relationship related to implications is the *contrapositive*. The contrapositive of $P \implies Q$ is $\neg Q \implies \neg P$. These two statements are logically equivalent:

$$P \implies Q \equiv \neg Q \implies \neg P \quad (1.2.2)$$

The equivalence implies that we can prove the original implication by instead proving its contrapositive. For example, consider the statement "If a number is divisible by 6, then it is divisible by 3." Instead of proving this directly, we could prove its contrapositive: "If a number is not divisible by 3, then it is not divisible by 6." To see this is true, note that if a number leaves a remainder when divided by 3, it must also leave a remainder when divided by 6, since 6 is a multiple of 3. The contrapositive approach is often clearer when the original statement involves complex conditions, and allows us to convert a statement into a logically equivalent form that might be easier to analyze.

Another way to understand the biconditional statement $P \iff Q$ is through the equivalence:

$$P \iff Q \equiv (P \implies Q) \wedge (\neg P \implies \neg Q) \quad (1.2.3)$$

In mathematical proofs, when we need to establish that two statements are equivalent, we prove both directions of the implication. First, we assume P and prove Q (the “if” part), and then we assume Q and prove P (the “only if” part). Only after both the directions are established can we conclude that $P \iff Q$.

Proof of the Statement: $A = B \iff A \subseteq B$ and $B \subseteq A$

To prove the fourth property, we need to show that for sets A and B :

- If $A = B$, then $A \subseteq B$ and $B \subseteq A$
- If $A \subseteq B$ and $B \subseteq A$, then $A = B$

Proof

Part 1: First, we prove that if $A = B$, then $A \subseteq B$ and $B \subseteq A$.

Since $A = B$ by our assumption, we have $\forall x \in A$ we have $x \in B$ (where the notation \forall means “for all”) which by definition gives us $A \subseteq B$. Similarly, let $y \in B$. Since $A = B$, we have $y \in A$. This gives us $B \subseteq A$.

Therefore, if $A = B$, then both $A \subseteq B$ and $B \subseteq A$ are true.

Part 2: Now, we prove that if $A \subseteq B$ and $B \subseteq A$, then $A = B$.

Suppose $A \subseteq B$ and $B \subseteq A$. Let x be any element. We need to show that $x \in A$ if and only if $x \in B$. If $x \in A$, then $x \in B$ (since $A \subseteq B$). Conversely, if $x \in B$, then $x \in A$ (since $B \subseteq A$). Thus, $x \in A$ if and only if $x \in B$, which implies $A = B$.

Therefore, $A = B$ if and only if $A \subseteq B$ and $B \subseteq A$.

□

Now, let us move on to the power set concept, which builds directly on our understanding of subsets.

1.2.2 Power Sets

Definition: Power Set

The power set of a set A , denoted by $\mathcal{P}(A)$ or 2^A , is the set of all subsets of A , including the empty set and A itself.

Example 1.2.1

Find the power set of $A = \{a, b, c\}$.

Solution: The power set $\mathcal{P}(A)$ contains all possible subsets of A :

$$\mathcal{P}(A) = \{\emptyset, \{a\}, \{b\}, \{c\}, \{a, b\}, \{a, c\}, \{b, c\}, \{a, b, c\}\}$$

Thus, A has 3 elements, and its power set $\mathcal{P}(A)$ has $2^3 = 8$ elements.

For any finite set A with n elements, its power set $\mathcal{P}(A)$ always contains exactly 2^n elements. To understand this, we need to consider how we form any subset of set A . For each element in A , we make an independent decision: either include it in our subset or exclude it. Since we have exactly 2 choices for each element, and these choices are independent of one another, the total number of possible subsets is 2^n .

For example, with set $A = \{a, b, c\}$ containing 3 elements, we have 8 possible subsets ($2^3 = 8$): the empty set, three 1-element subsets, three 2-element subsets, and the complete set itself. This exponential relationship explains why power sets grow so quickly. A 10-element set has $2^{10} = 1,024$ subsets, while a 20-element set has over a million.

It is important to note that given two sets A and B , $A = B$ if and only if $\mathcal{P}(A) = \mathcal{P}(B)$. If $\mathcal{P}(A) = \mathcal{P}(B)$, then every possible subset of A is also a subset of B , and vice versa. This means A and B must contain exactly the same elements and no more and no less. If they differed by even one element, their power sets would contain different subsets. The power set essentially captures the complete “fingerprint” of elements of a set.

1.3 Set Operations and Their Representations

Set operations provide powerful tools for constructing new sets from existing ones. In this section, we will explore the fundamental operations that allow us to analyze and manipulate relationships between sets, establishing the algebraic foundation of set theory.

1.3.1 Basic Set Operations and Venn Diagrams

Venn diagrams provide a visual representation of sets, making it easier to visualize them. Named after British mathematician John Venn who introduced them in 1880, these diagrams represent sets as circles within a rectangle that represents the universal set.

Union: The union of two sets A and B contains all elements that belong to either set A or set B (or both). Mathematically, this can be represented as: $A \cup B = \{x \mid x \in A \text{ or } x \in B\}$.

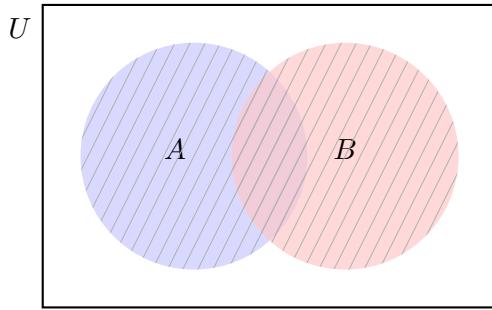


Figure 1.1: The cross-hatched region highlights the union $A \cup B$.

Intersection: The intersection of two sets A and B contains only the elements that are common to both the sets. Mathematically, this can be represented as: $A \cap B = \{x \mid x \in A \text{ and } x \in B\}$. Two sets A and B are disjoint if they have no elements in common, i.e., $A \cap B = \emptyset$.

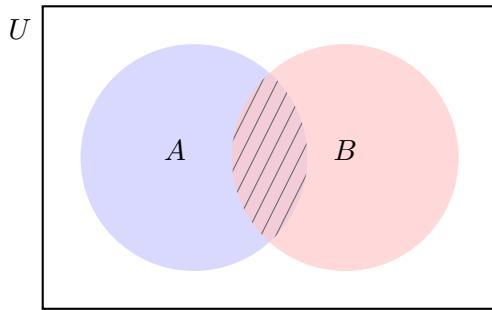


Figure 1.2: The highlighted region is $A \cap B$, which has the common elements of A and B .

Difference: The difference of two sets A and B contains only the elements that are exclusive to A . Mathematically, this can be represented as: $A \setminus B = \{x \mid x \in A \text{ and } x \notin B\}$.

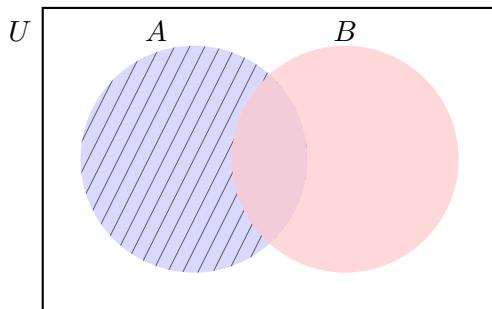


Figure 1.3: The highlighted region is $A \setminus B$, containing elements that are in A but not in B .

Complement: The complement of a set A contains all elements of universal set U that are not in set A . The complement of a set A is represented as $A^c = \{x \in U \mid x \notin A\}$.

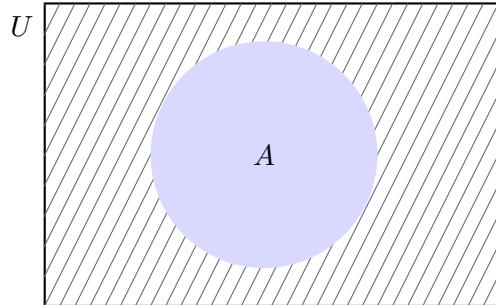


Figure 1.4: The highlighted component in the diagram represents the complement A^c , which includes all elements in the universal set that are not in set A .

Symmetric Difference: Symmetric difference of sets A and B contains all elements in either of the sets and not in their intersection; it is represented as: $A \Delta B = (A \setminus B) \cup (B \setminus A)$.

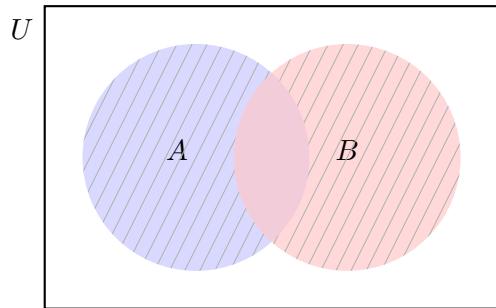


Figure 1.5: The highlighted region represents the symmetric difference $A \Delta B$, which includes all the elements that are either in set A or set B and not in both the sets.

Cartesian Product: The Cartesian product of sets A and B , denoted by $A \times B$, is the set of all ordered pairs (a, b) where $a \in A$ and $b \in B$:

$$A \times B = \{(a, b) \mid a \in A \text{ and } b \in B\}$$

The cardinality of a set, denoted by vertical bars around the set (e.g., $|A|$), represents the number of elements in the set. For finite sets, it simply counts how many distinct elements the set contains.

Example 1.3.1

Find the Cartesian product of $A = \{1, 2\}$ and $B = \{a, b, c\}$.

Solution: $A \times B = \{(1, a), (1, b), (1, c), (2, a), (2, b), (2, c)\}$

Note that $|A \times B| = |A| \times |B| = 2 \times 3 = 6$.

1.3.2 Algebraic Properties of Set Operations

Understanding the algebraic rules that govern set operations is essential for solving complex problems. We just discussed the five most commonly used set operations. In this part, we will explore the key properties of these operations, establishing important identities that will become powerful tools in our mathematical toolkit.

Theorem: Set Operation Properties

For any sets A , B , and C :

Commutative Laws:

- $A \cup B = B \cup A$
- $A \cap B = B \cap A$

Associative Laws:

- $A \cup (B \cup C) = (A \cup B) \cup C$
- $A \cap (B \cap C) = (A \cap B) \cap C$

Distributive Laws:

- $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$
- $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$

De Morgan's Laws:

- $(A \cup B)^c = A^c \cap B^c$
- $(A \cap B)^c = A^c \cup B^c$

Other Properties:

- $A \cup A = A$ (*idempotence*)
- $A \cap A = A$ (*idempotence*)
- $A \cup \emptyset = A$ (*identity for union*)

- $A \cap U = A$ (*identity for intersection*)
- $A \cup U = U$ (*domination for union*)
- $A \cap \emptyset = \emptyset$ (*domination for intersection*)
- $A \cup A^c = U$ (*complement law*)
- $A \cap A^c = \emptyset$ (*complement law*)
- $(A^c)^c = A$ (*double complement*)

While all these properties are fundamental to set theory, De Morgan's Laws deserve special attention as they provide powerful simplification techniques and appear frequently in entrance examinations. These powerful laws establish a fundamental relationship between unions, intersections, and complements, providing an elegant method for simplifying complex set expressions.

Example 1.3.2

Verify De Morgan's law $(A \cup B)^c = A^c \cap B^c$ for $U = \{1, 2, 3, 4, 5, 6\}$, $A = \{1, 2, 3\}$, and $B = \{2, 3, 4\}$.

Solution: First, we find $A \cup B = \{1, 2, 3, 4\}$. Then, $(A \cup B)^c = \{5, 6\}$.

Alternatively, we find: $A^c = \{4, 5, 6\}$ and $B^c = \{1, 5, 6\}$. Then, $A^c \cap B^c = \{5, 6\}$.

Since $(A \cup B)^c = A^c \cap B^c$, De Morgan's law is verified for this example.

De Morgan's Law also generalizes to multiple sets.

Theorem: De Morgan's Laws for Multiple Sets

For any collection of sets $\{A_1, A_2, \dots, A_n\}$:

$$\left(\bigcup_{i=1}^n A_i \right)^c = \bigcap_{i=1}^n A_i^c \quad (1.3.1)$$

$$\left(\bigcap_{i=1}^n A_i \right)^c = \bigcup_{i=1}^n A_i^c \quad (1.3.2)$$

In words:

- The complement of a union of sets equals the intersection of their complements.
- The complement of an intersection of sets equals the union of their complements.

1.4 Counting Techniques in Set Theory

In set theory, counting refers to the determination of cardinality—the number of elements contained within a set. While the concept appears straightforward for simple sets, it becomes increasingly sophisticated when dealing with set operations, combinations, and complex set expressions.

1.4.1 Fundamentals of Set Cardinality

The cardinality of a set A , denoted by $|A|$, represents its size or number of elements. For finite sets, this is simply a natural number. As we discussed previously, the basic set cardinality principles include:

- The empty set has cardinality zero: $|\emptyset| = 0$
- Singleton sets have cardinality one: $|\{a\}| = 1$
- The cardinality of power sets: $|\mathcal{P}(A)| = 2^{|A|}$
- For disjoint sets: $|A \cup B| = |A| + |B|$ when $A \cap B = \emptyset$
- For Cartesian products: $|A \times B| = |A| \cdot |B|$

These principles serve as the foundation for counting in set theory and provide straightforward methods for calculating cardinalities in simple scenarios. As set expressions become more intricate, direct counting becomes problematic. The primary challenge arises when sets overlap, making simple addition impossible due to the risk of counting elements multiple times.

When calculating the cardinality of a union, such as $|A \cup B|$, naively adding the individual cardinalities ($|A| + |B|$) would lead to inaccurate results, as elements in the intersection would be counted twice. This problem compounds significantly when dealing with three or more sets, where multiple overlapping regions create an exponential increase in complexity.

Real-world applications frequently present scenarios with multiple overlapping conditions represented as sets, such as counting students who satisfy various criteria or elements possessing multiple properties simultaneously. These intricate counting scenarios demand more sophisticated mathematical techniques that can systematically account for all possible overlaps while ensuring each element is counted exactly once in the final result.

1.4.2 The Inclusion-Exclusion Principle

The Inclusion-Exclusion Principle provides an elegant solution to the counting problem for unions of sets with overlaps. At its core, this principle establishes a formula for accurately determining the cardinality of a union by systematically accounting for all intersections.

For two sets, the principle states: $|A \cup B| = |A| + |B| - |A \cap B|$. For three sets, it expands to: $|A \cup B \cup C| = |A| + |B| + |C| - |A \cap B| - |A \cap C| - |B \cap C| + |A \cap B \cap C|$.

The general pattern alternates between adding and subtracting intersections of increasing order, ensuring each element is counted exactly once in the final tally.

Theorem: Inclusion-Exclusion Principle

For finite sets A and B :

$$|A \cup B| = |A| + |B| - |A \cap B|$$

For three sets A , B , and C :

$$\begin{aligned} |A \cup B \cup C| &= |A| + |B| + |C| \\ &\quad - |A \cap B| - |A \cap C| - |B \cap C| \\ &\quad + |A \cap B \cap C| \end{aligned}$$

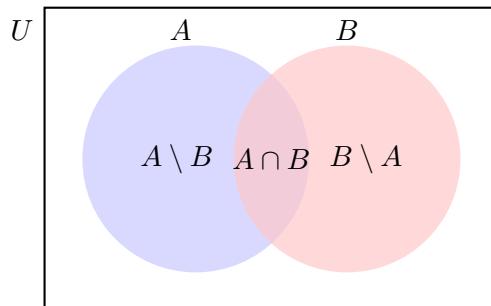
For n finite sets A_1, A_2, \dots, A_n :

$$\begin{aligned} \left| \bigcup_{i=1}^n A_i \right| &= \sum_{i=1}^n |A_i| - \sum_{1 \leq i < j \leq n} |A_i \cap A_j| \\ &\quad + \sum_{1 \leq i < j < k \leq n} |A_i \cap A_j \cap A_k| - \dots \\ &\quad + (-1)^{n+1} |A_1 \cap A_2 \cap \dots \cap A_n| \end{aligned}$$

In general, this can be written compactly as:

$$\left| \bigcup_{i=1}^n A_i \right| = \sum_{k=1}^n (-1)^{k-1} \sum_{1 \leq i_1 < i_2 < \dots < i_k \leq n} |A_{i_1} \cap A_{i_2} \cap \dots \cap A_{i_k}|$$

Let us develop an intuitive understanding of the inclusion-exclusion principle for two sets A and B . $|A \cup B|$ cannot be the sum of $|A|$ and $|B|$; the reason being $|A \cap B|$ is counted twice and hence needs to be deducted from the sum.



The union of sets A and B equals to the union of $A \setminus B$ and B . Thus, to calculate $|A \cup B|$, we need to get $|A \setminus B|$ and $|B|$. Typically, it is difficult to calculate $|A \setminus B|$ and is relatively easier

to get $|A \cap B|$. Hence, instead of calculating $|A \setminus B|$, we calculate $|A| - |A \cap B|$.

Example 1.4.1

In a class of 50 students, 30 study mathematics, 25 study physics, and 20 study both subjects. How many students study neither mathematics nor physics?

Solution: Let M be the set of students studying mathematics, and P be the set of students studying physics.

We know:

$$\begin{aligned}|M| &= 30 \\ |P| &= 25 \\ |M \cap P| &= 20\end{aligned}$$

Using the inclusion-exclusion principle:

$$\begin{aligned}|M \cup P| &= |M| + |P| - |M \cap P| \\ &= 30 + 25 - 20 \\ &= 35\end{aligned}$$

Therefore, the number of students studying at least one of the subjects is 35.

The number of students studying neither subject is:

$$50 - |M \cup P| = 50 - 35 = 15$$

Understanding set theory is crucial for the ISI entrance examination, as it forms the basis for many other mathematical concepts including functions, relations, probability, and combinatorics. The algebraic properties of sets provide a rigorous framework for logical reasoning and problem-solving across various domains of mathematics.

1.5 Practice Exercises

Exercise 1.1

Let $A = \{1, 2, 3, 4, 5\}$, $B = \{3, 4, 5, 6, 7\}$, and $C = \{1, 3, 5, 7, 9\}$. Find:

1. $A \cup B$
2. $A \cap B$
3. $A \setminus B$
4. $(A \cup B) \cap C$
5. $(A \cap C) \cup (B \cap C)$

Exercise 1.2

Prove that for any sets A , B , and C :

1. $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$
2. $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$

Exercise 1.3

Let $U = \{1, 2, 3, 4, 5, 6, 7, 8, 9, 10\}$ be the universal set, and let $A = \{1, 3, 5, 7, 9\}$ and $B = \{2, 4, 6, 8, 10\}$. Find:

1. A^c (the complement of A)
2. $(A \cup B)^c$
3. $A^c \cap B^c$
4. Verify De Morgan's Law: $(A \cup B)^c = A^c \cap B^c$

Exercise 1.4

In a class of 35 students, 20 study mathematics, 15 study physics, and 10 study both subjects.

1. How many students study mathematics or physics?
2. How many students study mathematics but not physics?
3. How many students study physics but not mathematics?
4. How many students study neither mathematics nor physics?

Exercise 1.5

Prove or disprove: If $A \subseteq B$ and $C \subseteq D$, then $A \times C \subseteq B \times D$.

Exercise 1.6

Let $\mathcal{P}(A)$ denote the power set of set A . If $A \subseteq B$, prove that $\mathcal{P}(A) \subseteq \mathcal{P}(B)$.

Exercise 1.7

Let A and B be sets. Prove that $A \subseteq B$ if and only if $A \cap B = A$.

Exercise 1.8

For sets A and B , define the symmetric difference as $A \Delta B = (A \setminus B) \cup (B \setminus A)$.

1. Prove that $A \Delta B = (A \cup B) \setminus (A \cap B)$
2. Show that $A \Delta B = B \Delta A$ (commutativity)
3. Prove that $(A \Delta B) \Delta C = A \Delta (B \Delta C)$ (associativity)

Exercise 1.9

Prove that for any sets A and B , $\mathcal{P}(A) \cap \mathcal{P}(B) = \mathcal{P}(A \cap B)$.

Exercise 1.10

For sets A, B, C , define the symmetric difference as $A \Delta B = (A \setminus B) \cup (B \setminus A)$ and consider the following statements:

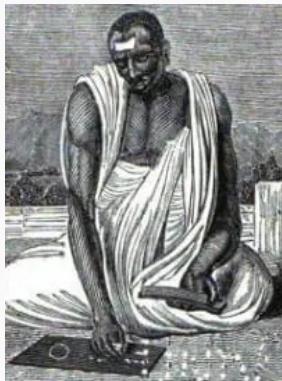
1. $|A \Delta B| = |A| + |B| - 2|A \cap B|$
2. $A \Delta B = \emptyset$ if and only if $A = B$
3. $(A \Delta B) \cap (B \Delta C) \subseteq A \Delta C$

Prove or disprove each statement.

Chapter 2

Number System

Brahmagupta (598-668 CE)



Brahmagupta was a pioneering Indian mathematician and astronomer whose work greatly influenced mathematics. As the head of the astronomical observatory at Ujjain, he authored the *Brāhma-sphuṭasiddhānta* (The Opening of the Universe), a landmark text containing groundbreaking mathematical concepts. He was the first to establish systematic rules for arithmetic with zero and negative numbers, treating them as legitimate numerical entities rather than mere symbols. Brahmagupta also provided methods for solving quadratic equations and formulated rules for mathematical operations involving zero, many of which remain fundamental today.

Brahmagupta states in his work “*Brahmasphutasiddhanta*” that dividing any number by zero results in “*khachheda*” (infinity). However, he also somewhat contradictorily claimed that zero divided by zero equals zero. It took centuries more for mathematicians to properly formalize the undefined nature of division by zero, making Brahmagupta’s early attempt to tackle this problem remarkably prescient despite its imperfections.

Number systems are foundational concepts we have worked with throughout our school years, often without realizing their full structure. The real number system provides a framework that connects the simple counting numbers we learn as children with more complex values like π and $\sqrt{2}$ that we encounter in advanced mathematics.

Apart from real numbers, we will also venture into the fascinating world of complex numbers. Complex numbers extend our number system to include the square roots of negative numbers and open doors to solving equations that have no real number solutions.

2.1 Real Numbers

The development of modern number system is one of the most profound intellectual achievements of humanity. Before delving deep into it, let us briefly trace their historical evolution.

2.1.1 Historical Development of Number Systems

When prehistoric humans first carved 29 notches into the Lebombo bone around 35,000 BCE, they couldn't have imagined how their simple counting marks would evolve into the sophisticated mathematical systems that now underpin our modern world. This journey from notched bones to abstract mathematical theory reveals humanity's remarkable capacity for intellectual growth.

The practical needs of early civilizations drove the first major innovations. In the fertile valleys of Egypt and Mesopotamia, administrators needed to count harvests, calculate taxes, and design monumental architecture. The Babylonian choice of a base-60 system, seemingly arbitrary, proved remarkably durable, persisting in our 60-minute hours and 360-degree circles. Each time we check our watch or use a protractor, we are connecting with mathematicians who lived 5,000 years ago.

The Indian contribution to our number system was truly revolutionary. Between the 1st and 6th centuries CE, Indian mathematicians developed what we now call the Hindu-Arabic numeral system. What made this system transformative was its elegant simplicity: using just ten symbols (0-9), humans could represent any number, no matter how large. The system's power comes from its positional nature, where the same symbol represents different values depending on its position. This innovation dramatically simplified calculation and created a system so efficient that it has become universal, displacing all other numerical notations across the globe.

As mathematical knowledge flowed westward through the Islamic world, scholars like Al-Khwarizmi (whose name gives us the word “algorithm”) refined these ideas, creating systematic approaches to solving equations that would eventually revolutionize European mathematics during the Renaissance. When Fibonacci introduced Hindu-Arabic numerals to Europe in his 1202 book *Liber Abaci*, he sparked a computational revolution. The computational advantage was so significant that despite initial resistance, European merchants and scientists gradually abandoned cumbersome Roman numerals in favor of the more practical “Arabic” digits.

The formalization of number systems reached its zenith in the 19th century, when mathematicians including Richard Dedekind, Georg Cantor, and Karl Weierstrass tackled a problem that had troubled mathematicians since ancient Greece: how to rigorously define the real number line. Through constructions like Dedekind cuts and Cauchy sequences, they provided precise mathematical foundations for the intuitive concept of continuity.

This formalization was far more than an academic exercise. By establishing the real number system on rigorous logical foundations, mathematicians created a framework that could be extended to increasingly abstract mathematical structures. This formalization allowed mathematicians to develop concepts that initially seemed purely theoretical but proved essential for quantum mechanics, relativity, and modern computing.

The story of complex numbers represents another fascinating chapter in this intellectual journey. Although negative numbers had slowly gained acceptance by the 16th century, the square roots of negative numbers remained deeply problematic. When Italian mathematicians developed formulas for solving cubic equations, they encountered expressions involving the square roots of negative numbers. Rather than dismissing these as nonsensical, Gerolamo Cardano and later mathematicians like Rafael Bombelli began manipulating these “impossible” quantities according to consistent rules. It was not until the late 18th century that a geometric interpretation emerged, with mathematicians like Caspar Wessel, Jean-Robert Argand, and Carl Friedrich Gauss representing complex numbers as points in a two-dimensional plane. This visualization transformed complex numbers from mysterious entities into a coherent mathematical system. By the 19th century, complex analysis had become one of mathematics’ most powerful tools, finding applications in fields from electrical engineering to quantum physics.

The story of numbers illustrates how mathematics grows through both practical necessity and abstract inquiry. From merchants tallying goods to philosophers contemplating infinity, the development of number systems represents one of humanity’s greatest intellectual achievements: a shared heritage that transcends cultural boundaries and continues to evolve today.

Having traced this remarkable intellectual journey, we now turn to the formal mathematical structures that emerged from these historical developments.

2.1.2 Definition and Properties of Real Numbers

Definition: Real Number System

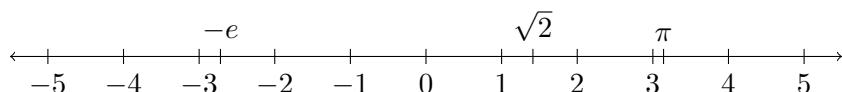
The set of real numbers, denoted by \mathbb{R} , consists of all numbers that can be represented as points on a number line, including:

- *Natural numbers (\mathbb{N}):* 1, 2, 3, ...
- *Whole numbers (\mathbb{N}_0):* 0, 1, 2, 3, ...
- *Integers (\mathbb{Z}):* ..., -3, -2, -1, 0, 1, 2, 3, ...

- *Rational numbers (\mathbb{Q}): numbers that can be written as $\frac{p}{q}$ where p and q are integers (with $q \neq 0$), such as $\frac{1}{2}$, 0.75, $-\frac{2}{3}$*
- *Irrational numbers ($\mathbb{R} \setminus \mathbb{Q}$): numbers with decimal expansions that never terminate or repeat, such as π , $\sqrt{2}$, e*

Real numbers form a continuous, unbroken line when graphed and can be positive, negative, or zero.

The real numbers can be visualized as points on an infinitely long continuous line.



Real numbers comprise rational and irrational numbers. In the previous figure, the labels above the line represent irrational numbers. This representation illustrates how rational and irrational numbers coexist on the number line, forming a complete continuum without gaps. We typically deal with real numbers as they have a closed form and desirable properties. Integers (\mathbb{Z}), whole numbers (\mathbb{N}_0), and natural numbers (\mathbb{N}) are all subsets of rational numbers (\mathbb{Q}).

2.1.3 Properties of Real Numbers

Real numbers possess several fundamental properties that make them particularly useful in mathematics.

Algebraic Properties

Theorem: Algebraic Properties of Real Numbers

For any real numbers a , b , and c :

- **Closure:** If $a, b \in \mathbb{R}$, then $a + b \in \mathbb{R}$ and $a \cdot b \in \mathbb{R}$
- **Commutativity:** $a + b = b + a$ and $a \cdot b = b \cdot a$
- **Associativity:** $(a + b) + c = a + (b + c)$ and $(a \cdot b) \cdot c = a \cdot (b \cdot c)$
- **Distributivity:** $a \cdot (b + c) = a \cdot b + a \cdot c$
- **Identity Elements:** $a + 0 = a$ and $a \cdot 1 = a$
- **Inverse Elements:** $a + (-a) = 0$ and $a \cdot \frac{1}{a} = 1$ for $a \neq 0$

Ordering Properties

The real numbers are *totally ordered*, meaning that for any two distinct real numbers, one is greater than the other.

Theorem: Order Properties of Real Numbers

- **Trichotomy:** For any $a, b \in \mathbb{R}$, exactly one of the following holds: $a < b$, $a = b$, or $a > b$
- **Transitivity:** If $a < b$ and $b < c$, then $a < c$
- **Preservation Under Addition:** If $a < b$, then $a + c < b + c$ for any c
- **Preservation Under Multiplication:** If $a < b$ and $c > 0$, then $a \cdot c < b \cdot c$
- **Sign Change Under Negative Multiplication:** If $a < b$ and $c < 0$, then $a \cdot c > b \cdot c$ (note the reversal of inequality)

Completeness Property

The defining characteristic of real numbers is their *completeness*, which fundamentally distinguishes them from both rational and irrational numbers alone. Unlike the set of rational numbers, which contains 'gaps' where irrational numbers would be, the real number line forms an unbroken continuum. This property creates a perfect mathematical harmony where, between any two distinct rational numbers, there are infinitely many irrational numbers, and similarly, between any two distinct irrational numbers, there are infinitely many rational numbers. The interweaving of these number types creates the seamless fabric of the real number line, leaving no spaces unfilled.

Theorem: Completeness Property

Every non-empty subset of real numbers that has an upper bound has a least upper bound (supremum) in \mathbb{R} . Similarly, every non-empty subset with a lower bound has a greatest lower bound (infimum) in \mathbb{R} .

Let us understand the completeness property via an example.

Example 2.1.1

Illustrate the completeness property of real numbers with an example.

Solution: Consider the set $S = \{x \in \mathbb{R} \mid x^2 < 2\}$, which consists of all real numbers

whose square is less than 2.

This set S is:

- Non-empty (for example, $1 \in S$ since $1^2 = 1 < 2$)
- Bounded above (for example, 2 is an upper bound since if $x \geq 2$, then $x^2 \geq 4 > 2$, so $x \notin S$)

According to the completeness property, S must have a least upper bound (supremum) in \mathbb{R} .

The supremum of S is $\sqrt{2}$. We can verify this:

- $\sqrt{2}$ is an upper bound for S because if $x > \sqrt{2}$, then $x^2 > 2$, meaning $x \notin S$.
- $\sqrt{2}$ is the least upper bound because for any $\varepsilon > 0$, $(\sqrt{2} - \varepsilon)^2 < 2$, which means $(\sqrt{2} - \varepsilon) \in S$.

Similarly, we can consider the set $T = \{x \in \mathbb{R} \mid x^2 < 2 \text{ and } x < 0\}$, consisting of all negative real numbers whose square is less than 2.

This set T is non-empty (for example, $-1 \in T$) and is bounded below (for example, -2 is a lower bound).

The infimum (greatest lower bound) of T is $-\sqrt{2}$, which is the greatest value that's still less than or equal to every element in T .

Interval Notation

In the previous example, we examined the set $S = \{x \in \mathbb{R} \mid x^2 < 2\}$ and found that its supremum is $\sqrt{2}$ and its infimum is $-\sqrt{2}$. This set can be expressed more concisely as $(-\sqrt{2}, \sqrt{2})$, using what we call interval notation. This notation provides a compact way to represent continuous segments of the real number line.

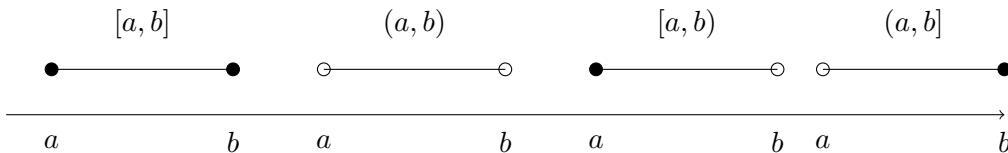
Definition: Interval Notation

Let $a, b \in \mathbb{R}$ with $a < b$. We define the following intervals:

- *Closed interval*: $[a, b] = \{x \in \mathbb{R} \mid a \leq x \leq b\}$ (includes both endpoints)
- *Open interval*: $(a, b) = \{x \in \mathbb{R} \mid a < x < b\}$ (excludes both endpoints)
- *Half-open intervals*:
 - *Left-closed, right-open interval*: $[a, b) = \{x \in \mathbb{R} \mid a \leq x < b\}$ (includes left endpoint only)

- *Left-open, right-closed interval:* $(a, b] = \{x \in \mathbb{R} \mid a < x \leq b\}$ (*includes right endpoint only*)
- *Unbounded intervals:*
 - *Right-unbounded, closed at left:* $[a, \infty) = \{x \in \mathbb{R} \mid x \geq a\}$ (*all reals greater than or equal to a*)
 - *Right-unbounded, open at left:* $(a, \infty) = \{x \in \mathbb{R} \mid x > a\}$ (*all reals greater than a*)
 - *Left-unbounded, closed at right:* $(-\infty, b] = \{x \in \mathbb{R} \mid x \leq b\}$ (*all reals less than or equal to b*)
 - *Left-unbounded, open at right:* $(-\infty, b) = \{x \in \mathbb{R} \mid x < b\}$ (*all reals less than b*)
- *The entire real line:* $(-\infty, \infty) = \mathbb{R}$

The notation uses square brackets [and] to indicate that an endpoint is included in the interval (closed), and parentheses (and) to indicate that an endpoint is excluded from the interval (open).



Intervals directly reflect the continuity property of the real numbers that we discussed earlier. Every interval represents a continuous segment of the real line with no gaps. This makes interval notation particularly suited to describing the real number system, where between any two numbers there are infinitely many others. This implies that every non-empty interval, no matter how small, contains uncountably many real numbers (the same cardinality as \mathbb{R} itself).

2.1.4 Rational Numbers

The set of rational numbers, denoted by \mathbb{Q} , consists of all numbers that can be expressed as the quotient of two integers $\frac{p}{q}$ where $q \neq 0$. In the vast ocean of real numbers that stretches infinitely along the number line, rational numbers represent surprisingly few points.

Although rational numbers may seem abundant in our everyday calculations, as we commonly work with fractions such as $\frac{1}{2}$ and $\frac{3}{4}$, the fractions actually form a minority within the real number system. Despite being infinitely many, rational numbers appear with much less frequency compared to irrational numbers. A remarkable mathematical fact is that if you were to randomly select a real number, the chance of choosing a rational number is exactly zero.

Despite their relative scarcity, rational numbers possess elegant properties and practical importance that make them fundamental to mathematics. They provide us with exact representations, predictable behavior, and serve as the foundation for our decimal number system. Understanding rational numbers provides crucial insight into the nature of number systems and prepares us for exploring the more complex terrain of irrational numbers.

Reducible Rational Numbers

Rational numbers can be classified based on whether they can be simplified further or not. This leads us to the concept of reducible and irreducible rational numbers.

Definition: Reducible Rational Numbers

A rational number $\frac{p}{q}$ (where p and q are integers, $q \neq 0$) is called reducible if p and q have a common factor greater than 1. In other words, there exists an integer $k > 1$ such that both p and q are divisible by k .

For example:

- $\frac{4}{6}$ is reducible because both 4 and 6 are divisible by 2, so $\frac{4}{6} = \frac{2}{3}$
- $\frac{-18}{24}$ is reducible because both -18 and 24 are divisible by 6, so $\frac{-18}{24} = \frac{-3}{4}$

In contrast, a rational number is called irreducible (or in lowest terms) if the numerator and denominator have no common factors other than 1.

2.1.5 Irrational Numbers

Irrational numbers, those real numbers that cannot be expressed as the ratio of two integers, possess several fundamental properties that distinguish them from their rational counterparts.

Theorem: Basic Properties of Irrational Numbers

Let x and y be real numbers, with $y \neq 0$. Then:

1. *If x is irrational and y is rational, then $x + y$ is irrational.*
2. *If x is irrational and y is rational ($y \neq 0$), then xy is irrational.*
3. *If x is irrational, then $-x$ is also irrational.*
4. *If x and y are both irrational, their sum $x + y$ may be rational or irrational.*
5. *If x and y are both irrational, their product xy may be rational or irrational.*

The last two properties highlight an interesting aspect of irrational numbers: combining two irrationals can sometimes yield a rational result. For instance, $\pi + (-\pi) = 0$ is rational, and $\sqrt{2} \cdot \sqrt{2} = 2$ is also rational.

Hippasus of Metapontum (500–450 BCE), a member of the Pythagorean school is often credited with the discovery of irrational numbers, specifically $\sqrt{2}$. The discovery reportedly caused such philosophical turmoil that, according to legend, Hippasus was drowned at sea by fellow Pythagoreans to preserve the school's fundamental worldview that all numbers could be expressed as ratios.

Example 2.1.2

Prove that $\sqrt{2}$ is an irrational number.

Solution: We will use proof by contradiction (an important technique for the ISI exam). Suppose $\sqrt{2}$ is rational. Then $\sqrt{2} = \frac{p}{q}$ where p and q are integers with no common factors and $q \neq 0$.

Squaring both sides:

$$2 = \frac{p^2}{q^2} \tag{2.1.1}$$

$$p^2 = 2q^2 \tag{2.1.2}$$

This means p^2 is even, which implies p is even (since the square of an odd number is odd). So $p = 2k$ for some integer k .

Substituting:

$$(2k)^2 = 2q^2 \tag{2.1.3}$$

$$4k^2 = 2q^2 \tag{2.1.4}$$

$$2k^2 = q^2 \tag{2.1.5}$$

Now q^2 is even, which means q is even. But this contradicts our assumption that p and q have no common factors (since they would both be even, and thus have 2 as a common factor).

Therefore, our assumption that $\sqrt{2}$ is rational must be false, so $\sqrt{2}$ is irrational.

It is believed that Hippasus may have discovered the irrationality geometrically, more of an intuitive understanding and without a rigorous proof in the modern sense. The provided proof is not because of him and likely evolved over time and may have been formalized later.

In India, the Sulba Sutras (circa 800-600 BCE) already demonstrated an understanding of incommensurable magnitudes (what we now call irrational numbers). Baudhāyana, author of

one of these texts, provided the following shloka to approximate $\sqrt{2}$ ($\approx 1 + \frac{1}{3} + \frac{1}{3 \times 4} - \frac{1}{3 \times 4 \times 34}$).

*Dvikaṇani vikaṇani ca yatkincit taddviṇam kuryat |
Tatpadena vibhajet pashchat tadardhena samyutam ||*

The approximation equals $\frac{577}{408} \approx 1.4142157$, accurate to five decimal places compared to the true value of $\sqrt{2} \approx 1.4142135$. This sophisticated approximation predates Greek discoveries and suggests an intuitive understanding of irrationality in ancient Indian mathematics.

The exact method used by Baudhāyana to derive the approximation is unknown, however, ancient mathematicians knew of a method that provides increasingly accurate rational approximations to square roots.

Theorem: Square Root Approximation Method

If $\frac{p}{q}$ is an approximation of \sqrt{n} , then a better approximation can be found using:

$$\frac{p+nq}{p+q}$$

Example 2.1.3

Apply the iterative approximation formula for $\sqrt{2}$ to generate a sequence of increasingly accurate rational approximations, and demonstrate the alternating property of these approximations.

Solution: we will use the formula: if $\frac{p}{q}$ is an approximation of $\sqrt{2}$, then $\frac{p+2q}{p+q}$ gives a better approximation.

Starting with the simplest approximation $\frac{1}{1} = 1$:

Step	Calculation	Approximation	Decimal Value	Comparison to $\sqrt{2} \approx 1.4142$
1	Initial value	$\frac{1}{1}$	1.0000	Too small
2	$\frac{1+2(1)}{1+1}$	$\frac{3}{2}$	1.5000	Too large
3	$\frac{3+2(2)}{3+2}$	$\frac{7}{5}$	1.4000	Too small
4	$\frac{7+2(5)}{7+5}$	$\frac{17}{12}$	1.4167	Too large
5	$\frac{17+2(12)}{17+12}$	$\frac{41}{29}$	1.4138	Too small
6	$\frac{41+2(29)}{41+29}$	$\frac{99}{70}$	1.4143	Too large

We observe:

- The approximations alternate between being less than and greater than $\sqrt{2}$
- Each approximation is closer to $\sqrt{2}$ than the previous one
- By the 6th iteration, we already have an approximation accurate to 4 decimal places

This alternating property allows us to establish increasingly tight bounds on the value of $\sqrt{2}$. After just 6 steps, we can conclude that $1.4138 < \sqrt{2} < 1.4143$.

If we were to continue this process, the 8th approximation would be $\frac{577}{408} \approx 1.4142157$, which is the approximation attributed to Baudhāyana in the Sulba Sutras, accurate to 5 decimal places.

2.2 Integer Functions

In our exploration of number systems, we've seen how real numbers form a continuous spectrum on the number line. Now, we will examine functions that create a bridge between the continuous realm of real numbers and the discrete world of integers. These functions, known as the Greatest Integer Function and the Smallest Integer Function, play an important role in various areas of mathematics.

2.2.1 Greatest Integer Function

Definition: Greatest Integer Function

The Greatest Integer Function, denoted by $\lfloor x \rfloor$ (also called the floor function), gives the largest integer less than or equal to x . For any real number x , $\lfloor x \rfloor$ is the unique integer n such that $n \leq x < n + 1$.

For instance, $\lfloor 3.7 \rfloor = 3$ and $\lfloor -2.3 \rfloor = -3$.

Geometrically, the graph of $y = \lfloor x \rfloor$ resembles a staircase, with jumps at integer values of x .

The graph of the Greatest Integer Function $y = \lfloor x \rfloor$ is a step function where each step is a horizontal line segment. For any real number x in the interval $[n, n + 1)$, where n is an integer, $\lfloor x \rfloor = n$. This creates a staircase pattern with jumps occurring exactly at integer values of x . The solid blue circles indicate points that belong to the function, while the hollow circles represent points just before the jumps where the function is discontinuous. For example: for $x \in [-1, 0)$, $\lfloor x \rfloor = -1$, while for $x \in [0, 1)$, $\lfloor x \rfloor = 0$. Note that $\lfloor n \rfloor = n$ for any integer n , which means that at integer values, the function takes the value of that integer.

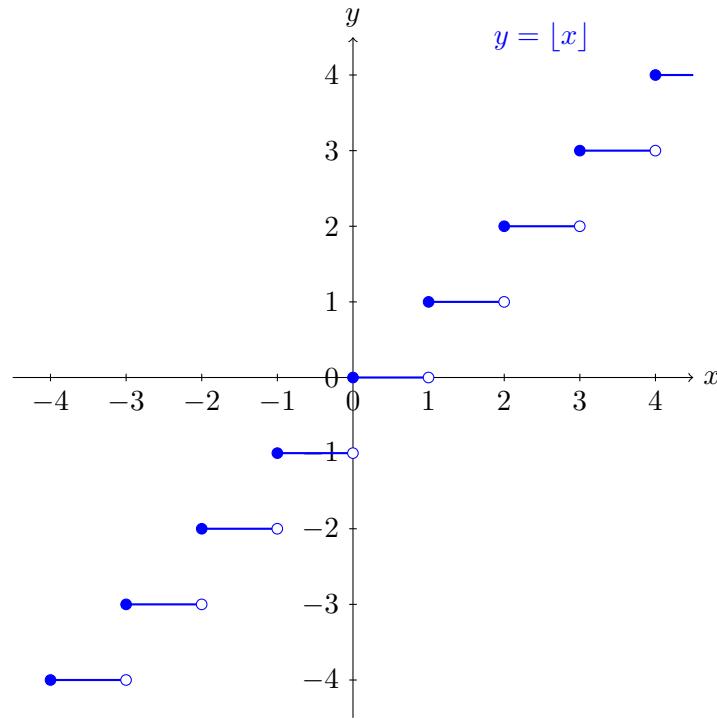


Figure 2.1: Graph of the Greatest Integer Function $y = \lfloor x \rfloor$

Properties of the Greatest Integer Function

The Greatest Integer Function has several important properties, as described in the next theorem.

Theorem: Properties of the Greatest Integer Function

For any real numbers x and y , and integer n :

1. $\lfloor x \rfloor = n$ if and only if $n \leq x < n + 1$
2. $\lfloor x + n \rfloor = \lfloor x \rfloor + n$
3. $\lfloor -x \rfloor = -\lceil x \rceil$ (relation to ceiling function)
4. $\lfloor x \rfloor + \lfloor -x \rfloor = \begin{cases} -1 & \text{if } x \notin \mathbb{Z} \\ 0 & \text{if } x \in \mathbb{Z} \end{cases}$
5. $\lfloor x \rfloor + \lfloor y \rfloor \leq \lfloor x + y \rfloor \leq \lfloor x \rfloor + \lfloor y \rfloor + 1$
6. $\lfloor \lfloor x \rfloor / n \rfloor = \lfloor x/n \rfloor$ for any integer $n > 0$

Example 2.2.1

Prove that for any real number x , $x - 1 < \lfloor x \rfloor \leq x$.

Solution: By definition, $\lfloor x \rfloor$ is the greatest integer not exceeding x . This means $\lfloor x \rfloor \leq x$.

Now, let $n = \lfloor x \rfloor$. By definition of the greatest integer function, $n \leq x < n + 1$. Taking the left inequality, $n \leq x$, we have $n > x - 1$, which means $\lfloor x \rfloor > x - 1$.

Combining both inequalities: $x - 1 < \lfloor x \rfloor \leq x$.

Example 2.2.2

Prove that for any real number x , we have

$$\lfloor 2x \rfloor = \lfloor x \rfloor + \lfloor x + \frac{1}{2} \rfloor$$

where $\lfloor \cdot \rfloor$ denotes the greatest integer function.

Solution: We need to prove that $\lfloor 2x \rfloor = \lfloor x \rfloor + \lfloor x + \frac{1}{2} \rfloor$ for any real x .

Let $x = n + f$ where $n = \lfloor x \rfloor$ is an integer and $0 \leq f < 1$ is the fractional part.

Then $x + \frac{1}{2} = n + f + \frac{1}{2}$ and $2x = 2n + 2f$.

We consider two cases based on the fractional part f :

Case 1: $0 \leq f < \frac{1}{2}$

In this case:

- $\lfloor x \rfloor = n$
- $x + \frac{1}{2} = n + f + \frac{1}{2}$ where $\frac{1}{2} \leq f + \frac{1}{2} < 1$, so $\lfloor x + \frac{1}{2} \rfloor = n$
- $2x = 2n + 2f$ where $0 \leq 2f < 1$, so $\lfloor 2x \rfloor = 2n$

Therefore: $\lfloor x \rfloor + \lfloor x + \frac{1}{2} \rfloor = n + n = 2n = \lfloor 2x \rfloor$

Case 2: $\frac{1}{2} \leq f < 1$

In this case:

- $\lfloor x \rfloor = n$
- $x + \frac{1}{2} = n + f + \frac{1}{2}$ where $1 \leq f + \frac{1}{2} < \frac{3}{2}$, so $\lfloor x + \frac{1}{2} \rfloor = n + 1$
- $2x = 2n + 2f$ where $1 \leq 2f < 2$, so $\lfloor 2x \rfloor = 2n + 1$

Therefore: $\lfloor x \rfloor + \lfloor x + \frac{1}{2} \rfloor = n + (n + 1) = 2n + 1 = \lfloor 2x \rfloor$

Since both cases yield the desired equality, the identity holds for all real numbers x .

2.2.2 Smallest Integer Function

Definition: Smallest Integer Function

The Smallest Integer Function, denoted by $\lceil x \rceil$ (also called the ceiling function), gives the smallest integer greater than or equal to x . For any real number x , $\lceil x \rceil$ is the unique integer n such that $n - 1 < x \leq n$.

For instance, $\lceil 3.7 \rceil = 4$ and $\lceil -2.3 \rceil = -2$.

Like the floor function, the ceiling function's graph also forms a staircase pattern, but with jumps occurring at different positions.

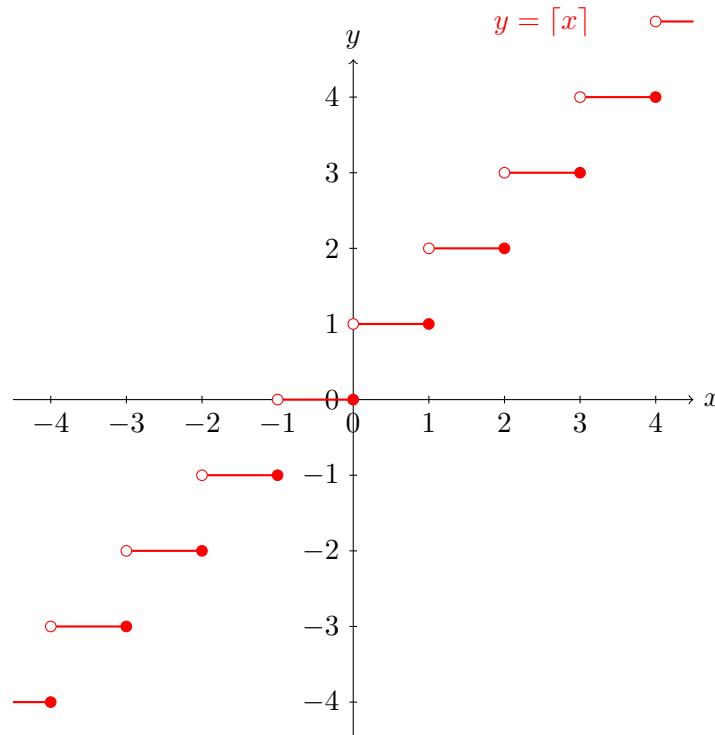


Figure 2.2: Graph of the Smallest Integer Function $y = \lceil x \rceil$

The graph of the Smallest Integer Function $y = \lceil x \rceil$ is a step function where each step is a horizontal line segment. For any real number x in the interval $(n - 1, n]$, where n is an integer, $\lceil x \rceil = n$. This creates a staircase pattern with jumps occurring exactly at integer values of x . The solid red circles indicate points that belong to the function, while the hollow circles represent points just before the jumps where the function is discontinuous. For example, for $x \in (-1, 0]$, $\lceil x \rceil = 0$, and for $x \in (0, 1]$, $\lceil x \rceil = 1$. Note that $\lceil n \rceil = n$ for any integer n , which means that at integer values, the function takes the value of that integer.

Properties of the Smallest Integer Function

The Smallest Integer Function has properties analogous to those of the Greatest Integer Function:

Theorem: Properties of the Smallest Integer Function

For any real numbers x and y , and integer n :

1. $\lceil x \rceil = n$ if and only if $n - 1 < x \leq n$
2. $\lceil x + n \rceil = \lceil x \rceil + n$
3. $\lceil -x \rceil = -\lfloor x \rfloor$ (relation to floor function)
4. $\lceil x \rceil + \lceil -x \rceil = \begin{cases} 1 & \text{if } x \notin \mathbb{Z} \\ 0 & \text{if } x \in \mathbb{Z} \end{cases}$
5. $\lceil x \rceil + \lceil y \rceil - 1 \leq \lceil x + y \rceil \leq \lceil x \rceil + \lceil y \rceil$
6. $\lceil \lceil x \rceil / n \rceil = \lceil x/n \rceil$ for any integer $n > 0$

To get a better understanding of the two integer functions, let us evaluate the numerical expression: $\lceil 3.8 \rceil - \lceil -2.5 \rceil$. As $\lceil 3.8 \rceil$ equals to 4 and $\lceil -2.5 \rceil$ equals to -2, hence the expression sums to 6.

Let us look at a more theoretical example to understand the smallest integer function.

Example 2.2.3

Prove that for any real number x , $x \leq \lceil x \rceil < x + 1$.

Solution: By definition, $\lceil x \rceil$ is the smallest integer not less than x . This means $\lceil x \rceil \geq x$, and since we're dealing with inequalities involving integers, this is equivalent to $\lceil x \rceil \geq x$.

Now, let $n = \lceil x \rceil$. By definition of the smallest integer function, $n - 1 < x \leq n$. Taking the right inequality, $x \leq n$, we have $x + 1 < n + 1$, which means $\lceil x \rceil + 1 > x + 1$. Therefore, $\lceil x \rceil < x + 1$.

Combining both inequalities: $x \leq \lceil x \rceil < x + 1$.

Relationship Between Floor and Ceiling Functions

The floor and ceiling functions are closely related. Here are some important relationships:

Theorem: Relationships Between Floor and Ceiling Functions

For any real number x :

1. $\lceil x \rceil = -\lfloor -x \rfloor$
2. $\lfloor x \rfloor = -\lceil -x \rceil$
3. $\lceil x \rceil = \lfloor x \rfloor + \begin{cases} 0 & \text{if } x \in \mathbb{Z} \\ 1 & \text{if } x \notin \mathbb{Z} \end{cases}$
4. $\lceil x \rceil - \lfloor x \rfloor = \begin{cases} 0 & \text{if } x \in \mathbb{Z} \\ 1 & \text{if } x \notin \mathbb{Z} \end{cases}$
5. $\lfloor x + \frac{1}{2} \rfloor = \lceil x - \frac{1}{2} \rceil$ (rounding to nearest integer)

2.2.3 The Fractional Part Function

Related to the floor and ceiling functions is the fractional part function, denoted by $\{x\}$ (not to be confused with set notation).

Definition: Fractional Part Function

The fractional part of a real number x , denoted by $\{x\}$ or sometimes $x - \lfloor x \rfloor$, is defined as:

$$\{x\} = x - \lfloor x \rfloor$$

For instance, $\{3.7\} = 3.7 - 3 = 0.7$ and $\{-2.3\} = -2.3 - (-3) = -2.3 + 3 = 0.7$.

Note that $\{x\} \in [0, 1)$ for all real numbers x . The fractional part function returns the portion of x to the right of the decimal point, adjusted for negative numbers so that the result is always non-negative and less than 1.

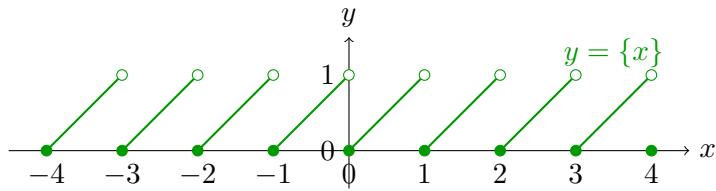


Figure 2.3: Graph of the Fractional Part Function $y = \{x\}$

The graph of the Fractional Part Function $y = \{x\}$ shows its sawtooth pattern. For each interval $[n, n + 1)$ where n is an integer, the function increases linearly from 0 to 1 (exclusive). At each integer $x = n$, the function takes the value 0 (shown by solid circles), and just before reaching the next integer, the function approaches 1 (shown by hollow circles) before abruptly dropping back to 0. This pattern illustrates that $\{x\} = x - \lfloor x \rfloor$, which extracts only the decimal portion of a number.

Properties of the Fractional Part Function

Theorem: Properties of the Fractional Part Function

For any real numbers x and y , and integer n :

1. $0 \leq \{x\} < 1$ for all $x \in \mathbb{R}$
2. $\{x\} = 0$ if and only if $x \in \mathbb{Z}$
3. $\{x + n\} = \{x\}$ for any integer n (periodicity)
4. $\{-x\} = \begin{cases} 0 & \text{if } x \in \mathbb{Z} \\ 1 - \{x\} & \text{if } x \notin \mathbb{Z} \end{cases}$
5. $\{x + y\} = \begin{cases} \{x\} + \{y\} & \text{if } \{x\} + \{y\} < 1 \\ \{x\} + \{y\} - 1 & \text{if } \{x\} + \{y\} \geq 1 \end{cases}$

Example 2.2.4

For any real number x , prove that $\lfloor 2x \rfloor \geq 2\lfloor x \rfloor$ and determine when equality holds.

Solution: Let $x = \lfloor x \rfloor + \{x\}$, where $\{x\}$ represents the fractional part of x such that $0 \leq \{x\} < 1$.

Then:

$$\lfloor 2x \rfloor = \lfloor 2(\lfloor x \rfloor + \{x\}) \rfloor \quad (2.2.1)$$

$$= \lfloor 2\lfloor x \rfloor + 2\{x\} \rfloor \quad (2.2.2)$$

$$(2.2.3)$$

Since $2\lfloor x \rfloor$ is an integer, and using the property $\lfloor a + b \rfloor = a + \lfloor b \rfloor$ when a is an integer:

$$\lfloor 2x \rfloor = 2\lfloor x \rfloor + \lfloor 2\{x\} \rfloor \quad (2.2.4)$$

Now, since $0 \leq \{x\} < 1$, we have $0 \leq 2\{x\} < 2$. If $0 \leq \{x\} < \frac{1}{2}$, then $\lfloor 2\{x\} \rfloor = 0$, and when $\frac{1}{2} \leq \{x\} < 1$, then $\lfloor 2\{x\} \rfloor = 1$.

Therefore:

$$\lfloor 2x \rfloor = 2\lfloor x \rfloor + \begin{cases} 0 & \text{if } 0 \leq \{x\} < \frac{1}{2} \\ 1 & \text{if } \frac{1}{2} \leq \{x\} < 1 \end{cases} \quad (2.2.5)$$

Since $\lfloor 2\{x\} \rfloor \geq 0$, we have $\lfloor 2x \rfloor \geq 2\lfloor x \rfloor$.

Equality holds when $\lfloor 2\{x\} \rfloor = 0$, which occurs when $0 \leq \{x\} < \frac{1}{2}$. In other words, equality holds when the fractional part of x is less than $\frac{1}{2}$.

2.2.4 The Factorial Function

While the functions we have explored so far transform continuous real numbers into discrete or bounded outputs, we now turn to a function that operates exclusively in the discrete realm of non-negative integers. The factorial function, though simple in definition, possesses remarkable mathematical depth and will appear throughout this book.

Definition: Factorial Function

The factorial of a non-negative integer n , denoted by $n!$, is defined as:

$$n! = \begin{cases} 1 & \text{if } n = 0 \\ n \cdot (n - 1) \cdot (n - 2) \cdots 2 \cdot 1 & \text{if } n \geq 1 \end{cases}$$

Alternatively, we can define it recursively: $0! = 1$ and $n! = n \cdot (n - 1)!$ for $n \geq 1$.

The definition $0! = 1$ might initially seem counterintuitive, but this convention ensures that many mathematical formulas involving factorials remain consistent and elegant.

Let's examine the first few factorial values:

- $0! = 1$
- $1! = 1$
- $2! = 2 \cdot 1 = 2$
- $3! = 3 \cdot 2 \cdot 1 = 6$
- $4! = 4 \cdot 3 \cdot 2 \cdot 1 = 24$
- $5! = 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1 = 120$
- $6! = 6 \cdot 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1 = 720$

Notice how rapidly factorial values grow. This explosive growth makes factorials particularly interesting and is used as an abbreviation to represent large numbers involving multiplication of consecutive integers. For instance, $998 \cdot 999 \cdot 1000 \cdot 1001$ can be represented as $\frac{1001!}{997!}$.

Properties of the Factorial Function

The factorial function possesses several fundamental properties that make it invaluable across mathematics:

Theorem: Properties of the Factorial Function

For non-negative integers n and m :

1. **Recursive Property:** $n! = n \cdot (n - 1)!$ for $n \geq 1$
2. **Growth Property:** $n! > 2^{n-1}$ for $n \geq 1$
3. **Divisibility:** If $m \leq n$, then $m!$ divides $n!$
4. **Stirling's Approximation:** For large n , $n! \approx \sqrt{2\pi n} \left(\frac{n}{e}\right)^n$

It is easy to see that $n! > 2^{n-1}$ for $n = 1$. For $n \geq 2$, we have $n! = n \cdot (n - 1) \dots 2$. As we have $n - 1$ elements in $n \cdot (n - 1) \dots 2$ and each element is greater than or equals to 2. Hence, $n! = n \cdot (n - 1) \dots 2 \geq 2^{n-1}$.

Stirling's Approximation states that for large n , $n! \approx \sqrt{2\pi n} \left(\frac{n}{e}\right)^n$. Stirling's approximation provides a remarkably accurate estimate for large factorials and reveals the asymptotic behavior of factorial growth. For practical calculations, it shows that $n!$ grows roughly like n^n divided by e^n , with an additional factor involving \sqrt{n} . This approximation becomes increasingly precise as n increases.

The factorial function appears in numerous mathematical contexts, making it one of the most practically important functions in discrete mathematics. Many important mathematical functions can be expressed using factorials. The exponential function has the beautiful expansion:

$$e^x = \sum_{n=0}^{\infty} \frac{x^n}{n!} = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \frac{x^4}{4!} + \dots$$

Similarly, the sine and cosine functions have factorial-based expansion that connects trigonometry with the factorial function.

Let us look at an interesting example that connects factorials to the greatest integer function.

Example 2.2.5

Show that $n!$ ends in exactly $\left\lfloor \frac{n}{5} \right\rfloor + \left\lfloor \frac{n}{25} \right\rfloor + \left\lfloor \frac{n}{125} \right\rfloor + \dots$ trailing zeros.

Solution: Trailing zeros in $n!$ are produced by factors of 10, and since $10 = 2 \times 5$, we need to count the pairs of factors 2 and 5 appearing in $n!$.

Since there are always more factors of 2 than factors of 5 in $n!$ (every even number contributes at least one factor of 2), the number of trailing zeros equals the number of factors of 5 in $n!$. For example, $100!$ has $\left\lfloor \frac{100}{5} \right\rfloor + \left\lfloor \frac{100}{25} \right\rfloor + \left\lfloor \frac{100}{125} \right\rfloor = 20 + 4 + 0 = 24$ trailing zeros.

In the expression, We need to consider higher powers of 5 because some numbers contribute more than one factor of 5 to $n!$. Consider the numbers from 1 to n :

- Numbers like 5, 10, 15, 20 each contribute exactly one factor of 5
- Numbers like 25, 50, 75 each contribute two factors of 5 (since $25 = 5^2$)
- Numbers like 125, 250, 375 each contribute three factors of 5 (since $125 = 5^3$)

The term $\left\lfloor \frac{n}{5} \right\rfloor$ counts all multiples of 5, giving us one factor from each. However, multiples of 25 actually contribute two factors of 5, so we need to add $\left\lfloor \frac{n}{25} \right\rfloor$ to account for the second factor from each multiple of 25. Similarly, multiples of 125 contribute a third factor of 5, requiring us to add $\left\lfloor \frac{n}{125} \right\rfloor$, and so on.

The expression for trailing zeroes is called Legendre's formula and for $p = 5$, we have:

$$v_5(n!) = \left\lfloor \frac{n}{5} \right\rfloor + \left\lfloor \frac{n}{25} \right\rfloor + \left\lfloor \frac{n}{125} \right\rfloor + \dots$$

For any integer k , the Legendre's formula can be generalized to:

$$v_k(n!) = \left\lfloor \frac{n}{k} \right\rfloor + \left\lfloor \frac{n}{k^2} \right\rfloor + \left\lfloor \frac{n}{k^3} \right\rfloor + \dots$$

Let us look at an another interesting property of the integer factorials.

Example 2.2.6

Show that $(n+1)! - n! = n \cdot n!$ and use this to find a pattern in factorial differences.

Solution: We can factor the left side:

$$(n+1)! - n! = (n+1) \cdot n! - n! \quad (2.2.6)$$

$$= n! \cdot [(n+1) - 1] \quad (2.2.7)$$

$$= n! \cdot n \quad (2.2.8)$$

$$= n \cdot n! \quad (2.2.9)$$

This identity reveals that the difference between consecutive factorials is always a multiple of the smaller factorial. This property is useful in simplifying expressions involving factorial differences and appears in various proofs.

We can observe that:

- $2! - 1! = 1 \cdot 1! = 1$
- $3! - 2! = 2 \cdot 2! = 4$
- $4! - 3! = 3 \cdot 3! = 18$
- $5! - 4! = 4 \cdot 4! = 96$

The pattern shows that consecutive factorial differences grow even more rapidly than the factorials themselves.

2.3 Complex Numbers

Complex numbers arise from the need to solve equations such as $x^2 + 1 = 0$ that do not have solutions in the real number system. Initially considered “impossible” or “imaginary” (hence the term “imaginary unit”), complex numbers were gradually accepted as legitimate mathematical objects via the work of mathematicians such as Bolzano, Cardano, Bombelli, Euler, and Gauss.

The complex number system extends the real number system by introducing the imaginary unit i , where $i^2 = -1$. The complex numbers are closed under addition, subtraction, multiplication, and division (except division by zero). This property ensures that when we perform arithmetic operations with complex numbers, the results remain within the complex number system.

Definition: Complex Numbers

A complex number is a number of the form $a + bi$, where a and b are real numbers and i is the imaginary unit defined by the property $i^2 = -1$. The set of all complex numbers is denoted by \mathbb{C} .

- a is called the real part of the complex number

- b is called the *imaginary part* of the complex number

2.3.1 Basic Operations with Complex Numbers

Complex numbers extend our numerical capabilities beyond the real number line, allowing us to solve equations that previously had no solutions. To work effectively with these numbers, we need to understand how to perform basic arithmetic operations on them.

Definition: Basic Operations of Complex Numbers

For two complex numbers $z_1 = a + bi$ and $z_2 = c + di$:

- **Addition:** $z_1 + z_2 = (a + c) + (b + d)i$
- **Subtraction:** $z_1 - z_2 = (a - c) + (b - d)i$
- **Multiplication:** $z_1 \cdot z_2 = (ac - bd) + (ad + bc)i$
- **Division:** $\frac{z_1}{z_2} = \frac{(a+bi)}{(c+di)} = \frac{(ac+bd)+(bc-ad)i}{c^2+d^2}$ for $z_2 \neq 0$

As discussed earlier, the outcome of all these operations is another complex number. The multiplication formula can be derived by using the distributive property and substituting $i^2 = -1$:

$$(a + bi)(c + di) = ac + adi + bci + bd i^2 \quad (2.3.1)$$

$$= ac + adi + bci + bd(-1) \quad (2.3.2)$$

$$= ac - bd + (ad + bc)i \quad (2.3.3)$$

The division formula is obtained by multiplying both numerator and denominator by the complex conjugate of the denominator, $c - di$:

$$\frac{a + bi}{c + di} = \frac{(a + bi)(c - di)}{(c + di)(c - di)} \quad (2.3.4)$$

$$= \frac{ac - adi + bci - bdi^2}{c^2 - (di)^2} \quad (2.3.5)$$

$$= \frac{ac - adi + bci + bd}{c^2 + d^2} \quad (2.3.6)$$

$$= \frac{(ac + bd) + (bc - ad)i}{c^2 + d^2} \quad (2.3.7)$$

2.3.2 Complex Conjugate

The complex conjugate serves as a kind of inverse to a complex number, with the special property that multiplying a complex number by its conjugate always yields a real value. It reflects a

complex number across the real axis, preserving the real part while reversing the sign of the imaginary part, and plays a crucial role in division and many other applications of complex numbers.

Definition: Complex Conjugate

The complex conjugate of a complex number $z = a + bi$ is defined as $\bar{z} = a - bi$.

The complex conjugate has several important properties:

- $\bar{\bar{z}} = z$ (conjugate of conjugate returns the original number)
- $\overline{z_1 + z_2} = \bar{z}_1 + \bar{z}_2$ (conjugate distributes over addition)
- $\overline{z_1 \cdot z_2} = \bar{z}_1 \cdot \bar{z}_2$ (conjugate distributes over multiplication)
- $\overline{\left(\frac{z_1}{z_2}\right)} = \frac{\bar{z}_1}{\bar{z}_2}$ for $z_2 \neq 0$ (conjugate distributes over division)
- If z is real, then $\bar{z} = z$
- If z is purely imaginary, then $\bar{z} = -z$

The property for multiplication can be proven as follows:

$$\overline{z_1 \cdot z_2} = \overline{(a + bi)(c + di)} \quad (2.3.8)$$

$$= \overline{ac + adi + bci + bdi^2} \quad (2.3.9)$$

$$= \overline{ac + adi + bci - bd} \quad (2.3.10)$$

$$= \overline{(ac - bd) + (ad + bc)i} \quad (2.3.11)$$

$$= (ac - bd) - (ad + bc)i \quad (2.3.12)$$

$$= a(c - di) - bi(c - di) \quad (2.3.13)$$

$$= (a - bi)(c - di) \quad (2.3.14)$$

$$= \bar{z}_1 \cdot \bar{z}_2 \quad (2.3.15)$$

The complex conjugate is particularly useful in division, as we saw earlier, since multiplying both numerator and denominator by the conjugate of the denominator converts the denominator to a real number:

$$\frac{z_1}{z_2} = \frac{z_1 \cdot \bar{z}_2}{z_2 \cdot \bar{z}_2} = \frac{z_1 \cdot \bar{z}_2}{|z_2|^2} \quad (2.3.16)$$

Example 2.3.1

Find the complex conjugate of $z = 3 - 4i$ and show that $z \cdot \bar{z} = |z|^2$.

Solution: The complex conjugate of $z = 3 - 4i$ is $\bar{z} = 3 + 4i$. Let us verify that

$$z \cdot \bar{z} = |z|^2:$$

$$z \cdot \bar{z} = (3 - 4i)(3 + 4i) \quad (2.3.17)$$

$$= 9 + 12i - 12i - 16i^2 \quad (2.3.18)$$

$$= 9 - 16(-1) \quad (2.3.19)$$

$$= 9 + 16 \quad (2.3.20)$$

$$= 25 \quad (2.3.21)$$

And indeed:

$$|z|^2 = |3 - 4i|^2 = 3^2 + (-4)^2 = 9 + 16 = 25 \quad (2.3.22)$$

Thus, $z \cdot \bar{z} = |z|^2$ is verified.

2.3.3 Modulus of Complex Numbers

Definition: Modulus

The modulus (or absolute value) of a complex number $z = a + bi$ is defined as $|z| = \sqrt{a^2 + b^2}$. Geometrically, it represents the distance from the origin to the point z in the complex plane.

The modulus has several important properties:

- $|z| \geq 0$ for all $z \in \mathbb{C}$, and $|z| = 0$ if and only if $z = 0$
- $|z_1 \cdot z_2| = |z_1| \cdot |z_2|$ (modulus of a product equals product of moduli)
- $\left| \frac{z_1}{z_2} \right| = \frac{|z_1|}{|z_2|}$ for $z_2 \neq 0$ (modulus of a quotient equals quotient of moduli)
- $|z + w| \leq |z| + |w|$ (triangle inequality)
- $|z - w| \geq ||z| - |w||$ (reverse triangle inequality)
- For any $z \in \mathbb{C}$, $z \cdot \bar{z} = |z|^2$

The complex conjugate and modulus are closely related through several important identities:

- $z \cdot \bar{z} = |z|^2 = a^2 + b^2$
- $z + \bar{z} = 2\operatorname{Re}(z) = 2a$
- $z - \bar{z} = 2i\operatorname{Im}(z) = 2bi$
- $\frac{z + \bar{z}}{2} = \operatorname{Re}(z) = a$

- $\frac{z-\bar{z}}{2i} = \operatorname{Im}(z) = b$
- $|z| = |\bar{z}|$ (the modulus is invariant under conjugation)

These relationships make the complex conjugate and modulus powerful tools in complex analysis and its applications.

2.3.4 Polar Form

Any complex number can also be written in polar form:

$$z = a + bi = r(\cos \theta + i \sin \theta) = re^{i\theta} \quad (2.3.23)$$

where $r = |z| = \sqrt{a^2 + b^2}$ is the modulus and $\theta = \arg(z)$ is the argument.

The expression $e^{i\theta} = \cos \theta + i \sin \theta$ is known as Euler's formula and provides a beautiful connection between complex exponentials and trigonometric functions.

Polar form is particularly useful for multiplication and division:

- Multiplication: $z_1 \cdot z_2 = r_1 r_2 e^{i(\theta_1 + \theta_2)}$
- Division: $\frac{z_1}{z_2} = \frac{r_1}{r_2} e^{i(\theta_1 - \theta_2)}$

Example 2.3.2

Convert the complex number $z = 1 - i$ to polar form.

Solution: First, we find the modulus:

$$|z| = \sqrt{1^2 + (-1)^2} = \sqrt{2} \quad (2.3.24)$$

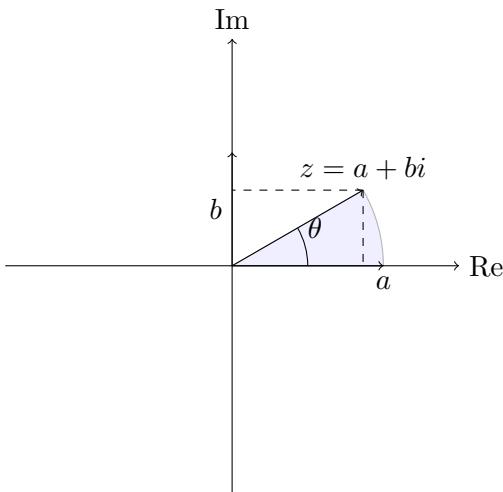
Next, we find the argument. Since the real part is positive and the imaginary part is negative, the point is in the fourth quadrant:

$$\theta = \tan^{-1} \left(\frac{-1}{1} \right) = -\frac{\pi}{4} \quad (2.3.25)$$

Therefore, the polar form is:

$$z = 1 - i = \sqrt{2}e^{-i\pi/4} = \sqrt{2} \left(\cos \left(-\frac{\pi}{4} \right) + i \sin \left(-\frac{\pi}{4} \right) \right) = \sqrt{2} \left(\frac{\sqrt{2}}{2} - i \frac{\sqrt{2}}{2} \right) \quad (2.3.26)$$

Complex numbers can be represented geometrically in the complex plane, where the horizontal axis represents the real part and the vertical axis represents the imaginary part. Each complex number $z = a + bi$ corresponds to the point (a, b) in this plane.



Here the modulus $|z|$ represents the distance from the origin to the point z . The argument θ is the angle between the positive real axis and the line segment from the origin to z .

2.3.5 De Moivre's Theorem

De Moivre's theorem provides a powerful formula for calculating powers of complex numbers in polar form.

Theorem: De Moivre's Theorem

For any complex number $z = r(\cos \theta + i \sin \theta)$ and any integer n , we have:

$$z^n = r^n(\cos(n\theta) + i \sin(n\theta)) = r^n e^{in\theta}$$

This theorem can be proven by induction and is a direct consequence of the properties of complex multiplication in polar form.

Example 2.3.3

Compute $(1 + i)^6$ using De Moivre's theorem.

Solution: Let us convert $1 + i$ to polar form first:

$$|1 + i| = \sqrt{1^2 + 1^2} = \sqrt{2} \quad (2.3.27)$$

$$\theta = \tan^{-1}\left(\frac{1}{1}\right) = \frac{\pi}{4} \quad (2.3.28)$$

So $1 + i = \sqrt{2}e^{i\pi/4} = \sqrt{2}(\cos(\pi/4) + i \sin(\pi/4))$.

Using De Moivre's theorem:

$$(1+i)^6 = (\sqrt{2})^6 \cdot (\cos(6\pi/4) + i \sin(6\pi/4)) \quad (2.3.29)$$

$$= 8 \cdot (\cos(6\pi/4) + i \sin(6\pi/4)) \quad (2.3.30)$$

$$= 8 \cdot (\cos(3\pi/2) + i \sin(3\pi/2)) \quad (2.3.31)$$

$$= 8 \cdot (0 - i) \quad (2.3.32)$$

$$= -8i \quad (2.3.33)$$

2.3.6 Roots of Complex Numbers

De Moivre's theorem can be extended to find the n -th roots of a complex number. The n -th roots of a complex number $z = re^{i\theta}$ are given by:

$$w_k = r^{1/n} e^{i(\theta+2\pi k)/n} \quad (2.3.34)$$

where $k = 0, 1, 2, \dots, n - 1$.

This means every complex number has exactly n distinct n -th roots, evenly spaced in a circle of radius $r^{1/n}$ in the complex plane.

Example 2.3.4

Find all the cube roots of -8 .

Solution: We can write $-8 = 8e^{i\pi}$ (since -8 lies on the negative real axis).

The three cube roots are:

$$w_k = 8^{1/3} e^{i(\pi+2\pi k)/3} \quad (2.3.35)$$

$$= 2e^{i(\pi+2\pi k)/3} \quad (2.3.36)$$

for $k = 0, 1, 2$.

$$\text{For } k = 0: w_0 = 2e^{i\pi/3} = 2 \left(\frac{1}{2} + i \frac{\sqrt{3}}{2} \right) = 1 + i\sqrt{3}$$

$$\text{For } k = 1: w_1 = 2e^{i\pi} = -2$$

$$\text{For } k = 2: w_2 = 2e^{i5\pi/3} = 2 \left(\frac{1}{2} - i \frac{\sqrt{3}}{2} \right) = 1 - i\sqrt{3}$$

Therefore, the three cube roots of -8 are $1 + i\sqrt{3}$, -2 , and $1 - i\sqrt{3}$.

Notice that for a real number (like -8 in the example above), the non-real roots appear as conjugate pairs. In this case, $1 + i\sqrt{3}$ and $1 - i\sqrt{3}$ are complex conjugates of each other. This

pattern holds generally: when finding roots of real numbers, the complex roots will appear as conjugate pairs, while for complex numbers with non-zero imaginary parts, this symmetry does not necessarily hold.

2.3.7 Geometric Interpretation of n -th Roots

To understand why a complex number has exactly n distinct n -th roots, we need to explore the polar form of complex numbers and examine how exponentiation affects them geometrically.

When we represent a complex number $z = re^{i\theta}$ in polar form, r represents the distance from the origin (magnitude), and θ represents the angle from the positive real axis (argument). Taking the n -th root involves two operations:

1. Finding the n -th root of the magnitude: This gives us $r^{1/n}$, which determines the radius of a circle centered at the origin.
2. Dividing the argument by n : This gives us $\frac{\theta}{n}$, which determines the angle of the first root.

The key insight is that angles in the complex plane are periodic with period 2π . When we compute $z^{1/n}$, we need to account for this periodicity. If $w^n = z$, then $(we^{i2\pi k/n})^n = z$ for any integer k , because:

$$(we^{i2\pi k/n})^n = w^n \cdot e^{i2\pi k} = w^n \cdot 1 = z$$

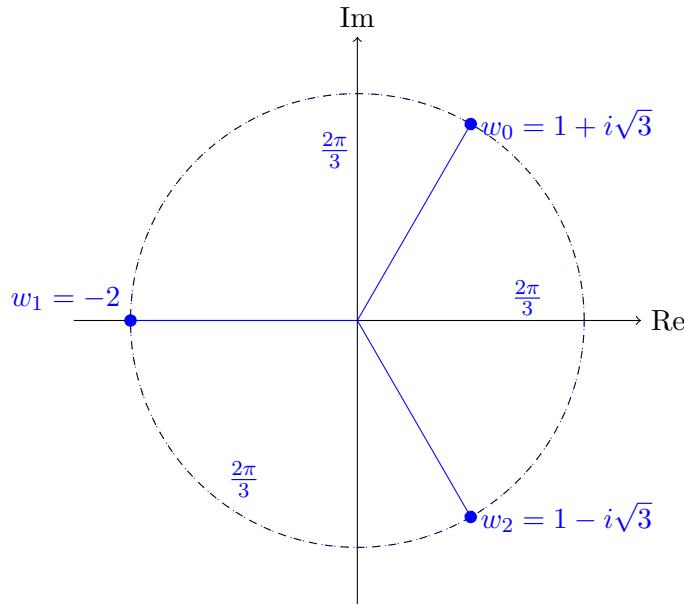
This is why the formula for the n -th roots has the form:

$$w_k = r^{1/n} e^{i(\theta+2\pi k)/n} \quad \text{for } k = 0, 1, 2, \dots, n-1$$

Geometrically, we can visualize this as follows:

1. Start with the original complex number $z = re^{i\theta}$ at distance r from the origin and angle θ .
2. Compute the radius of the circle where all n -th roots will lie: $r^{1/n}$.
3. Place the first root ($k = 0$) at angle $\frac{\theta}{n}$ on this circle.
4. Place the remaining roots by incrementing the angle by $\frac{2\pi}{n}$ each time.

This gives us n distinct points, evenly spaced around a circle of radius $r^{1/n}$, with angular separation $\frac{2\pi}{n}$ between consecutive roots.

Figure 2.4: Cube roots of -8 in the complex plane

We can understand why there are exactly n roots by considering what happens when $k = n$:

$$w_n = r^{1/n} e^{i(\theta+2\pi n)/n} = r^{1/n} e^{i\theta/n} e^{i2\pi} = r^{1/n} e^{i\theta/n} = w_0$$

Since $e^{i2\pi} = 1$, the pattern repeats after n roots. This cycle continues indefinitely, which is why we only consider $k = 0, 1, 2, \dots, n - 1$.

This geometric perspective about complex roots also explains a fundamental result about real numbers: why even-powered roots of positive real numbers always yield exactly two real solutions, while odd-powered roots yield exactly one.

The core idea of the proof lies in that a real number $a > 0$ can be expressed as $a = re^{i \cdot 0}$ (since it lies on the positive real axis). Its n -th roots are positioned at angles $\frac{2\pi k}{n}$ for $k = 0, 1, 2, \dots, n - 1$.

A root is real precisely when its angle is a multiple of π —that is, when it lies on the real axis. For even values of n , exactly two values of k (namely, $k = 0$ and $k = \frac{n}{2}$) produce angles that are multiples of π , resulting in two real roots. For odd values of n , only $k = 0$ produces a real root. This connection between complex analysis and real number properties demonstrates the elegant unity of mathematics across seemingly distinct domains.

Theorem: Number of Real Roots

A positive real number has exactly two real n -th roots if n is even, and exactly one real n -th root if n is odd. A negative real number has no real n -th roots if n is even, and exactly one real n -th root if n is odd.

Proof

Let's consider a real number a and examine its n -th roots in the complex plane.

Case 1: $a > 0$. We can write $a = |a|e^{i \cdot 0}$ since a lies on the positive real axis.

Using the formula for n -th roots:

$$w_k = |a|^{1/n} e^{i(0+2\pi k)/n} = |a|^{1/n} e^{i2\pi k/n}$$

for $k = 0, 1, 2, \dots, n - 1$.

A complex number is real if and only if its imaginary part is zero, which occurs when the angle is a multiple of π . Therefore, we need to find values of k such that $\frac{2\pi k}{n}$ is a multiple of π .

This happens when $\frac{2\pi k}{n} = m\pi$ for some integer m , which simplifies to $\frac{2k}{n} = m$.

For k in the range $\{0, 1, 2, \dots, n - 1\}$:

- When $k = 0$, we get $\frac{2 \cdot 0}{n} = 0$, which is a multiple of π . This corresponds to the principal n -th root $|a|^{1/n}$ on the positive real axis.
- When $k = \frac{n}{2}$ (possible only if n is even), we get $\frac{2 \cdot \frac{n}{2}}{n} = 1$, giving an angle of π . This corresponds to the root $-|a|^{1/n}$ on the negative real axis.

No other values of k in our range yield angles that are multiples of π . Therefore, a positive real number has exactly one real n -th root if n is odd (the principal root $|a|^{1/n}$) and exactly two real n -th roots if n is even ($|a|^{1/n}$ and $-|a|^{1/n}$).

Case 2: $a < 0$. We can write $a = |a|e^{i\pi}$ since a lies on the negative real axis.

Using the formula for n -th roots:

$$w_k = |a|^{1/n} e^{i(\pi+2\pi k)/n} = |a|^{1/n} e^{i\pi(1+2k)/n}$$

for $k = 0, 1, 2, \dots, n - 1$.

As before, a root is real when its angle is a multiple of π , which happens when $\frac{\pi(1+2k)}{n} = m\pi$ for some integer m , or equivalently, when $\frac{1+2k}{n} = m$.

For k in the range $\{0, 1, 2, \dots, n - 1\}$:

- When n is odd, $k = \frac{n-1}{2}$ gives us $\frac{1+2 \cdot \frac{n-1}{2}}{n} = \frac{1+n-1}{n} = 1$, corresponding to a real root at angle π , which is $-|a|^{1/n}$ on the negative real axis.
- When n is even, there is no value of k in our range that makes $\frac{1+2k}{n}$ an integer, so there are no real n -th roots.

Therefore, a negative real number has exactly one real n -th root if n is odd and no real n -th roots if n is even.

This completes the proof.



2.4 Practice Exercises

Exercise 2.1

Using the method of proof by contradiction, prove that $\sqrt{3} + \sqrt{5}$ is an irrational number.

Exercise 2.2

If x is an irrational number and y is a non-zero rational number, prove that $x \cdot y$ is irrational.

Exercise 2.3

Using the iterative approximation formula for square roots, find the first four rational approximations for $\sqrt{3}$ starting with $\frac{1}{1}$.

Exercise 2.4

If n is a positive integer, prove that $\lfloor \frac{n}{2} \rfloor + \lceil \frac{n}{2} \rceil = n$.

Exercise 2.5

Prove that for any real numbers x and y , $\lfloor x + y \rfloor \geq \lfloor x \rfloor + \lfloor y \rfloor$.

Exercise 2.6

Find the sum $\sum_{k=1}^{10} \lfloor \sqrt{k} \rfloor$.

Exercise 2.7

Find all complex numbers z such that $z^3 = \bar{z}$, where \bar{z} represents the complex conjugate of z .

Exercise 2.8

Using De Moivre's theorem, express $\cos(5\theta)$ in terms of powers of $\cos(\theta)$ and $\sin(\theta)$.

Exercise 2.9

For what values of x does $\{x\} + \{2x\} + \{3x\} = 1$ hold?

Exercise 2.10

Find all solutions to the equation $z^4 + 16 = 0$ in the complex plane, and represent them geometrically.

Chapter 3

Foundational Mathematics



Peter Gustav Lejeune Dirichlet (1805-1859)

Dirichlet is considered one of the founders of modern number theory and a pioneer of analytic methods in number theory. He formalized the Pigeonhole Principle (originally called “Schubfachprinzip” or drawer principle) and proved several foundational results, including his famous theorem on primes in arithmetic progressions. His work on the distribution of prime numbers was revolutionary, proving that there are infinitely many primes in any arithmetic sequence where the first term and common difference have no common factors. A student of Gauss, he later succeeded him at Göttingen, cementing his place among the mathematical elite of his era.

Dirichlet had such difficulty learning Latin in school that his teacher once proclaimed he would never amount to anything academically. This same “academically hopeless” student later became one of history’s greatest mathematicians and mastered multiple languages to read mathematical works in their original form, proving his early teacher wrong.

This chapter is designed to provide conceptual results that are needed for a comprehensive study of number theory. While fundamental results such as the Pigeonhole Principle, Binomial Theorem, and Fundamental Theorem of Algebra are not exclusive to number theory, they serve as essential tools that we will employ in subsequent chapters. Understanding these foundational results will allow us to delve deeper into number theory beginning with the next chapter, where we will explore core number-theoretic concepts.

We will begin our mathematical journey with Pigeonhole Principle, one of the simplest results in mathematics and yet one so powerful that it never ceases to amaze me.

3.1 The Pigeonhole Principle

Theorem: Pigeonhole Principle

If n pigeons are present in m pigeonholes, with $n > m$, then at least one pigeonhole must contain more than one pigeon.

The pigeonhole principle may seem obvious at the first blush, but its applications can be surprisingly elegant and powerful. Even though Dirichlet was the first to formally introduce it in 1834, the concept may have been understood intuitively before him. The modern name "Pigeonhole Principle" emerged later as the concept spread to English-speaking mathematics communities. The principle has since become an indispensable tool to solve mathematical results. Let's look at a few applications of this principle.

Example 3.1.1

In a group of 367 people, prove that at least two people have the same birthday.

Solution: There are at most 366 possible birthdays (including February 29). With 367 people, by the pigeonhole principle, at least two people must share the same birthday.

Example 3.1.2

Prove that among any $n + 1$ integers, there are always two whose difference is divisible by n .

Solution: Consider the remainders when each of the $n + 1$ integers is divided by n . There are only n possible remainders (0 to $n - 1$). By the pigeonhole principle, at least

two of the $n + 1$ integers must have the same remainder when divided by n .

Let these two integers be a and b . Then $a = nq_1 + r$ and $b = nq_2 + r$ for some integers q_1 , q_2 , and r (where r is the common remainder). Their difference is $a - b = n(q_1 - q_2)$, which is clearly divisible by n .

3.2 Mathematical Induction

Theorem: Principle of Mathematical Induction

Let $P(n)$ be a statement about the natural number n . If:

1. *$P(1)$ is true (base case), and*
2. *For any $k \geq 1$, if $P(k)$ is true, then $P(k + 1)$ is also true (inductive step),*

then $P(n)$ is true for all natural numbers n .

Mathematical induction is another extremely powerful fundamental method and has a wide range of applications. Mathematical induction is typically used to establish that a given statement is true for all natural numbers. Let's look at an example problem.

Example 3.2.1

Prove that $\sum_{i=1}^n i = \frac{n(n+1)}{2}$ for all natural numbers n .

Definition: Summation Notation

$\sum_{i=1}^n i$ is the sum of all the natural numbers less than or equal to n . For example:
 $\sum_{i=1}^5 i = 1 + 2 + 3 + 4 + 5 = 15$.

Solution:

1. Base case: For $n = 1$,

$$\begin{aligned}\text{LHS} &= 1 \\ \text{RHS} &= \frac{1(1+1)}{2} = 1\end{aligned}$$

The statement holds for $n = 1$.

2. Inductive step: Assume the statement is true for some k . We need to prove it for $k + 1$. Assume $\sum_{i=1}^k i = \frac{k(k+1)}{2}$

Now,

$$\begin{aligned}\sum_{i=1}^{k+1} i &= \left(\sum_{i=1}^k i \right) + (k+1) \\ &= \frac{k(k+1)}{2} + (k+1) \\ &= \frac{k(k+1) + 2(k+1)}{2} \\ &= \frac{(k+1)(k+2)}{2} \\ &= \frac{(k+1)((k+1)+1)}{2}\end{aligned}$$

This is exactly the formula for $n = k + 1$.

Therefore, by mathematical induction, the formula is true for all natural numbers n .

The core logic behind induction is to first prove the statement for the base case. In the previous example, the base case was $n = 1$. But there is no requirement that the base case has to be 1. It can be any integer.

Once the base case has been proved, we prove that the statement holds for $n = k + 1$, assuming that the statement holds for $n = k$. If the base statement was proven for $n = 1$, then this second step implies that the statement is true for $n = 2$. Now, because the statement is true for $n = 2$, it follows from the second step that the statement is true for $n = 3$ and so on. In other words, the two steps of mathematical induction imply that the statement is true for all integers starting from the base case integer. The statement might not hold for integers smaller than the base case integer.

Let's look at two more examples to engrain the concepts behind the mathematical induction.

Example 3.2.2

Prove that $3^n - 1$ is divisible by 2 for all positive integers n .

Solution:

1. Base case: For $n = 1$, $3^1 - 1 = 2$, which is divisible by 2.
2. Inductive step: Assume the statement is true for some k . We need to prove it for $k + 1$.

Assume $3^k - 1 = 2m$ for some integer m .

Now, consider $3^{k+1} - 1$:

$$\begin{aligned} 3^{k+1} - 1 &= 3 \cdot 3^k - 1 \\ &= 3(3^k - 1) + 2 \\ &= 3(2m) + 2 \\ &= 6m + 2 \\ &= 2(3m + 1) \end{aligned}$$

This is clearly divisible by 2.

Therefore, by mathematical induction, $3^n - 1$ is divisible by 2 for all positive integers n .

Example 3.2.3

Prove that $n! < n^n$ for all integers $n \geq 2$.

Solution:

1. Base case: For $n = 2$, $2! = 2 < 2^2 = 4$. The statement holds.
2. Inductive step: Assume the statement is true for some $k \geq 2$. We need to prove it for $k + 1$.

Assume $k! < k^k$. Now, consider $(k + 1)!$:

$$\begin{aligned} (k + 1)! &= (k + 1) \cdot k! \\ &< (k + 1) \cdot k^k \quad (\text{by inductive hypothesis}) \\ &< (k + 1) \cdot (k + 1)^k \quad (\text{since } k + 1 > k) \\ &= (k + 1)^{k+1} \end{aligned}$$

Therefore, by mathematical induction, $n! < n^n$ for all integers $n \geq 2$.

3.2.1 Strong Induction

In our previous examples, we used what mathematicians call "Weak Induction" (or standard induction), where we prove the base case $P(1)$ and then assume $P(k)$ to show $P(k+1)$. However, there is an another variant called "Strong Induction" that is sometimes necessary for solving certain types of problems.

Theorem: Principle of Strong Induction

Let $P(n)$ be a statement about positive integers. If:

1. $P(1)$ is true (base case), and
2. For any integer $k \geq 1$, if $P(1), P(2), \dots, P(k)$ are all true, then $P(k+1)$ is also true (inductive step)

Then $P(n)$ is true for all positive integers n .

Strong induction is particularly useful when the statement about $P(k + 1)$ depends not just on $P(k)$, but potentially on $P(j)$ for some $j < k$. The key difference is that with weak induction, we only assume $P(k)$ is true, while with strong induction, we assume $P(1), P(2), \dots, P(k)$ are all true when proving $P(k + 1)$.

Example 3.2.4

Prove that every integer greater than 1 can be written as a product of prime numbers.

Definition: Prime Numbers

A prime number is a natural number greater than 1 that cannot be formed by multiplying two smaller natural numbers. Equivalently, a natural number $p > 1$ is prime if and only if it has exactly two positive divisors: 1 and itself.

The first few prime numbers are: 2, 3, 5, 7, 11, 13, 17, 19, 23, 29, ...

Solution:

1. Base case: For $n = 2$, the statement is true because 2 is itself a prime number, so it's trivially a product of primes (just one prime).
 2. Inductive step: Assume that for all integers m with $2 \leq m \leq k$, the statement $P(m)$ is true: m can be written as a product of primes.
- Now, we need to prove $P(k+1)$: the integer $k+1$ can be written as a product of primes. There are two cases to consider:
- Case 1: If $k+1$ is a prime number, then it is already expressed as a product of primes (itself), so $P(k+1)$ is true.
 - Case 2: If $k+1$ is not a prime number, then it has a factorization $k+1 = a \cdot b$, where a and b are integers with $2 \leq a, b \leq k$. By our inductive hypothesis, both a and b can be written as products of primes: $a = p_1 p_2 \cdots p_r$, $b = q_1 q_2 \cdots q_s$.

Therefore: $k + 1 = ab = p_1 p_2 \cdots p_r \cdot q_1 q_2 \cdots q_s$. This expresses $k+1$ as a product of primes, so $P(k+1)$ is true.

Thus every integer greater than 1 can be written as a product of primes. This proof requires strong induction because when factoring $k+1$ as $a \cdot b$, we need to use the fact that a and b (which are less than $k+1$) can be written as products of primes. Weak induction wouldn't work directly since we need the result for numbers less than k , not just for k itself.

Example 3.2.5

The Fibonacci sequence is defined recursively as:

$$F_1 = 1 \quad (3.2.1)$$

$$F_2 = 1 \quad (3.2.2)$$

$$F_n = F_{n-1} + F_{n-2} \quad \text{for } n \geq 3 \quad (3.2.3)$$

Prove that for all $n \geq 1$, we have $F_n < 2^n$.

Solution:

1. Base cases:

- For $n = 1$: $F_1 = 1 < 2^1 = 2$. True.
- For $n = 2$: $F_2 = 1 < 2^2 = 4$. True.

2. Inductive step: Assume that for all integers m with $1 \leq m \leq k$, where $k \geq 2$, we have $F_m < 2^m$. Now, we need to prove $F_{k+1} < 2^{k+1}$.

We know that $F_{k+1} = F_k + F_{k-1}$.

By our inductive hypothesis:

$$F_k < 2^k \quad (3.2.4)$$

$$F_{k-1} < 2^{k-1} \quad (3.2.5)$$

Therefore:

$$F_{k+1} = F_k + F_{k-1} \quad (3.2.6)$$

$$< 2^k + 2^{k-1} \quad (3.2.7)$$

$$= 2^k + \frac{2^k}{2} \quad (3.2.8)$$

$$= 2^k(1 + \frac{1}{2}) \quad (3.2.9)$$

$$= 2^k \cdot \frac{3}{2} \quad (3.2.10)$$

$$< 2^k \cdot 2 \quad (3.2.11)$$

$$= 2^{k+1} \quad (3.2.12)$$

Thus, $F_{k+1} < 2^{k+1}$, completing the inductive step. By the principle of strong induction, $F_n < 2^n$ for all $n \geq 1$.

This problem requires strong induction because the definition of F_{k+1} depends on both F_k and F_{k-1} , not just on F_k . We need to use our hypothesis for both of these previous terms.

While the strong induction might seem more powerful than weak induction, the two forms are actually logically equivalent. Any proof that can be done with strong induction can theoretically be restructured to use weak induction, and vice versa, though the strong induction form is often more natural and clearer for certain problems.

Having mastered both forms of mathematical induction, we now turn to one of its most sophisticated applications: proving the Arithmetic Mean-Geometric Mean inequality (commonly abbreviated as AM-GM inequality).

3.3 The Arithmetic Mean-Geometric Mean Inequality

The AM-GM inequality provides a beautiful example of how mathematical induction can establish a result that might otherwise be difficult to prove directly. Moreover, once established, this inequality becomes a powerful tool that we can apply to solve a variety of challenging problems.

Theorem: Arithmetic Mean-Geometric Mean Inequality

For any n non-negative real numbers a_1, a_2, \dots, a_n , we have:

$$\frac{a_1 + a_2 + \cdots + a_n}{n} \geq \sqrt[n]{a_1 \cdot a_2 \cdots a_n} \quad (3.3.1)$$

with equality if and only if $a_1 = a_2 = \cdots = a_n$.

In words, this theorem states that the arithmetic mean (average) of a set of non-negative real numbers is always greater than or equal to their geometric mean (the n -th root of their product). Equality occurs only when all the numbers are equal.

Proof

We will prove this result using forward-backward induction, an elegant technique that combines two types of inductive steps. Let $P(n)$ be the statement that the AM-GM inequality holds for n non-negative numbers.

1. *Base case:* For $n = 2$, we need to prove that for any non-negative real numbers a and b :

$$\frac{a+b}{2} \geq \sqrt{ab} \quad (3.3.2)$$

Squaring both sides (which is valid since both sides are non-negative):

$$\left(\frac{a+b}{2}\right)^2 \geq ab \quad (3.3.3)$$

$$\frac{(a+b)^2}{4} \geq ab \quad (3.3.4)$$

$$\frac{a^2 + 2ab + b^2}{4} \geq ab \quad (3.3.5)$$

$$a^2 + 2ab + b^2 \geq 4ab \quad (3.3.6)$$

$$a^2 - 2ab + b^2 \geq 0 \quad (3.3.7)$$

$$(a-b)^2 \geq 0 \quad (3.3.8)$$

The last inequality is always true for any real numbers a and b , with equality if and only if $a = b$. This proves the base case.

2. *Forward Inductive Step:* We will first prove that $(P(k) \implies P(2k))$, before proving $(P(k) \implies P(k+1))$.

Assume $P(k)$ is true, meaning for any k non-negative numbers b_1, b_2, \dots, b_k :

$$\frac{b_1 + b_2 + \dots + b_k}{k} \geq \sqrt[k]{b_1 \cdot b_2 \cdots b_k} \quad (3.3.9)$$

Now consider $2k$ non-negative numbers a_1, a_2, \dots, a_{2k} . We divide them into two groups and define:

$$A_1 = \frac{a_1 + a_2 + \dots + a_k}{k} \quad (3.3.10)$$

$$A_2 = \frac{a_{k+1} + a_{k+2} + \dots + a_{2k}}{k} \quad (3.3.11)$$

The arithmetic mean of all $2k$ numbers can be written as:

$$\frac{a_1 + a_2 + \dots + a_{2k}}{2k} = \frac{(a_1 + \dots + a_k) + (a_{k+1} + \dots + a_{2k})}{2k} \quad (3.3.12)$$

$$= \frac{kA_1 + kA_2}{2k} \quad (3.3.13)$$

$$= \frac{A_1 + A_2}{2} \quad (3.3.14)$$

By the base case $P(2)$ applied to A_1 and A_2 :

$$\frac{A_1 + A_2}{2} \geq \sqrt{A_1 A_2} \quad (3.3.15)$$

By our induction hypothesis $P(k)$ applied to each group:

$$A_1 \geq \sqrt[k]{a_1 \cdot a_2 \cdots a_k} = G_1 \quad (3.3.16)$$

$$A_2 \geq \sqrt[k]{a_{k+1} \cdot a_{k+2} \cdots a_{2k}} = G_2 \quad (3.3.17)$$

Since these inequalities imply $A_1 \geq G_1$ and $A_2 \geq G_2$, we have $\sqrt{A_1 A_2} \geq \sqrt{G_1 G_2}$. Therefore:

$$\frac{a_1 + a_2 + \cdots + a_{2k}}{2k} \geq \sqrt{G_1 G_2} \quad (3.3.18)$$

$$= \sqrt{\sqrt[k]{a_1 \cdots a_k} \sqrt[k]{a_{k+1} \cdots a_{2k}}} \quad (3.3.19)$$

$$= \sqrt{\left(\prod_{i=1}^k a_i\right)^{1/k} \left(\prod_{i=k+1}^{2k} a_i\right)^{1/k}} \quad (3.3.20)$$

$$= \sqrt{\left(\prod_{i=1}^{2k} a_i\right)^{1/k}} \quad (3.3.21)$$

$$= \left(\prod_{i=1}^{2k} a_i\right)^{1/(2k)} \quad (3.3.22)$$

$$= \sqrt[2k]{a_1 a_2 \cdots a_{2k}} \quad (3.3.23)$$

This proves that $P(k) \implies P(2k)$. Starting with $P(2)$, we can establish $P(4)$, $P(8)$, $P(16)$, and generally $P(2^m)$ for all integers $m \geq 1$.

3. Backward Inductive Step: we prove $(P(k) \implies P(k-1))$.

Let us assume $P(k)$ is true for some $k \geq 3$. We need to show $P(k-1)$ holds.

Let a_1, a_2, \dots, a_{k-1} be $k-1$ non-negative numbers. Define their arithmetic mean:

$$A_{k-1} = \frac{a_1 + a_2 + \cdots + a_{k-1}}{k-1} \quad (3.3.24)$$

We construct a clever extension by adding a k -th number $a_k = A_{k-1}$ and applying the known inequality $P(k)$ to all k numbers:

$$\frac{a_1 + a_2 + \cdots + a_{k-1} + a_k}{k} \geq \sqrt[k]{a_1 \cdot a_2 \cdots a_{k-1} \cdot a_k} \quad (3.3.25)$$

Substituting $a_k = A_{k-1}$ and noting that $a_1 + a_2 + \cdots + a_{k-1} = (k-1)A_{k-1}$:

$$\frac{(k-1)A_{k-1} + A_{k-1}}{k} \geq \sqrt[k]{a_1 \cdot a_2 \cdots a_{k-1} \cdot A_{k-1}} \quad (3.3.26)$$

$$\frac{kA_{k-1}}{k} \geq \sqrt[k]{\left(\prod_{i=1}^{k-1} a_i\right) \cdot A_{k-1}} \quad (3.3.27)$$

$$A_{k-1} \geq \sqrt[k]{\left(\prod_{i=1}^{k-1} a_i\right) \cdot A_{k-1}} \quad (3.3.28)$$

Let $G_{k-1} = \sqrt[k-1]{a_1 \cdot a_2 \cdots a_{k-1}}$, which gives $\prod_{i=1}^{k-1} a_i = G_{k-1}^{k-1}$. Substituting:

$$A_{k-1} \geq \sqrt[k]{G_{k-1}^{k-1} \cdot A_{k-1}} \quad (3.3.29)$$

Raising both sides to the power of k :

$$A_{k-1}^k \geq G_{k-1}^{k-1} \cdot A_{k-1} \quad (3.3.30)$$

If $A_{k-1} = 0$, then all a_1, \dots, a_{k-1} must be zero, which means $G_{k-1} = 0$, and the inequality holds trivially.

If $A_{k-1} > 0$, we can divide both sides by A_{k-1} :

$$A_{k-1}^{k-1} \geq G_{k-1}^{k-1} \quad (3.3.31)$$

Taking the $(k-1)$ -th root of both sides (which preserves the inequality for non-negative values):

$$A_{k-1} \geq G_{k-1} \quad (3.3.32)$$

$$\frac{a_1 + a_2 + \cdots + a_{k-1}}{k-1} \geq \sqrt[k-1]{a_1 \cdot a_2 \cdots a_{k-1}} \quad (3.3.33)$$

This establishes $P(k-1)$, completing the backward induction step.

Thus, We have shown:

- $P(2)$ is true (base case)
- $P(k) \Rightarrow P(2k)$ (forward step) $\Rightarrow P(2^m)$ is true for all $m \geq 1$
- $P(k) \Rightarrow P(k-1)$ (backward step)

For any positive integer n , we can find an integer m such that $2^m \geq n$.

Let $k = 2^m$. Since $P(k)$ is true, we can apply the backward induction step repeatedly:

$P(k) \Rightarrow P(k-1) \Rightarrow P(k-2) \Rightarrow \cdots \Rightarrow P(n)$

Therefore, $P(n)$ is true for all positive integers n .

For the equality case, we need equality at every step of the induction. Tracing through:

- In the base case, equality requires $a_1 = a_2$
- In the forward step, equality requires $A_1 = A_2$ and equality in both applications of $P(k)$
- In the backward step, equality requires $a_1 = a_2 = \dots = a_{k-1} = a_k = A_{k-1}$

These conditions together imply that equality holds in the AM-GM inequality if and only if all the numbers are equal: $a_1 = a_2 = \dots = a_n$.

□

You might have noticed that our proof of the AM-GM inequality using forward-backward induction is quite elaborate. This was intentional for two important pedagogical reasons.

First, this proof offers deeper insight into the mechanics and versatility of induction as a proof technique. The forward-backward approach demonstrates how induction can be applied in creative ways beyond the standard technique we've seen in earlier examples. By working through this more sophisticated induction strategy, we gain a richer understanding of how mathematical induction operates.

Second, this example serves as an important cautionary tale. While induction is undoubtedly a powerful method, it sometimes produces proofs that are considerably more complex than necessary. The AM-GM inequality actually has several simpler proofs using other techniques. This illustrates an important principle in mathematics: the most direct proof method for a given theorem isn't always immediately obvious, and sometimes an alternative approach may yield more elegant results.

This realization should encourage us to approach problem-solving with flexibility. Mathematical induction is an essential tool, but it should be used judiciously, especially when other methods might lead to more straightforward solutions.

Now that we have established this powerful inequality, let's explore various ways it can be applied to solve practical problems.

Example 3.3.1

Use the AM-GM inequality to find the minimum value of the expression $f(x) = x + \frac{9}{x}$ for $x > 0$.

Solution: We can apply the AM-GM inequality directly to the two terms x and $\frac{9}{x}$:

$$\frac{x + \frac{9}{x}}{2} \geq \sqrt{x \cdot \frac{9}{x}} \quad (3.3.34)$$

$$\frac{x + \frac{9}{x}}{2} \geq \sqrt{9} \quad (3.3.35)$$

$$\frac{x + \frac{9}{x}}{2} \geq 3 \quad (3.3.36)$$

Therefore:

$$x + \frac{9}{x} \geq 6 \quad (3.3.37)$$

This gives us the minimum value of $f(x)$, which is 6. To find the value of x at which this minimum occurs, we use the equality condition of the AM-GM inequality, which states that equality holds if and only if all terms are equal. So:

$$x = \frac{9}{x} \quad (3.3.38)$$

$$x^2 = 9 \quad (3.3.39)$$

$$x = 3 \quad (3.3.40)$$

Therefore, the minimum value of $f(x) = x + \frac{9}{x}$ for $x > 0$ is 6, occurring at $x = 3$.

Example 3.3.2

Prove that for any three positive real numbers a , b , and c , we have:

$$\frac{a}{b+c} + \frac{b}{a+c} + \frac{c}{a+b} \geq \frac{3}{2} \quad (3.3.41)$$

Solution: We'll apply the AM-GM inequality to carefully chosen terms. First, let's multiply each fraction by an appropriate expression:

$$\frac{a}{b+c} = \frac{a(a+b)(a+c)}{(b+c)(a+b)(a+c)} \quad (3.3.42)$$

$$\frac{b}{a+c} = \frac{b(a+b)(b+c)}{(a+c)(a+b)(b+c)} \quad (3.3.43)$$

$$\frac{c}{a+b} = \frac{c(a+c)(b+c)}{(a+b)(a+c)(b+c)} \quad (3.3.44)$$

Adding these fractions:

$$\frac{a}{b+c} + \frac{b}{a+c} + \frac{c}{a+b} = \frac{a(a+b)(a+c) + b(a+b)(b+c) + c(a+c)(b+c)}{(a+b)(b+c)(a+c)} \quad (3.3.45)$$

Let's define:

$$x = a(a+b)(a+c) \quad (3.3.46)$$

$$y = b(a+b)(b+c) \quad (3.3.47)$$

$$z = c(a+c)(b+c) \quad (3.3.48)$$

By the AM-GM inequality for three positive numbers:

$$\frac{x+y+z}{3} \geq \sqrt[3]{xyz} \quad (3.3.49)$$

It can be shown through algebraic manipulation that $\sqrt[3]{xyz} \geq \frac{3}{2}(a+b)(b+c)(a+c)$.

Therefore:

$$\frac{a}{b+c} + \frac{b}{a+c} + \frac{c}{a+b} = \frac{x+y+z}{(a+b)(b+c)(a+c)} \geq \frac{3\sqrt[3]{xyz}}{(a+b)(b+c)(a+c)} \geq \frac{3}{2} \quad (3.3.50)$$

Thus, we've proven the inequality:

$$\frac{a}{b+c} + \frac{b}{a+c} + \frac{c}{a+b} \geq \frac{3}{2} \quad (3.3.51)$$

Example 3.3.3

For positive real numbers a , b , and c , prove that:

$$a^3 + b^3 + c^3 \geq 3abc \quad (3.3.52)$$

Solution: We'll use the AM-GM inequality for three numbers. For any three positive real numbers, we know:

$$\frac{x+y+z}{3} \geq \sqrt[3]{xyz} \quad (3.3.53)$$

Let's set $x = a^3$, $y = b^3$, and $z = c^3$. Then:

$$\frac{a^3 + b^3 + c^3}{3} \geq \sqrt[3]{a^3 \cdot b^3 \cdot c^3} \quad (3.3.54)$$

$$\frac{a^3 + b^3 + c^3}{3} \geq \sqrt[3]{a^3 b^3 c^3} \quad (3.3.55)$$

$$\frac{a^3 + b^3 + c^3}{3} \geq \sqrt[3]{(abc)^3} \quad (3.3.56)$$

$$\frac{a^3 + b^3 + c^3}{3} \geq abc \quad (3.3.57)$$

Therefore:

$$a^3 + b^3 + c^3 \geq 3abc \quad (3.3.58)$$

Equality holds if and only if $a^3 = b^3 = c^3$, which means $a = b = c$.

The AM-GM inequality is particularly useful in minimization problems, as we saw in our examples. However, its applications extend far beyond optimization. It is used in proving other inequalities, solving geometrical problems, and has connections to many areas of mathematics including calculus, probability, and number theory.

As we continue our study of mathematics, we will encounter many opportunities to apply this powerful result. The AM-GM inequality exemplifies how a theorem proven through induction can become a fundamental tool for solving diverse mathematical problems.

3.4 Mathematical Functions

While induction gives us a powerful tool for proving statements about natural numbers, we now turn our attention to mathematical functions, which allow us to describe relationships between sets of numbers or other mathematical objects. Functions are fundamental to virtually every branch of mathematics, and will be particularly important in subsequent chapters.

3.4.1 Introduction to Functions

A function f from set A to set B (written as $f : A \rightarrow B$) is a rule that assigns to each element $x \in A$ exactly one element $y \in B$, denoted by $f(x) = y$. The set A is called the domain and B is called the codomain of the function.

A domain represents the set of all permissible input values for a function. A codomain plays a crucial role in completely defining a function. While a range (the set of actual outputs) is a subset of the codomain and tells us which values are actually produced by the function, the codomain provides the context or universe in which the function operates. For example, a function $f(x) = x^2$ with domain \mathbb{R} could have codomain \mathbb{R} or \mathbb{R}^+ , and this distinction affects the nature of the function.

The image of an element x under function f is simply the output value $f(x)$. Conversely, for a value y in the range, the preimage (or inverse image) is the set of all elements x in the domain such that $f(x) = y$. This concept becomes particularly important when we study functions that are not one-to-one, where multiple inputs may yield the same output.

3.4.2 Types of Functions Based on Mapping

One-to-One (Injective) Functions

Definition:

A function $f : A \rightarrow B$ is called one-to-one or injective if for any $x_1, x_2 \in A$ with $x_1 \neq x_2$, we have $f(x_1) \neq f(x_2)$. In other words,

$$f \text{ is injective} \iff \forall x_1, x_2 \in A, x_1 \neq x_2 \Rightarrow f(x_1) \neq f(x_2) \quad (3.4.1)$$

$$\iff \forall x_1, x_2 \in A, f(x_1) = f(x_2) \Rightarrow x_1 = x_2 \quad (3.4.2)$$

Intuitively, an injective function maps distinct elements in the domain to distinct elements in the codomain. This means no two different input values can produce the same output value. The horizontal line test offers a geometric interpretation: a function is injective if any horizontal line intersects its graph at most once. For finite sets, an injective function $f : A \rightarrow B$ requires that $|A| \leq |B|$, as each element in A needs a unique “target” in B .

Example 3.4.1

Prove that the function $f(x) = 2x + 3$ is injective.

Solution: We need to show that for any $x_1, x_2 \in \mathbb{R}$, if $f(x_1) = f(x_2)$, then $x_1 = x_2$.

Let $x_1, x_2 \in \mathbb{R}$ such that $f(x_1) = f(x_2)$. Then $2x_1 + 3 = 2x_2 + 3$ Subtracting 3 from both sides: $2x_1 = 2x_2$ Dividing both sides by 2: $x_1 = x_2$

Therefore, f is injective by the definition.

Example 3.4.2

Show that while the function $g(x) = x^2$ with domain \mathbb{R} is not injective, the same function with domain restricted to \mathbb{R}^+ (positive real numbers) is injective.

Solution: To show that g is not injective on the domain \mathbb{R} , we need to find two distinct values that map to the same value in the codomain.

Consider $x_1 = 2$ and $x_2 = -2$. Clearly, $x_1 \neq x_2$. But $g(x_1) = g(2) = 2^2 = 4$ and $g(x_2) = g(-2) = (-2)^2 = 4$. Thus, $g(x_1) = g(x_2)$ while $x_1 \neq x_2$.

This contradicts the definition of injectivity, implying that g is not injective. Now, let's

show that $h(x) = x^2$ with domain \mathbb{R}^+ is injective. To prove injectivity, we need to show that for any $x_1, x_2 \in \mathbb{R}^+$, if $h(x_1) = h(x_2)$ then $x_1 = x_2$.

Let $x_1, x_2 \in \mathbb{R}^+$ and assume that $h(x_1) = h(x_2)$. Then:

$$x_1^2 = x_2^2 \quad (3.4.3)$$

Taking the square root of both sides:

$$\sqrt{x_1^2} = \sqrt{x_2^2} \quad (3.4.4)$$

Since $x_1, x_2 \in \mathbb{R}^+$ (both are positive), we have $\sqrt{x_1^2} = |x_1| = x_1$ and $\sqrt{x_2^2} = |x_2| = x_2$.

Therefore:

$$x_1 = x_2 \quad (3.4.5)$$

This proves that $h(x) = x^2$ with domain \mathbb{R}^+ is injective.

This example illustrates an important principle in the study of functions: restricting the domain of a function can change its fundamental properties. The function $f(x) = x^2$ fails to be injective on \mathbb{R} because pairs of opposite numbers (like 2 and -2) map to the same output. By restricting the domain to only positive numbers, we eliminate this possibility, making the function injective.

Many-to-One Functions

A function that is not injective is called many-to-one. In such functions, at least two distinct elements in the domain map to the same element in the codomain.

Definition:

f is many-to-one if and only if $\exists x_1, x_2 \in A$ such that $x_1 \neq x_2$ and $f(x_1) = f(x_2)$

The existence of a single pair of distinct inputs mapping to the same output is sufficient to classify a function as many-to-one. These functions are common in mathematics and real-world applications where information gets "compressed" or where multiple scenarios lead to the same outcome.

Example 3.4.3

Demonstrate why the function $f(x) = x^2$ with domain \mathbb{R} is many-to-one.

Solution: For any non-zero real number x , we have $f(x) = f(-x) = x^2$. This is because the squaring operation removes the sign information.

Specifically: $f(1) = 1^2 = 1$ and $f(-1) = (-1)^2 = 1$ $f(2) = 2^2 = 4$ and $f(-2) = (-2)^2 = 4$

Since for any $x \neq 0$, we have two distinct inputs x and $-x$ that map to the same output x^2 , the function $f(x) = x^2$ is many-to-one.

Similar to function $f(x) = x^2$, one can also show that the function $f(x) = |x|$ is also many-to-one.

Onto (Surjective) Functions

A function $f : A \rightarrow B$ is called onto or surjective if for every element $y \in B$, there exists at least one element $x \in A$ such that $f(x) = y$.

Definition:

f is surjective iff $\forall y \in B, \exists x \in A$ such that $f(x) = y$

In a surjective function, the range equals the codomain; every possible output value is actually realized by at least one input value. This concept becomes particularly important in establishing relationships between sets and in the study of invertible functions. For finite sets, if $f : A \rightarrow B$ is surjective, then $|A| \geq |B|$, as each element in B must be “reached” by at least one element from A .

Example 3.4.4

Show that $f(x) = \sin(x)$ with domain \mathbb{R} and codomain $[-1, 1]$ is surjective.

Solution: For a function to be surjective, every element in the codomain must be the image of at least one element in the domain.

For any $y \in [-1, 1]$, we need to find at least one $x \in \mathbb{R}$ such that $\sin(x) = y$.

Since the range of the sine function is precisely $[-1, 1]$, for any y in this interval, there exists some angle θ such that $\sin(\theta) = y$. More specifically, we can take $\theta = \arcsin(y)$,

which is well-defined for all $y \in [-1, 1]$.

Therefore, for every $y \in [-1, 1]$, there exists at least one $x \in \mathbb{R}$ (specifically, $x = \arcsin(y)$) such that $f(x) = \sin(x) = y$. This confirms that f is surjective.

One can similarly prove that $f(x) = x^3$ with domain and codomain \mathbb{R} is surjective.

Bijective (One-to-One and Onto) Functions

Definition:

A function $f : A \rightarrow B$ is bijective if it is both injective and surjective.

Bijective functions establish a perfect one-to-one correspondence between the domain and codomain; each element in the domain maps to exactly one element in the codomain, and every element in the codomain is mapped to by exactly one element in the domain. For finite sets, a bijection between sets A and B exists if and only if $|A| = |B|$.

Example 3.4.5

Prove that the function $f(x) = 3x + 2$ with domain and codomain \mathbb{R} is bijective.

Solution: To prove that f is bijective, we need to show that it is both injective and surjective.

First, let's prove that f is injective: Let $x_1, x_2 \in \mathbb{R}$ such that $f(x_1) = f(x_2)$. Then $3x_1 + 2 = 3x_2 + 2$ Subtracting 2 from both sides: $3x_1 = 3x_2$ Dividing both sides by 3: $x_1 = x_2$ Therefore, f is injective.

Now, let's prove that f is surjective: For any $y \in \mathbb{R}$, we need to find an $x \in \mathbb{R}$ such that $f(x) = y$. We need to solve: $3x + 2 = y$. Subtracting 2 from both sides: $3x = y - 2$ Dividing both sides by 3: $x = \frac{y-2}{3}$

For any given $y \in \mathbb{R}$, $x = \frac{y-2}{3} \in \mathbb{R}$ is a valid input, and $f(x) = 3(\frac{y-2}{3}) + 2 = y - 2 + 2 = y$. Therefore, f is surjective.

Since f is both injective and surjective, it is bijective.

Inverse Functions and Composition

One of the most important properties of bijective functions is that they possess an inverse. If $f : A \rightarrow B$ is bijective, then for each $y \in B$, there exists exactly one $x \in A$ such that $f(x) = y$. This allows us to define the inverse function $f^{-1} : B \rightarrow A$ by setting $f^{-1}(y) = x$ where x is the unique element of A satisfying $f(x) = y$.

The inverse function has the fundamental property that:

$$f^{-1}(f(x)) = x \quad \text{for all } x \in A \quad (3.4.6)$$

$$f(f^{-1}(y)) = y \quad \text{for all } y \in B \quad (3.4.7)$$

This means that applying f followed by f^{-1} (or vice versa) returns us to the original input. This operation where one function is applied after another leads to the concept of function composition.

Given two functions $f : A \rightarrow B$ and $g : B \rightarrow C$, their composition $(g \circ f) : A \rightarrow C$ is defined by:

$$(g \circ f)(x) = g(f(x)) \quad \text{for all } x \in A \quad (3.4.8)$$

Using this notation, the fundamental property of inverse functions can be expressed as:

$$f^{-1} \circ f = \text{id}_A \quad \text{and} \quad f \circ f^{-1} = \text{id}_B \quad (3.4.9)$$

where $\text{id}_A : A \rightarrow A$ and $\text{id}_B : B \rightarrow B$ are the identity functions on sets A and B respectively, defined by $\text{id}_A(x) = x$ for all $x \in A$ and $\text{id}_B(y) = y$ for all $y \in B$. The identity function simply returns its input unchanged.

Example 3.4.6

Given $f(x) = 3x - 2$, find $f^{-1}(x)$ and verify the inverse property by computing $(f^{-1} \circ f)(x)$ and $(f \circ f^{-1})(x)$.

Solution: To find the inverse of $f(x) = 3x - 2$, we solve for x in terms of y :

Let $y = f(x) = 3x - 2$

$$y = 3x - 2 \quad (3.4.10)$$

$$y + 2 = 3x \quad (3.4.11)$$

$$\frac{y+2}{3} = x \quad (3.4.12)$$

Therefore, $f^{-1}(y) = \frac{y+2}{3}$. Replacing y with x , we have $f^{-1}(x) = \frac{x+2}{3}$.

Now let's verify the inverse property:

$$(f^{-1} \circ f)(x) = f^{-1}(f(x)) \quad (3.4.13)$$

$$= f^{-1}(3x - 2) \quad (3.4.14)$$

$$= \frac{(3x - 2) + 2}{3} \quad (3.4.15)$$

$$= \frac{3x}{3} \quad (3.4.16)$$

$$= x \quad (3.4.17)$$

Similarly:

$$(f \circ f^{-1})(x) = f(f^{-1}(x)) \quad (3.4.18)$$

$$= f\left(\frac{x+2}{3}\right) \quad (3.4.19)$$

$$= 3\left(\frac{x+2}{3}\right) - 2 \quad (3.4.20)$$

$$= (x+2) - 2 \quad (3.4.21)$$

$$= x \quad (3.4.22)$$

This confirms that f^{-1} is indeed the inverse of f .

Function composition has several important properties:

- It is associative: $h \circ (g \circ f) = (h \circ g) \circ f$
- If f and g are both injective, then $g \circ f$ is injective
- If f and g are both surjective, then $g \circ f$ is surjective
- If f and g are both bijective, then $g \circ f$ is bijective and $(g \circ f)^{-1} = f^{-1} \circ g^{-1}$

These concepts of inverse functions and composition form a bridge between the structural properties of functions (injectivity, surjectivity, bijectivity) and their analytical properties such as monotonicity and convexity, which we will explore next.

3.4.3 Monotonic Functions

Another important classification of functions is based on their monotonicity - whether they consistently increase or decrease over their domain.

Definition: Monotonic Functions

Let $f : A \rightarrow B$ be a function where $A \subseteq \mathbb{R}$ and $B \subseteq \mathbb{R}$. Then:

- f is monotonically increasing (or non-decreasing) if for all $x_1, x_2 \in A$ with $x_1 \leq x_2$, we have $f(x_1) \leq f(x_2)$.
- f is strictly increasing if for all $x_1, x_2 \in A$ with $x_1 < x_2$, we have $f(x_1) < f(x_2)$.
- f is monotonically decreasing (or non-increasing) if for all $x_1, x_2 \in A$ with $x_1 \leq x_2$, we have $f(x_1) \geq f(x_2)$.
- f is strictly decreasing if for all $x_1, x_2 \in A$ with $x_1 < x_2$, we have $f(x_1) > f(x_2)$.

A function is monotonic if it is either monotonically increasing or monotonically decreasing.

The distinction between “monotonically increasing” and “strictly increasing” is important. A monotonically increasing function allows for flat sections where $f(x_1) = f(x_2)$ for some $x_1 \neq x_2$, while a strictly increasing function requires that the function value must increase as the input increases.

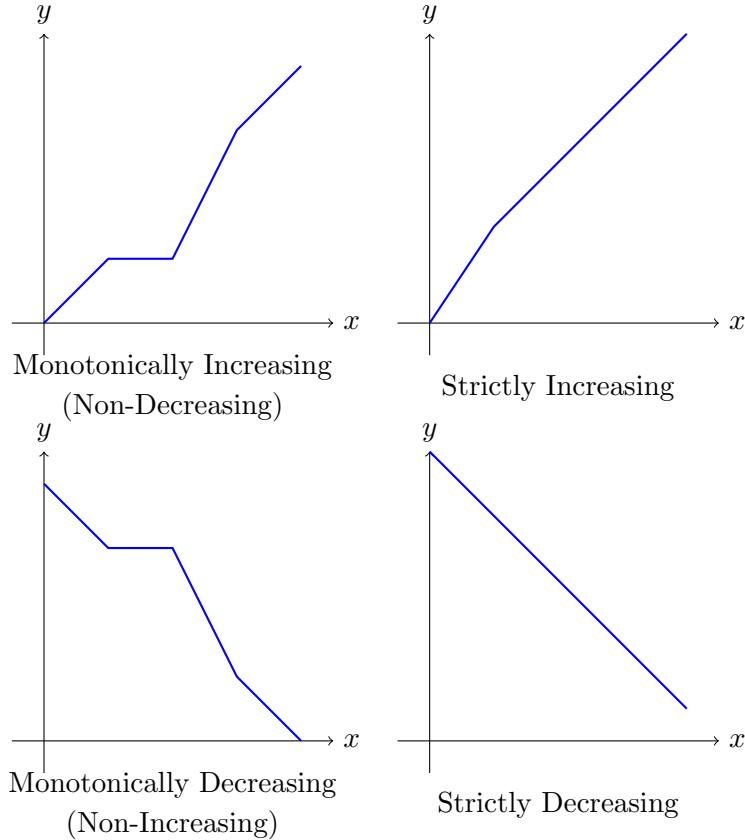


Figure 3.1: Examples of monotonic functions

Example 3.4.7

Prove that $f(x) = 3x + 2$ is strictly increasing on \mathbb{R} .

Solution: To prove that $f(x) = 3x + 2$ is strictly increasing, we need to show that for any $x_1, x_2 \in \mathbb{R}$ with $x_1 < x_2$, we have $f(x_1) < f(x_2)$.

Let $x_1, x_2 \in \mathbb{R}$ with $x_1 < x_2$. Then:

$$f(x_1) = 3x_1 + 2 \quad (3.4.23)$$

$$f(x_2) = 3x_2 + 2 \quad (3.4.24)$$

Taking the difference:

$$f(x_2) - f(x_1) = (3x_2 + 2) - (3x_1 + 2) \quad (3.4.25)$$

$$= 3x_2 - 3x_1 \quad (3.4.26)$$

$$= 3(x_2 - x_1) \quad (3.4.27)$$

Since $x_1 < x_2$, we have $x_2 - x_1 > 0$. Also, the constant 3 is positive. Therefore:

$$3(x_2 - x_1) > 0 \quad (3.4.28)$$

This means $f(x_2) - f(x_1) > 0$, which implies $f(x_2) > f(x_1)$.

Therefore, $f(x) = 3x + 2$ is strictly increasing on \mathbb{R} .

Example 3.4.8

Determine the monotonicity of $f(x) = x^2$ on different domains.

Solution: Let's examine the monotonicity of $f(x) = x^2$ on different intervals:

1. On the domain $[0, \infty)$:

For any $x_1, x_2 \in [0, \infty)$ with $x_1 < x_2$, we have:

$$f(x_1) = x_1^2 < x_2^2 = f(x_2) \quad (3.4.29)$$

This is true because if $0 \leq x_1 < x_2$, then $x_1^2 < x_2^2$. Therefore, $f(x) = x^2$ is strictly increasing on $[0, \infty)$.

2. On the domain $(-\infty, 0]$:

For any $x_1, x_2 \in (-\infty, 0]$ with $x_1 < x_2$, we have:

$$f(x_1) = x_1^2 > x_2^2 = f(x_2) \quad (3.4.30)$$

This holds because if $x_1 < x_2 \leq 0$, then $|x_1| > |x_2|$, and hence $x_1^2 > x_2^2$. Therefore, $f(x) = x^2$ is strictly decreasing on $(-\infty, 0]$.

3. On the entire real line \mathbb{R} :

From the above analysis, we see that $f(x) = x^2$ is not monotonic on \mathbb{R} since it decreases on $(-\infty, 0]$ and increases on $[0, \infty)$.

Relationship to Injectivity

Monotonicity has a direct relationship with injectivity. Specifically:

Theorem:

If a function $f : A \rightarrow B$ is strictly monotonic (either strictly increasing or strictly decreasing), then f is injective.

Proof

Let's prove this for a strictly increasing function f . The proof for strictly decreasing functions is similar.

Suppose f is strictly increasing, and consider any two distinct points $x_1, x_2 \in A$ with $x_1 \neq x_2$.

Without loss of generality, assume $x_1 < x_2$. Since f is strictly increasing, we have:

$$f(x_1) < f(x_2) \quad (3.4.31)$$

This means $f(x_1) \neq f(x_2)$, which is precisely the definition of injectivity.

Similarly, if f is strictly decreasing and $x_1 < x_2$, then $f(x_1) > f(x_2)$, again implying $f(x_1) \neq f(x_2)$.

Therefore, any strictly monotonic function is injective.

□

Monotonicity is a fundamental property of functions that plays a crucial role in various areas

of mathematics, including optimization, calculus, and as we'll see later, number theory. Understanding when a function consistently increases or decreases helps us predict its behavior and solve equations involving such functions efficiently.

3.4.4 Convex Functions

Moving beyond classification based on mapping properties, we now explore functions classified by their geometric behavior.

Definition:

A function $f : \mathbb{R} \rightarrow \mathbb{R}$ is convex if for any two points x_1, x_2 in its domain and any $t \in [0, 1]$, the following inequality holds:

$$f(tx_1 + (1 - t)x_2) \leq tf(x_1) + (1 - t)f(x_2) \quad (3.4.32)$$

Geometrically, this definition means that the line segment connecting any two points on the graph of the function lies above or on the graph itself. This captures the intuitive notion of a function that “curves upward.”

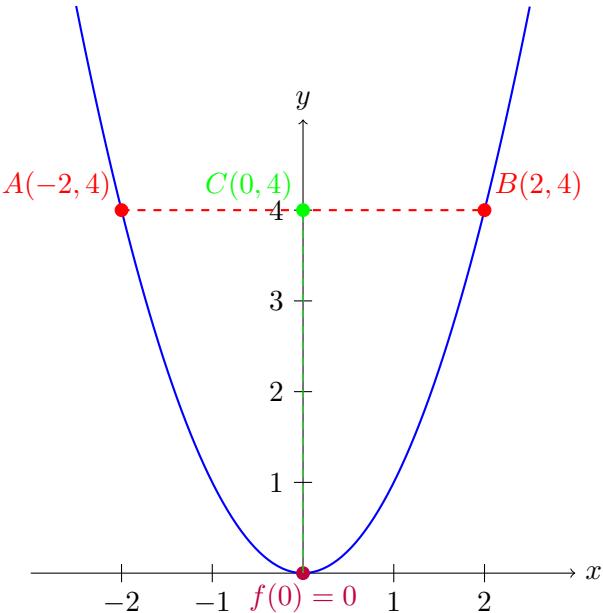


Figure 3.2: Illustration of convexity for $f(x) = x^2$. The line segment connecting points A and B lies above the curve.

The figure above illustrates the geometric interpretation of convexity. For the function $f(x) = x^2$, we can observe that the line segment (shown in red) connecting any two points on the curve lies entirely above the curve itself. This is particularly evident at the midpoint: when $x = 0$,

the point on the curve is at $(0, 0)$, while the midpoint of the line segment is at $(0, 4)$, showing a clear gap of 4 units.

This visual demonstration helps us understand intuitively what the mathematical definition of convexity means. The vertical green dashed line highlights this difference between the function value and the value on the line segment. For convex functions, this gap (or the distance between the curve and the line segment) is always non-negative, which gives convex functions their characteristic “cup-like” shape.

Now, let's verify this property rigorously using the mathematical definition of convexity.

Example 3.4.9

Prove that $f(x) = x^2$ is a convex function.

Solution: For any $x_1, x_2 \in \mathbb{R}$ and $t \in [0, 1]$, we need to show:

$$f(tx_1 + (1 - t)x_2) \leq tf(x_1) + (1 - t)f(x_2) \quad (3.4.33)$$

Let's compute the left-hand side:

$$f(tx_1 + (1 - t)x_2) = (tx_1 + (1 - t)x_2)^2 \quad (3.4.34)$$

$$= t^2x_1^2 + 2t(1 - t)x_1x_2 + (1 - t)^2x_2^2 \quad (3.4.35)$$

And the right-hand side:

$$tf(x_1) + (1 - t)f(x_2) = tx_1^2 + (1 - t)x_2^2 \quad (3.4.36)$$

To prove convexity, we need to show:

$$t^2x_1^2 + 2t(1 - t)x_1x_2 + (1 - t)^2x_2^2 \leq tx_1^2 + (1 - t)x_2^2 \quad (3.4.37)$$

Rearranging terms:

$$t^2x_1^2 - tx_1^2 + 2t(1 - t)x_1x_2 + (1 - t)^2x_2^2 - (1 - t)x_2^2 \leq 0 \quad (3.4.38)$$

$$t(t - 1)x_1^2 + 2t(1 - t)x_1x_2 + (1 - t)((1 - t) - 1)x_2^2 \leq 0 \quad (3.4.39)$$

$$t(t - 1)x_1^2 + 2t(1 - t)x_1x_2 + (1 - t)(-t)x_2^2 \leq 0 \quad (3.4.40)$$

$$t(t - 1)x_1^2 + 2t(1 - t)x_1x_2 - t(1 - t)x_2^2 \leq 0 \quad (3.4.41)$$

$$t(1 - t)(-x_1^2 + 2x_1x_2 - x_2^2) \leq 0 \quad (3.4.42)$$

$$-t(1 - t)(x_1 - x_2)^2 \leq 0 \quad (3.4.43)$$

Since $t \in [0, 1]$, we have $t(1 - t) \geq 0$ and $(x_1 - x_2)^2 \geq 0$, so:

$$-t(1 - t)(x_1 - x_2)^2 \leq 0 \quad (3.4.44)$$

Thus, the inequality holds, and $f(x) = x^2$ is convex.

Strictly Convex Functions

A function f is strictly convex if for any two distinct points x_1, x_2 in its domain and any $t \in (0, 1)$, the following strict inequality holds:

$$f(tx_1 + (1 - t)x_2) < tf(x_1) + (1 - t)f(x_2) \quad (3.4.45)$$

The key difference is the strict inequality, which ensures that the line segment connecting any two points on the graph lies strictly above (not on) the graph, except at the endpoints.

Concave Functions

A function f is concave if $-f$ is convex. Equivalently, for any x_1, x_2 in the domain and any $t \in [0, 1]$:

$$f(tx_1 + (1 - t)x_2) \geq tf(x_1) + (1 - t)f(x_2) \quad (3.4.46)$$

Geometrically, a concave function “curves downward,” and the line segment connecting any two points on its graph lies below or on the graph.

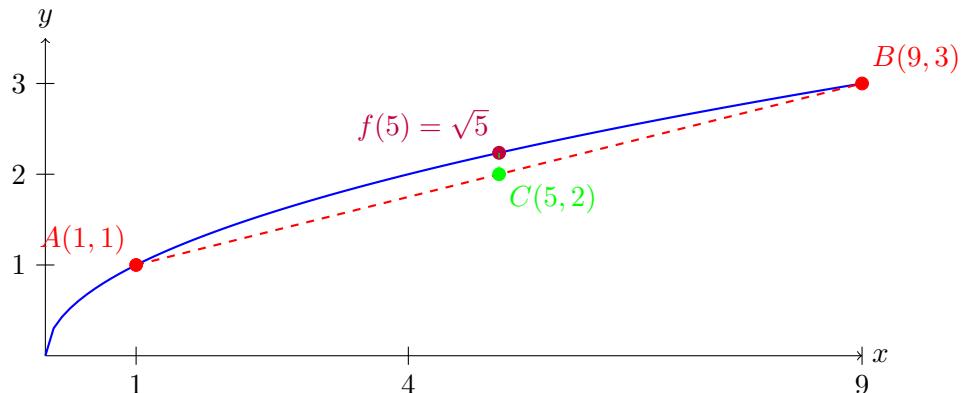


Figure 3.3: Illustration of concavity for $f(x) = \sqrt{x}$. The line segment connecting points A and B lies below the curve.

The figure above illustrates the geometric interpretation of concavity using the square root function $f(x) = \sqrt{x}$. Unlike convex functions where line segments lie above the curve, for concave functions like \sqrt{x} , the line segment connecting any two points on the curve lies entirely below the curve (or on it at the endpoints).

In the illustration, we've chosen two points on the square root curve: $A(1, 1)$ where $\sqrt{1} = 1$, and $B(9, 3)$ where $\sqrt{9} = 3$. The red dashed line represents the line segment connecting these points. The midpoint of this line segment is at $C(5, 2)$, while the actual function value at $x = 5$ is $\sqrt{5} \approx 2.236$. The vertical green dashed line highlights this difference, showing that the function value is greater than the corresponding point on the line segment.

This visualization demonstrates the essential property of concave functions: when we take a weighted average of inputs (in this case, the midpoint between 1 and 9), the function value at this average is greater than or equal to the weighted average of the function values (the height of the line segment at that point).

Example 3.4.10

Prove that $f(x) = \sqrt{x}$ is a concave function.

Solution: To prove that $f(x) = \sqrt{x}$ is concave, we need to verify that for any $x_1, x_2 > 0$ (since the domain of \sqrt{x} is $[0, \infty)$) and any $t \in [0, 1]$, the following inequality holds:

$$f(tx_1 + (1 - t)x_2) \geq tf(x_1) + (1 - t)f(x_2) \quad (3.4.47)$$

Substituting $f(x) = \sqrt{x}$:

$$\sqrt{tx_1 + (1 - t)x_2} \geq t\sqrt{x_1} + (1 - t)\sqrt{x_2} \quad (3.4.48)$$

We can approach this proof in two ways:

Method 1 (Using the definition directly): Let's square both sides (which won't change the direction of the inequality since both sides are non-negative):

$$tx_1 + (1 - t)x_2 \geq (t\sqrt{x_1} + (1 - t)\sqrt{x_2})^2 \quad (3.4.49)$$

$$= t^2x_1 + 2t(1 - t)\sqrt{x_1x_2} + (1 - t)^2x_2 \quad (3.4.50)$$

Rearranging:

$$tx_1 + (1 - t)x_2 - t^2x_1 - (1 - t)^2x_2 \geq 2t(1 - t)\sqrt{x_1x_2} \quad (3.4.51)$$

$$t(1 - t)x_1 + t(1 - t)x_2 \geq 2t(1 - t)\sqrt{x_1x_2} \quad (3.4.52)$$

$$t(1 - t)(x_1 + x_2) \geq 2t(1 - t)\sqrt{x_1x_2} \quad (3.4.53)$$

If $t = 0$ or $t = 1$, both sides equal 0, and the inequality holds. For $t \in (0, 1)$, we can divide by $t(1 - t)$:

$$x_1 + x_2 \geq 2\sqrt{x_1x_2} \quad (3.4.54)$$

This is the AM-GM inequality for two numbers, which we know is true. Therefore, $f(x) = \sqrt{x}$ is concave.

In our exploration of mathematical functions, we've covered a wide range of concepts that are fundamental to mathematics in general and number theory in particular. The concepts and examples presented in this section will serve as valuable tools as we progress further into number

theory.

3.5 Polynomials and Their Roots

Having established a solid understanding of functions and their properties, we now turn our attention to one of the most important classes of functions in mathematics: polynomials. Polynomials represent a natural bridge between our study of general functions and the more specific realm of number theory.

A polynomial with real coefficients is an expression of the form:

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

where $a_n, a_{n-1}, \dots, a_1, a_0$ are real numbers and $a_n \neq 0$. The highest power n is called the degree of the polynomial, represented as $\deg(P)$. For instance, $(x - 3)$ is a polynomial of degree 1.

Definition: Monic Polynomial

A polynomial is called monic if the coefficient of its highest-degree term (the leading coefficient) is equal to 1.

For example, the polynomial $P(x) = x^3 + 2x^2 - 5x + 7$ is monic because the coefficient of x^3 (the highest-degree term) is 1.

Monic polynomials hold special significance in mathematics as they provide a canonical form for polynomials. Any non-zero polynomial can be converted to monic form by dividing all terms by the leading coefficient, which preserves the roots while simplifying theoretical analysis. This standardization is particularly valuable when studying polynomial properties and their related theorems.

Many of the properties of mathematical functions we just examined can be directly applied to the study of polynomials. For instance, a polynomial of odd degree is always surjective over the real numbers, while a polynomial of degree greater than one is never injective over the complex numbers (a result connected to the Fundamental Theorem of Algebra, which we will explore).

3.5.1 Polynomial Division

Just as the division of integers produces quotients and remainders, polynomial division follows a similar principle, allowing us to express one polynomial in terms of another through quotients and remainders.

Example 3.5.1

Find the remainder when $P(x) = x^3 - 2x^2 + 4x - 7$ is divided by $(x - 3)$.

Solution:

$$\begin{array}{r} x^2 + x + 7 \\ x - 3 \overline{) x^3 - 2x^2 + 4x - 7} \\ \underline{x^3 - 3x^2} \\ \quad + x^2 + 4x - 7 \\ \underline{\quad + x^2 - 3x} \\ \quad + 7x - 7 \\ \underline{\quad + 7x - 21} \\ \quad + 14 \end{array}$$

The quotient is $x^2 + x + 7$ and the remainder is 14.

It is important to note that the degree of a polynomial remainder is always less than the degree of the divisor. For instance, if a linear polynomial (polynomials of the form $(x - a)$ with $\deg = 1$) divides another polynomial, the remainder will always be a scalar.

Even though the remainder calculation in the previous example is correct, the method is somewhat slow and cumbersome. Let's look at a more elegant manner of calculating polynomial remainders.

Theorem: Remainder Theorem

If a polynomial $P(x)$ is divided by $(x - a)$, then the remainder equals $P(a)$.

Proof

When we divide a polynomial $P(x)$ by $(x - a)$, we can express it in the form:

$$P(x) = Q(x)(x - a) + R$$

where $Q(x)$ is the quotient polynomial and R is the remainder.

Since we're dividing by a linear term $(x - a)$, the remainder R must be a constant (a polynomial of degree 0). This is because the degree of the remainder must be less than the degree of the divisor, which is 1 for $(x - a)$.

Now, let's evaluate this equation at $x = a$:

$$P(a) = Q(a)(a - a) + R \quad (3.5.1)$$

$$= Q(a) \cdot 0 + R \quad (3.5.2)$$

$$= R \quad (3.5.3)$$

Therefore, $P(a) = R$, which means the remainder when $P(x)$ is divided by $(x - a)$ equals $P(a)$.

□

Let us illustrate the Remainder Theorem via an example.

Example 3.5.2

Find the remainder when $P(x) = x^3 - 2x^2 + 4x - 7$ is divided by $(x - 3)$.

Solution: According to the Remainder Theorem, the remainder equals $P(3)$.

$$P(3) = 3^3 - 2(3)^2 + 4(3) - 7 \quad (3.5.4)$$

$$= 27 - 2(9) + 4(3) - 7 \quad (3.5.5)$$

$$= 27 - 18 + 12 - 7 \quad (3.5.6)$$

$$= 14 \quad (3.5.7)$$

The remainder is 14, which is the same as our earlier computation.

If we subtract 14 from $P(x) = x^3 - 2x^2 + 4x - 7$, we get another polynomial $Q(x) = x^3 - 2x^2 + 4x - 21$. It follows from the previous discussion that $Q(x) = (x^2 + x + 7)(x - 3)$.

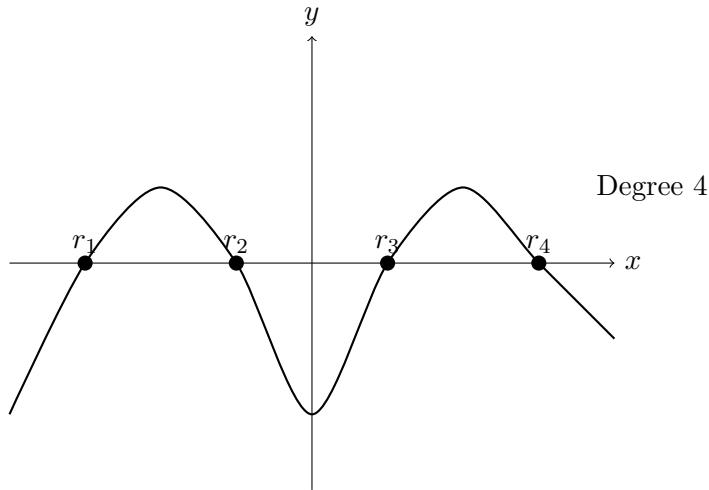
3.5.2 Polynomial Roots

From the remainder theorem, if a linear polynomial $(x - a)$ divides an another polynomial $P(x)$ of higher degree, then the remainder equals $P(a)$. What if the remainder is zero? In the special case, a is a root of $P(x)$, and the linear polynomial $(x - a)$ is a factor of $P(x)$.

Definition: Root of a Polynomial

A root (or zero) of a polynomial $P(x)$ is a value r such that $P(r) = 0$.

For instance, $Q(3) = (x^2 + x + 7)(3 - 3) = 0$ implies that 3 is a root of the polynomial $Q(x) = x^3 - 2x^2 + 4x - 21$. The roots of a polynomial correspond to the x-intercepts of its graph. This gives us a visual way to understand the behavior of polynomial functions:



In the diagram above, the polynomial crosses the x-axis at four places, indicating that it has four real roots.

Definition: Factor of a Polynomial

A polynomial $Q(x)$ is a factor of polynomial $P(x)$ if there exists another polynomial $R(x)$ such that $P(x) = Q(x) \cdot R(x)$. In particular, if r is a root of $P(x)$, then $(x - r)$ is a factor of $P(x)$.

From the definition, it follows that $(x - 3)$ is a factor of the polynomial $Q(x) = x^3 - 2x^2 + 4x - 21$. Because $(x - 3)$ is of degree 1, $(x - 3)$ is a linear factor of the polynomial $Q(x)$. Similarly, $x^2 + x + 7$ is a quadratic (polynomial of degree 2) factor of $Q(x)$.

Every polynomial has 1 and itself as its factors. Our discussion brings us to the factor theorem.

Theorem: Factor Theorem

$(x - a)$ is a factor of polynomial $P(x)$ if and only if $P(a) = 0$, i.e., a is a root of $P(x)$.

As discussed in previous chapters, the phrase “if and only if” (often abbreviated as “iff”) indicates a biconditional relationship between two statements. This means both statements are equivalent. In this theorem, it establishes that:

- If $(x - a)$ is a factor of $P(x)$, then $P(a) = 0$ (“only if” statement)

- If $P(a) = 0$, then $(x - a)$ is a factor of $P(x)$ (“if” statement)

This powerful logical connection tells us that we can conclude if $(x - a)$ is a factor of a polynomial simply by evaluating the polynomial at $x = a$, without needing to perform polynomial division. Conversely, knowing that a is a root of the polynomial immediately tells us that $(x - a)$ must be a factor.

Proof

From the Remainder Theorem, we know that when $P(x)$ is divided by $(x - a)$, the remainder is $P(a)$.

So we can write:

$$P(x) = Q(x)(x - a) + P(a)$$

Now:

- *If $P(a) = 0$, then $P(x) = Q(x)(x - a)$, which means $(x - a)$ is a factor of $P(x)$.*
- *Conversely, if $(x - a)$ is a factor of $P(x)$, then $P(x) = Q(x)(x - a)$ for some polynomial $Q(x)$, which means the remainder $P(a) = 0$.*

Therefore, $(x - a)$ is a factor of $P(x)$ if and only if $P(a) = 0$.

□

The two theorems gives us powerful tools to factorize polynomials. While Remainder Theorem allows us to find remainders without performing long divisions, Factor Theorem helps us find roots and factors of polynomials.

Definition: Multiplicity of a Root

The multiplicity of a root r of a polynomial $P(x)$ is the highest power m such that $(x - r)^m$ divides $P(x)$. A root with multiplicity greater than 1 is called a multiple root.

Example 3.5.3

Find all roots of $P(x) = x^3 - 4x^2 + 5x - 2$ and determine their multiplicities.

Solution: Let's try some values to find a root. After a few rounds of hit and trial, we

see that:

$$P(1) = 1 - 4 + 5 - 2 = 0 \quad (3.5.8)$$

So $x = 1$ is a root. We can factor out $(x - 1)$:

$$P(x) = (x - 1)(x^2 - 3x + 2) \quad (3.5.9)$$

We can further factor the quadratic term:

$$x^2 - 3x + 2 = (x - 1)(x - 2) \quad (3.5.10)$$

Therefore:

$$P(x) = (x - 1)(x - 1)(x - 2) \quad (3.5.11)$$

$$= (x - 1)^2(x - 2) \quad (3.5.12)$$

The roots are:

- $x = 1$ with multiplicity 2
- $x = 2$ with multiplicity 1

The total number of roots counting multiplicities is 3, which matches the degree of the polynomial.

In the previous example, we identified $x = 1$ as a root of $P(x) = x^3 - 4x^2 + 5x - 2$ via hit and trial, which is not an efficient process. The rational root theorem helps us find potential rational roots, when working with polynomials that have integer coefficients.

Theorem: Rational Root Theorem

If $P(x) = a_n x^n + a_{n-1} x^{n-1} + \cdots + a_1 x + a_0$ is a polynomial with integer coefficients, then any rational root $\frac{p}{q}$ (in lowest terms) must satisfy:

- p divides a_0 (the constant term)
- q divides a_n (the leading coefficient)

A formal proof of the rational root theorem exceeds this book's scope, but we can appreciate its practical value. This theorem efficiently identifies rational roots of polynomials with integer coefficients by limiting the number of candidates to check. Let's explore an example to demonstrate the theorem's effectiveness in action.

Example 3.5.4

Find all rational roots of $P(x) = 2x^3 - 7x^2 - 3x + 18$.

Solution: Using the Rational Root Theorem:

- Possible values for p (factors of 18): $\pm 1, \pm 2, \pm 3, \pm 6, \pm 9, \pm 18$
- Possible values for q (factors of 2): $\pm 1, \pm 2$

Possible rational roots: $\pm 1, \pm 2, \pm 3, \pm 6, \pm 9, \pm 18, \pm \frac{1}{2}, \pm \frac{3}{2}, \pm \frac{9}{2}$

Let's test some of these values:

$$P(1) = 2(1)^3 - 7(1)^2 - 3(1) + 18 \quad (3.5.13)$$

$$= 2 - 7 - 3 + 18 \quad (3.5.14)$$

$$= 10 \quad (3.5.15)$$

Since $P(1) \neq 0$, $x = 1$ is not a root.

$$P(2) = 2(2)^3 - 7(2)^2 - 3(2) + 18 \quad (3.5.16)$$

$$= 2(8) - 7(4) - 3(2) + 18 \quad (3.5.17)$$

$$= 16 - 28 - 6 + 18 \quad (3.5.18)$$

$$= 0 \quad (3.5.19)$$

So $x = 2$ is a root. We can factor out $(x - 2)$:

$$P(x) = (x - 2)(2x^2 - 3x - 9)$$

For the quadratic factor, we can use the quadratic formula:

$$x = \frac{3 \pm \sqrt{9 + 72}}{4} \quad (3.5.20)$$

$$= \frac{3 \pm \sqrt{81}}{4} \quad (3.5.21)$$

$$= \frac{3 \pm 9}{4} \quad (3.5.22)$$

This gives $x = 3$ and $x = -\frac{3}{2}$. Therefore, the roots of $P(x)$ are $x = 2$, $x = 3$, and $x = -\frac{3}{2}$.

3.5.3 The Fundamental Theorem of Algebra

The Fundamental Theorem of Algebra, first proven by Carl Friedrich Gauss in 1799, states that every non-constant polynomial with complex coefficients has at least one complex root. This means that a polynomial of degree n has exactly n complex roots when counted with multiplicity. This powerful result connects algebra with complex analysis and ensures that every polynomial can be completely factored into linear terms over the complex numbers, allowing mathematicians to understand the behavior of polynomials through their roots.

Theorem: Fundamental Theorem of Algebra

Every polynomial equation of degree $n \geq 1$ with complex coefficients has exactly n complex roots when counted with multiplicity.

This profound result guarantees that any polynomial $p(x) = a_nx^n + a_{n-1}x^{n-1} + \cdots + a_1x + a_0$, where $a_n \neq 0$ and the coefficients are complex numbers, can be completely factored as:

$$p(x) = a_n(x - r_1)(x - r_2) \cdots (x - r_n)$$

where r_1, r_2, \dots, r_n are the roots of the polynomial.

For polynomials with real coefficients, a key corollary emerges: while such polynomials still have exactly n roots, these roots may include both real and complex values. Since real numbers are special cases of complex numbers (with zero imaginary part), the Fundamental Theorem applies seamlessly. Consequently, a polynomial of degree n with real coefficients can have at most n real roots, counting multiplicities.

Theorem: Number of Real Roots

A polynomial of degree n with real coefficients can have at most n real roots, counting multiplicities.

This theorem tells us that a polynomial with real coefficients might have fewer real roots than its degree. We have previously worked with the polynomial $Q(x) = x^3 - 2x^2 + 4x - 21$. We have shown that $Q(x) = (x^2 + x + 7)(x - 3)$. As the quadratic polynomial $x^2 + x + 7$ has no real roots, $Q(x)$ has only one real root. The two complex roots of $Q(x)$ are $-\frac{1}{2} + \frac{3\sqrt{3}}{2}i$ and $-\frac{1}{2} - \frac{3\sqrt{3}}{2}i$. Notice that complex roots of polynomials with real coefficients always appear in conjugate pairs.

Definition: Reducibility of Polynomials

A polynomial $P(x)$ with real coefficients is said to be reducible over \mathbb{R} if it can be expressed as a product of two polynomials $Q(x)$ and $R(x)$, both with real coefficients and both of degree at least 1. That is, $P(x) = Q(x) \cdot R(x)$ where $\deg(Q) \geq 1$ and $\deg(R) \geq 1$. If a polynomial cannot be expressed as such a product, it is called irreducible over \mathbb{R} .

The polynomial $Q(x) = x^3 - 2x^2 + 4x - 21$ is reducible over \mathbb{R} and equates to $Q(x) = (x^2 + x + 7)(x - 3)$. However, the polynomial factor $x^2 + x + 7$ is irreducible over \mathbb{R} , as it doesn't have any real roots.

3.5.4 The Complete Factorization Theorem**Theorem: Complete Factorization Theorem**

Every polynomial $p(x)$ of degree $n \geq 1$ with complex coefficients can be expressed as:

$$p(x) = a(x - r_1)^{m_1}(x - r_2)^{m_2} \cdots (x - r_k)^{m_k}$$

where a is a non-zero constant, r_1, r_2, \dots, r_k are distinct roots of $p(x)$, and m_1, m_2, \dots, m_k are their respective multiplicities, with $m_1 + m_2 + \cdots + m_k = n$.

This theorem extends the Fundamental Theorem of Algebra by explicitly accounting for repeated roots. It guarantees that any polynomial with complex coefficients can be completely decomposed into linear factors, each corresponding to a root of the polynomial.

When working with polynomials that have real coefficients, we can express the factorization in a form that highlights the distinction between real and complex roots:

Theorem: Complete Factorization over the Reals

Every polynomial $P(x)$ of degree $n \geq 1$ with real coefficients can be factored as:

$$P(x) = a(x - r_1)^{m_1}(x - r_2)^{m_2} \cdots (x - r_k)^{m_k}[Q_1(x)]^{n_1}[Q_2(x)]^{n_2} \cdots [Q_j(x)]^{n_j}$$

where:

- a is a non-zero real constant
- r_1, r_2, \dots, r_k are the real roots with multiplicities m_1, m_2, \dots, m_k
- $Q_1(x), Q_2(x), \dots, Q_j(x)$ are irreducible quadratic polynomials of the form $x^2 + bx + c$ where $b^2 < 4c$
- $m_1 + m_2 + \cdots + m_k + 2(n_1 + n_2 + \cdots + n_j) = n$

The elegant connection between these two theorems lies in how complex roots manifest in polynomials with real coefficients. Complex roots always appear in conjugate pairs—if $a + bi$ is a root, then $a - bi$ must also be a root. When we multiply the corresponding linear factors:

$$(x - (a + bi))(x - (a - bi)) = x^2 - 2ax + (a^2 + b^2)$$

This product creates an irreducible quadratic polynomial with real coefficients. The condition $b^2 < 4c$ ensures that the quadratic has no real roots.

Thus, the real factorization theorem reflects the fundamental structure of polynomials with real coefficients: they can be completely factored into a product of terms representing real roots (linear factors) and complex conjugate pairs (irreducible quadratic factors).

Example 3.5.5

Consider the polynomial $P(x) = x^4 - 1$.

Over the complex numbers, it factors completely into linear terms:

$$P(x) = (x - 1)(x + 1)(x - i)(x + i)$$

Over the real numbers, we get:

$$P(x) = (x - 1)(x + 1)(x^2 + 1)$$

Notice that the pair of complex roots i and $-i$ combines to form the irreducible quadratic factor $x^2 + 1$, which cannot be factored further using only real numbers.

Example 3.5.6

Factor $P(x) = x^5 - 3x^4 + 4x^3 - 4x^2 + 3x - 1$ over the real numbers, given that $x = 1$ is a root of multiplicity 2.

Solution: Since $x = 1$ is a root of multiplicity 2, we know $(x - 1)^2$ is a factor.

$$P(x) = (x - 1)^2 \cdot Q(x)$$

Using polynomial division, we get, $Q(x) = x^3 - x^2 + x - 1$. Let's check if this cubic has any rational roots. Using the Rational Root Theorem, potential roots include ± 1 .

Testing $x = 1$:

$$(1)^3 - (1)^2 + (1) - 1 = 1 - 1 + 1 - 1 = 0$$

Testing $x = -1$:

$$(-1)^3 - (-1)^2 + (-1) - 1 = -1 - 1 - 1 - 1 = -4$$

Therefore:

$$\begin{aligned} P(x) &= (x - 1)^2 \cdot (x - 1)(x^2 + 1) \\ &= (x - 1)^3(x^2 + 1) \end{aligned}$$

The complete factorization over the real numbers is:

$$P(x) = (x - 1)^3(x^2 + 1)$$

The five roots are:

- $x = 1$ with multiplicity 3
- $x = i$ and $x = -i$ (appearing as the irreducible quadratic factor $x^2 + 1$)

3.5.5 Vieta's Formulas

Vieta's formulas, named after François Viète (1540-1603), establish elegant relationships between the coefficients of a polynomial and the elementary symmetric functions of its roots. These formulas provide powerful insights into polynomial structure without requiring explicit knowledge of the roots themselves.

Consider a monic polynomial of degree n :

$$P(x) = x^n + a_{n-1}x^{n-1} + a_{n-2}x^{n-2} + \cdots + a_1x + a_0$$

By the Fundamental Theorem of Algebra, we know that $P(x)$ has exactly n complex roots (counting multiplicities). If we denote these roots as r_1, r_2, \dots, r_n , then the polynomial can be factored as:

$$P(x) = (x - r_1)(x - r_2) \cdots (x - r_n)$$

When we expand this product and compare coefficients with the original form of $P(x)$, we obtain Vieta's formulas, which establish relationships between the coefficients of the polynomial and the elementary symmetric functions of its roots - whether or not we know the explicit values of those roots.

$$r_1 + r_2 + \cdots + r_n = -a_{n-1} \quad (3.5.23)$$

$$r_1 r_2 + r_1 r_3 + \cdots + r_{n-1} r_n = a_{n-2} \quad (3.5.24)$$

$$r_1 r_2 r_3 + \cdots = -a_{n-3} \quad (3.5.25)$$

$$\vdots \quad (3.5.26)$$

$$r_1 r_2 \cdots r_n = (-1)^n a_0 \quad (3.5.27)$$

In simpler terms:

- The sum of all roots equals $-a_{n-1}$ (the negative of the coefficient of x^{n-1})
- The sum of products of roots taken two at a time equals a_{n-2} (the coefficient of x^{n-2})
- The sum of products of roots taken three at a time equals $-a_{n-3}$ (the negative of the coefficient of x^{n-3})
- This pattern continues, alternating signs
- The product of all roots equals $(-1)^n a_0$ (either a_0 or $-a_0$ depending on whether n is even or odd)

For non-monic polynomials, we can divide throughout by the leading coefficient to apply these formulas.

Example 3.5.7

If r and s are the roots of the quadratic equation $x^2 + 5x + 6 = 0$, find:

- (a) $r + s$
- (b) rs
- (c) $r^2 + s^2$
- (d) $\frac{1}{r} + \frac{1}{s}$

Solution: For a quadratic equation $ax^2 + bx + c = 0$ with roots r and s , Vieta's formulas give:

$$r + s = -\frac{b}{a} \quad (3.5.28)$$

$$rs = \frac{c}{a} \quad (3.5.29)$$

For the equation $x^2 + 5x + 6 = 0$, we have $a = 1$, $b = 5$, and $c = 6$.

- (a) $r + s = -\frac{b}{a} = -\frac{5}{1} = -5$
- (b) $rs = \frac{c}{a} = \frac{6}{1} = 6$

(c) To find $r^2 + s^2$, we can use the identity:

$$\begin{aligned}(r+s)^2 &= r^2 + 2rs + s^2 \\ \Rightarrow r^2 + s^2 &= (r+s)^2 - 2rs \\ &= (-5)^2 - 2(6) \\ &= 25 - 12 \\ &= 13\end{aligned}$$

(d) To find $\frac{1}{r} + \frac{1}{s}$, we can use:

$$\begin{aligned}\frac{1}{r} + \frac{1}{s} &= \frac{s+r}{rs} \\ &= \frac{-5}{6} \\ &= -\frac{5}{6}\end{aligned}$$

Example 3.5.8

If $P(x) = x^3 - 6x^2 + 11x - 6$ has roots r , s , and t , find the value of $r^2s + rs^2 + s^2t + st^2 + t^2r + tr^2$.

Solution: From Vieta's formulas, for the cubic polynomial $P(x) = x^3 - 6x^2 + 11x - 6$, we have:

$$\begin{aligned}r + s + t &= 6 \\ rs + st + tr &= 11 \\ rst &= 6\end{aligned}$$

We need to find $r^2s + rs^2 + s^2t + st^2 + t^2r + tr^2$. Let's manipulate this expression:

$$\begin{aligned}r^2s + rs^2 + s^2t + st^2 + t^2r + tr^2 &= rs(r+s) + st(s+t) + tr(t+r) \\ &= rs(r+s+t-t) + st(r+s+t-r) + tr(r+s+t-s) \\ &= rs(r+s+t) - rst + st(r+s+t) - rst + tr(r+s+t) - rst \\ &= (rs+st+tr)(r+s+t) - 3rst \\ &= 11 \cdot 6 - 3 \cdot 6 \\ &= 66 - 18 \\ &= 48\end{aligned}$$

Example 3.5.9

If the polynomial $P(x) = x^4 + px^3 + qx^2 + rx + s$ has roots 2, -1, 3, and -2, find the values of p , q , r , and s .

Solution: Let's denote the roots as $r_1 = 2$, $r_2 = -1$, $r_3 = 3$, and $r_4 = -2$.

Using Vieta's formulas:

$$r_1 + r_2 + r_3 + r_4 = -p \quad (3.5.30)$$

$$2 + (-1) + 3 + (-2) = -p \quad (3.5.31)$$

$$2 = -p \quad (3.5.32)$$

$$\Rightarrow p = -2 \quad (3.5.33)$$

For the coefficient q :

$$r_1r_2 + r_1r_3 + r_1r_4 + r_2r_3 + r_2r_4 + r_3r_4 = q \quad (3.5.34)$$

$$2(-1) + 2(3) + 2(-2) + (-1)(3) + (-1)(-2) + 3(-2) = q \quad (3.5.35)$$

$$-2 + 6 - 4 - 3 + 2 - 6 = q \quad (3.5.36)$$

$$-7 = q \quad (3.5.37)$$

For the coefficient r :

$$r_1r_2r_3 + r_1r_2r_4 + r_1r_3r_4 + r_2r_3r_4 = -r \quad (3.5.38)$$

$$2(-1)(3) + 2(-1)(-2) + 2(3)(-2) + (-1)(3)(-2) = -r \quad (3.5.39)$$

$$-6 + 4 - 12 + 6 = -r \quad (3.5.40)$$

$$-8 = -r \quad (3.5.41)$$

$$\Rightarrow r = 8 \quad (3.5.42)$$

For the coefficient s :

$$r_1r_2r_3r_4 = s \quad (3.5.43)$$

$$2(-1)(3)(-2) = s \quad (3.5.44)$$

$$12 = s \quad (3.5.45)$$

Therefore, $P(x) = x^4 - 2x^3 - 7x^2 + 8x + 12$.

We can verify this by factoring $P(x) = (x - 2)(x + 1)(x - 3)(x + 2)$.

Vieta's formulas are particularly useful in solving problems where the coefficients of a polynomial are known, but the roots are not, or vice versa. They provide a bridge between the algebraic

structure of polynomials and their geometric interpretation as points on the complex plane.

3.6 Binomial Theorem

The binomial theorem is one of the most elegant and useful results in algebra, allowing us to expand expressions of the form $(x + y)^n$ without tedious multiplication. Beyond its computational value, this theorem reveals beautiful patterns in combinatorics and forms the foundation for many advanced mathematical concepts. We begin by understanding the building blocks of this theorem: the binomial coefficients.

3.6.1 Binomial Coefficients

Theorem: Binomial Coefficient

For non-negative integers n and k with $k \leq n$, the binomial coefficient $\binom{n}{k}$ (read as “ n choose k ”) represents the number of ways to select k objects from a set of n distinct objects, where the order of selection does not matter. The binomial coefficient is defined as:

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}$$

where $n!$ denotes the factorial of n .

Special cases:

- $\binom{n}{0} = \binom{n}{n} = 1$ for any $n \geq 0$
- $\binom{n}{k} = 0$ for $k > n$
- $\binom{n}{k} = \binom{n}{n-k}$ (symmetry property)

The symmetry property is particularly interesting, as it tells us that choosing k objects from n is equivalent to choosing the $n - k$ objects to leave out. This insight becomes valuable when calculating binomial coefficients in practice—always compute the smaller of k or $n - k$ to minimize calculations.

With these coefficients in hand, we can now explore the central result—the binomial expansion.

3.6.2 The Binomial Expansion

Theorem: Binomial Theorem

For any real numbers x and y and a non-negative integer n :

$$(x + y)^n = \sum_{k=0}^n \binom{n}{k} x^{n-k} y^k$$

where $\binom{n}{k}$ is the binomial coefficient.

This powerful formula tells us exactly how to expand $(x + y)^n$ for any non-negative integer n . Each term in the expansion has a specific pattern: it consists of a binomial coefficient multiplied by powers of x and y , where the exponents sum to n . Let's see this in action with some examples.

Example 3.6.1

Use the Binomial Theorem to expand $(1 + x)^4$.

Solution: Using the formula with $n = 4$:

$$\begin{aligned} (1 + x)^4 &= \sum_{k=0}^4 \binom{4}{k} 1^{4-k} x^k \\ &= \binom{4}{0} 1^4 x^0 + \binom{4}{1} 1^3 x^1 + \binom{4}{2} 1^2 x^2 + \binom{4}{3} 1^1 x^3 + \binom{4}{4} 1^0 x^4 \\ &= 1 + 4x + 6x^2 + 4x^3 + x^4 \end{aligned}$$

Therefore, $(1 + x)^4 = 1 + 4x + 6x^2 + 4x^3 + x^4$.

Notice how the coefficients in this expansion (1, 4, 6, 4, 1) display the symmetry we mentioned earlier. This pattern appears in many contexts throughout mathematics.

The binomial theorem isn't just useful for simple expressions like $(1 + x)^n$. It can handle more complex cases as well.

Example 3.6.2

Find the coefficient of x^3y^2 in the expansion of $(2x + y)^5$.

Solution: In the expansion $(2x + y)^5 = \sum_{k=0}^5 \binom{5}{k} (2x)^{5-k} y^k$, the term involving x^3y^2

corresponds to $k = 2$ (the power of y).

For this term, we need $(2x)^{5-2} = (2x)^3 = 8x^3$

The coefficient of x^3y^2 is:

$$\begin{aligned}\binom{5}{2} \cdot 8 &= 10 \cdot 8 \\ &= 80\end{aligned}$$

Therefore, the coefficient of x^3y^2 in the expansion of $(2x + y)^5$ is 80.

Beyond just calculating expansions, the binomial theorem leads to several important results and identities that have wide-ranging applications in combinatorics, probability, and other areas of mathematics.

3.6.3 Important Related Results

The binomial theorem yields several elegant identities that reveal deeper patterns in combinatorial mathematics. We will explore several of the most important ones here.

Theorem: Sum of Binomial Coefficients

For any non-negative integer n :

$$\sum_{k=0}^n \binom{n}{k} = 2^n$$

Proof

This follows directly from the Binomial Theorem. Setting $x = y = 1$, we get:

$$(1+1)^n = \sum_{k=0}^n \binom{n}{k} 1^{n-k} 1^k \tag{3.6.1}$$

$$2^n = \sum_{k=0}^n \binom{n}{k} \tag{3.6.2}$$

□

The sum of binomial coefficients identity has a beautiful combinatorial interpretation: 2^n represents the number of possible subsets of a set with n elements, which matches the sum of the number of ways to choose $0, 1, 2, \dots, n$ elements from the set. It is important to realize that

given any set of size n , the number of ways to select even and odd elements from the set remains the same, irrespective of if n itself is even or not.

Theorem:

For any positive integer n , we have

$$\sum_{\substack{k=0 \\ k \text{ even}}}^n \binom{n}{k} = \sum_{\substack{k=0 \\ k \text{ odd}}}^n \binom{n}{k} = 2^{n-1}$$

That is, the sum of binomial coefficients with even indices equals the sum of binomial coefficients with odd indices, and both equal 2^{n-1} .

Proof

Let us define:

$$S_{even} = \sum_{\substack{k=0 \\ k \text{ even}}}^n \binom{n}{k}$$

and

$$S_{odd} = \sum_{\substack{k=0 \\ k \text{ odd}}}^n \binom{n}{k}$$

We know that $(1+x)^n = \sum_{k=0}^n \binom{n}{k} x^k$. Setting $x = 1$, we obtain sum of binomial coefficients identity:

$$(1+1)^n = 2^n = \sum_{k=0}^n \binom{n}{k} = S_{even} + S_{odd} \quad (3.6.3)$$

Similarly, for $x = -1$, we have:

$$(1-1)^n = 0 = \sum_{k=0}^n \binom{n}{k} (-1)^k = \sum_{\substack{k=0 \\ k \text{ even}}}^n \binom{n}{k} - \sum_{\substack{k=0 \\ k \text{ odd}}}^n \binom{n}{k} = S_{even} - S_{odd} \quad (3.6.4)$$

Adding the last two equations, we have $2S_{even} = 2^n$, which implies $S_{even} = 2^{n-1}$. On subtraction, we get $2S_{odd} = 2^n$, which leads to $S_{odd} = 2^{n-1}$.

Therefore, both sums equal 2^{n-1} , confirming that binomial coefficients with even and odd indices contribute equally to the total sum 2^n .

□

Next, we explore Pascal's Identity, which forms the foundation of the famous Pascal's Triangle. To motivate this identity, let us consider the set $S = \{1, 2, 3, \dots, n+1\}$ and ask: in how many ways can we select a k -element subset from S ?

The answer is clearly $\binom{n+1}{k}$. However, we can partition all possible k -element subsets of S based on whether they contain the element $n+1$ or not:

- **Case 1:** Subsets that contain $n+1$. To form such a subset, we must choose the remaining $k-1$ elements from $\{1, 2, \dots, n\}$. There are $\binom{n}{k-1}$ such subsets.
- **Case 2:** Subsets that do not contain $n+1$. These are simply k -element subsets of $\{1, 2, \dots, n\}$. There are $\binom{n}{k}$ such subsets.

Since every k -element subset of S falls into exactly one of these two disjoint cases, we have:

$$\binom{n+1}{k} = \binom{n}{k-1} + \binom{n}{k}$$

This set-theoretic argument shows that Pascal's Identity is a natural consequence of how we can systematically count subsets by partitioning them based on the presence or absence of a distinguished element.

Theorem: Pascal's Identity

For any non-negative integers n and k where $0 \leq k \leq n+1$:

$$\binom{n+1}{k} = \binom{n}{k-1} + \binom{n}{k}$$

Proof

Consider the binomial expansion of $(1+x)^{n+1}$:

$$(1+x)^{n+1} = (1+x)(1+x)^n$$

Using the Binomial Theorem on both sides:

$$\begin{aligned} \sum_{k=0}^{n+1} \binom{n+1}{k} x^k &= (1+x) \sum_{k=0}^n \binom{n}{k} x^k \\ &= \sum_{k=0}^n \binom{n}{k} x^k + \sum_{k=0}^n \binom{n}{k} x^{k+1} \end{aligned}$$

$$\begin{aligned}
&= \binom{n}{0} + \sum_{k=1}^n \binom{n}{k} x^k + \sum_{k=1}^{n+1} \binom{n}{k-1} x^k \\
&= \binom{n}{0} + \sum_{k=1}^n (\binom{n}{k} + \binom{n}{k-1}) x^k + \binom{n}{n} x^{n+1}
\end{aligned}$$

Comparing coefficients of x^k on both sides:

1. For $k = 0$:

$$\binom{n+1}{0} = \binom{n}{0} = 1$$

2. For $1 \leq k \leq n$:

$$\binom{n+1}{k} = \binom{n}{k} + \binom{n}{k-1}$$

3. For $k = n+1$:

$$\binom{n+1}{n+1} = \binom{n}{n} = 1$$

Thus, we have proved Pascal's Identity for all cases where $0 \leq k \leq n+1$.

□

Finally, we examine a remarkable result that reveals deep connections between binomial coefficients and provides a powerful tool for combinatorial counting.

Let us ask a question: if we have two disjoint sets of sizes m and n , in how many ways can we choose exactly r objects from their union?

The direct answer is $\binom{m+n}{r}$. However, we can also approach this systematically by considering how many objects we select from each set. If we choose k objects from the first set (of size m), then we must choose exactly $r-k$ objects from the second set (of size n). The number of ways to do this is $\binom{m}{k} \cdot \binom{n}{r-k}$.

Since k can range from 0 to r (subject to the constraints that $k \leq m$ and $r-k \leq n$), we obtain:

$$\binom{m+n}{r} = \sum_{k=0}^r \binom{m}{k} \binom{n}{r-k}$$

This elegant identity, known as Vandermonde's convolution or the Chu-Vandermonde identity, demonstrates how the “global” counting problem $\binom{m+n}{r}$ can be decomposed into a sum of “local” products $\binom{m}{k} \binom{n}{r-k}$.

Theorem: Chu-Vandermonde Identity

For any non-negative integers m , n , and r :

$$\binom{m+n}{r} = \sum_{k=0}^r \binom{m}{k} \binom{n}{r-k}$$

Proof

Consider the expansion of $(1+x)^{m+n}$ using the Binomial Theorem:

$$(1+x)^{m+n} = \sum_{r=0}^{m+n} \binom{m+n}{r} x^r \quad (1)$$

Now, let's consider $(1+x)^m \cdot (1+x)^n$. By the Binomial Theorem, we have:

$$(1+x)^m = \sum_{i=0}^m \binom{m}{i} x^i \quad \text{and} \quad (1+x)^n = \sum_{j=0}^n \binom{n}{j} x^j$$

Multiplying these expansions:

$$\begin{aligned} (1+x)^m \cdot (1+x)^n &= \left(\sum_{i=0}^m \binom{m}{i} x^i \right) \cdot \left(\sum_{j=0}^n \binom{n}{j} x^j \right) \\ &= \sum_{r=0}^{m+n} \left(\sum_{k=0}^r \binom{m}{k} \binom{n}{r-k} \right) x^r \quad (2) \end{aligned}$$

The last step comes from collecting terms with the same power of x . Here, $r = i + j$ and $k = i$.

Since $(1+x)^{m+n} = (1+x)^m \cdot (1+x)^n$, the right-hand sides of equations (1) and (2) must be equal. Comparing coefficients of x^r on both sides:

$$\binom{m+n}{r} = \sum_{k=0}^r \binom{m}{k} \binom{n}{r-k}$$

Thus, we have proved Vandermonde's Identity using the Binomial Theorem.

□

The identities we have established reveal the elegant interconnectedness of binomial coefficients.

In fact, Pascal's identity can be derived as a special case of Vandermonde's identity by setting $m = 1$ and $n = p$:

$$\binom{1+p}{r} = \sum_{k=0}^r \binom{1}{k} \binom{p}{r-k}$$

Since $\binom{1}{k} = 0$ for $k > 1$, and $\binom{1}{0} = \binom{1}{1} = 1$, this simplifies to:

$$\binom{p+1}{r} = \binom{1}{0} \binom{p}{r} + \binom{1}{1} \binom{p}{r-1} = \binom{p}{r} + \binom{p}{r-1}$$

which is exactly Pascal's identity. This connection highlights the beautiful hierarchy of relationships in combinatorial mathematics, with simpler patterns emerging as special cases of more general principles.

With these foundational mathematical tools now established, from the Pigeonhole Principle and mathematical induction to functions, polynomials, and the binomial theorem, we are well-equipped to begin our exploration of number theory in the following chapters. These concepts will serve as the essential building blocks for understanding the patterns that define the fascinating world of numbers.

3.7 Practice Exercises

3.8 Practice Exercises

Exercise 3.1

Let $P(z) = z^3 - 3z^2 + 4z - 2$ be a polynomial with complex coefficients. If one of the roots of $P(z)$ is $1 + i$, determine all roots of $P(z)$ and express the polynomial in factored form.

Exercise 3.2

If $P(x) = x^3 - 6x^2 + 11x - 6$, find $P(x + 1)$ and use this to find all the roots of $P(x)$.

Exercise 3.3

Prove by induction that $1^2 + 2^2 + \dots + n^2 = \frac{n(n+1)(2n+1)}{6}$ for all natural numbers n .

Exercise 3.4

Prove that if $n^2 + 1$ points are placed inside a square with side length n , then there exist at least two points whose distance is less than or equal to $\sqrt{2}$.

Exercise 3.5

Prove that $\sum_{k=0}^n \binom{n}{k}^2 = \binom{2n}{n}$.

Exercise 3.6

Prove that $\binom{2n}{n} \leq \frac{4^n}{\sqrt{\pi n}}$ for all positive integers n .

Exercise 3.7

If $P(x) = x^3 + ax^2 + bx + c$ has three real roots α , β , and γ such that $\alpha + \beta + \gamma = 0$ and $\alpha\beta + \beta\gamma + \gamma\alpha = -3$, find the value of $\alpha\beta\gamma$.

Exercise 3.8

Find the coefficient of x^{10} in the expansion of $(1 + x + x^2)^{10}$.

Exercise 3.9

Prove that among any set of six integers, there are two whose sum or difference is divisible by 8.

Exercise 3.10

Let $S = \{z \in \mathbb{C} : |z| = 1\}$ be the set of complex numbers with modulus 1. Prove that for any $n + 1$ distinct points z_1, z_2, \dots, z_{n+1} on S , there must exist at least one pair z_i, z_j with $i \neq j$ such that $|z_i - z_j| < 2 \sin \frac{\pi}{n}$.

Part II

Introduction to Number Theory

Chapter 4

Divisibility



Euclid of Alexandria (c. 300 BCE) Euclid, often called the "Father of Geometry," was an ancient Greek mathematician whose textbook *Elements* became so influential that it was the standard math textbook for over 2,000 years. While we know almost nothing about Euclid's personal life, his work changed mathematics forever. He transformed geometry from practical rules into a beautiful system of proofs and theorems, introducing the axiomatic method that forms the backbone of modern mathematics. In Books VII-IX of *Elements*, Euclid tackled number theory, formalizing ideas about divisibility that earlier civilizations like the Babylonians and Egyptians had used mainly as calculation tools. His algorithm for finding the greatest common divisor (GCD) remains one of the oldest continuously used algorithms in mathematics—over 2,300 years old and still running on your computers today!

According to a famous story, when Egypt's King Ptolemy I asked Euclid for an easier way to learn mathematics, Euclid boldly replied, "There is no royal road to geometry," essentially telling the king that even rulers can't take shortcuts in learning math—something your teachers might agree with!

The concept of divisibility is perhaps the most fundamental idea in number theory. This chapter explores the elegant patterns that emerge when we study how integers divide each other.

The basic properties of divisibility will lead us to the greatest common divisor (GCD). The GCD represents the largest integer that divides two numbers perfectly, a concept with wide ranging applications. We will discuss Euclidean algorithm for computing GCDs, a method so elegant that it has remained essentially unchanged since Euclid formalized it over 2,300 years ago.

We'll also encounter Bézout's identity, which reveals the surprising fact that GCDs can always be expressed as linear combinations of the original numbers. This profound result connects divisibility to the broader landscape of number theory and lays groundwork for modular arithmetic, which we'll explore in later chapters.

4.1 Divisibility

Definition: Divisibility

An integer a is said to divide another integer b , denoted by $a \mid b$, if there exists an integer k such that $b = ak$. In this case, we say " a divides b " or " a is a divisor of b " or " b is a multiple of a ".

If a does not divide b , we write $a \nmid b$.

For example, $3 \mid 12$ because $12 = 3 \cdot 4$. However, $5 \nmid 12$ because there is no integer k such that $12 = 5k$.

4.1.1 Basic Properties of Divisibility

Theorem: Properties of Divisibility

Let a , b , and c be integers. Then:

1. $a \mid a$ (reflexive property)
2. If $a \mid b$ and $b \mid c$, then $a \mid c$ (transitive property)
3. If $a \mid b$ and $a \mid c$, then $a \mid (bx + cy)$ for any integers x and y (linear combination property)
4. If $a \mid b$ and $b \mid a$, then $a = \pm b$
5. If $a \mid b$ and $b \neq 0$, then $|a| \leq |b|$

Proof

We'll prove each property:

1. For the reflexive property: $a = a \cdot 1$, so $a | a$.
2. For the transitive property: If $a | b$ and $b | c$, then $b = ak$ and $c = bm$ for some integers k and m . Substituting, we get $c = akm$, which means $a | c$.
3. For the linear combination property: If $a | b$ and $a | c$, then $b = ar$ and $c = as$ for some integers r and s . Consider $bx + cy = arx + asy = a(rx + sy)$. Since $(rx + sy)$ is an integer, $a | (bx + cy)$.
4. If $a | b$ and $b | a$, then $b = ak$ and $a = bm$ for some integers k and m . Substituting, we get $a = akm$, which gives $a(1 - km) = 0$. Since $a \neq 0$ (because $a | b$), we must have $km = 1$. For integers, this means $k = m = 1$ or $k = m = -1$, which implies $a = \pm b$.
5. If $a | b$ and $b \neq 0$, then $b = ak$ for some integer $k \neq 0$. Taking absolute values, $|b| = |a| \cdot |k|$. Since $|k| \geq 1$ for non-zero integers, we have $|b| \geq |a|$.

□

4.1.2 Division Algorithm

The division algorithm is a fundamental result that formalizes the process of division with remainder.

Theorem: Division Algorithm

Given integers a and b with $b > 0$, there exist unique integers q (quotient) and r (remainder) such that:

$$a = bq + r, \quad \text{where } 0 \leq r < b$$

Proof

Let's construct the quotient q directly. For any integer a and positive integer b , we define:

$$q = \lfloor \frac{a}{b} \rfloor$$

where $\lfloor x \rfloor$ denotes the greatest integer less than or equal to x .

Now, let's define $r = a - bq$. We need to show that $0 \leq r < b$ and that q and r are unique.

First, let's verify that $0 \leq r < b$:

$$q = \lfloor \frac{a}{b} \rfloor \Rightarrow q \leq \frac{a}{b} < q + 1 \quad (4.1.1)$$

Multiplying throughout by b (which is positive):

$$bq \leq a < b(q + 1) = bq + b \quad (4.1.2)$$

Subtracting bq from all parts:

$$0 \leq a - bq < b \quad (4.1.3)$$

Since $r = a - bq$, we have $0 \leq r < b$, as required.

For uniqueness, suppose there are two pairs (q_1, r_1) and (q_2, r_2) such that:

$$a = bq_1 + r_1 \quad \text{and} \quad a = bq_2 + r_2$$

$$\text{where } 0 \leq r_1 < b \text{ and } 0 \leq r_2 < b$$

Subtracting, we get $b(q_1 - q_2) = r_2 - r_1$.

If $q_1 \neq q_2$, then $|b(q_1 - q_2)| \geq b$ (since $b > 0$ and q_1, q_2 are integers).

However, since $0 \leq r_1, r_2 < b$, we have $|r_2 - r_1| < b$.

This contradiction means $q_1 = q_2$, which then implies $r_1 = r_2$.

Therefore, the integers q and r are uniquely determined.

□

Example 4.1.1

Find the quotient and remainder when -17 is divided by 5 .

Solution: Using the division algorithm, we need to find integers q and r such that $-17 = 5q + r$ with $0 \leq r < 5$.

One approach is to first find the largest multiple of 5 less than or equal to -17 :

$$-17 = -20 + 3 \quad (4.1.4)$$

$$= 5(-4) + 3 \quad (4.1.5)$$

So the quotient is $q = -4$ and the remainder is $r = 3$.

Note that we cannot have $q = -3$ and $r = -2$, even though $-17 = 5(-3) - 2$, because the remainder must be non-negative.

4.2 Greatest Common Divisor

4.2.1 Greatest Common Divisor (GCD)

Definition: Greatest Common Divisor (GCD)

The greatest common divisor of two integers a and b , not both zero, denoted by $\gcd(a, b)$, is the largest positive integer that divides both a and b .

Formally, $d = \gcd(a, b)$ if and only if:

1. $d | a$ and $d | b$ (i.e., d is a common divisor of a and b)
2. For any integer c , if $c | a$ and $c | b$, then $c \leq d$ (i.e., d is the greatest of all common divisors)

Based on the definition, $\gcd(12, 18) = 6$ because 6 is the largest positive integer that divides both 12 and 18. Let us have a look at the properties of the greatest common divisor function.

Properties of Greatest Common Divisor

Theorem: Properties of GCD

For integers a , b , and c :

1. $\gcd(0, 0)$ is undefined
2. $\gcd(a, b) = \gcd(b, a)$ (commutative property)
3. $\gcd(a, b) = \gcd(a, -b) = \gcd(-a, b) = \gcd(-a, -b)$
4. $\gcd(a, 0) = |a|$ for $a \neq 0$
5. $\gcd(a, b) = \gcd(a, b - qa)$ for any integer q

6. If $c \mid a$ and $c \mid b$, then $c \mid \gcd(a, b)$
7. If $a \mid bc$ and $\gcd(a, b) = 1$, then $a \mid c$

The case $\gcd(0, 0)$ is undefined because every integer divides 0. Since there is no largest positive integer, we cannot identify a “greatest” common divisor. This is why our definition of GCD requires that the integers not both be zero.

The property $\gcd(a, b) = \gcd(a, b - qa)$ is particularly important as it forms the basis of the Euclidean algorithm. Let us prove this result and take the rest as an exercise.

Theorem: GCD Linear Combination Property

For any integers a , b , and q : $\gcd(a, b) = \gcd(a, b - qa)$

Proof

We need to show that the set of common divisors of a and b is identical to the set of common divisors of a and $b - qa$.

First, let d be any common divisor of a and b . Then $d \mid a$ and $d \mid b$.

Since $d \mid a$, we have $a = dk$ for some integer k . Since $d \mid b$, we have $b = dm$ for some integer m .

Now consider $b - qa$:

$$b - qa = dm - q(dk) \quad (4.2.1)$$

$$= dm - dqk \quad (4.2.2)$$

$$= d(m - qk) \quad (4.2.3)$$

Since $m - qk$ is an integer, we have $d \mid (b - qa)$.

Therefore, d is a common divisor of a and $b - qa$.

Conversely, let d be any common divisor of a and $b - qa$. Then $d \mid a$ and $d \mid (b - qa)$.

Since $d \mid a$, we have $a = dp$ for some integer p . Since $d \mid (b - qa)$, we have $b - qa = dn$ for some integer n .

Now consider b :

$$b = (b - qa) + qa \quad (4.2.4)$$

$$= dn + q(dp) \quad (4.2.5)$$

$$= dn + dpq \quad (4.2.6)$$

$$= d(n + pq) \quad (4.2.7)$$

Since $n + pq$ is an integer, we have $d \mid b$.

Therefore, d is a common divisor of a and b .

Since the common divisors of a and b are exactly the same as the common divisors of a and $b - qa$, we conclude that $\gcd(a, b) = \gcd(a, b - qa)$.

□

Algorithm: Euclidean Algorithm

To find $\gcd(a, b)$ for integers a and b (not both zero):

Step 1: If $b = 0$, then $\gcd(a, b) = |a|$ and the algorithm stops.

Step 2: Otherwise, divide a by b to get quotient q and remainder r such that $a = bq + r$ with $0 \leq r < |b|$.

Step 3: Replace (a, b) with (b, r) and return to Step 1.

The Euclidean algorithm works by repeatedly applying the division algorithm and using the property that $\gcd(a, b) = \gcd(b, r)$ where r is the remainder when a is divided by b .

Example 4.2.1

Find $\gcd(48, 18)$ using the Euclidean algorithm.

Solution: We apply the Euclidean algorithm step by step:

$$48 = 18 \cdot 2 + 12 \quad (4.2.8)$$

$$18 = 12 \cdot 1 + 6 \quad (4.2.9)$$

$$12 = 6 \cdot 2 + 0 \quad (4.2.10)$$

Since the remainder is now 0, the GCD is the last non-zero remainder, which is 6.

Therefore, $\gcd(48, 18) = 6$.

4.2.2 Bézout's Identity

A remarkable property of the GCD is that it can be expressed as a linear combination of the original numbers.

Theorem: Bézout's Identity

If a and b are integers, not both zero, then there exist integers x and y such that:

$$\gcd(a, b) = ax + by$$

Proof

We will prove this using the Euclidean algorithm. Let's apply the algorithm to find $\gcd(a, b)$, tracking how each remainder can be expressed as a linear combination of a and b .

When we apply the Euclidean algorithm to find $\gcd(a, b)$, we get a sequence of equations:

$$a = bq_1 + r_1 \quad \text{where } 0 < r_1 < |b| \quad (4.2.11)$$

$$b = r_1q_2 + r_2 \quad \text{where } 0 < r_2 < r_1 \quad (4.2.12)$$

$$r_1 = r_2q_3 + r_3 \quad \text{where } 0 < r_3 < r_2 \quad (4.2.13)$$

$$\vdots \quad (4.2.14)$$

$$r_{n-2} = r_{n-1}q_n + r_n \quad \text{where } 0 < r_n < r_{n-1} \quad (4.2.15)$$

$$r_{n-1} = r_nq_{n+1} + 0 \quad (4.2.16)$$

The GCD is r_n , the last non-zero remainder.

Now, we can rewrite each remainder as a linear combination of a and b . Starting from the first equation:

$$r_1 = a - bq_1 \quad (4.2.17)$$

This expresses r_1 as a linear combination of a and b with coefficients $s_1 = 1$ and $t_1 = -q_1$, so $r_1 = s_1a + t_1b$.

From the second equation:

$$r_2 = b - r_1q_2 \quad (4.2.18)$$

$$= b - (a - bq_1)q_2 \quad (4.2.19)$$

$$= b - aq_2 + bq_1q_2 \quad (4.2.20)$$

$$= b(1 + q_1q_2) - aq_2 \quad (4.2.21)$$

So $r_2 = s_2a + t_2b$ where $s_2 = -q_2$ and $t_2 = 1 + q_1q_2$.

From the third equation:

$$r_3 = r_1 - r_2q_3 \quad (4.2.22)$$

$$= (s_1a + t_1b) - (s_2a + t_2b)q_3 \quad (4.2.23)$$

$$= s_1a + t_1b - s_2q_3a - t_2q_3b \quad (4.2.24)$$

$$= a(s_1 - s_2q_3) + b(t_1 - t_2q_3) \quad (4.2.25)$$

So $r_3 = s_3a + t_3b$ where $s_3 = s_1 - s_2q_3$ and $t_3 = t_1 - t_2q_3$.

Continuing this process for each remainder, we can express each r_i as $r_i = s_i a + t_i b$ for some integers s_i and t_i .

Eventually, we reach the last non-zero remainder $r_n = s_n a + t_n b$, which is the GCD.

Therefore, $\gcd(a, b) = r_n = s_n a + t_n b$, which proves Bézout's Identity with $x = s_n$ and $y = t_n$.

□

The proof given above establishes the existence of Bézout coefficients. Let's demonstrate the backward substitution approach with a concrete example.

Example 4.2.2

Find integers x and y such that $\gcd(48, 18) = 48x + 18y$ using the backward substitution method.

Solution: We have already computed that $\gcd(48, 18) = 6$ using the Euclidean algorithm with the following equations:

$$48 = 18 \cdot 2 + 12 \quad (1)$$

$$18 = 12 \cdot 1 + 6 \quad (2)$$

$$12 = 6 \cdot 2 + 0 \quad (3)$$

Now we work backward to express 6 as a linear combination of 48 and 18.

Step 1: From equation (2), express the GCD in terms of the previous remainders:

$$6 = 18 - 12 \cdot 1 \quad (4)$$

Step 2: From equation (1), express 12 in terms of 48 and 18:

$$12 = 48 - 18 \cdot 2 \quad (5)$$

Step 3: Substitute equation (5) into equation (4):

$$6 = 18 - (48 - 18 \cdot 2) \cdot 1 \quad (4.2.26)$$

$$= 18 - 48 + 18 \cdot 2 \quad (4.2.27)$$

$$= 18 + 18 \cdot 2 - 48 \quad (4.2.28)$$

$$= 18(1 + 2) - 48 \cdot 1 \quad (4.2.29)$$

$$= 18 \cdot 3 - 48 \cdot 1 \quad (4.2.30)$$

$$= 48 \cdot (-1) + 18 \cdot 3 \quad (4.2.31)$$

Therefore, $x = -1$ and $y = 3$, giving us:

$$\gcd(48, 18) = 6 = 48 \cdot (-1) + 18 \cdot 3$$

As we saw in the example above, finding the Bézout coefficients via backward substitution requires us to:

1. First compute the GCD using the standard Euclidean algorithm and store all intermediate equations
2. Then work backward through these equations, making successive substitutions
3. Keep track of increasingly complex algebraic manipulations as we substitute one remainder into another

This backward substitution process becomes lengthy and error-prone, especially for large numbers where the Euclidean algorithm may require many steps. Each substitution step introduces more terms and coefficients to manage, making it easy to make arithmetic errors.

This motivates the development of a more efficient algorithm that computes both the GCD and the Bézout coefficients simultaneously in a forward manner, eliminating the need for storing intermediate steps and performing backward substitution.

Algorithm: Extended Euclidean Algorithm

To find integers x and y such that $\gcd(a, b) = ax + by$ for integers a and b (not both zero):

Initialization:

- Set $r_1 = a$, $r_2 = b$

- Set $x_1 = 1, x_2 = 0$ (coefficients of a)
- Set $y_1 = 0, y_2 = 1$ (coefficients of b)

Main Loop: While $r_2 \neq 0$, do:

1. Compute quotient $q = \lfloor r_1/r_2 \rfloor$ and remainder $r = r_1 - qr_2$
2. Compute new coefficient $x = x_1 - qx_2$
3. Compute new coefficient $y = y_1 - qy_2$
4. Update: $(r_1, x_1, y_1) \leftarrow (r_2, x_2, y_2)$
5. Update: $(r_2, x_2, y_2) \leftarrow (r, x, y)$

Output:

- The GCD is r_1
- The coefficients are $x = x_1$ and $y = y_1$, satisfying $ax + by = \gcd(a, b)$

The Extended Euclidean Algorithm is significantly more efficient than the method described in the proof of Bézout's Identity because it computes both the GCD and the Bézout coefficients simultaneously in a forward manner. The algorithm also eliminates the need to store all intermediate steps for backward substitution and reduces the computational complexity by maintaining and updating the coefficients at each step.

These efficiency advantages make the Extended Euclidean Algorithm the preferred method in practice, especially for large numbers where manual computation of the backward substitution would be impractical. Let's understand why this algorithm works. At each step, we ensure that:

$$r_1 = ax_1 + by_1 \quad (4.2.32)$$

$$r_2 = ax_2 + by_2 \quad (4.2.33)$$

When we compute $r = r_1 - qr_2$, we get:

$$r = r_1 - qr_2 \quad (4.2.34)$$

$$= (ax_1 + by_1) - q(ax_2 + by_2) \quad (4.2.35)$$

$$= a(x_1 - qx_2) + b(y_1 - qy_2) \quad (4.2.36)$$

$$= ax + by \quad (4.2.37)$$

This shows that the new remainder r can be expressed as $ax + by$ where $x = x_1 - qx_2$ and $y = y_1 - qy_2$.

By updating the values as specified in the algorithm, we maintain the invariant relationships until $r_2 = 0$, at which point r_1 is the GCD and (x_1, y_1) are the desired coefficients.

Example 4.2.3

Find integers x and y such that $\gcd(48, 18) = 48x + 18y$.

Solution: We apply the Extended Euclidean Algorithm:

Initialization: $r_1 = 48, r_2 = 18, x_1 = 1, y_1 = 0, x_2 = 0, y_2 = 1$.

First iteration:

$$q = \lfloor 48/18 \rfloor = 2 \quad (4.2.38)$$

$$r = 48 - 18 \cdot 2 = 12 \quad (4.2.39)$$

$$x = x_1 - qx_2 = 1 - 2 \cdot 0 = 1 \quad (4.2.40)$$

$$y = y_1 - qy_2 = 0 - 2 \cdot 1 = -2 \quad (4.2.41)$$

Update: $(r_1, x_1, y_1) = (18, 0, 1)$ and $(r_2, x_2, y_2) = (12, 1, -2)$.

Second iteration:

$$q = \lfloor 18/12 \rfloor = 1 \quad (4.2.42)$$

$$r = 18 - 12 \cdot 1 = 6 \quad (4.2.43)$$

$$x = x_1 - qx_2 = 0 - 1 \cdot 1 = -1 \quad (4.2.44)$$

$$y = y_1 - qy_2 = 1 - 1 \cdot (-2) = 3 \quad (4.2.45)$$

Update: $(r_1, x_1, y_1) = (12, 1, -2)$ and $(r_2, x_2, y_2) = (6, -1, 3)$.

Third iteration:

$$q = \lfloor 12/6 \rfloor = 2 \quad (4.2.46)$$

$$r = 12 - 6 \cdot 2 = 0 \quad (4.2.47)$$

$$x = x_1 - qx_2 = 1 - 2 \cdot (-1) = 3 \quad (4.2.48)$$

$$y = y_1 - qy_2 = -1 - 2 \cdot 3 = -8 \quad (4.2.49)$$

Update: $(r_1, x_1, y_1) = (6, -1, 3)$ and $(r_2, x_2, y_2) = (0, 3, -8)$.

Since $r_2 = 0$, the algorithm terminates.

Therefore, $\gcd(48, 18) = 6 = 48 \cdot (-1) + 18 \cdot 3$.

We can verify: $48 \cdot (-1) + 18 \cdot 3 = -48 + 54 = 6$.

This example clearly demonstrates the efficiency of the Extended Euclidean Algorithm compared to the backward substitution method. We directly computed both the GCD and the Bézout coefficients in just three forward iterations, without having to backtrack through the calculations.

4.3 Least Common Multiple

Definition: Least Common Multiple (LCM)

The least common multiple of two integers a and b , not both zero, denoted by $\text{lcm}(a, b)$, is the smallest positive integer that is divisible by both a and b .

For example, $\text{lcm}(12, 18) = 36$ because 36 is the smallest positive integer that is divisible by both 12 and 18.

Theorem: Relationship between GCD and LCM

For integers a and b , not both zero:

$$\gcd(a, b) \cdot \text{lcm}(a, b) = |ab|$$

Proof

Let $d = \gcd(a, b)$ and $l = \text{lcm}(a, b)$.

We can write $a = da'$ and $b = db'$, where $\gcd(a', b') = 1$.

Since l is divisible by both a and b , we have $l = ak = bm$ for some positive integers k and m .

This gives us:

$$l = da'k = db'm$$

Dividing by d :

$$\frac{l}{d} = a'k = b'm$$

Since $\gcd(a', b') = 1$, we know $a' \mid m$ and $b' \mid k$.

Let $m = a't$ and $k = b't$ for some positive integer t .

Then:

$$\frac{l}{d} = a'b't$$

The smallest positive value for t is 1, giving:

$$\frac{l}{d} = a'b' = \frac{a}{d} \cdot \frac{b}{d} = \frac{ab}{d^2}$$

Rearranging:

$$ld = ab$$

Taking absolute values:

$$\gcd(a, b) \cdot \text{lcm}(a, b) = |ab|$$

□

This relationship provides a simple way to compute the LCM once the GCD is known:

$$\text{lcm}(a, b) = \frac{|ab|}{\gcd(a, b)}$$

Example 4.3.1

Find $\text{lcm}(48, 18)$ using the GCD-LCM relationship.

Solution: We previously computed $\gcd(48, 18) = 6$.

Using the relationship:

$$\text{lcm}(48, 18) = \frac{|48 \cdot 18|}{\gcd(48, 18)} = \frac{864}{6} = 144$$

Therefore, $\text{lcm}(48, 18) = 144$.

Example 4.3.2

Find the number of positive integers less than or equal to 1000 that are not divisible by 5, 7, or 11.

Solution: We'll use the inclusion-exclusion principle. Let:

- A = set of integers ≤ 1000 divisible by 5
- B = set of integers ≤ 1000 divisible by 7
- C = set of integers ≤ 1000 divisible by 11

We want to find $|U \setminus (A \cup B \cup C)|$ where $U = \{1, 2, 3, \dots, 1000\}$.

By inclusion-exclusion:

$$|A \cup B \cup C| = |A| + |B| + |C| - |A \cap B| - |A \cap C| - |B \cap C| + |A \cap B \cap C|$$

Computing each term:

$$|A| = \lfloor 1000/5 \rfloor = 200 \quad (4.3.1)$$

$$|B| = \lfloor 1000/7 \rfloor = 142 \quad (4.3.2)$$

$$|C| = \lfloor 1000/11 \rfloor = 90 \quad (4.3.3)$$

For intersections, we need the LCM of pairs. Since 5, 7, and 11 are pairwise coprime:

$$|A \cap B| = \lfloor 1000/\text{lcm}(5, 7) \rfloor = \lfloor 1000/35 \rfloor = 28 \quad (4.3.4)$$

$$|A \cap C| = \lfloor 1000/\text{lcm}(5, 11) \rfloor = \lfloor 1000/55 \rfloor = 18 \quad (4.3.5)$$

$$|B \cap C| = \lfloor 1000/\text{lcm}(7, 11) \rfloor = \lfloor 1000/77 \rfloor = 12 \quad (4.3.6)$$

$$|A \cap B \cap C| = \lfloor 1000/\text{lcm}(5, 7, 11) \rfloor = \lfloor 1000/385 \rfloor = 2 \quad (4.3.7)$$

Therefore:

$$|A \cup B \cup C| = 200 + 142 + 90 - 28 - 18 - 12 + 2 \quad (4.3.8)$$

$$= 432 - 58 + 2 \quad (4.3.9)$$

$$= 376 \quad (4.3.10)$$

The number of integers ≤ 1000 not divisible by 5, 7, or 11 is:

$$1000 - 376 = 624$$

This example demonstrates how divisibility properties and the relationship between GCD and LCM are essential for solving counting problems. Since $\gcd(a, b) = 1$ for distinct primes, we have $\text{lcm}(a, b) = ab$, which simplifies our calculations significantly.

4.4 Linear Diophantine Equations

A natural extension of our study of divisibility and Bézout's identity is the concept of Linear Diophantine Equations.

Definition: Diophantine Equation

A Diophantine equation is an equation where we seek integer solutions. A linear Diophantine equation in two variables takes the form:

$$ax + by = c$$

where a , b , and c are integers, and we seek integer solutions for x and y .

4.4.1 Existence of Solutions

The first question we might ask is: when does a linear Diophantine equation have solutions? The answer is closely tied to the concept of GCD.

Theorem: Solvability of Linear Diophantine Equations

The linear Diophantine equation $ax + by = c$ has integer solutions if and only if $\gcd(a, b) \mid c$.

Proof

Let $d = \gcd(a, b)$.

(\Rightarrow) Suppose there exist integers x_0 and y_0 such that $ax_0 + by_0 = c$.

Since $d \mid a$ and $d \mid b$, we know that $a = da'$ and $b = db'$ for some integers a' and b' .

Substituting, we get:

$$c = ax_0 + by_0 \quad (4.4.1)$$

$$= da'x_0 + db'y_0 \quad (4.4.2)$$

$$= d(a'x_0 + b'y_0) \quad (4.4.3)$$

Therefore, $d \mid c$.

(\Leftarrow) Suppose $d \mid c$, which means $c = dc'$ for some integer c' .

By Bézout's identity, there exist integers s and t such that $as + bt = d$.

Multiplying both sides by c' :

$$asc' + btc' = dc' \quad (4.4.4)$$

$$asc' + btc' = c \quad (4.4.5)$$

Therefore, $x_0 = sc'$ and $y_0 = tc'$ is a solution to the original equation.

□

4.4.2 Finding All Solutions

Once we know that solutions exist, the next step is to find all possible solutions.

Theorem: General Solution of Linear Diophantine Equations

If $ax + by = c$ has a particular solution (x_0, y_0) , then all solutions are given by:

$$x = x_0 + a't, \quad y = y_0 - b't$$

where $d = \gcd(a, b)$, $a' = \frac{a}{d}$, $b' = \frac{b}{d}$ (note that a' and b' are integers since d divides both a and b), and t is an arbitrary integer parameter.

Equivalently, this can be written as:

$$x = x_0 + \frac{b}{d}t, \quad y = y_0 - \frac{a}{d}t$$

Proof

Let (x, y) be any solution to $ax + by = c$, and let (x_0, y_0) be a particular solution. Then:

$$ax + by = c \tag{4.4.6}$$

$$ax_0 + by_0 = c \tag{4.4.7}$$

Subtracting, we get:

$$a(x - x_0) + b(y - y_0) = 0 \tag{4.4.8}$$

$$a(x - x_0) = -b(y - y_0) \tag{4.4.9}$$

Let $d = \gcd(a, b)$, and define $a' = \frac{a}{d}$ and $b' = \frac{b}{d}$. Note that a' and b' are integers since d divides both a and b . Moreover, $\gcd(a', b') = 1$ since we have factored out the greatest common divisor.

Substituting $a = da'$ and $b = db'$ into our equation:

$$da'(x - x_0) = -db'(y - y_0) \tag{4.4.10}$$

$$a'(x - x_0) = -b'(y - y_0) \quad (\text{dividing by } d)$$

Since $\gcd(a', b') = 1$, we can apply the fundamental property that if $a' \mid b'k$ and $\gcd(a', b') = 1$, then $a' \mid k$.

From $a'(x - x_0) = -b'(y - y_0)$, we see that:

- $a' \mid b'(y - y_0)$, and since $\gcd(a', b') = 1$, we have $a' \mid (y - y_0)$
- $b' \mid a'(x - x_0)$, and since $\gcd(a', b') = 1$, we have $b' \mid (x - x_0)$

Therefore, there exists an integer t such that:

$$y - y_0 = a't \quad (4.4.11)$$

$$x - x_0 = -b't \quad (4.4.12)$$

Solving for x and y :

$$x = x_0 - b't = x_0 + b'(-t) \quad (4.4.13)$$

$$y = y_0 + a't = y_0 - a'(-t) \quad (4.4.14)$$

Since t is an arbitrary integer, we can replace $(-t)$ with t to get:

$$x = x_0 + b't = x_0 + \frac{b}{d}t \quad (4.4.15)$$

$$y = y_0 - a't = y_0 - \frac{a}{d}t \quad (4.4.16)$$

where t is an arbitrary integer parameter.

Conversely, it can be verified that any pair (x, y) of this form satisfies the original equation $ax + by = c$.

□

4.4.3 Finding a Particular Solution

To find a particular solution to $ax + by = c$, we use the Extended Euclidean Algorithm to compute integers s and t such that $as + bt = \gcd(a, b)$.

Let $d = \gcd(a, b)$. Since the equation $ax + by = c$ has solutions if and only if $d \mid c$, we can write $c = dk$ for some integer k .

From Bézout's identity, we have $as + bt = d$. Multiplying both sides by k :

$$a(sk) + b(tk) = dk \quad (4.4.17)$$

$$a(sk) + b(tk) = c \quad (4.4.18)$$

Therefore, a particular solution is:

$$x_0 = sk, \quad y_0 = tk$$

where $k = \frac{c}{d}$.

Example 4.4.1

Find all integer solutions to the Diophantine equation $48x + 18y = 30$.

Solution: We previously computed Bézout's coefficients for $\gcd(48, 18)$, obtaining:

$$6 = 48 \cdot (-1) + 18 \cdot 3$$

Since $30 = 6 \cdot 5$, we can simply multiply this identity by 5 to get:

$$30 = 48 \cdot (-5) + 18 \cdot 15 \quad (4.4.19)$$

Therefore, a particular solution is:

$$x_0 = -5 \quad (4.4.20)$$

$$y_0 = 15 \quad (4.4.21)$$

The general solution is:

$$x = x_0 + \frac{b}{d}t = -5 + \frac{18}{6}t = -5 + 3t \quad (4.4.22)$$

$$y = y_0 - \frac{a}{d}t = 15 - \frac{48}{6}t = 15 - 8t \quad (4.4.23)$$

Therefore, all integer solutions to $48x + 18y = 30$ are given by:

$$x = -5 + 3t, \quad y = 15 - 8t$$

where t is an arbitrary integer.

4.4.4 Diophantine Equations with Constraints

In many applications, we may need solutions that satisfy additional constraints, such as positive solutions.

Example 4.4.2

Find all positive integer solutions to the Diophantine equation $5x + 3y = 31$.

Solution: At first glance one could tell that $\gcd(5, 3) = 1$ and because $1|31$, hence a solution exists to the given Diophantine equation. Let's apply the Extended Euclidean

Algorithm to find the Bézout coefficients directly,

Initialize: $r_1 = 5, r_2 = 3, x_1 = 1, y_1 = 0, x_2 = 0, y_2 = 1$.

First iteration:

$$5 = 3 \cdot 1 + 2 \quad (4.4.24)$$

$$q = 1, r = 2 \quad (4.4.25)$$

Compute new coefficients:

$$x = x_1 - qx_2 = 1 - 1 \cdot 0 = 1 \quad (4.4.26)$$

$$y = y_1 - qy_2 = 0 - 1 \cdot 1 = -1 \quad (4.4.27)$$

Update: $(r_1, x_1, y_1) = (3, 0, 1)$ and $(r_2, x_2, y_2) = (2, 1, -1)$.

Second iteration:

$$3 = 2 \cdot 1 + 1 \quad (4.4.28)$$

$$q = 1, r = 1 \quad (4.4.29)$$

Compute new coefficients:

$$x = x_1 - qx_2 = 0 - 1 \cdot 1 = -1 \quad (4.4.30)$$

$$y = y_1 - qy_2 = 1 - 1 \cdot (-1) = 2 \quad (4.4.31)$$

Update: $(r_1, x_1, y_1) = (2, 1, -1)$ and $(r_2, x_2, y_2) = (1, -1, 2)$.

Third iteration:

$$2 = 1 \cdot 2 + 0 \quad (4.4.32)$$

$$q = 2, r = 0 \quad (4.4.33)$$

Since $r = 0$, the algorithm terminates with $\gcd(5, 3) = 1$ and Bézout coefficients $x = -1$ and $y = 2$, giving us:

$$5 \cdot (-1) + 3 \cdot 2 = -5 + 6 = 1$$

Since $31 = 1 \cdot 31$, a particular solution is:

$$x_0 = (-1) \cdot 31 = -31 \quad (4.4.34)$$

$$y_0 = 2 \cdot 31 = 62 \quad (4.4.35)$$

The general solution is:

$$x = -31 + 3t \quad (4.4.36)$$

$$y = 62 - 5t \quad (4.4.37)$$

For positive integer solutions, we need:

$$x = -31 + 3t > 0 \quad (4.4.38)$$

$$\Rightarrow t > \frac{31}{3} \Rightarrow t \geq 11 \quad (4.4.39)$$

And:

$$y = 62 - 5t > 0 \quad (4.4.40)$$

$$\Rightarrow t < \frac{62}{5} \Rightarrow t \leq 12 \quad (4.4.41)$$

Therefore, the only values of t that give positive integer solutions are $t = 11$ and $t = 12$.

For $t = 11$:

$$x = -31 + 3 \cdot 11 = -31 + 33 = 2 \quad (4.4.42)$$

$$y = 62 - 5 \cdot 11 = 62 - 55 = 7 \quad (4.4.43)$$

For $t = 12$:

$$x = -31 + 3 \cdot 12 = -31 + 36 = 5 \quad (4.4.44)$$

$$y = 62 - 5 \cdot 12 = 62 - 60 = 2 \quad (4.4.45)$$

Therefore, the positive integer solutions to $5x + 3y = 31$ are $(2, 7)$ and $(5, 2)$.

4.5 Practice Exercises

Exercise 4.1

Find the value of $\gcd(2^n - 1, 2^m - 1)$ where n and m are positive integers.

Exercise 4.2

Find all pairs of positive integers (a, b) such that $\gcd(a, b) = 12$ and $\text{lcm}(a, b) = 60$.

Exercise 4.3

Find all non-negative integer solutions to the equation $7x + 11y = 100$.

Exercise 4.4

If a and b are positive integers, prove that $\text{lcm}(a, b) \leq ab$.

Exercise 4.5

Find the remainder when 2^{100} is divided by 7.

Exercise 4.6

Determine the smallest positive integer solution (x, y) to the equation $91x - 42y = 1$.

Exercise 4.7

Prove that if a and b are positive integers with $\gcd(a, b) = d$, then $\gcd(\frac{a}{d}, \frac{b}{d}) = 1$.

Exercise 4.8

Let a , b , and c be positive integers. Prove that $\gcd(a, b, c) \cdot \text{lcm}(a, b, c) \neq abc$ in general.
Provide a counterexample.

Exercise 4.9

Find the largest positive integer that divides $n^3 - n$ for all positive integers n .

Exercise 4.10

Let a and b be positive integers with $\gcd(a, b) = d > 1$. Prove that there exist infinitely many positive integers n such that $\gcd(a + n, b + n) > 1$.

Chapter 5

Prime Numbers



Eratosthenes of Cyrene (c. 276-194 BCE)

Eratosthenes served as the chief librarian of the Library of Alexandria, one of the ancient world's greatest center of learning. His most enduring contribution to number theory is the "Sieve of Eratosthenes," an ingenious and efficient algorithm for finding all prime numbers up to any given limit. His sieve method revolutionized how mathematicians approached the identification of prime numbers, establishing a systematic approach that would influence mathematical thinking for centuries. This method continues to be fundamental to computational number theory today. A polymath of extraordinary range, he also calculated the Earth's circumference with remarkable accuracy, created the first global projection of the world, and invented a system for finding prime numbers.

Eratosthenes was nicknamed "Beta" (the second letter of the Greek alphabet) because scholars considered him second-best in many fields but first in none—despite his remarkable achievements, he lived in the shadow of other specialists while never specializing himself.

5.1 Introduction to Prime Numbers

Prime numbers are fundamental to the study of number theory and appear frequently in the ISI entrance examination. Understanding their properties and distributions will provide you with a powerful toolkit for tackling complex mathematical problems.

Definition: Prime Number

A prime number p is a natural number greater than 1 that has exactly two distinct divisors: 1 and p itself.

The first few prime numbers are:

2, 3, 5, 7, 11, 13, 17, 19, 23, 29, 31, 37, 41, 43, 47, ...

Note that 1 is not considered a prime number since it has only one divisor (itself). The number 2 is the only even prime number; all other prime numbers are odd. When dealing with prime numbers, one of the most fundamental tasks is identifying which numbers are prime. While we can check each number individually using divisibility tests, this becomes impractical for larger ranges. The development of sieve methods represents one of the earliest examples of algorithmic thinking in mathematics.

Sieve of Eratosthenes

The Sieve of Eratosthenes is an elegant and efficient algorithm for finding all prime numbers up to a specified limit. The core insight behind this method is logical: if a number is not prime, it must be divisible by at least one prime number smaller than itself. The algorithm systematically eliminates composite numbers by marking all multiples of each prime, ensuring that only prime numbers remain unmarked. This approach transforms the problem of primality testing into a simpler process of elimination.

Algorithm: Sieve of Eratosthenes

To find all prime numbers up to a limit n :

1. Create a list of consecutive integers from 2 to n : $(2, 3, 4, \dots, n)$.
2. Start with the first prime number, $p = 2$.
3. Mark all multiples of p from p^2 to n as composite.
4. Find the next unmarked number after p , which is the next prime.
5. Repeat steps 3 and 4 until $p^2 > n$.
6. All unmarked numbers in the list are prime.

Note that we start marking from p^2 in step 3 because any smaller multiple of p would have already been marked as a multiple of a smaller prime. For example, when $p = 5$, we don't need to mark 10, 15, or 20, as they would have been marked already as multiples of 2 or 3.

Example 5.1.1

Use the Sieve of Eratosthenes to find all prime numbers less than 30.

Solution: Let's apply the Sieve of Eratosthenes algorithm step by step:

1. Create a list of consecutive integers from 2 to 29:

2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29

2. The first number, 2, is prime. Mark all multiples of 2 (except 2 itself):

2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29

3. The next unmarked number is 3. It's prime. Mark all multiples of 3 (except 3 itself), starting from $3^2 = 9$:

2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29

4. The next unmarked number is 5. It's prime. Mark all multiples of 5 (except 5 itself), starting from $5^2 = 25$:

2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29

5. Since the next unmarked number is 7 and $7^2 = 49 > 30$, we're done.
6. The unmarked numbers are the primes less than 30: 2, 3, 5, 7, 11, 13, 17, 19, 23, 29.

While the basic Sieve of Eratosthenes is already efficient, several optimizations can make it even more powerful:

Algorithm: Optimized Sieve of Eratosthenes

1. Only consider odd numbers in the initial list (except 2), since all even numbers greater than 2 are composite.
2. When marking multiples, use steps of size $2p$ instead of p for odd primes, since every other multiple would be even and already excluded.
3. For very large ranges, implement a segmented sieve that processes the range in smaller blocks.

The optimized version significantly reduces both time and memory requirements, especially for large ranges.

5.2 Fundamental Properties of Prime Numbers

The study of prime numbers dates back to ancient times. Around 300 BCE, the Greek mathematician Euclid proved that there are infinitely many prime numbers. This was one of the first major results in number theory and remains one of the most elegant proofs in mathematics.

Theorem: Infinity of Primes

There are infinitely many prime numbers.

Proof

We will prove this by contradiction. Assume that there are only finitely many primes, which we can list as p_1, p_2, \dots, p_k .

Now, consider the number $N = (p_1 \cdot p_2 \cdot \dots \cdot p_k) + 1$.

This number N is either prime or composite. If N is prime, then we have found a prime number that is not in our list, contradicting our assumption.

Let's consider the alternative case where N is composite. Then N must have a prime factor. Could this prime factor be one of the primes in our list?

For any prime p_i in our list, if p_i divides N , then: $p_i \mid (p_1 \cdot p_2 \cdot \dots \cdot p_k + 1)$

Since p_i divides the product $p_1 \cdot p_2 \cdot \dots \cdot p_k$, this would mean $p_i \mid 1$, which is impossible since $p_i > 1$.

Therefore, N must either be a prime not in our list, or have a prime factor not in our list. In either case, our assumption that we had listed all primes is contradicted, proving that there are infinitely many primes.

□

5.2.1 Prime Factorization

One of the most fundamental concepts in number theory is the idea that prime numbers serve as the building blocks for all natural numbers. This principle is formalized in the Fundamental Theorem of Arithmetic.

Definition: Prime Factorization

The prime factorization of a natural number $n > 1$ is the representation of n as a product of prime numbers:

$$n = p_1^{a_1} \cdot p_2^{a_2} \cdots \cdot p_k^{a_k}$$

where p_1, p_2, \dots, p_k are distinct prime numbers and a_1, a_2, \dots, a_k are positive integers representing the exponents of the respective primes.

Theorem: Fundamental Theorem of Arithmetic

Every natural number greater than 1 can be represented uniquely as a product of prime numbers, up to the order of factors.

The Fundamental Theorem of Arithmetic establishes the theoretical foundation that every natural number greater than 1 has a unique prime factorization. This remarkable property makes prime numbers the elementary building blocks of the natural number system. However, to make practical use of this theorem, we need an efficient method to find these prime factorizations.

The algorithm for finding prime factorization provides a systematic approach to decompose any natural number into its constituent prime factors. This process begins with the smallest prime number and methodically tests divisibility, capturing each prime factor and its multiplicity along the way. The algorithm leverages an important mathematical insight: if a number n has no prime factors less than or equal to \sqrt{n} , then n itself must be prime. This optimization significantly reduces the computational work needed, especially for large numbers.

Algorithm: Finding Prime Factorization

To find the prime factorization of a number n :

1. Start with the smallest prime number $p = 2$.
2. Check if p divides n . If it does, divide n by p and repeat this step.
3. If p does not divide n , move to the next prime number.
4. Continue until n becomes 1 or until $p > \sqrt{n}$. In the latter case, if $n > 1$, then n itself is prime.

Through this algorithm, we can verify the Fundamental Theorem of Arithmetic empirically by finding the unique prime factorization of any given number.

Example 5.2.1

Find the prime factorization of 84

Solution: Apply the algorithm systematically:

- $p = 2$: Since $84 = 2 \cdot 42$, we have one factor of 2. Continue with 42.
- $p = 2$: Since $42 = 2 \cdot 21$, we have another factor of 2. Continue with 21.
- $p = 2$: Since 21 is odd, 2 doesn't divide it. Move to $p = 3$.
- $p = 3$: Since $21 = 3 \cdot 7$, we have one factor of 3. Continue with 7.
- $p = 3$: Since 3 doesn't divide 7, move to $p = 5$.
- $p = 5$: Since 5 doesn't divide 7, move to $p = 7$.
- $p = 7$: Since $7 = 7 \cdot 1$, we're done.

Therefore, $84 = 2^2 \cdot 3 \cdot 7$.

5.2.2 Prime Factorization and its Relation to GCD and LCM

Prime factorization provides an elegant and systematic approach to computing both the greatest common divisor (GCD) and the least common multiple (LCM) of integers. When we express integers in their prime factorized form, calculating GCD and LCM becomes remarkably straightforward.

Theorem: GCD using Prime Factorization

If $a = p_1^{\alpha_1} \cdot p_2^{\alpha_2} \cdots p_k^{\alpha_k}$ and $b = p_1^{\beta_1} \cdot p_2^{\beta_2} \cdots p_k^{\beta_k}$ are the prime factorizations of a and b , respectively, then:

$$\gcd(a, b) = p_1^{\min(\alpha_1, \beta_1)} p_2^{\min(\alpha_2, \beta_2)} \cdots p_k^{\min(\alpha_k, \beta_k)}$$

Theorem: LCM using Prime Factorization

If $a = p_1^{\alpha_1} \cdot p_2^{\alpha_2} \cdots p_k^{\alpha_k}$ and $b = p_1^{\beta_1} \cdot p_2^{\beta_2} \cdots p_k^{\beta_k}$ are the prime factorizations of a and b , respectively, then:

$$\text{lcm}(a, b) = p_1^{\max(\alpha_1, \beta_1)} \cdot p_2^{\max(\alpha_2, \beta_2)} \cdots p_k^{\max(\alpha_k, \beta_k)}$$

These formulations reveal an elegant relationship: the GCD takes the minimum powers of each prime factor, while the LCM takes the maximum powers. This complementary nature is further

exemplified by the fundamental identity, which we discussed in the previous chapter.

Theorem: GCD-LCM Identity

For any two positive integers a and b :

$$\gcd(a, b) \cdot \text{lcm}(a, b) = a \cdot b$$

Example 5.2.2

Find the GCD and LCM of 60 and 84 using prime factorization

Solution: First, we find the prime factorizations:

$$\begin{aligned} 60 &= 2^2 \cdot 3 \cdot 5 \\ 84 &= 2^2 \cdot 3 \cdot 7 \end{aligned}$$

Then, applying the theorems:

$$\begin{aligned} \gcd(60, 84) &= 2^{\min(2,2)} \cdot 3^{\min(1,1)} \cdot 5^{\min(1,0)} \cdot 7^{\min(0,1)} \\ &= 2^2 \cdot 3^1 \cdot 5^0 \cdot 7^0 \\ &= 4 \cdot 3 \cdot 1 \cdot 1 \\ &= 12 \end{aligned}$$

$$\begin{aligned} \text{lcm}(60, 84) &= 2^{\max(2,2)} \cdot 3^{\max(1,1)} \cdot 5^{\max(1,0)} \cdot 7^{\max(0,1)} \\ &= 2^2 \cdot 3^1 \cdot 5^1 \cdot 7^1 \\ &= 4 \cdot 3 \cdot 5 \cdot 7 \\ &= 420 \end{aligned}$$

We can verify the GCD-LCM identity:

$$\gcd(60, 84) \cdot \text{lcm}(60, 84) = 12 \cdot 420 = 5040 = 60 \cdot 84$$

Since $\gcd(60, 84) = 12 \neq 1$, the numbers 60 and 84 are not coprime. However, consider the numbers 35 and 48:

$$\begin{aligned} 35 &= 5 \cdot 7 \\ 48 &= 2^4 \cdot 3 \end{aligned}$$

These numbers have no common prime factors in their factorizations, so $\gcd(35, 48) = 1$, meaning 35 and 48 are coprime. Consequently, $\text{lcm}(35, 48) = 35 \cdot 48 = 1680$.

Definition: Coprime / Relatively Prime

Two integers a and b are said to be coprime (or relatively prime) if their greatest common divisor is 1, that is, $\gcd(a, b) = 1$. This means that the only positive integer that divides both a and b is 1.

Theorem: Characterization of Coprime Integers

From the perspective of prime factorization, two integers a and b are coprime if and only if their prime factorizations share no common prime factors. That is, if:

$$a = p_1^{\alpha_1} \cdot p_2^{\alpha_2} \cdots \cdot p_r^{\alpha_r}$$

$$b = q_1^{\beta_1} \cdot q_2^{\beta_2} \cdots \cdot q_s^{\beta_s}$$

then a and b are coprime if and only if $p_i \neq q_j$ for all i, j .

When two integers a and b are coprime:

1. $\gcd(a, b) = 1$
2. $\text{lcm}(a, b) = a \cdot b$
3. There exist integers x and y such that $ax + by = 1$ (Bézout's identity)
4. If $a | c$ and $b | c$, then $ab | c$

These properties make coprime integers particularly important in number theory and its applications.

5.3 Divisor Function

We can use prime factorization to calculate the total number of divisors of a natural number n .

Definition: Divisor Function

The divisor function $\tau(n)$ counts the total number of positive divisors of a natural number n , including 1 and n itself. If $n = p_1^{a_1} \cdot p_2^{a_2} \cdots \cdot p_k^{a_k}$ is the prime factorization of n , then:

$$\tau(n) = (a_1 + 1) \cdot (a_2 + 1) \cdots \cdot (a_k + 1)$$

For 84, the number of divisors is:

$$(2 + 1) \cdot (1 + 1) \cdot (1 + 1) = 3 \cdot 2 \cdot 2 = 12$$

So 84 has 12 divisors, including 1 and 84 itself, i.e., $\tau(84) = 12$.

To understand why $\tau(n) = (a_1 + 1) \cdot (a_2 + 1) \cdots \cdots (a_k + 1)$, consider how divisors are constructed from prime factors:

For a number with prime factorization $n = p_1^{a_1} \cdot p_2^{a_2} \cdots \cdots p_k^{a_k}$, any divisor must be formed by taking some subset of these prime factors, with each prime p_i appearing with an exponent between 0 and a_i .

For $84 = 2^2 \cdot 3^1 \cdot 7^1$, we have the following options:

- For the prime factor 2: we can use 2^0 (no factor of 2), 2^1 , or 2^2 (3 possibilities)
- For the prime factor 3: we can use 3^0 (no factor of 3) or 3^1 (2 possibilities)
- For the prime factor 7: we can use 7^0 (no factor of 7) or 7^1 (2 possibilities)

Any divisor of 84 is formed by making exactly one choice from each option set and multiplying them:

$$\begin{aligned} 2^0 \cdot 3^0 \cdot 7^0 &= 1 \\ 2^1 \cdot 3^0 \cdot 7^0 &= 2 \\ 2^2 \cdot 3^0 \cdot 7^0 &= 4 \\ 2^0 \cdot 3^1 \cdot 7^0 &= 3 \\ 2^1 \cdot 3^1 \cdot 7^0 &= 6 \\ &\vdots \end{aligned}$$

The total count of all possible combinations equals the product of the number of options for each prime:

$$(a_1 + 1) \cdot (a_2 + 1) \cdots \cdots (a_k + 1)$$

For 84, that's $(2 + 1) \cdot (1 + 1) \cdot (1 + 1) = 3 \cdot 2 \cdot 2 = 12$ distinct divisors.

The formula is essentially a direct application of the multiplication principle from combinatorics: when we have $a_1 + 1$ ways to choose the power of p_1 , $a_2 + 1$ ways to choose the power of p_2 , and so on, the total number of different divisors we can construct is the product of these choice counts.

Theorem: Multiplicativity of the Divisor Function

The divisor function $\tau(n)$ is multiplicative, meaning that if $\gcd(m, n) = 1$ (i.e., m and n are coprime), then:

$$\tau(m \cdot n) = \tau(m) \cdot \tau(n)$$

Proof

To prove that $\tau(n)$ is multiplicative, we need to show that $\tau(m \cdot n) = \tau(m) \cdot \tau(n)$ whenever $\gcd(m, n) = 1$.

When $\gcd(m, n) = 1$, the prime factorizations of m and n have no common prime factors. Let:

$$\begin{aligned} m &= p_1^{a_1} \cdot p_2^{a_2} \cdots \cdot p_r^{a_r} \\ n &= q_1^{b_1} \cdot q_2^{b_2} \cdots \cdot q_s^{b_s} \end{aligned}$$

where $p_i \neq q_j$ for all i, j since m and n are coprime.

Then:

$$m \cdot n = p_1^{a_1} \cdot p_2^{a_2} \cdots \cdot p_r^{a_r} \cdot q_1^{b_1} \cdot q_2^{b_2} \cdots \cdot q_s^{b_s}$$

By the definition of the divisor function:

$$\tau(m) = (a_1 + 1) \cdot (a_2 + 1) \cdots \cdot (a_r + 1)$$

$$\tau(n) = (b_1 + 1) \cdot (b_2 + 1) \cdots \cdot (b_s + 1)$$

$$\tau(m \cdot n) = (a_1 + 1) \cdots \cdot (a_r + 1) \cdot (b_1 + 1) \cdots \cdot (b_s + 1)$$

Therefore:

$$\tau(m \cdot n) = \tau(m) \cdot \tau(n)$$

This proves that $\tau(n)$ is multiplicative. □

The multiplicative property of the divisor function makes it particularly useful in number theory. It allows us to compute $\tau(n)$ efficiently for large numbers by breaking them down into coprime factors and multiplying the results. This property extends to many other number-theoretic functions and is fundamental to the study of arithmetic functions.

Let us apply the multiplicative property of $\tau(n)$ in practice for $m = 15$ and $n = 8$. First, we confirm that $\gcd(15, 8) = 1$, so they are indeed coprime.

The prime factorization for 15 is $3 \cdot 5$ and for 8 is 2^3 . We can calculate $\tau(15)$ as :

$$\tau(15) = (1 + 1) \cdot (1 + 1) = 2 \cdot 2 = 4$$

Indeed, the divisors of 15 are $\{1, 3, 5, 15\}$.

Calculating $\tau(8)$:

$$\tau(8) = (3 + 1) = 4$$

The divisors of 8 are $\{1, 2, 4, 8\}$. Therefore, $\tau(15) \cdot \tau(8) = 4 \cdot 4 = 16$.

Now, let us calculate $\tau(15 \cdot 8) = \tau(120)$ directly. The prime factorization of 120 is:

$$120 = 2^3 \cdot 3 \cdot 5$$

Therefore:

$$\tau(120) = (3 + 1) \cdot (1 + 1) \cdot (1 + 1) = 4 \cdot 2 \cdot 2 = 16$$

We can verify that the divisors of 120 are: $\{1, 2, 3, 4, 5, 6, 8, 10, 12, 15, 20, 24, 30, 40, 60, 120\}$, which indeed has 16 elements.

5.3.1 Sum of Divisors

Definition: Sum of Divisors

The sum of divisors function, denoted $\sigma(n)$, is an arithmetic function that gives the sum of all positive divisors of a natural number n , including 1 and n itself. Mathematically:

$$\sigma(n) = \sum_{d|n} d$$

where the sum is taken over all positive divisors d of n .

Theorem: Formula for Sum of Divisors

If $n = p_1^{a_1} \cdot p_2^{a_2} \cdots \cdot p_k^{a_k}$ is the prime factorization of n , then the sum of divisors is given by:

$$\sigma(n) = \prod_{i=1}^k \frac{p_i^{a_i+1} - 1}{p_i - 1}$$

Proof

For a prime power p^a , the divisors are $1, p, p^2, \dots, p^a$. Therefore:

$$\sigma(p^a) = 1 + p + p^2 + \cdots + p^a = \frac{p^{a+1} - 1}{p - 1}$$

where the last step uses the formula for the sum of a geometric series.

Since σ is a multiplicative function (which can be proven similarly to the divisor counting

function), for $n = p_1^{a_1} \cdot p_2^{a_2} \cdots \cdot p_k^{a_k}$ with distinct primes p_i , we have:

$$\sigma(n) = \sigma(p_1^{a_1}) \cdot \sigma(p_2^{a_2}) \cdots \cdot \sigma(p_k^{a_k})$$

Therefore:

$$\sigma(n) = \prod_{i=1}^k \frac{p_i^{a_i+1} - 1}{p_i - 1}$$

□

Example 5.3.1

Find the sum of divisors of 12.

Solution: First, we determine the prime factorization of 12:

$$12 = 2^2 \cdot 3^1$$

Using the formula:

$$\begin{aligned}\sigma(12) &= \frac{2^{2+1} - 1}{2 - 1} \cdot \frac{3^{1+1} - 1}{3 - 1} \\ &= \frac{2^3 - 1}{1} \cdot \frac{3^2 - 1}{2} \\ &= \frac{8 - 1}{1} \cdot \frac{9 - 1}{2} \\ &= 7 \cdot 4 \\ &= 28\end{aligned}$$

Indeed, the divisors of 12 are $\{1, 2, 3, 4, 6, 12\}$, and their sum is $1 + 2 + 3 + 4 + 6 + 12 = 28$.

The sum of divisors function exhibits several interesting characteristics that help us understand its behavior. For any prime number p , the function yields $\sigma(p) = p + 1$, which makes intuitive sense as a prime number has exactly two divisors: 1 and the number itself. This property extends elegantly to prime powers, where $\sigma(p^k) = 1 + p + p^2 + \cdots + p^k$. This summation can be rewritten using the formula for geometric series as $\frac{p^{k+1} - 1}{p - 1}$, providing a compact representation for calculating the sum of divisors of any prime power. This is an important observation as it also relates to multiplicative nature of the function.

Theorem: Multiplicativity of the Sum of Divisors Function

The sum of divisors function $\sigma(n)$ is multiplicative. That is, if $\gcd(m, n) = 1$, then $\sigma(mn) = \sigma(m) \cdot \sigma(n)$.

Proof

To prove that $\sigma(n)$ is multiplicative, we need to show that $\sigma(mn) = \sigma(m) \cdot \sigma(n)$ whenever $\gcd(m, n) = 1$.

Let D_m be the set of all divisors of m , and D_n be the set of all divisors of n . When $\gcd(m, n) = 1$, any divisor of mn can be uniquely expressed as a product $d_1 \cdot d_2$, where $d_1 \in D_m$ and $d_2 \in D_n$.

This is because:

1. If d divides mn , then $d = d_1 \cdot d_2$ where $d_1|m$ and $d_2|n$.
2. This decomposition is unique when $\gcd(m, n) = 1$ by the fundamental theorem of arithmetic.

Therefore, the set of all divisors of mn is precisely $\{d_1 \cdot d_2 : d_1 \in D_m, d_2 \in D_n\}$.

Now we can compute $\sigma(mn)$:

$$\begin{aligned}\sigma(mn) &= \sum_{d|mn} d \\ &= \sum_{d_1 \in D_m} \sum_{d_2 \in D_n} d_1 \cdot d_2 \\ &= \left(\sum_{d_1 \in D_m} d_1 \right) \cdot \left(\sum_{d_2 \in D_n} d_2 \right) \\ &= \sigma(m) \cdot \sigma(n)\end{aligned}$$

The crucial step is the third line, where we use the distributive property of multiplication over addition to factor the double sum. This step is valid precisely because every divisor of mn appears exactly once in the double sum when $\gcd(m, n) = 1$.

Thus, $\sigma(n)$ is indeed multiplicative.

□

Theorem: Relation Between $\sigma(n)$ and $\tau(n)$

For any positive integer $n > 1$:

$$\sigma(n) \geq \tau(n) \cdot \sqrt{n}$$

with equality if and only if all divisors of n are equal, which is impossible for $n > 1$.

Proof

Let $n > 1$ be a positive integer, and let $\{d_1, d_2, \dots, d_{\tau(n)}\}$ be the set of all positive divisors of n , where $\tau(n)$ is the number of divisors of n .

The sum of divisors is given by:

$$\sigma(n) = \sum_{i=1}^{\tau(n)} d_i$$

By the arithmetic mean-geometric mean inequality, we know that for any set of positive real numbers, their arithmetic mean is greater than or equal to their geometric mean, with equality if and only if all numbers are equal. Applied to our divisors:

$$\frac{\sigma(n)}{\tau(n)} \geq \sqrt[\tau(n)]{d_1 \cdot d_2 \cdot \dots \cdot d_{\tau(n)}}$$

Now, observe that the product of all divisors of n can be expressed in a special way. For each divisor d of n , we know that $\frac{n}{d}$ is also a divisor of n . Thus, the divisors can be paired such that each pair multiplies to n .

If $\tau(n)$ is even, then all divisors can be paired, and the product of all divisors is $n^{\tau(n)/2}$.

If $\tau(n)$ is odd, then there is one divisor (specifically, \sqrt{n}) that pairs with itself, and the remaining divisors can be paired. In this case, the product of all divisors is $n^{(\tau(n)-1)/2} \cdot \sqrt{n} = n^{\tau(n)/2}$.

Therefore, in either case:

$$\sqrt[\tau(n)]{d_1 \cdot d_2 \cdot \dots \cdot d_{\tau(n)}} = \sqrt[n^{\tau(n)/2}]{} = \sqrt{n}$$

Substituting this back into our inequality:

$$\frac{\sigma(n)}{\tau(n)} \geq \sqrt{n}$$

Multiplying both sides by $\tau(n)$:

$$\sigma(n) \geq \tau(n) \cdot \sqrt{n}$$

Equality holds if and only if all divisors are equal. However, for any integer $n > 1$, the divisors always include at least 1 and n itself, which are distinct. Therefore, strict inequality holds for all $n > 1$.

□

5.3.2 Mersenne Primes

Definition: Mersenne Prime

A Mersenne prime is a prime number of the form $M_p = 2^p - 1$, where p is also a prime.

Theorem: Mersenne Prime Property

If $2^p - 1$ is prime, then p must be prime. The converse is not always true.

Proof

We'll prove this by contrapositive, showing that if p is not prime, then $2^p - 1$ is not prime.

If p is not prime, then $p = a \cdot b$ for some integers $a, b > 1$.

Now, $2^p - 1 = 2^{a \cdot b} - 1 = (2^a)^b - 1$.

Using the formula for the difference of powers: $x^n - 1 = (x - 1)(x^{n-1} + x^{n-2} + \dots + x + 1)$, we get:

$$\begin{aligned} 2^p - 1 &= (2^a)^b - 1 \\ &= (2^a - 1)((2^a)^{b-1} + (2^a)^{b-2} + \dots + 2^a + 1) \end{aligned}$$

Since $a > 1$, we have $2^a - 1 > 1$. Also, the second factor is greater than 1. Therefore, $2^p - 1$ has factors other than 1 and itself, so it's not prime.

This proves that if $2^p - 1$ is prime, then p must be prime.

For the converse, consider $p = 11$, which is prime. However, $2^{11} - 1 = 2047 = 23 \cdot 89$, which is not prime.

□

Example 5.3.2

Verify whether $M_7 = 2^7 - 1 = 127$ is a Mersenne prime.

Solution: First, we need to check if 127 is prime. We only need to check divisibility by primes up to $\sqrt{127} \approx 11.27$.

$127 \div 2 = 63$ remainder 1 (not divisible by 2) $127 \div 3 = 42$ remainder 1 (not divisible by 3) $127 \div 5 = 25$ remainder 2 (not divisible by 5) $127 \div 7 = 18$ remainder 1 (not divisible by 7) $127 \div 11 = 11$ remainder 6 (not divisible by 11)

Since 127 is not divisible by any prime up to $\sqrt{127}$, it is prime.

Therefore, $M_7 = 2^7 - 1 = 127$ is indeed a Mersenne prime.

Mersenne primes are important because they are related to $\sigma(n)$, the sum of the divisors function for integer n . It is trivial to notice that for any number $n > 1$, $\sigma(n) > n$. This follows naturally from the definition, as $\sigma(n)$ includes n among its terms, along with at least one additional positive divisor (namely 1). This property forms the foundation for classifying numbers as deficient, perfect, or abundant based on how their sum of proper divisors compares to the number itself.

Definition: Perfect Numbers

A positive integer n is called a perfect number if it equals the sum of its proper divisors (all positive divisors excluding n itself). In other words, n is perfect if:

$$\sigma(n) - n = n$$

or equivalently:

$$\sigma(n) = 2n$$

The first perfect number is 6. The divisors of 6 are $\{1, 2, 3, 6\}$, and their sum is $1 + 2 + 3 + 6 = 12 = 2 \cdot 6$. One can also use the multiplicativity property of the sum of divisors function to check that:

$$\sigma(6) = \sigma(2) \cdot \sigma(3) = (1 + 2) \cdot (1 + 3) = 3 \cdot 4 = 12$$

.

Perfect numbers are exceedingly rare: only 51 are known to date. This scarcity prompted early mathematicians to search for patterns and formulas to generate these special numbers. The breakthrough came through the combined insights of Euclid (circa 300 BCE) and Leonhard Euler (1707-1783), separated by two millennia.

Euclid first showed in his Elements that numbers of the form $2^{p-1}(2^p - 1)$ are perfect when $2^p - 1$ is prime. Euler later completed the characterization by proving the converse: every even perfect number must have Euclid's form. This combined result, known as the Euclid-Euler Theorem, completely characterizes all even perfect numbers.

Theorem: Euclid-Euler Theorem

An even number is perfect if and only if it has the form $2^{p-1}(2^p - 1)$, where p is a prime number and $2^p - 1$ is a Mersenne prime.

Proof

We will prove that if p is prime and $M_p = 2^p - 1$ is also prime (a Mersenne prime), then $N = 2^{p-1}M_p$ is perfect.

A number is perfect when it equals the sum of its proper divisors. Let's calculate the sum of all divisors of $N = 2^{p-1}M_p$ and show that it equals $2N$.

First, observe that $N = 2^{p-1}M_p$ has the prime factorization $2^{p-1} \cdot M_p$ since M_p is prime.

Using the formula for the sum of divisors:

$$\begin{aligned}\sigma(N) &= \sigma(2^{p-1}) \cdot \sigma(M_p) \\ &= \frac{2^p - 1}{2 - 1} \cdot (1 + M_p) \\ &= (2^p - 1) \cdot (1 + (2^p - 1)) \\ &= (2^p - 1) \cdot 2^p \\ &= 2^p(2^p - 1) \\ &= 2 \cdot 2^{p-1}(2^p - 1) \\ &= 2N\end{aligned}$$

Since the sum of all divisors $\sigma(N) = 2N$, and the sum of all proper divisors equals $\sigma(N) - N = 2N - N = N$, the number N is perfect.

□

An interesting consequence of the Euclid-Euler theorem is that finding perfect numbers is equivalent to finding Mersenne primes. Each new Mersenne prime discovery leads directly to a new

perfect number. The largest known perfect number, corresponding to the 51st Mersenne prime discovered in 2018, has over 49 million digits!

5.4 Distribution of Prime Numbers

The study of how prime numbers are distributed among the integers represents one of the most profound inquiries in number theory. Unlike the rational numbers \mathbb{Q} , which maintain a constant density in the real number line \mathbb{R} (between any two real numbers, we can always find a rational number), prime numbers become increasingly sparse as we move further along the number line.

The Prime Number Theorem gives an asymptotic formula for the number of primes up to a given value. This remarkable theorem, independently proved by Jacques Hadamard and Charles Jean de la Vallée Poussin in 1896, provides a rigorous mathematical framework for understanding the distribution of primes.

The theorem answers one of the most fundamental questions in number theory: how frequently do prime numbers occur? Although Euclid proved over two millennia ago that there are infinitely many primes, understanding their distribution pattern remained elusive until the late 19th century. The Prime Number Theorem not only confirmed the intuitive observation that primes become increasingly rare as numbers grow larger, but it also precisely quantified this scarcity with the elegant approximation $\pi(x) \sim x/\ln(x)$.

Theorem: Prime Number Theorem

Let $\pi(x)$ denote the number of primes less than or equal to x . Then:

$$\pi(x) \approx \frac{x}{\ln(x)}$$

The function $x/\ln(x)$ approximates the number of primes up to x , and the approximation gets better as x gets larger.

Example 5.4.1

Use the Prime Number Theorem to estimate the number of primes less than 1,000,000 and compare with the actual value.

Solution: According to the Prime Number Theorem, the number of primes less than 1,000,000 can be approximated by:

$$\frac{1,000,000}{\ln(1,000,000)} = \frac{1,000,000}{13.82} \approx 72,382$$

The actual number of primes less than 1,000,000 is 78,498.

The relative error is:

$$\frac{|78,498 - 72,382|}{78,498} \cdot 100\% \approx 7.8\%$$

This demonstrates how the approximation works reasonably well even for moderate values.

An interesting question in number theory concerns how far apart consecutive primes can be. The prime gaps theorem describes that in detail.

Theorem: Prime Gaps

For any positive integer n , there exist n consecutive composite integers.

Proof

Consider the sequence of n consecutive integers:

$$(n+1)! + 2, (n+1)! + 3, \dots, (n+1)! + (n+1)$$

For each k where $2 \leq k \leq n+1$, we have:

$$\begin{aligned} (n+1)! + k &= k \cdot \frac{(n+1)!}{k} + k \\ &= k \cdot \left(\frac{(n+1)!}{k} + 1 \right) \end{aligned}$$

Since k divides $(n+1)!$ (as $k \leq n+1$), we can see that k divides $(n+1)! + k$. Thus, each number in our sequence is composite, as each is divisible by at least one integer k where $2 \leq k \leq n+1$.

Therefore, we have constructed n consecutive composite integers.

□

This theorem shows that prime gaps can be arbitrarily large, even though there are infinitely many primes. Conversely, the famous unsolved Twin Prime Conjecture asks whether there are infinitely many pairs of primes that differ by only 2, such as (3,5), (11,13), and (17,19).

5.5 Practice Exercises

Exercise 5.1

Let p be a prime number greater than 3. Prove that $p^2 - 1$ is always divisible by 24.

Exercise 5.2

Find the smallest positive integer n such that $n!$ has exactly 20 trailing zeros.

Exercise 5.3

Twin primes are pairs of prime numbers that differ by exactly 2. Examples include (3, 5), (5, 7), (11, 13), (17, 19), and (41, 43).

Prove that for any twin prime pair $(p, p + 2)$ where $p > 3$, the sum $p + (p + 2) = 2p + 2$ is always divisible by 12.

Exercise 5.4

Determine all pairs of positive integers (a, b) such that $a^2 - b^2 = 2022$.

Exercise 5.5

Given the fact that $2^{31} - 1$ is a prime number, find the number of divisors of $2^{30}(2^{31} - 1)$.

Exercise 5.6

Demonstrate that for any prime number $p > 5$, the number $p^2 + 2$ is composite.

Exercise 5.7

Let p_1, p_2, p_3 be primes with $p_2 \neq p_3$, such that $4 + p_1p_2 = x^2$ and $4 + p_1p_3 = y^2$ for some integers x, y . Find all possible values of p_1, p_2, p_3 .

Exercise 5.8

Prove that there are infinitely many odd natural numbers n such that n , $n + 2$, and $n + 4$ are all composite numbers.

Exercise 5.9

Find the smallest positive integer n such that n has exactly 2024 positive divisors.

Exercise 5.10

Let $m = 2 \times 3 \times 5 \times 7 \times 11 = 2310$. Prove that there does not exist any positive integer $n < 2310$ such that $n(2310 - n)$ is a multiple of 2310.

Chapter 6

Modular Arithmetic

Carl Friedrich Gauss (1777-1855)

Carl Friedrich Gauss, often referred to as the “Prince of Mathematicians,” revolutionized number theory with his seminal work “Disquisitiones Arithmeticae” published in 1801 when he was just 24 years old. In this masterpiece, Gauss formally introduced modular arithmetic, using the congruence notation “ $a \equiv b \pmod{n}$ ” that we still use today. His systematic treatment of modular arithmetic transformed it from a collection of computational techniques into a rigorous mathematical framework.



Gauss’s work on modular arithmetic led to profound results in number theory, including his proof of the law of quadratic reciprocity, which he called the “fundamental theorem” of number theory. Gauss proved that the regular 17-sided polygon could be constructed with straightedge and compass—solving a problem that had remained open since antiquity.

Despite holding the position of director at the Göttingen Observatory, Gauss published relatively little, adhering to his personal motto ”pauca sed matura” (few, but ripe), preferring to release only completely polished work. His personal mathematical diary, discovered after his death, revealed that he had privately discovered many important results years or decades before they were published by others.

In previous chapters, we explored the fundamental concepts in number theory, namely divisibility and prime numbers. When one integer a divides another integer b , written as $a|b$, we know that $b = ak$ for some integer k . This leaves us with a binary relation: either a divides b or it does not.

Modular arithmetic extends this concept by formalizing what happens when division isn't exact. Rather than simply stating that a doesn't divide b , we can quantify the "remainder" of this attempted division. This approach creates a richer mathematical structure that preserves many properties of ordinary arithmetic while introducing powerful new tools for solving number-theoretic problems.

6.1 Introduction to Modular Arithmetic

One might ask: why develop a formal system of modular arithmetic when the concepts of remainders dates back to ancient mathematics? The simple act of treating remainders as mathematical objects in their own right, rather than just the leftovers from division, unlocks remarkable insights and problem-solving capabilities.

Consider trying to find the last digit of 7^{100} without a calculator. This seems daunting, but with modular arithmetic, we can reason that we only care about behavior modulo 10. Let us look at the remainders of powers of 7 when divided by 10: $7^1 = 7$ (last digit 7), $7^2 = 49$ (last digit 9), $7^3 = 343$ (last digit 3), $7^4 = 2401$ (last digit 1), $7^5 = 16807$ (last digit 7).

By examining patterns in the last digits of powers of 7, we discover a cycle of length 4. Thus, the last digit of 7^{100} can be calculated by realizing that $7^{100} = (7^4)^{25}$. This implies that the cycle of 4 repeats 25 times and hence the last digit of 7^{100} is the same as 7^0 , which equals 1.

This example illustrates how modular arithmetic provides not just a language for discussing remainders, but a coherent system where complex calculations become tractable and patterns reveal deep mathematical structures. By formalizing the ancient concept of remainders, Gauss gave us a powerful lens through which number-theoretic problems can be viewed with remarkable clarity.

Modular arithmetic is often referred to as "clock arithmetic" because it resembles how hours wrap around on a 12-hour clock: after 12 comes 1 again, not 13. In modular arithmetic, numbers "wrap around" after reaching a certain value — the modulus.

Definition:

Two integers a and b are congruent modulo n , written as $a \equiv b \pmod{n}$, if n divides their difference $(a - b)$. Equivalently, a and b leave the same remainder when divided by n .

This definition directly connects to our study of divisibility. When we write $a \equiv b \pmod{n}$,

we're stating that $n \mid (a - b)$, or equivalently, that a and b differ by a multiple of n . This creates equivalence classes of integers based on their remainders when divided by n .

For example, $17 \equiv 5 \pmod{12}$ because $17 = 12 \cdot 1 + 5$ and $5 = 12 \cdot 0 + 5$. Both 17 and 5 leave the same remainder (5) when divided by 12. Similarly, we can identify the complete set of equivalence classes modulo 12 as $\{0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11\}$, where each class represents all integers that leave a particular remainder when divided by 12.

This formalization of the remainder concept allows us to study arithmetic operations in a new light, preserving many familiar properties while revealing deeper structures in the integers.

Theorem: Properties of Modular Congruence

Modular congruence behaves like an equivalence relation. For any integers a, b, c and positive integer n :

1. **Reflexivity:** $a \equiv a \pmod{n}$ for any integer a .
2. **Symmetry:** If $a \equiv b \pmod{n}$, then $b \equiv a \pmod{n}$.
3. **Transitivity:** If $a \equiv b \pmod{n}$ and $b \equiv c \pmod{n}$, then $a \equiv c \pmod{n}$.

Modular arithmetic preserves basic operations. For any integers a, b, c, d and positive integer n :

1. **Addition:** If $a \equiv b \pmod{n}$ and $c \equiv d \pmod{n}$, then $a + c \equiv b + d \pmod{n}$.
2. **Subtraction:** If $a \equiv b \pmod{n}$ and $c \equiv d \pmod{n}$, then $a - c \equiv b - d \pmod{n}$.
3. **Multiplication:** If $a \equiv b \pmod{n}$ and $c \equiv d \pmod{n}$, then $a \cdot c \equiv b \cdot d \pmod{n}$.

Division under congruence involves additional subtleties compared to other mathematical operations. Unlike addition, subtraction, and multiplication, we cannot always “divide” by a number in modular arithmetic. For example, while $6 \equiv 2 \pmod{4}$, we cannot conclude that $3 \equiv 1 \pmod{4}$ by “dividing by 2.” Division is only possible when the divisor has a multiplicative inverse, which leads us to our next topic.

6.2 Multiplicative Inverses in Modular Arithmetic

In regular arithmetic, dividing by a number a is equivalent to multiplying by its reciprocal $\frac{1}{a}$. The reciprocal $\frac{1}{a}$ is called the multiplicative inverse of a because it satisfies the fundamental property $a \cdot \frac{1}{a} = 1$. This relationship allows us to transform any division into an equivalent multiplication.

Similarly, in modular arithmetic, “division” by a modulo n corresponds to multiplication by a special number called the *multiplicative inverse* of a modulo n . We denote this inverse as a^{-1} in the modular context. Just as $\frac{1}{a}$ gives 1 when multiplied by a in regular arithmetic, the

multiplicative inverse satisfies $a \cdot a^{-1} \equiv 1 \pmod{n}$.

However, unlike regular arithmetic where every non-zero real number has an inverse, not every number in modular arithmetic has a multiplicative inverse. This distinction introduces a fascinating subtlety into modular systems and leads us to investigate precisely when these inverses exist and how to compute them.

6.2.1 Definition and Existence

Definition:

A number b is called the multiplicative inverse of a modulo n if:

$$a \cdot b \equiv 1 \pmod{n}$$

We denote this inverse as $a^{-1} \pmod{n}$.

Example 6.2.1

Let's find the multiplicative inverse of 3 modulo 7.

We need to find a number b such that $3 \cdot b \equiv 1 \pmod{7}$. Testing values:

$$\begin{aligned} 3 \cdot 1 &= 3 \equiv 3 \pmod{7} \\ 3 \cdot 2 &= 6 \equiv 6 \pmod{7} \\ 3 \cdot 3 &= 9 \equiv 2 \pmod{7} \\ 3 \cdot 4 &= 12 \equiv 5 \pmod{7} \\ 3 \cdot 5 &= 15 \equiv 1 \pmod{7} \end{aligned}$$

So $3^{-1} \equiv 5 \pmod{7}$.

Not every number has a multiplicative inverse modulo n . The question of when a multiplicative inverse exists leads us to an important theorem.

Theorem: Existence Theorem for Multiplicative Inverse

Let a and n be positive integers. A multiplicative inverse of a modulo n exists if and only if a and n are coprime (i.e., $\gcd(a, n) = 1$). Moreover, when the inverse exists, it is unique modulo n .

Proof

First, let's prove that if $\gcd(a, n) = 1$, then a multiplicative inverse exists. By Bézout's identity, when $\gcd(a, n) = 1$, there exist integers x and y such that:

$$ax + ny = 1$$

Taking this equation modulo n :

$$\begin{aligned} ax + ny &\equiv 1 \pmod{n} \\ ax &\equiv 1 \pmod{n} \end{aligned}$$

Since $ny \equiv 0 \pmod{n}$. This shows that x is a multiplicative inverse of a modulo n .

For the converse, suppose a has a multiplicative inverse x modulo n , but $\gcd(a, n) = d > 1$. Then:

$$ax \equiv 1 \pmod{n}$$

This means $ax = kn + 1$ for some integer k . But since d divides both a and n , the left side of this equation must be divisible by d . This would imply that d divides 1, which is a contradiction.

For uniqueness, suppose x_1 and x_2 are both inverses of a modulo n :

$$\begin{aligned} ax_1 &\equiv 1 \pmod{n} \\ ax_2 &\equiv 1 \pmod{n} \end{aligned}$$

Subtracting: $a(x_1 - x_2) \equiv 0 \pmod{n}$. Since $\gcd(a, n) = 1$, we must have $x_1 - x_2 \equiv 0 \pmod{n}$, which means $x_1 \equiv x_2 \pmod{n}$, proving uniqueness.

□

Example 6.2.2

Let's try to find the multiplicative inverse of 6 modulo 14.

Solution: Extended Euclidean Algorithm gives us $\gcd(6, 14) = 2 \neq 1$ and hence no multiplicative inverse exists.

The existence of a multiplicative inverse modulo n for a number is directly connected to whether that number is coprime to n . In fact, the complete set of numbers less than n that have multiplicative inverses modulo n consists precisely of those numbers that are coprime to n . This

fundamental observation—that counting numbers with multiplicative inverses means counting numbers coprime to the modulus—leads us naturally to Euler’s totient function, which we will explore in the next section.

6.3 Euler’s Totient Function

Now that we understand the importance of coprime numbers in modular arithmetic, we naturally ask: how many integers in the range $[1, n]$ are coprime to n ? This count is given by a fundamental function in number theory.

6.3.1 Definition

Definition:

For a positive integer n , Euler’s totient function $\phi(n)$ counts the positive integers up to n that are coprime to n . Formally:

$$\phi(n) = |\{k : 1 \leq k \leq n - 1, \gcd(k, n) = 1\}|$$

Example 6.3.1

Let’s calculate $\phi(12)$.

Solution: The numbers from 1 to 12 are: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12. Of these, the ones coprime to 12 are: 1, 5, 7, 11.

So $\phi(12) = 4$.

6.3.2 Properties of Euler’s Totient Function

Euler’s totient function has several important properties:

Theorem: Properties of Euler’s Totient Function

For Euler’s totient function ϕ :

1. **For prime p :** $\phi(p) = p - 1$ This is because every number from 1 to $p - 1$ is

coprime to p .

2. **For prime power p^k :** $\phi(p^k) = p^k - p^{k-1} = p^k(1 - \frac{1}{p})$

This counts numbers from 1 to p^k that are not divisible by p .

3. **Multiplicative property:** If $\gcd(a, b) = 1$, then $\phi(ab) = \phi(a) \cdot \phi(b)$

This is a powerful property that allows us to compute $\phi(n)$ for any n by breaking it down into prime factors.

4. **Sum over divisors:** For any positive integer n :

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

This elegant identity relates the totient values of all divisors of n to n itself.

Let us develop an intuitive understanding of Euler's totient function. For a prime number p , it is clear that $\phi(p) = p - 1$, since every number from 1 to $p - 1$ is coprime to p . This insight can be extended to prime powers.

For p^k where p is prime and $k \geq 1$, let's examine which numbers between 1 and p^k are not coprime to p^k . These are precisely the multiples of p :

$$p, 2p, 3p, \dots, (p^{k-1})p$$

There are exactly p^{k-1} such numbers. Therefore:

$$\begin{aligned} \phi(p^k) &= p^k - p^{k-1} \\ &= p^k \left(1 - \frac{1}{p}\right) \end{aligned}$$

We can visualize this by arranging all numbers from 1 to p^k in p^{k-1} rows of p numbers each:

$$\begin{aligned} &1, 2, \dots, 0 \times p + p \\ &p + 1, p + 2, \dots, 1 \times p + p \\ &\vdots \\ &(p^{k-1} - 1)p + 1, (p^{k-1} - 1)p + 2, \dots, (p^{k-1} - 1) \times p + p \end{aligned}$$

In each row, exactly one number (the multiple of p) is not coprime to p^k . With p^{k-1} rows, we have p^{k-1} numbers that are not coprime to p^k , confirming our formula.

The multiplicative property of Euler's totient function, that $\phi(ab) = \phi(a)\phi(b)$ when $\gcd(a, b) = 1$, stems from the fact that when a and b are coprime, being coprime to ab is equivalent to

being simultaneously coprime to a and to b . This fundamental insight explains why ϕ behaves multiplicatively for coprime inputs.

Using these properties, we can calculate $\phi(n)$ for any positive integer n as follows:

1. Factor n into its prime powers: $n = p_1^{a_1} \cdot p_2^{a_2} \cdot \dots \cdot p_k^{a_k}$
2. Apply the formula:

$$\begin{aligned}\phi(n) &= \phi(p_1^{a_1}) \cdot \phi(p_2^{a_2}) \cdot \dots \cdot \phi(p_k^{a_k}) \\ &= p_1^{a_1} \left(1 - \frac{1}{p_1}\right) \cdot p_2^{a_2} \left(1 - \frac{1}{p_2}\right) \cdot \dots \cdot p_k^{a_k} \left(1 - \frac{1}{p_k}\right) \\ &= n \cdot \prod_{i=1}^k \left(1 - \frac{1}{p_i}\right)\end{aligned}$$

Example 6.3.2

Calculate $\phi(36)$.

1. Prime factorization: $36 = 2^2 \cdot 3^2$
2. Apply the formula:

$$\begin{aligned}\phi(36) &= \phi(2^2) \cdot \phi(3^2) \\ &= 2^2 \left(1 - \frac{1}{2}\right) \cdot 3^2 \left(1 - \frac{1}{3}\right) \\ &= 4 \cdot \frac{1}{2} \cdot 9 \cdot \frac{2}{3} \\ &= 2 \cdot 6 = 12\end{aligned}$$

We can verify this directly: the numbers coprime to 36 are 1, 5, 7, 11, 13, 17, 19, 23, 25, 29, 31, 35, which total 12 numbers.

Example 6.3.3

Find the number of positive integers less than or equal to 1000 that are not divisible by any of 5, 7, or 11.

Solution: The key insight is that a positive integer x is not divisible by any of 5, 7, or 11 if and only if $\gcd(x, 385) = 1$, where $385 = 5 \times 7 \times 11$.

Moreover, this property is periodic with period 385: if x is not divisible by any of 5, 7, or 11, then $x + 385$ is also not divisible by any of these primes. This is because:

- If $5 \nmid x$, then $5 \nmid (x + 385)$ since $5 \mid 385$
- If $7 \nmid x$, then $7 \nmid (x + 385)$ since $7 \mid 385$
- If $11 \nmid x$, then $11 \nmid (x + 385)$ since $11 \mid 385$

This periodicity means that in every block of 385 consecutive integers, the same number are coprime to 385. This count is exactly $\phi(385)$.

Since $385 = 5 \times 7 \times 11$ where 5, 7, and 11 are distinct primes:

$$\phi(385) = 4 \times 6 \times 10 = 240$$

Since $1000 = 2 \times 385 + 230$, we have $2 \times 240 = 480$ positive integers co-prime to 5, 7, and 11 among the first 770 integers.

The remaining 230 integers $\{771, 772, \dots, 1000\}$ is equivalent to considering $\{1, 2, \dots, 230\}$ due to periodicity.

For $\{1, 2, \dots, 230\}$, using inclusion-exclusion:

Divisible by 5:	$\lfloor 230/5 \rfloor = 46$
Divisible by 7:	$\lfloor 230/7 \rfloor = 32$
Divisible by 11:	$\lfloor 230/11 \rfloor = 20$
Divisible by both 5 and 7:	$\lfloor 230/35 \rfloor = 6$
Divisible by both 5 and 11:	$\lfloor 230/55 \rfloor = 4$
Divisible by both 7 and 11:	$\lfloor 230/77 \rfloor = 2$
Divisible by 5, 7, and 11:	$\lfloor 230/385 \rfloor = 0$

Therefore:

$$230 - 46 - 32 - 20 + 6 + 4 + 2 - 0 = 144$$

Therefore, the total count is: $480 + 144 = 624$.

We had solved this example previously using inclusion-exclusion. This totient function solution provides deeper insight by exposing the underlying periodic structure of the problem.

6.3.3 Sum of Euler's Totient Function Over Divisors

Having discussed the first three properties of Euler's totient function, let us discuss the identity $\sum_{d|n} \phi(d) = n$, which states that the sum of $\phi(d)$ over all divisors d of n equals n itself. This result may seem surprising at first, but it has a beautiful intuitive explanation.

The key insight is to realize that:

$$\sum_{d|n} \phi(d) = \sum_{d|n} \phi\left(\frac{n}{d}\right)$$

The equality holds because $d \cdot \frac{n}{d} = n$, which implies that if d is a divisor of n , then $\frac{n}{d}$ is also a divisor of n .

Consider all the integers from 1 to n . Each integer in this range has a unique greatest common divisor with n . This GCD must be a divisor of n . We can use this fact to partition the set $\{1, 2, \dots, n\}$ according to the GCD that each number shares with n .

Let's visualize this with $n = 12$. We can organize all numbers from 1 to 12 according to their GCD with 12.

GCD with 12	Numbers	Count	Corresponds to
1	1, 5, 7, 11	4	$\phi(12)$
2	2, 10	2	$\phi(6)$
3	3, 9	2	$\phi(4)$
4	4, 8	2	$\phi(3)$
6	6	1	$\phi(2)$
12	12	1	$\phi(1)$

Every integer k from 1 to n falls into exactly one of these partitions:

- If $\gcd(k, n) = 1$, then k contributes to the count $\phi(n)$
- If $\gcd(k, n) = d$ (where $d > 1$ is a divisor of n), then k contributes to the count $\phi(n/d)$

The last statement holds because: if $\gcd(k, n) = d$, then $\gcd(\frac{k}{d}, \frac{n}{d}) = 1$. In other words, $\frac{k}{d}$ is coprime to $\frac{n}{d}$ and contributes to $\phi(\frac{n}{d})$.

Let's illustrate this with a concrete example. Consider the number $k = 8$ with $n = 12$. As $\gcd(8, 12) = 4$, so $k = 8$ contributes to $\phi(12/4) = \phi(3)$ via $\frac{8}{4} = 2$. The reduced numbers 2 and 3 are coprime.

This pattern holds for all numbers from 1 to n . Each number k gets counted in exactly one $\phi(d)$ term in our sum, specifically the term $\phi\left(\frac{n}{\gcd(k,n)}\right)$. Since every number from 1 to n falls into exactly one of these partitions (having a unique GCD with n), and we've accounted for all possible GCDs (which are the divisors of n), we have:

$$n = \sum_{d|n} \phi\left(\frac{n}{d}\right)$$

For $n = 12$:

$$\begin{aligned}\phi(12) + \phi(6) + \phi(4) + \phi(3) + \phi(2) + \phi(1) &= 4 + 2 + 2 + 2 + 1 + 1 \\ &= 12\end{aligned}$$

This elegant relationship shows why $\sum_{d|n} \phi(d) = n$ holds for any positive integer n . Through a change of variables (setting $\frac{n}{d} = m$, which means as d runs through the divisors of n , so does m), we obtain the elegant formula:

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

Theorem: Sum of Euler's Totient Function Over Divisors

For any positive integer n :

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

where the sum is taken over all positive divisors d of n .

Proof

We will prove this theorem by counting the numbers from 1 to n based on their greatest common divisor (\gcd) with n .

For each divisor d of n , let's define the set:

$$A_d = \{k \in \{1, 2, \dots, n\} : \gcd(k, n) = d\}$$

These sets partition the integers from 1 to n , since every integer k in this range has a unique greatest common divisor with n , and this \gcd must be a divisor of n .

Now, we'll show that $|A_d| = \phi(n/d)$ for each divisor d of n . Consider any number k in A_d . Since $\gcd(k, n) = d$, we can write $k = d \cdot k'$ for some integer k' . Because $k \leq n$, we must have $k' \leq n/d$.

The key insight: $\gcd(k, n) = d$ if and only if $\gcd(k', n/d) = 1$. This gives us a one-to-one correspondence between numbers $k \leq n$ with $\gcd(k, n) = d$ and numbers $k' \leq n/d$ that are coprime to n/d .

By the definition of Euler's totient function, the count of such k' is exactly $\phi(n/d)$. Therefore, $|A_d| = \phi(n/d)$.

Since the sets A_d partition our original set, we can write:

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

Now comes an elegant change of variables. For each divisor d of n , let's set $m = n/d$. As d runs through all the divisors of n , so does m . This means we can rewrite our sum as:

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

Thus, we have proven that:

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

□

This elegant theorem connects the totient values of all divisors of n to n itself, revealing a beautiful structure underlying the distribution of coprime numbers.

6.4 Euler's Theorem

Having explored the elegant properties of Euler's totient function, we now turn to one of the most powerful results in number theory that builds directly upon it: Euler's theorem. This theorem establishes a profound connection between Euler's totient function and modular exponentiation.

Recall that Euler's totient function $\phi(n)$ counts the positive integers up to n that are coprime to n . We've seen that these numbers have special properties—they are precisely the numbers that have multiplicative inverses modulo n . This connection between coprime numbers and their multiplicative inverses leads us to Euler's theorem.

6.4.1 Statement and Motivation

Euler's theorem tells us something remarkable about what happens when we repeatedly multiply a number by itself modulo n . Specifically, it reveals a cyclical pattern that emerges when we work with numbers coprime to the modulus.

Theorem: Euler's Theorem

If a and n are coprime positive integers, then:

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

This means that if we take any number a that is coprime to n , and raise it to the power of $\phi(n)$, the result will be congruent to 1 modulo n . Euler's Theorem implies that $a \cdot a^{\phi(n)-1} \equiv 1 \pmod{n}$, in other words, $a^{\phi(n)-1}$ is a multiplicative inverse of a modulo n .

6.4.2 Intuitive Understanding of Euler's Theorem

Euler's theorem states that if a and n are coprime positive integers, then $a^{\phi(n)} \equiv 1 \pmod{n}$. While the formal proof is elegant, developing an intuitive understanding helps us grasp the deeper meaning behind this powerful result.

The Cyclic Nature of Powers

Let's build our intuition through a simple example. Consider $n = 8$. The numbers coprime to 8 are 1, 3, 5, and 7, so $\phi(8) = 4$. Let's examine what happens when we take successive powers of $a = 3$ modulo 8:

$$\begin{aligned} 3^1 &= 3 \equiv 3 \pmod{8} \\ 3^2 &= 9 \equiv 1 \pmod{8} \\ 3^3 &= 27 \equiv 3 \pmod{8} \\ 3^4 &= 81 \equiv 1 \pmod{8} \end{aligned}$$

Notice the cyclical pattern: 3, 1, 3, 1, ... with period 2. While $\phi(8) = 4$, we find that $3^2 \equiv 1 \pmod{8}$, showing that the cycle length can be a divisor of $\phi(n)$.

Now let's try $a = 5$ modulo 8:

$$\begin{aligned} 5^1 &= 5 \pmod{8} \\ 5^2 &= 25 \equiv 1 \pmod{8} \\ 5^3 &= 125 \equiv 5 \pmod{8} \\ 5^4 &= 625 \equiv 1 \pmod{8} \end{aligned}$$

Again, we see a cycle with period 2: 5, 1, 5, 1, ...

For $a = 7$ modulo 8:

$$\begin{aligned} 7^1 &= 7 \pmod{8} \\ 7^2 &= 49 \equiv 1 \pmod{8} \\ 7^3 &= 343 \equiv 7 \pmod{8} \\ 7^4 &= 2401 \equiv 1 \pmod{8} \end{aligned}$$

We observe the same cyclic behavior with period 2: 7, 1, 7, 1, ...

This cyclic nature is at the heart of Euler's theorem. The theorem guarantees that for any number a coprime to n , repeatedly multiplying by a modulo n will eventually return to 1, and this must happen after at most $\phi(n)$ steps.

Before delving further into Euler's Theorem, let us provide a formal definition for order in modular arithmetic.

Definition: Cycle and Order in Modular Arithmetic

*The **order** of a modulo n , denoted $\text{ord}_n(a)$, is the smallest positive integer k such that $a^k \equiv 1 \pmod{n}$. This represents the length of the cycle of powers of a modulo n .*

The Rearrangement Perspective

Another insightful way to understand Euler's theorem comes from examining what happens when we multiply all numbers coprime to n by a fixed number a that is also coprime to n .

For example, with $n = 8$, the set of numbers coprime to 8 is $\{1, 3, 5, 7\}$. Let's multiply each by $a = 3$ and reduce modulo 8:

$$\begin{aligned} 1 \cdot 3 &\equiv 3 \pmod{8} \\ 3 \cdot 3 &\equiv 9 \equiv 1 \pmod{8} \\ 5 \cdot 3 &\equiv 15 \equiv 7 \pmod{8} \\ 7 \cdot 3 &\equiv 21 \equiv 5 \pmod{8} \end{aligned}$$

Remarkably, we get $\{3, 1, 7, 5\}$, which is just a rearrangement of the original set $\{1, 3, 5, 7\}$. This key insight shows that multiplication by a coprime number creates a permutation of the residue classes.

This isn't a coincidence! When we multiply all residues coprime to n by a (which is also coprime to n), we simply get a rearrangement of these residues modulo n . This happens because:

- Multiplication by a preserves coprimality with n , so each product is still coprime to n
- The products must all be distinct modulo n (if two were congruent, for example if $5 \cdot 3 \equiv 7 \cdot 3 \pmod{8}$, then we could multiply both sides by 3^{-1} since $\gcd(3, 8) = 1$ to get $5 \equiv 7 \pmod{8}$, which is false)

This rearrangement property leads directly to Euler's theorem. If we multiply all numbers in the set, we get:

$$(1 \cdot 3 \cdot 5 \cdot 7) \cdot a^{\phi(8)} \equiv (3 \cdot 1 \cdot 7 \cdot 5) \pmod{8}$$

$$(1 \cdot 3 \cdot 5 \cdot 7) \cdot a^{\phi(8)} \equiv (1 \cdot 3 \cdot 5 \cdot 7) \pmod{8}$$

Since the product $(1 \cdot 3 \cdot 5 \cdot 7)$ is coprime to 8, we can divide both sides by it to obtain:

$$a^{\phi(8)} \equiv 1 \pmod{8}$$

Why the Cycle Length Divides $\phi(n)$

Our exploration of Euler's theorem has revealed two complementary perspectives: the cyclic nature of powers and the rearrangement of residues. These perspectives are intimately connected, and together they explain why the cycle length must divide $\phi(n)$.

We've seen that after reaching this point, the sequence of powers repeats:

$$a^{\text{ord}_n(a)} \equiv 1 \pmod{n} \tag{6.4.1}$$

$$a^{\text{ord}_n(a)+1} \equiv a \pmod{n} \tag{6.4.2}$$

$$a^{\text{ord}_n(a)+2} \equiv a^2 \pmod{n} \tag{6.4.3}$$

and so on, creating a cycle of length $\text{ord}_n(a)$.

Using the rearrangement principle, we saw earlier that $a^{\phi(n)} \equiv 1 \pmod{n}$. As $\text{ord}_n(a)$ is the smallest integer such that $a^{\text{ord}_n(a)} \equiv 1 \pmod{n}$, hence $\text{ord}_n(a) \mid \phi(n)$.

Proof of Euler's Theorem

Having developed an intuition behind Euler's Theorem, let us provide a formal proof to the theorem.

Proof

Let $R = \{r_1, r_2, \dots, r_{\phi(n)}\}$ be the complete set of residues modulo n that are coprime to n .

Consider the set $S = \{ar_1 \bmod n, ar_2 \bmod n, \dots, ar_{\phi(n)} \bmod n\}$.

We claim that S is a rearrangement of R modulo n . To prove this:

1. Each element in S is coprime to n (since multiplication by a preserves coprimality when $\gcd(a, n) = 1$).
2. The elements of S are distinct modulo n . If $ar_i \equiv ar_j \pmod{n}$ for some $i \neq j$, then multiplying both sides by a^{-1} would give $r_i \equiv r_j \pmod{n}$, which contradicts the definition of R .

Therefore, S must be a rearrangement of R modulo n . Multiplying all congruences:

$$\begin{aligned} a \cdot r_1 &\equiv r'_1 \pmod{n} \\ a \cdot r_2 &\equiv r'_2 \pmod{n} \\ &\vdots \\ a \cdot r_{\phi(n)} &\equiv r'_{\phi(n)} \pmod{n} \end{aligned}$$

We get:

$$a^{\phi(n)} \cdot (r_1 \cdot r_2 \cdot \dots \cdot r_{\phi(n)}) \equiv r_1 \cdot r_2 \cdot \dots \cdot r_{\phi(n)} \pmod{n}$$

Since the product of the residues is coprime to n , we can divide both sides by this product to obtain:

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

□

As discussed previously, Euler's Theorem can be used to get multiplicative inverse of an integer a modulo n .

Example 6.4.1

Let's find the multiplicative inverse of 7 modulo 15 using Euler's Theorem.

Solution: First we verify that 7 and 15 are coprime, i.e., $\gcd(7, 15) = 1$. Also, calculate $\phi(15) = \phi(3) \times \phi(5) = 2 \times 4 = 8$.

According to Euler's Theorem, when $\gcd(a, n) = 1$, we have $a^{\phi(n)} \equiv 1 \pmod{n}$. Multiplying both sides by a^{-1} , we get $a^{\phi(n)-1} \equiv a^{-1} \pmod{n}$. This implies that $7^{-1} \equiv 7^{\phi(15)-1} \equiv 7^{8-1} \equiv 7^7 \pmod{15}$.

We need to calculate $7^7 \pmod{15}$ efficiently using exponentiation by squaring.

$$7^1 = 7$$

$$7^2 = 49 \equiv 4 \pmod{15}$$

$$7^4 = (7^2)^2 = 4^2 = 16 \equiv 1 \pmod{15}$$

$$7^7 = 7^4 \times 7^2 \times 7^1 = 1 \times 4 \times 7 = 28 \equiv 13 \pmod{15}$$

Therefore, $7^{-1} \equiv 13 \pmod{15}$. In this example, we could have stopped when we discovered that $7^4 \equiv 1 \pmod{15}$. This tells us that $\text{ord}_{15}(7) = 4$, which divides $\phi(15) = 8$ as expected. We could then focus directly on 7^3 to find the inverse as $7^{-1} \equiv 7^3 \equiv 13 \pmod{15}$.

6.4.3 Fermat's Little Theorem as a Special Case

Theorem: Fermat's Little Theorem

If p is a prime number and a is not divisible by p , then:

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

This is a direct consequence of Euler's theorem since for a prime p , $\phi(p) = p - 1$.

Fermat's Little Theorem was discovered by Pierre de Fermat around 1636, nearly a century before Euler's more general result. In his famous marginal note, Fermat claimed to have a proof but did not provide it. The first known published proof was given by Leibniz in 1683. Euler later discovered his more general theorem, which includes Fermat's result as a special case.

Another useful formulation of Fermat's Little Theorem states that for any integer a and prime p :

$$a^p \equiv a \pmod{p}$$

This version applies to all integers a , not just those coprime to p . When p divides a , both sides are congruent to 0 modulo p . When p does not divide a , we can divide both sides by a to obtain the original statement.

Example 6.4.2

Let's verify Fermat's Little Theorem with $p = 11$ and $a = 7$.

Solution: The theorem states that $7^{10} \equiv 1 \pmod{11}$.

We'll compute this step-by-step using modular exponentiation:

$$\begin{aligned} 7^1 &\equiv 7 & (\text{mod } 11) \\ 7^2 &\equiv 49 \equiv 5 & (\text{mod } 11) \\ 7^4 &\equiv 5^2 \equiv 25 \equiv 3 & (\text{mod } 11) \\ 7^8 &\equiv 3^2 \equiv 9 & (\text{mod } 11) \\ 7^{10} &\equiv 7^8 \cdot 7^2 \equiv 9 \cdot 5 \equiv 45 \equiv 1 & (\text{mod } 11) \end{aligned}$$

Indeed, $7^{10} \equiv 1 \pmod{11}$, confirming Fermat's Little Theorem.

This result makes calculations modulo 11 much easier. For example, to compute $7^{123} \pmod{11}$, we can use:

$$\begin{aligned} 7^{123} &= 7^{12 \cdot 10 + 3} = (7^{10})^{12} \cdot 7^3 \equiv 1^{12} \cdot 7^3 & (\text{mod } 11) \\ &\equiv 7^3 & (\text{mod } 11) \\ &\equiv 7 \cdot 7^2 & (\text{mod } 11) \\ &\equiv 7 \cdot 5 & (\text{mod } 11) \\ &\equiv 35 & (\text{mod } 11) \\ &\equiv 2 & (\text{mod } 11) \end{aligned}$$

Without Fermat's Little Theorem, computing 7^{123} directly would be far more challenging.

Fermat's Little Theorem forms the basis of the Fermat primality test, which works as follows: given an integer n to test for primality, choose a random integer a with $1 < a < n$ and compute $a^{n-1} \pmod{n}$. If the result is not 1, then n is definitely composite. If the result is 1, then n may be prime, or it may be a pseudoprime to base a .

This test is probabilistic in nature but can be repeated with different bases to increase confidence. While there exist composite numbers (called Carmichael numbers) that pass the test for all bases coprime to themselves, these are relatively rare, making the test useful in practice when combined with other methods.

6.5 Wilson's Theorem

Wilson's theorem gives us another fascinating property about prime numbers in the context of modular arithmetic.

Theorem: Wilson's Theorem

A positive integer $p > 1$ is prime if and only if:

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

Before presenting a formal proof, let's develop an intuitive understanding of why Wilson's Theorem holds. This beautiful result has a surprisingly elegant explanation that makes it feel almost obvious once you see it.

Let's explore what happens when we multiply all numbers from 1 to $p - 1$ modulo p :

$$(p - 1)! = 1 \cdot 2 \cdot 3 \cdot \dots \cdot (p - 1)$$

The magic appears when we recognize that most of these numbers can be paired up in a special way. For any number a between 1 and $p - 1$, there exists its modular multiplicative inverse a^{-1} such that:

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

Importantly, this inverse a^{-1} is also a number between 1 and $p - 1$. This means that most of the factors in $(p - 1)!$ can be arranged in pairs, with each pair multiplying to give 1 modulo p .

Example 6.5.1

Find the multiplicative inverse of all the natural numbers modulo 7.

Solution: Let's examine this with $p = 7$. The numbers from 1 to 6 pair up as follows:

Number	Its Inverse	Their Product mod 7
1	1	$1 \equiv 1 \pmod{7}$
2	4	$2 \cdot 4 = 8 \equiv 1 \pmod{7}$
3	5	$3 \cdot 5 = 15 \equiv 1 \pmod{7}$
6	6	$6 \cdot 6 = 36 \equiv 1 \pmod{7}$

Notice how each number pairs with its multiplicative inverse, and each pair multiplies to 1 modulo 7.

As we see in the previous example, some numbers are their own multiplicative inverses—specifically those that satisfy $a^2 \equiv 1 \pmod{p}$. These are precisely 1 and $p - 1$.

- $1 \cdot 1 = 1$, so 1 is its own inverse.

- $(p - 1)^2 = p^2 - 2p + 1 \equiv 1 \pmod{p}$, so $p - 1$ is its own inverse.

For any prime $p > 2$, it can be proven that these are the only two self-inverse elements. For all other numbers from 2 to $p - 2$, each pairs with a distinct number different from itself.

Now, when we multiply all numbers from 1 to $p - 1$, every pair of distinct inverses contributes a factor of 1 to the product. The only elements that don't pair up with distinct numbers are 1 and $p - 1$. Therefore:

$$(p - 1)! \equiv 1 \cdot (p - 1) \cdot \prod_{\text{pairs}} 1 \equiv p - 1 \equiv -1 \pmod{p}$$

It's like a dance where everyone finds a partner, leaving only 1 and $p - 1$ unpaired. These two "leftover" elements multiply to give us our result of -1 .

Example 6.5.2

Confirm Wilson's Theorem for $p = 7$.

Solution: For $p = 7$, we have:

$$\begin{aligned} 6! &= 1 \cdot 2 \cdot 3 \cdot 4 \cdot 5 \cdot 6 \\ &= 1 \cdot (2 \cdot 4) \cdot (3 \cdot 5) \cdot 6 \\ &\equiv 1 \cdot 1 \cdot 1 \cdot 6 \pmod{7} \\ &\equiv 6 \pmod{7} \\ &\equiv -1 \pmod{7} \end{aligned}$$

This confirms Wilson's Theorem for $p = 7$.

The beauty of Wilson's Theorem is that it works for any prime p . The pattern is always the same: numbers between 2 and $p - 2$ pair up to give products congruent to 1, while 1 and $p - 1$ remain, multiplying to give $p - 1 \equiv -1 \pmod{p}$.

For the case where $p = 2$, we have $(2 - 1)! = 1! = 1 \equiv -1 \pmod{2}$, so the theorem holds in this case as well. Now that we have an intuitive understanding of why Wilson's Theorem holds, let's proceed to a more formal proof.

Proof

Part 1: Assume p is prime. Show that $(p - 1)! \equiv -1 \pmod{p}$.

First, let's test a small prime. If $p = 2$, $(2 - 1)! = 1! = 1$. Since $1 \equiv -1 \pmod{2}$ (because $1 - (-1) = 2$ is divisible by 2), the theorem holds for $p = 2$.

Now, let p be a prime number greater than 2. Consider the set of numbers $S = \{1, 2, 3, \dots, p - 1\}$. The existence theorem for multiplicative inverse tells that each of the integers in set S has a unique multiplicative inverse that also exists in the set.

Let's start with finding the integers in S which are their own multiplicative inverse. This means we have to find a in S that satisfy $a^2 \equiv 1 \pmod{p}$. This congruence means $a^2 - 1$ must be divisible by p . Factoring gives $(a + 1)(a - 1)$. So, we need p to divide the product $(a + 1)(a - 1)$. Since p is prime, if it divides a product, it must divide at least one of the factors. So, either p divides $(a - 1)$ or p divides $(a + 1)$.

- If $p \mid (a - 1)$: Since a is in $S = \{1, 2, \dots, p - 1\}$, we know $0 \leq a - 1 \leq p - 2$. The only multiple of p in this range is 0. So, $a - 1 = 0$, which gives $a = 1$.

- If $p \mid (a + 1)$: Since a is in S , we know $2 \leq a + 1 \leq p$. The only multiple of p in this range is p itself. So, $a + 1 = p$, which gives $a = p - 1$.

Therefore, the only numbers in the set $\{1, 2, \dots, p - 1\}$ that are their own multiplicative inverses modulo p are 1 and $p - 1$.

Now look at the factorial $(p - 1)! = 1 \times 2 \times 3 \times \dots \times (p - 1)$.

The numbers 1 and $p - 1$ are special (they are their own inverses). All the other numbers in the list, from 2 up to $p - 2$, can be paired up. Each number a in this middle section $\{2, 3, \dots, p - 2\}$ has a unique inverse b ($a \neq b$) which is also in this section. The product of each such pair $a \times b$ is congruent to 1 modulo p .

So, when we calculate the product $(p - 1)!$ modulo p , we have:

$$(p - 1)! = 1 \times (p - 1) \times (\text{product of all numbers from 2 to } p - 2)$$

Modulo p , the product of all numbers from 2 to $p - 2$ becomes a product of pairs, where each pair gives 1:

$$(p - 1)! \equiv 1 \times (p - 1) \times (\underbrace{1 \times 1 \times \dots \times 1}_{\text{from the pairs}}) \pmod{p}$$

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

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

This proves the first direction.

Part 2: Assume $(p - 1)! \equiv -1 \pmod{p}$. **Show that p must be prime.**

We start by assuming the condition $(p - 1)! \equiv -1 \pmod{p}$ holds for some integer $p > 1$. We want to prove that p must be prime.

We use proof by contradiction. We assume that p is composite. If p is composite, then it must have a divisor d such that $1 < d < p$. Since d is a divisor of p and is smaller than p , d must be one of the numbers in the list $\{1, 2, \dots, p - 1\}$. This means $(p - 1)!$ is divisible by d . We write this as $d | (p - 1)!$.

Now, we use the given information: $(p - 1)! \equiv -1 \pmod{p}$. This means that $(p - 1)! + 1$ is divisible by p . We write this as $p | ((p - 1)! + 1)$. If $d | p$ and $p | (p - 1)! + 1$, implies $d | ((p - 1)! + 1)$.

Thus, we have $d | (p - 1)!$ and $d | (p - 1)! + 1$. If a number d divides two integers, it must also divide their difference. So, d must divide 1, which implies $d = 1$.

But this contradicts our choice of d . We chose d to be a divisor of p such that $1 < d < p$. The conclusion $d = 1$ contradicts $d > 1$.

Since our assumption that “ p is composite” leads to a contradiction, this assumption must be false. Therefore, p cannot be composite. Since $p > 1$, it must be prime. This proves the second direction.

□

6.6 Wilson’s Theorem for Composite Number

Wilson’s Theorem can be generalized to composite numbers. Let us look at the theorem statement.

Theorem: Wilson’s Theorem for Composite Numbers

Let $n > 1$ be a composite integer.

1. If $n = 4$, then $(n - 1)! \equiv 2 \pmod{4}$.
2. If $n > 4$, then $(n - 1)! \equiv 0 \pmod{n}$.

When n is composite and greater than 4, we can write $n = ab$ where $1 < a \leq b < n$. The factorial $(n - 1)!$ includes all integers from 1 to $(n - 1)$ as factors.

Case 1: $n = ab$ where $a \neq b$

Since $1 < a < b < n$, both a and b appear as factors in the product:

$$(n - 1)! = 1 \times 2 \times \cdots \times a \times \cdots \times b \times \cdots \times (n - 1)$$

Therefore, $(n - 1)!$ is divisible by $a \times b = n$, making $(n - 1)! \equiv 0 \pmod{n}$.

Case 2: $n = a^2$ where $a > 2$

When $n = a^2$ and $a > 2$, we need to verify that $(n - 1)!$ is divisible by $n = a^2$.

For $a > 2$, we can show that both a and $2a$ appear in the factorial:

$$a < a^2 - 1 \quad (\text{since } a > 1) \quad (6.6.1)$$

$$2a < a^2 \quad (\text{since } a > 2 \text{ implies } a^2 - 2a > 0) \quad (6.6.2)$$

Since both a and $2a$ are factors in $(n - 1)!$, the factorial is divisible by $a \times 2a = 2a^2 = 2n$. This means $(n - 1)!$ is certainly divisible by n , so $(n - 1)! \equiv 0 \pmod{n}$.

Why $n = 4$ is special

The case $n = 4 = 2^2$ is special because when $a = 2$, we have $2a = 4 = n$, which means $2a$ does *not* appear as a factor in $(n - 1)! = 3!$. The factorial only contains $a = 2$ once, not $a^2 = 4$ times, so it isn't divisible by $n = 4$.

Instead, we get $(4 - 1)! = 3! = 6 \equiv 2 \pmod{4}$.

The complete picture

Wilson's Theorem and its composite counterpart give us an elegant characterization of $(n - 1)!$ \pmod{n} :

$$(n - 1)! \equiv \begin{cases} -1 \pmod{n} & \text{if } n \text{ is prime} \\ 2 \pmod{n} & \text{if } n = 4 \\ 0 \pmod{n} & \text{if } n > 4 \text{ is composite} \end{cases}$$

This provides a theoretical (though computationally impractical for large n) primality test: calculate $(n - 1)!$ \pmod{n} and check if it equals -1 . The insights we developed while decoding Wilson's Theorem can be used to derive interesting results. We will look at one such result.

Example 6.6.1

Let $p > 3$ be a prime number. Prove that the sum of multiplicative inverse modulo p satisfies:

$$\sum_{i=1}^{p-1} i^{-1} \equiv 0 \pmod{p}$$

Solution: For each integer i such that $1 \leq i \leq p - 1$, since p is prime, i is not divisible by p . Therefore, i has a unique multiplicative inverse modulo p , denoted by i^{-1} , such that $i \cdot i^{-1} \equiv 1 \pmod{p}$. This inverse i^{-1} will also be an integer in the set $\{1, 2, \dots, p - 1\}$.

Consider the set of integers $A = \{1, 2, \dots, p - 1\}$. Consider the set of their multiplicative inverses modulo p , $B = \{1^{-1}, 2^{-1}, \dots, (p - 1)^{-1}\}$.

We claim that the set B is simply a rearrangement of the set A . To show this, we need to show that the mapping $f : A \rightarrow A$ defined by $f(i) = i^{-1} \pmod{p}$ is a bijection (a one-to-one correspondence).

1. **The map f sends elements of A to elements of A :** If $i \in A$, then $i \not\equiv 0 \pmod{p}$. Its inverse i^{-1} exists and $i^{-1} \not\equiv 0 \pmod{p}$. So i^{-1} must be congruent to one of the integers in $\{1, 2, \dots, p-1\}$, which is set A .
2. **The map f is one-to-one (injective):** Suppose $f(i) \equiv f(j) \pmod{p}$ for some $i, j \in A$. This means $i^{-1} \equiv j^{-1} \pmod{p}$. Multiplying both sides by $i \cdot j$ (which is not divisible by p), we get:

$$(i \cdot j) \cdot i^{-1} \equiv (i \cdot j) \cdot j^{-1} \pmod{p} \quad (6.6.3)$$

$$(j \cdot i) \cdot i^{-1} \equiv (i \cdot j) \cdot j^{-1} \pmod{p} \quad (6.6.4)$$

$$j \cdot (i \cdot i^{-1}) \equiv i \cdot (j \cdot j^{-1}) \pmod{p} \quad (6.6.5)$$

$$j \cdot 1 \equiv i \cdot 1 \pmod{p} \quad (6.6.6)$$

$$j \equiv i \pmod{p} \quad (6.6.7)$$

Since $i, j \in \{1, 2, \dots, p-1\}$, this implies $i = j$. Thus, the map is one-to-one.

Since f is a one-to-one map from the finite set A to itself, it must also be onto (surjective), and therefore it is a bijection. This means that the set of inverses $B = \{1^{-1}, 2^{-1}, \dots, (p-1)^{-1}\}$ contains exactly the same elements as the set $A = \{1, 2, \dots, p-1\}$, just possibly in a different order.

Therefore, the sum of the elements in set B must be congruent to the sum of the elements in set A , modulo p :

$$\sum_{i=1}^{p-1} i^{-1} \equiv \sum_{i=1}^{p-1} i \pmod{p}$$

Now, we evaluate the sum $\sum_{i=1}^{p-1} i$. This is the sum of an arithmetic progression:

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

Since $p > 3$, p is an odd prime. This means $p-1$ is an even integer. Let $p-1 = 2k$ for some integer $k = \frac{p-1}{2}$. Substituting this into the sum:

$$\sum_{i=1}^{p-1} i = \frac{(2k)p}{2} = kp$$

Since k is an integer, the sum $1 + 2 + \cdots + (p - 1)$ is an integer multiple of p . Therefore, the sum is congruent to 0 modulo p :

$$\sum_{i=1}^{p-1} i \equiv 0 \pmod{p}$$

Combining our findings, we have shown that for any prime $p > 3$:

$$\sum_{i=1}^{p-1} i^{-1} \equiv \sum_{i=1}^{p-1} i \equiv 0 \pmod{p}$$

From Euler's foundational work on the phi function through Wilson's elegant characterization of primality via factorials, this chapter has explored the beautiful interplay between multiplicative patterns and number-theoretic structures. These theorems not only provide powerful tools for solving congruences but also reveal the hidden symmetries that make number theory so fascinating. As we continue our journey, these results will serve as building blocks for more advanced topics in the mathematics of patterns and primes.

6.7 Polynomial Congruences

In our study of modular arithmetic, we frequently encounter equations of the form $f(x) \equiv 0 \pmod{n}$, where $f(x)$ is a polynomial. These are called polynomial congruences, and they form a critical area of number theory.

Definition: Polynomial Congruence

A polynomial congruence is an equation of the form

$$f(x) \equiv 0 \pmod{n} \tag{6.7.1}$$

where $f(x) = a_n x^n + a_{n-1} x^{n-1} + \cdots + a_1 x + a_0$ is a polynomial with integer coefficients, and n is a positive integer.

A solution to this congruence is an integer a such that $f(a) \equiv 0 \pmod{n}$. Since we're working in modular arithmetic, we consider solutions modulo n , meaning that if a is a solution, then any $b \equiv a \pmod{n}$ is also a solution.

More generally, a polynomial congruence can be written as $f(x) \equiv g(x) \pmod{n}$, which is equivalent to $f(x) - g(x) \equiv 0 \pmod{n}$. Before we delve into the core results of polynomial congruences, let us look at the following proposition.

Proposition 6.1. *If $f(x) \equiv g(x) \pmod{n}$ for all x (meaning the coefficients are congruent modulo n), then $f(a) \equiv g(a) \pmod{n}$ for any integer a .*

This might seem obvious, but it establishes an important connection between polynomial congruence and congruence of their values. A more interesting question arises: how many solutions can a polynomial congruence have? The answer depends critically on the modulus, particularly whether it's prime or composite.

6.7.1 Solutions to Polynomial Congruences Modulo a Prime

We now present one of the most fundamental results regarding polynomial congruences with prime moduli.

Theorem: Number of Solutions to Polynomial Congruences

Let p be prime and consider a polynomial $f(x) = a_nx^n + a_{n-1}x^{n-1} + \cdots + a_1x + a_0$ where the coefficients are integers. If $a_n \not\equiv 0 \pmod{p}$ (so that $f(x)$ has degree exactly n when considered modulo p), then the congruence $f(x) \equiv 0 \pmod{p}$ has at most n solutions modulo p .

This theorem is analogous to the Fundamental Theorem of Algebra, which states that a polynomial of degree n over the complex numbers has exactly n roots (counting multiplicities). In the modular setting with prime modulus, we get a similar bound, although we may have fewer than n solutions.

Proof

Let's prove this by induction on the degree n of the polynomial.

Base Case: For $n = 1$, we have a linear polynomial $f(x) = a_1x + a_0$ where $a_1 \not\equiv 0 \pmod{p}$. Since p is prime, a_1 has a multiplicative inverse modulo p . Therefore, we can solve for x :

$$a_1x + a_0 \equiv 0 \pmod{p} \quad (6.7.2)$$

$$a_1x \equiv -a_0 \pmod{p} \quad (6.7.3)$$

$$x \equiv -a_0 \cdot a_1^{-1} \pmod{p} \quad (6.7.4)$$

Thus, a linear polynomial has exactly one solution modulo p , which is at most $n = 1$, confirming our base case.

Inductive Hypothesis: Assume that for all polynomials of degree less than n , the theorem holds (i.e., a polynomial of degree $k < n$ has at most k solutions modulo p).

Inductive Step: Now consider a polynomial $f(x)$ of degree n where $a_n \not\equiv 0 \pmod{p}$. Suppose, for contradiction, that $f(x) \equiv 0 \pmod{p}$ has more than n solutions. Let r_1, r_2, \dots, r_{n+1} be $n+1$ distinct solutions.

For a root r_1 of $f(x)$, using remainder theorem, we can rewrite $f(x)$ as:

$$f(x) = (x - r_1)g(x) + f(r_1) \quad (6.7.5)$$

where $g(x)$ is a polynomial of degree $n-1$ with integer coefficients.

Because $f(r_1) \equiv 0 \pmod{p}$:

$$f(x) \equiv (x - r_1)g(x) \pmod{p} \quad (6.7.6)$$

For any other root r_j where $j > 1$, we have:

$$f(r_j) \equiv 0 \pmod{p} \quad (6.7.7)$$

$$(r_j - r_1)g(r_j) \equiv 0 \pmod{p} \quad (6.7.8)$$

Since $r_j \not\equiv r_1 \pmod{p}$ (as they are distinct solutions), and p is prime, we must have $g(r_j) \equiv 0 \pmod{p} \forall j = 2, 3, \dots, n+1$.

Therefore, r_2, r_3, \dots, r_{n+1} are all solutions to $g(x) \equiv 0 \pmod{p}$. This gives us n distinct solutions to a polynomial of degree $n-1$, which contradicts our inductive hypothesis.

Hence, our assumption was wrong, and $f(x) \equiv 0 \pmod{p}$ can have at most n solutions modulo p .

□

Let's illustrate this theorem with a concrete example.

Example 6.7.1

Consider the polynomial $f(x) = x^3 - x + 1$ modulo 7. Let's find all solutions to $f(x) \equiv 0 \pmod{7}$.

Solution: We can compute $f(x)$ for each $x \in \{0, 1, 2, 3, 4, 5, 6\}$:

$$f(0) = 0^3 - 0 + 1 \equiv 1 \pmod{7} \quad (6.7.9)$$

$$f(1) = 1^3 - 1 + 1 \equiv 1 \pmod{7} \quad (6.7.10)$$

$$f(2) = 2^3 - 2 + 1 \equiv 8 - 2 + 1 \equiv 0 \pmod{7} \quad (6.7.11)$$

$$f(3) = 3^3 - 3 + 1 \equiv 27 - 3 + 1 \equiv 25 \equiv 4 \pmod{7} \quad (6.7.12)$$

$$f(4) = 4^3 - 4 + 1 \equiv 64 - 4 + 1 \equiv 61 \equiv 5 \pmod{7} \quad (6.7.13)$$

$$f(5) = 5^3 - 5 + 1 \equiv 125 - 5 + 1 \equiv 121 \equiv 2 \pmod{7} \quad (6.7.14)$$

$$f(6) = 6^3 - 6 + 1 \equiv 216 - 6 + 1 \equiv 211 \equiv 1 \pmod{7} \quad (6.7.15)$$

We find that $f(2) \equiv 0 \pmod{7}$, so $x = 2$ is the only solution. This aligns with our theorem: a polynomial of degree 3 can have at most 3 solutions modulo a prime, but may have fewer.

An important application of this theorem is the following proposition.

Proposition 6.2. *For any prime p and integer $d \geq 1$, the congruence $x^d \equiv 1 \pmod{p}$ has at most d distinct solutions modulo p .*

The proposition is a direct corollary of the previous theorem. We consider the polynomial $f(x) = x^d - 1$. Since $f(x)$ has degree d and the leading coefficient is 1 (which is non-zero modulo p), the congruence $f(x) \equiv 0 \pmod{p}$ has at most d solutions. But this congruence is equivalent to $x^d \equiv 1 \pmod{p}$, establishing the proposition.

6.8 Practice Exercises

Exercise 6.1

Consider the following statement regarding systems of linear congruences.

Let n_1, n_2, \dots, n_k be positive integers that are pairwise coprime (i.e., $\gcd(n_i, n_j) = 1$ for any $i \neq j$). Then for any integers a_1, a_2, \dots, a_k , the system of simultaneous congruences:

$$x \equiv a_1 \pmod{n_1} \tag{6.8.1}$$

$$x \equiv a_2 \pmod{n_2} \tag{6.8.2}$$

$$\vdots \tag{6.8.3}$$

$$x \equiv a_k \pmod{n_k} \tag{6.8.4}$$

has a unique solution for x modulo the product $N = n_1 n_2 \cdots n_k$.

This fundamental result in number theory is known as the **Chinese Remainder Theorem**. Provide a proof for this theorem.

Exercise 6.2

Prove that for coprime positive integers a and b , $\phi(ab) = \phi(a) \cdot \phi(b)$.

Exercise 6.3

Find all values of $n > 1$ for which $\phi(n) = n - 1$. Prove your answer.

Exercise 6.4

Let p be a prime number. Prove that for any integer a coprime to p :

$$a^{p-1} - 1 \equiv 0 \pmod{p^2}$$

if and only if $a^p - a \equiv 0 \pmod{p^2}$.

Exercise 6.5

Prove or disprove the following statement for any positive integer $n > 2$:

$$\sum_{d|n} \frac{\phi(d)}{d} = \prod_{p|n} \left(1 + \frac{1}{p}\right)$$

where the product is taken over all distinct prime divisors p of n .

Exercise 6.6

Prove that for any positive integer n :

$$\sum_{k=1}^n \phi(k) \left\lfloor \frac{n}{k} \right\rfloor = \frac{n(n+1)}{2}$$

Exercise 6.7

Prove that an integer n is prime if and only if $\sigma(n) + \phi(n) = n\tau(n)$, where $\sigma(n)$ is the sum of divisors of n and $\tau(n)$ is the number of divisors of n .

Exercise 6.8

Prove that for any odd prime p : $\left(\left(\frac{p-1}{2}\right)!\right)^2 \equiv 1 \pmod{p}$ if and only if $p \equiv 3 \pmod{4}$.

Exercise 6.9

Prove that if p is a prime number and a is an integer not divisible by p , then

$$a^{\phi(p^k)} \equiv 1 \pmod{p^k}$$

for any positive integer k . Use this to find the last two digits of 3^{1000} .

Exercise 6.10

Consider the polynomial $f(x) = x^p - x$ where p is a prime number. Prove that $f(x) \equiv 0 \pmod{p}$ has exactly p solutions, and these solutions are precisely the residue classes

$\{0, 1, 2, \dots, p - 1\}$ modulo p .

Chapter 7

The Möbius Function



August Ferdinand Möbius (1790-1868)

August Ferdinand Möbius was a German mathematician and astronomer who made fundamental contributions to number theory, geometry, and topology that continue to influence mathematics today. Born in Schulpforta, Prussia, Möbius studied under Carl Friedrich Gauss at the University of Göttingen. Möbius spent most of his career as a professor at the University of Leipzig, where in 1832 he discovered the elegant inversion formula that allows one to systematically reverse summation processes involving divisors. This work laid crucial groundwork for modern analytic number theory.

Möbius is known for discovering the one-sided surface known as the Möbius strip in 1858, a remarkable object that has only one side and one edge. Despite discovering the famous strip, he was reportedly quite terrible at paper crafts and hands-on demonstrations. His students often had to help him construct the very geometric models he had theoretically discovered, leading to the amusing sight of a brilliant mathematician being tutored in paper-folding by his undergraduate students.

In this chapter, we explore one of the most elegant tools in number theory: the Möbius function and its associated inversion formula. We have seen in the previous chapter that a function when summed over multiple values (divisors of a natural number, for instance) can lead to interesting properties. For instance, the identity function $f(n) = n$ can be broken down into a sum of Euler's totient function over the divisors of n . A natural question arises: can the summation be reversed so that Euler's totient function can be written as a sum of the identity function? More generally, if $F(n) = \sum_{d|n} f(d)$, can $f(n)$ be “inverted” as a sum of various values of $F(n)$?

The inversion gives us a deeper understanding of the relationship between the two functions. The Möbius function is a powerful intermediary that helps us “inverse” f as a sum of various values of F .

Prior to defining the Möbius function, let us first understand convolutions, which helps one to form new mathematical functions based on two previously defined functions.

7.1 Dirichlet Convolution

Definition: Dirichlet Convolution

Let f and g be arithmetic functions defined over natural numbers. The **Dirichlet convolution** of f and g , denoted $f * g$, is the arithmetic function defined by:

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

where the sum is over all positive divisors d of n .

Equivalently, we can write this as:

$$(f * g)(n) = \sum_{ab=n} f(a)g(b)$$

where the sum is over all ordered pairs (a, b) of positive integers whose product is n .

The key insight is that convolution combines information from all the ways to factor n into two parts, weighting each factorization by the corresponding function values.

Example 7.1.1

Let $f(n) = n$ and $g(n) = 1$ for all positive integers n . Compute $(f * g)(12)$.

Solution: We need to compute:

$$(f * g)(12) = \sum_{d|12} f(d)g\left(\frac{12}{d}\right) = \sum_{d|12} d \cdot 1 = \sum_{d|12} d$$

First, let's find all divisors of 12 by considering its prime factorization: $12 = 2^2 \cdot 3$. The divisors are: 1, 2, 3, 4, 6, 12.

Now we can compute:

$$(f * g)(12) = 1 + 2 + 3 + 4 + 6 + 12 = 28$$

Note that this is $\sigma(12)$, the sum of divisors function. This shows how convolution naturally gives rise to well-known arithmetic functions.

7.1.1 Properties of Dirichlet Convolution

Theorem: Properties of Dirichlet Convolution

Dirichlet convolution satisfies the following properties:

1. **Commutativity:** $f * g = g * f$
2. **Associativity:** $(f * g) * h = f * (g * h)$
3. **Distributivity:** $f * (g + h) = f * g + f * h$
4. **Identity element:** There exists a function ε such that $f * \varepsilon = f$ for all f
5. **Multiplicativity:** If f and g are multiplicative, then $f * g$ is multiplicative

Understanding Commutativity

Let's examine why Dirichlet convolution is commutative in detail. We start with the definition:

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

The key insight is that as d runs through all divisors of n , the expression $\frac{n}{d}$ also runs through all divisors of n , but in reverse order. More precisely, if d_1, d_2, \dots, d_k are all the divisors of n listed in increasing order, then $\frac{n}{d_1}, \frac{n}{d_2}, \dots, \frac{n}{d_k}$ are the same divisors listed in decreasing order.

For example, if $n = 12$, the divisors are $1, 2, 3, 4, 6, 12$. Then $\frac{12}{d}$ gives us $12, 6, 4, 3, 2, 1$, the same set of divisors.

Let's make a substitution: let $d' = \frac{n}{d}$. As d runs through all divisors of n , so does d' . We can rewrite our sum as:

$$(f * g)(n) = \sum_{d|n} f(d)g\left(\frac{n}{d}\right) \quad (7.1.1)$$

$$= \sum_{d'|n} f\left(\frac{n}{d'}\right)g(d') \quad (7.1.2)$$

$$= \sum_{d'|n} g(d')f\left(\frac{n}{d'}\right) \quad (7.1.3)$$

$$= (g * f)(n) \quad (7.1.4)$$

This shows that $(f * g)(n) = (g * f)(n)$ for all n , proving commutativity.

The Identity Element

The identity element for Dirichlet convolution is the function ε defined by:

$$\varepsilon(n) = \begin{cases} 1 & \text{if } n = 1 \\ 0 & \text{if } n > 1 \end{cases}$$

Let's verify that ε acts as an identity. For any arithmetic function f :

$$(f * \varepsilon)(n) = \sum_{d|n} f(d)\varepsilon\left(\frac{n}{d}\right) \quad (7.1.5)$$

Now, $\varepsilon\left(\frac{n}{d}\right) = 1$ only when $\frac{n}{d} = 1$, which means $d = n$. For all other divisors d of n , we have $\varepsilon\left(\frac{n}{d}\right) = 0$. Therefore:

$$(f * \varepsilon)(n) = f(n)\varepsilon(1) = f(n) \cdot 1 = f(n) \quad (7.1.6)$$

This confirms that $f * \varepsilon = f$ for any arithmetic function f .

Convolution Inverses

Given an arithmetic function f , we say that g is the **convolution inverse** of f if $f * g = \varepsilon$. Not every function has an inverse, but when an inverse exists, it is unique. To see why, suppose both g_1 and g_2 are inverses of f . Then:

$$g_1 = g_1 * \varepsilon = g_1 * (f * g_2) = (g_1 * f) * g_2 = \varepsilon * g_2 = g_2 \quad (7.1.7)$$

where we used associativity and commutativity of convolution.

The key requirement for a function f to have a convolution inverse is that $f(1) \neq 0$. When this condition is met, we can construct the inverse f^{-1} recursively:

$$f^{-1}(1) = \frac{1}{f(1)} \quad (7.1.8)$$

$$f^{-1}(n) = -\frac{1}{f(1)} \sum_{\substack{d|n \\ d < n}} f\left(\frac{n}{d}\right) f^{-1}(d) \quad \text{for } n > 1 \quad (7.1.9)$$

Theorem: Convolution Inverse

*If f^{-1} is defined by the recursive formula above, then $f * f^{-1} = \varepsilon$.*

Proof

We need to show that $(f * f^{-1})(n) = \varepsilon(n)$ for all positive integers n , given the recursive definition of f^{-1} .

Base case: For $n = 1$:

$$(f * f^{-1})(1) = \sum_{d|1} f(d) f^{-1}\left(\frac{1}{d}\right) \quad (7.1.10)$$

$$= f(1) f^{-1}(1) \quad (7.1.11)$$

$$= f(1) \cdot \frac{1}{f(1)} \quad (\text{by definition of } f^{-1}(1)) \quad (7.1.12)$$

$$= 1 \quad (7.1.13)$$

$$= \varepsilon(1) \quad (7.1.14)$$

The base case holds.

Inductive step: For $n > 1$, we want to show $(f * f^{-1})(n) = 0$. We start with the definition of Dirichlet convolution and split off the term where the divisor $d = 1$:

$$(f * f^{-1})(n) = \sum_{d|n} f(d) f^{-1}\left(\frac{n}{d}\right) \quad (7.1.15)$$

$$= f(1) f^{-1}(n) + \sum_{\substack{d|n \\ d>1}} f(d) f^{-1}\left(\frac{n}{d}\right) \quad (7.1.16)$$

Now, substitute the recursive definition of $f^{-1}(n)$:

$$f^{-1}(n) = -\frac{1}{f(1)} \sum_{\substack{d'|n \\ d' < n}} f\left(\frac{n}{d'}\right) f^{-1}(d')$$

Plugging this into our expression for the convolution:

$$(f * f^{-1})(n) = f(1) \left[-\frac{1}{f(1)} \sum_{\substack{d'|n \\ d' < n}} f\left(\frac{n}{d'}\right) f^{-1}(d') \right] + \sum_{\substack{d|n \\ d > 1}} f(d) f^{-1}\left(\frac{n}{d}\right) \quad (7.1.17)$$

$$= - \sum_{\substack{d'|n \\ d' < n}} f\left(\frac{n}{d'}\right) f^{-1}(d') + \sum_{\substack{d|n \\ d > 1}} f(d) f^{-1}\left(\frac{n}{d}\right) \quad (7.1.18)$$

Now we show that the second sum equals the first sum. In the second sum, make the substitution $d' = \frac{n}{d}$. As d runs through all divisors of n with $d > 1$, we have $d' = \frac{n}{d}$ running through all divisors of n with $d' < n$. Therefore:

$$\sum_{\substack{d|n \\ d > 1}} f(d) f^{-1}\left(\frac{n}{d}\right) = \sum_{\substack{d'|n \\ d' < n}} f\left(\frac{n}{d'}\right) f^{-1}(d')$$

Substituting this back into our expression:

$$(f * f^{-1})(n) = - \sum_{\substack{d'|n \\ d' < n}} f\left(\frac{n}{d'}\right) f^{-1}(d') + \sum_{\substack{d'|n \\ d' < n}} f\left(\frac{n}{d'}\right) f^{-1}(d') \quad (7.1.19)$$

$$= 0 \quad (7.1.20)$$

Since $\varepsilon(n) = 0$ for $n > 1$, we have shown $(f * f^{-1})(n) = \varepsilon(n)$.

The proof holds for any $n > 1$ because the recursive definition of $f^{-1}(n)$ depends only on the values $f^{-1}(d)$ where d is a proper divisor of n (i.e., $d < n$).

□

Realize that the proof uses strong induction rather than ordinary induction. Recall that in strong induction, we assume the statement holds for all positive integers less than n , and then prove it for n .

Strong induction is necessary here because the recursive definition of $f^{-1}(n)$ depends on the values $f^{-1}(d)$ for all proper divisors d of n (where $d < n$), not just for $n - 1$. For example, when computing $f^{-1}(12)$, we need the values $f^{-1}(1)$, $f^{-1}(2)$, $f^{-1}(3)$, $f^{-1}(4)$, and $f^{-1}(6)$, but

not $f^{-1}(11)$.

This recursive structure, where each step depends on multiple previous values rather than just the immediately preceding one, naturally leads to strong induction.

Multiplicativity of Convolution Functions

Recall that an arithmetic function h is multiplicative if $h(mn) = h(m)h(n)$ whenever $\gcd(m, n) = 1$. A function defined as a convolution of two multiplicative functions is also multiplicative.

Theorem: Multiplicativity of Dirichlet Convolution

*If f and g are multiplicative arithmetic functions, then their Dirichlet convolution $f * g$ is also multiplicative.*

Proof

Let m and n be positive integers with $\gcd(m, n) = 1$. We need to prove that $(f * g)(mn) = (f * g)(m) \cdot (f * g)(n)$.

First, observe that if $\gcd(m, n) = 1$ and $d \mid mn$, then d can be uniquely written as $d = d_1d_2$ where $d_1 \mid m$, $d_2 \mid n$, and $\gcd(d_1, d_2) = 1$. This follows from the fundamental theorem of arithmetic.

Now:

$$(f * g)(mn) = \sum_{d \mid mn} f(d)g\left(\frac{mn}{d}\right) \quad (7.1.21)$$

$$= \sum_{\substack{d_1 \mid m \\ d_2 \mid n}} f(d_1d_2)g\left(\frac{mn}{d_1d_2}\right) \quad (7.1.22)$$

$$= \sum_{\substack{d_1 \mid m \\ d_2 \mid n}} f(d_1d_2)g\left(\frac{m}{d_1} \cdot \frac{n}{d_2}\right) \quad (7.1.23)$$

Since f and g are multiplicative and $\gcd(d_1, d_2) = 1$, $\gcd\left(\frac{m}{d_1}, \frac{n}{d_2}\right) = 1$:

$$(f * g)(mn) = \sum_{\substack{d_1 \mid m \\ d_2 \mid n}} f(d_1)f(d_2)g\left(\frac{m}{d_1}\right)g\left(\frac{n}{d_2}\right) \quad (7.1.24)$$

$$= \sum_{d_1 \mid m} \sum_{d_2 \mid n} f(d_1)g\left(\frac{m}{d_1}\right)f(d_2)g\left(\frac{n}{d_2}\right) \quad (7.1.25)$$

$$= \left(\sum_{d_1|m} f(d_1)g\left(\frac{m}{d_1}\right) \right) \left(\sum_{d_2|n} f(d_2)g\left(\frac{n}{d_2}\right) \right) \quad (7.1.26)$$

$$= (f * g)(m) \cdot (f * g)(n) \quad (7.1.27)$$

Therefore, $f * g$ is multiplicative.

□

This result is extremely powerful because it allows us to understand the behavior of convolutions by studying their values on prime powers and then extending multiplicatively to all integers.

7.1.2 Geometric Visualization

Consider the positive integer lattice points (i, j) where $i, j \geq 1$. Each point (i, j) naturally corresponds to the product ij . This representation allows us to visualize divisibility relationships geometrically.

For a fixed positive integer n , the equation $xy = n$ defines a hyperbola in the first quadrant. The lattice points on this hyperbola correspond exactly to the ordered factorizations of n .

Key Observations from the Lattice:

- The hyperbola $xy = 6$ contains $\tau(6) = 4$ lattice points: $(1, 6), (2, 3), (3, 2), (6, 1)$
- The hyperbola $xy = 18$ contains $\tau(18) = 6$ lattice points: $(1, 18), (2, 9), (3, 6), (6, 3), (9, 2), (18, 1)$
- Each point $(d, \frac{n}{d})$ corresponds to a divisor d of n

Convolution as Hyperbola Summation

The Dirichlet convolution $(f * g)(n)$ has a beautiful geometric interpretation: it's the sum of $f(x) \cdot g(y)$ over all lattice points (x, y) on the hyperbola $xy = n$.

This example demonstrates convolution for both $n = 6$, where $f(n) = n$ and $g(n) = 1$. This implies that:

$$\sigma(6) = (f * g)(6) \quad (7.1.28)$$

$$= f(1)g(6) + f(2)g(3) + f(3)g(2) + f(6)g(1) \quad (7.1.29)$$

$$= 1 + 2 + 3 + 6 = 12 \quad (7.1.30)$$

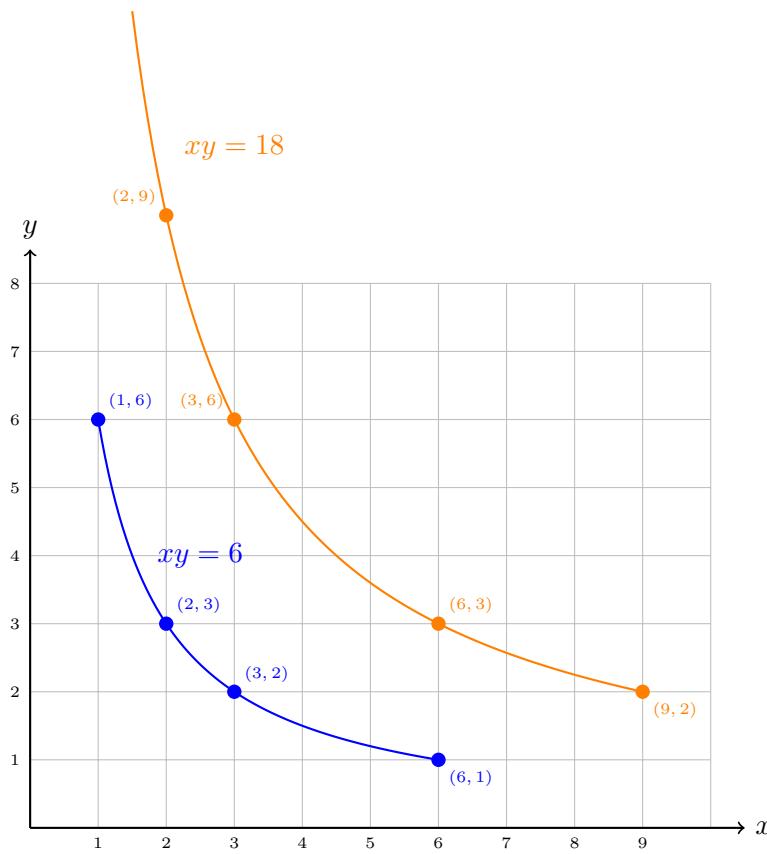


Figure 7.1: Lattice representation of divisor relationships. The hyperbola $xy = 6$ passes through 4 lattice points corresponding to the divisors of 6. The hyperbola $xy = 18$ passes through 6 lattice points (some outside the visible grid).

This perspective reveals why convolution appears naturally in number theory; the sum of divisors function $\sigma(n)$ is simply the convolution of the identity function with the constant function.

7.2 The Möbius Function

Definition: Square-Free Numbers

A positive integer n is called **square-free** if it is not divisible by any perfect square other than 1. Equivalently, n is square-free if and only if no prime appears more than once in its prime factorization.

A positive integer that is not square-free is called **non-square-free**.

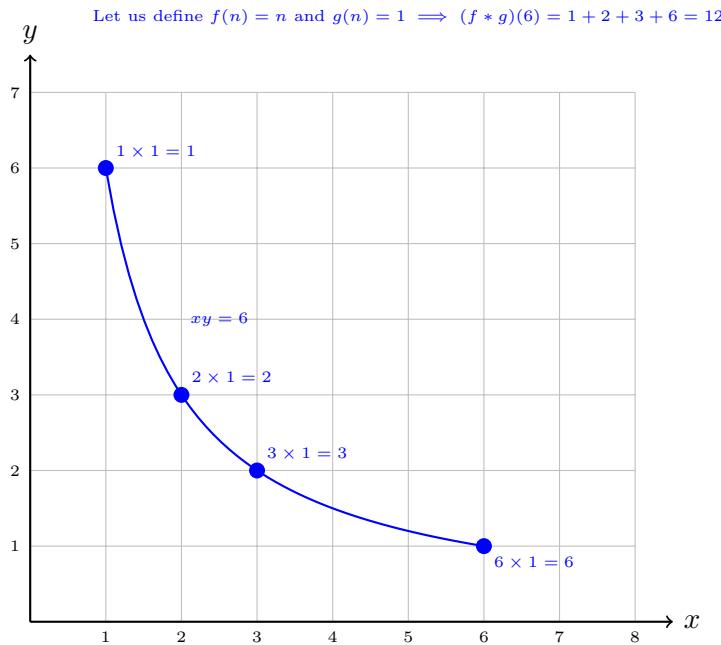


Figure 7.2: Dirichlet convolution as summation along hyperbolas. Each lattice point $(d, \frac{n}{d})$ on a hyperbola $xy = n$ contributes $f(d) \cdot g(\frac{n}{d})$ to the convolution. Here we show $(f * g)(6) = \sigma(6) = 12$

For example:

$$6 = 2 \times 3 \quad (\text{square-free: no repeated primes}) \quad (7.2.1)$$

$$12 = 2^2 \times 3 \quad (\text{not square-free: contains } 2^2) \quad (7.2.2)$$

$$15 = 3 \times 5 \quad (\text{square-free: no repeated primes}) \quad (7.2.3)$$

$$18 = 2 \times 3^2 \quad (\text{not square-free: contains } 3^2) \quad (7.2.4)$$

Therefore, 6 and 15 are square-free, while 12 and 18 are not square-free. Having defined square-free integers, we can now use this concept to define the Möbius function.

Definition: Möbius Function

The **Möbius function** $\mu(n)$ is defined for positive integers n as:

$$\mu(n) = \begin{cases} 1 & \text{if } n = 1 \\ (-1)^k & \text{if } n \text{ is the product of } k \text{ distinct primes} \\ 0 & \text{if } n \text{ has a squared prime factor} \end{cases}$$

Example 7.2.1

Compute $\mu(n)$ for $n = 1, 2, 3, \dots, 6$.

Solution: Let us compute each value systematically:

$$\begin{aligned}\mu(1) &= 1 \quad (\text{by definition}) \\ \mu(2) &= (-1)^1 = -1 \quad (\text{one prime: } 2) \\ \mu(3) &= (-1)^1 = -1 \quad (\text{one prime: } 3) \\ \mu(4) &= 0 \quad (\text{has squared factor: } 2^2) \\ \mu(5) &= (-1)^1 = -1 \quad (\text{one prime: } 5) \\ \mu(6) &= (-1)^2 = 1 \quad (\text{two distinct primes: } 2, 3)\end{aligned}$$

Thus $\mu(n) \neq 0$ only for square-free numbers, and among these, $\mu(n)$ alternates between 1 and -1 based on the number of prime factors.

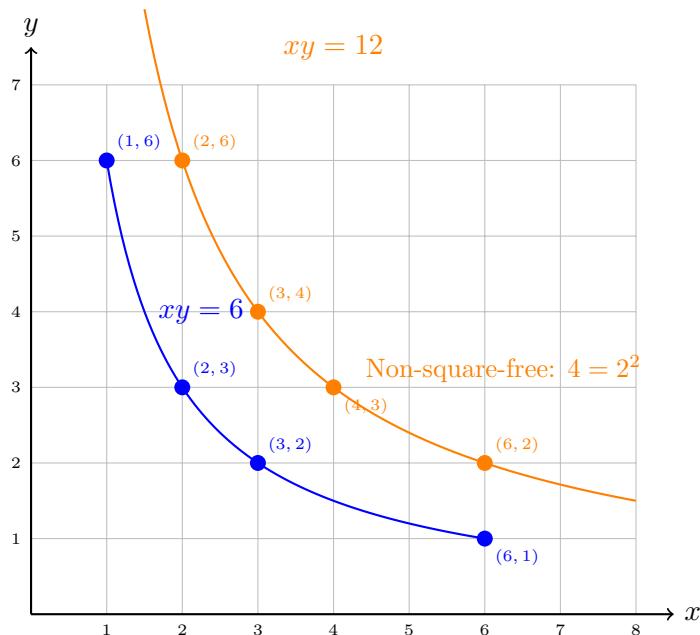
7.2.1 Geometric Visualization

Figure 7.3: Lattice representation showing square-free vs. non-square-free numbers. The hyperbola $xy = 6$ corresponds to the square-free number 6, while $xy = 12$ corresponds to the non-square-free number 12. The Möbius function values reflect this distinction.

The Möbius function has a geometric interpretation on a lattice that helps us understand the difference between square-free and non-square-free numbers. The lattice visualization shows that while hyperbolas organize divisor relationships geometrically, the Möbius function captures the deeper arithmetic structure:

- Both $n = 6$ and $n = 12$ create well-defined hyperbolas with their divisor points
- However, only $n = 6$ is square-free, giving $\mu(6) = 1$
- The non-square-free $n = 12$ gives $\mu(12) = 0$

7.2.2 Properties of the Möbius Function

Theorem: Multiplicativity of the Möbius Function

The Möbius function is multiplicative: If $\gcd(m, n) = 1$, then:

$$\mu(mn) = \mu(m)\mu(n)$$

Proof

Since $\gcd(m, n) = 1$, the integers m and n share no common prime factors.

Case 1: If either m or n has a squared prime factor, then mn also has that squared prime factor. Thus $\mu(m) = 0$ or $\mu(n) = 0$, and $\mu(mn) = 0$, making both sides equal to 0.

Case 2: If both m and n are square-free, let $m = p_1 p_2 \cdots p_r$ and $n = q_1 q_2 \cdots q_s$ where all primes are distinct (since $\gcd(m, n) = 1$).

Then $mn = p_1 p_2 \cdots p_r q_1 q_2 \cdots q_s$ is square-free with $r + s$ distinct prime factors.

Therefore:

$$\mu(mn) = (-1)^{r+s} = (-1)^r \cdot (-1)^s = \mu(m) \cdot \mu(n)$$

□

The most important property of the Möbius function is the following fundamental identity.

Theorem: Fundamental Identity of the Möbius Function

For any positive integer n :

$$\sum_{d|n} \mu(d) = \varepsilon(n)$$

where $\varepsilon(n)$ is the identity element for Dirichlet convolution, defined by:

$$\varepsilon(n) = \begin{cases} 1 & \text{if } n = 1 \\ 0 & \text{if } n > 1 \end{cases}$$

In terms of convolution, this can be written as: $\mu * 1 = \varepsilon$, where $1(n) = 1$ for all n .

Proof

We prove this by considering two cases.

Case 1: For $n = 1$, the only divisor is $d = 1$, so $\sum_{d|1} \mu(d) = \mu(1) = 1 = \varepsilon(1)$.

Case 2: For $n > 1$, we want to show that $\sum_{d|n} \mu(d) = 0$. Let $n = p_1^{a_1} p_2^{a_2} \cdots p_k^{a_k}$ be the prime factorization where $k \geq 1$. The key insight is to focus on the square-free divisors, since only these contribute non-zero terms to the sum.

The divisors d of n with $\mu(d) \neq 0$ are exactly those of the form $d = p_{i_1} p_{i_2} \cdots p_{i_j}$ where $\{i_1, i_2, \dots, i_j\} \subseteq \{1, 2, \dots, k\}$ (i.e., square-free divisors).

For each subset $S \subseteq \{1, 2, \dots, k\}$ with $|S| = j$, we get a divisor $d = \prod_{i \in S} p_i$ with $\mu(d) = (-1)^j$.

Therefore:

$$\sum_{d|n} \mu(d) = \sum_{j=0}^k \sum_{|S|=j} (-1)^j \quad (7.2.5)$$

$$= \sum_{j=0}^k \binom{k}{j} (-1)^j \quad (7.2.6)$$

$$= (1 + (-1))^k \quad (\text{by the binomial theorem}) \quad (7.2.7)$$

$$= 0^k \quad (7.2.8)$$

$$= 0 \quad (7.2.9)$$

since $k \geq 1$.

□

Even though we have given an elegant proof of the fundamental identity using the binomial theorem, the intuition behind the result might not be immediate. The fundamental result holds because for any natural number j (where j represents the number of square-free unique primes in n), the number of subsets of $\{1, 2, \dots, j\}$ with an even number of elements equals the number of subsets with an odd number of elements.

Mathematically:

$$\sum_{\text{even } k} \binom{j}{k} = \sum_{\text{odd } k} \binom{j}{k} = 2^{j-1} \text{ for } j \geq 1$$

Therefore:

$$\sum_{d|n} \mu(d) = \sum_{\text{even } j} \binom{k}{j} (+1) + \sum_{\text{odd } j} \binom{k}{j} (-1) = 2^{k-1} - 2^{k-1} = 0$$

The result would have been more intuitive if we could give an argument around cancellation of elements in $\sum_{d|n} \mu(d)$. More specifically, if we could take a divisor d_1 and find another divisor d_2 ($d_1 \neq d_2$) such that $\mu(d_1)$ and $\mu(d_2)$ are of opposite signs.

We can provide such an argument only for the case when n is square-free and j is odd.

Consider the pairing of divisors: when divisor $d = p_i$ for some $i \in \{1, 2, \dots, j\}$, then $\frac{n}{d} = p_1 p_2 \cdots p_{i-1} p_{i+1} \cdots p_k$. Recall that $\frac{n}{d}$ is also a divisor of n . We have $\mu(d) = \mu(p_i) = -1$ and $\mu\left(\frac{n}{d}\right) = (-1)^{k-1} = (-1)^{\text{even}} = 1$, so they cancel each other out.

The Mertens Function

The fundamental identity describes the nature of the Möbius function over the sum of divisors for a given natural number n . But what about the sum over all the integers less than or equal to n ? Such a function is called the Mertens function and is defined as:

$$M(n) = \sum_{k=1}^n \mu(k)$$

For a given natural number n , one can recursively calculate the Mertens formula, as described in the next theorem.

Theorem: Recursion for $M(n)$

For any integer $n > 1$:

$$M(n) = 1 - \sum_{k=2}^n M\left(\left\lfloor \frac{n}{k} \right\rfloor\right)$$

where $M(n) = \sum_{i=1}^n \mu(i)$ is the Mertens function.

Proof

Let us first try to prove the following identity:

$$\sum_{d=1}^n \mu(d) \left\lfloor \frac{n}{d} \right\rfloor = 1$$

We start with the fundamental property of the Möbius function:

$$\sum_{d|k} \mu(d) = \begin{cases} 1 & \text{if } k = 1 \\ 0 & \text{if } k > 1 \end{cases}$$

Now, let's sum this property for all integers k from 1 to n :

$$\sum_{k=1}^n \left(\sum_{d|k} \mu(d) \right) = \left(\sum_{d|1} \mu(d) \right) + \sum_{k=2}^n \left(\sum_{d|k} \mu(d) \right) = 1 + \sum_{k=2}^n 0 = 1$$

So we have established:

$$\sum_{k=1}^n \sum_{d|k} \mu(d) = 1$$

Next, we change the order of summation.

$$\sum_{k=1}^n \sum_{d|k} \mu(d) = \sum_{d=1}^n \sum_{\substack{k=1 \\ d|k}}^n \mu(d)$$

For the inner sum, $\mu(d)$ is a constant. We are just adding it up for each multiple of d up to n . The number of such multiples is exactly $\lfloor n/d \rfloor$. So the expression becomes:

$$\sum_{d=1}^n \mu(d) \left\lfloor \frac{n}{d} \right\rfloor$$

By equating our two results, we have proven the identity:

$$\sum_{d=1}^n \mu(d) \left\lfloor \frac{n}{d} \right\rfloor = 1$$

Now, let us derive the recursion for $M(n)$. The identity we just proved can be related to the Mertens function. Consider the sum $\sum_{k=1}^n M(\lfloor n/k \rfloor)$. Let's expand the definition

of M and reorder the summation as we did in Part 1:

$$\sum_{k=1}^n M\left(\left\lfloor \frac{n}{k} \right\rfloor\right) = \sum_{k=1}^n \sum_{d=1}^{\lfloor n/k \rfloor} \mu(d) \quad (7.2.10)$$

$$= \sum_{d=1}^n \sum_{k=1}^{\lfloor n/d \rfloor} \mu(d) \quad (7.2.11)$$

$$= \sum_{d=1}^n \mu(d) \left\lfloor \frac{n}{d} \right\rfloor \quad (7.2.12)$$

$$= 1 \quad (7.2.13)$$

This holds for any $n \geq 1$. For $n > 1$, we can split the term for $k = 1$ from the rest of the sum:

$$M\left(\left\lfloor \frac{n}{1} \right\rfloor\right) + \sum_{k=2}^n M\left(\left\lfloor \frac{n}{k} \right\rfloor\right) = 1$$

$$M(n) + \sum_{k=2}^n M\left(\left\lfloor \frac{n}{k} \right\rfloor\right) = 1$$

Rearranging the terms to isolate $M(n)$ gives us our final theorem:

$$M(n) = 1 - \sum_{k=2}^n M\left(\left\lfloor \frac{n}{k} \right\rfloor\right)$$

□

7.3 Möbius Inversion Formula

Before we jump into the main result, let us first introduce an important theorem related to convolution inverses.

Theorem: Convolution Inversion Principle

If f and g are arithmetic functions such that $g * h = f$ for some arithmetic function h , and if h has a convolution inverse h^{-1} (meaning $h * h^{-1} = \varepsilon$), then:

$$g = f * h^{-1}$$

Proof

Given $g * h = f$ and $h * h^{-1} = \varepsilon$, we have:

$$g = g * \varepsilon \quad (\text{identity property of convolution}) \quad (7.3.1)$$

$$= g * (h * h^{-1}) \quad (\text{substitute the inverse relationship}) \quad (7.3.2)$$

$$= (g * h) * h^{-1} \quad (\text{associativity of convolution}) \quad (7.3.3)$$

$$= f * h^{-1} \quad (\text{substitute the given condition } g * h = f) \quad (7.3.4)$$

□

Since $\mu * 1 = \varepsilon$, we know that μ is the convolution inverse of the constant function 1. This immediately gives us the Möbius inversion formula.

Theorem: Möbius Inversion Formula

Let f and g be arithmetic functions. Then:

$$g(n) = \sum_{d|n} f(d) \Leftrightarrow f(n) = \sum_{d|n} \mu\left(\frac{n}{d}\right) g(d)$$

Equivalently:

$$g(n) = \sum_{d|n} f(d) \Leftrightarrow f(n) = \sum_{d|n} \mu(d) g\left(\frac{n}{d}\right)$$

In convolution notation: $g = f * 1 \Leftrightarrow f = g * \mu$.

As mentioned earlier, the proof is trivial in convolution notation. Let us have a quick discussion of that.

Forward direction (\Rightarrow): We are given $g = f * 1$. Since we know that $\mu * 1 = \varepsilon$ (the fundamental identity), μ is the convolution inverse of 1.

By the Convolution Inversion Principle with $h = 1$ and $h^{-1} = \mu$:

$$f = g * \mu$$

Reverse direction (\Leftarrow): Given $f = g * \mu$ and $1 * \mu = \varepsilon$ (commutativity of the fundamental identity), 1 is the convolution inverse of μ .

By the Convolution Inversion Principle with $h = \mu$ and $h^{-1} = 1$:

$$g = f * 1$$

We have already provided a formal proof for Möbius Inversion Formula. However, for completion, let us also have at a more direct proof without using convolutions.

Proof

Forward direction (\Rightarrow): Suppose $g(n) = \sum_{d|n} f(d)$. We want to show that $f(n) = \sum_{d|n} \mu\left(\frac{n}{d}\right) g(d)$.

Consider:

$$\sum_{d|n} \mu\left(\frac{n}{d}\right) g(d) = \sum_{d|n} \mu\left(\frac{n}{d}\right) \sum_{e|d} f(e) \quad (7.3.5)$$

$$= \sum_{d|n} \sum_{e|d} \mu\left(\frac{n}{d}\right) f(e) \quad (7.3.6)$$

We change the order of summation. The condition “ $d | n$ and $e | d$ ” is equivalent to “ $e | n$ and d is a multiple of e dividing n ”.

Setting $d = ek$ where $k | \frac{n}{e}$ (as d runs through multiples of e dividing n , the parameter $k = \frac{d}{e}$ runs through all divisors of $\frac{n}{e}$):

$$\sum_{d|n} \sum_{e|d} \mu\left(\frac{n}{d}\right) f(e) = \sum_{e|n} f(e) \sum_{k|\frac{n}{e}} \mu\left(\frac{n}{ek}\right) \quad (7.3.7)$$

$$= \sum_{e|n} f(e) \sum_{k|\frac{n}{e}} \mu\left(\frac{\frac{n}{e}}{k}\right) \quad (7.3.8)$$

By the fundamental identity of the Möbius function:

$$\sum_{k|m} \mu\left(\frac{m}{k}\right) = \varepsilon(m) = \begin{cases} 1 & \text{if } m = 1 \\ 0 & \text{if } m > 1 \end{cases}$$

Therefore:

$$\sum_{k|\frac{n}{e}} \mu\left(\frac{\frac{n}{e}}{k}\right) = \begin{cases} 1 & \text{if } \frac{n}{e} = 1, \text{ i.e., } e = n \\ 0 & \text{if } \frac{n}{e} > 1, \text{ i.e., } e < n \end{cases}$$

Thus:

$$\sum_{d|n} \mu\left(\frac{n}{d}\right) g(d) = f(n) \cdot 1 = f(n)$$

Reverse direction (\Leftarrow): Suppose $f(n) = \sum_{d|n} \mu\left(\frac{n}{d}\right) g(d)$. We want to show that $g(n) = \sum_{d|n} f(d)$.

Consider:

$$\sum_{d|n} f(d) = \sum_{d|n} \sum_{e|d} \mu\left(\frac{d}{e}\right) g(e) \quad (7.3.9)$$

$$= \sum_{d|n} \sum_{e|d} \mu\left(\frac{d}{e}\right) g(e) \quad (7.3.10)$$

Using the same change of variables technique, setting $d = ek$ where $k \mid \frac{n}{e}$:

$$\sum_{d|n} \sum_{e|d} \mu\left(\frac{d}{e}\right) g(e) = \sum_{e|n} g(e) \sum_{k|\frac{n}{e}} \mu\left(\frac{ek}{e}\right) \quad (7.3.11)$$

$$= \sum_{e|n} g(e) \sum_{k|\frac{n}{e}} \mu(k) \quad (7.3.12)$$

By the fundamental identity:

$$\sum_{k|\frac{n}{e}} \mu(k) = \begin{cases} 1 & \text{if } \frac{n}{e} = 1, \text{ i.e., } e = n \\ 0 & \text{if } \frac{n}{e} > 1, \text{ i.e., } e < n \end{cases}$$

Therefore:

$$\sum_{d|n} f(d) = g(n) \cdot 1 = g(n)$$

This completes the algebraic proof of both directions.

□

Let us try to gain a more intuitive understanding of the Möbius function and the inversion formula. Imagine we have two related arithmetic functions $f(n)$ and $g(n)$ where:

- $f(n)$ represents some fundamental property of the number n itself
- $g(n)$ is defined by summing f over all divisors: $g(n) = \sum_{d|n} f(d)$

The Möbius inversion formula answers the fundamental question: if we know $g(n)$ (the divisor sums), how can we recover the original function $f(n)$?

The theorem provides us a process to unravel. We know that $g(n) = \sum_{d|n} f(d)$, which means each $f(d)$ contributes to multiple $g(n)$ values. This creates a complex web of dependencies. For example:

- $f(1)$ appears in $g(1), g(2), g(3), g(4), g(5), g(6), \dots$ (all numbers)

- $f(2)$ appears in $g(2), g(4), g(6), g(8), \dots$ (all even numbers)
- $f(3)$ appears in $g(3), g(6), g(9), g(12), \dots$ (all multiples of 3)

Möbius inversion works by applying carefully chosen “correction factors” $\mu(d)$ that systematically cancel out the unwanted contributions:

$$f(6) = \mu(1)g(6) + \mu(2)g(3) + \mu(3)g(2) + \mu(6)g(1) \quad (7.3.13)$$

$$= 1 \cdot g(6) + (-1) \cdot g(3) + (-1) \cdot g(2) + 1 \cdot g(1) \quad (7.3.14)$$

The fundamental identity ensures that when we apply these corrections, all the “contamination” from other $f(d)$ values cancels out perfectly, leaving only the desired $f(6)$.

Why Simple Subtraction Fails

Let us see why we need a systematic approach. Suppose we want to find $f(12)$ and we know:

$$g(12) = f(1) + f(2) + f(3) + f(4) + f(6) + f(12)$$

A naive approach might be: $f(12) = g(12) - [f(1) + f(2) + f(3) + f(4) + f(6)]$

But we don’t know the individual f values! We only know g values. Let’s try expressing the bracket using g values:

$$g(6) = f(1) + f(2) + f(3) + f(6) \quad (7.3.15)$$

$$g(4) = f(1) + f(2) + f(4) \quad (7.3.16)$$

$$g(3) = f(1) + f(3) \quad (7.3.17)$$

$$g(2) = f(1) + f(2) \quad (7.3.18)$$

$$g(1) = f(1) \quad (7.3.19)$$

If we try $g(12) - g(6) - g(4)$, we get:

$$[f(1) + f(2) + f(3) + f(4) + f(6) + f(12)] - [f(1) + f(2) + f(3) + f(6)] - [f(1) + f(2) + f(4)]$$

Notice that $f(1)$ appears $+1 - 1 - 1 = -1$ times and $f(2)$ appears $+1 - 1 - 1 = -1$ times. We have over-subtracted! We need to add them back in. Since $g(2) = f(1) + f(2)$, we have:

$$f(12) = g(12) - g(6) - g(4) + g(2)$$

Möbius Function as Coefficients

The Möbius Inversion Formula states:

$$\text{if } g(n) = \sum_{d|n} f(d), \text{ then } f(n) = \sum_{d|n} \mu(d)g\left(\frac{n}{d}\right)$$

Let's unpack the right side for $n = 12$:

$$f(12) = \mu(1)g(12) + \mu(2)g(6) + \mu(3)g(4) + \mu(4)g(3) + \mu(6)g(2) + \mu(12)g(1) \quad (7.3.20)$$

Let's substitute the μ values:

- $\mu(1) = 1$
- $\mu(2) = -1$ (1 prime factor)
- $\mu(3) = -1$ (1 prime factor)
- $\mu(4) = 0$ (has squared factor 2^2)
- $\mu(6) = 1$ (2 prime factors: 2, 3)
- $\mu(12) = 0$ (has squared factor 2^2)

So the formula becomes:

$$f(12) = (1)g(12) + (-1)g(6) + (-1)g(4) + (0)g(3) + (1)g(2) + (0)g(1) \quad (7.3.21)$$

$$= g(12) - g(6) - g(4) + g(2) \quad (7.3.22)$$

This is exactly our previous subtraction argument had suggested! The formula automatically calculates the correct combination of adding and subtracting the g values to perfectly isolate $f(12)$.

7.4 Applications to Arithmetic Functions

7.4.1 Euler's Totient Function

We can use Möbius inversion to derive the explicit formula for Euler's totient function.

Theorem: Euler's Totient Function

For any positive integer n with prime factorization $n = p_1^{a_1} p_2^{a_2} \cdots p_k^{a_k}$, the Euler totient function $\phi(n)$ can be computed using the formula:

$$\phi(n) = n \prod_{p|n} \left(1 - \frac{1}{p}\right)$$

where the product is taken over all distinct prime divisors of n .

Proof

We are aware of the following identity:

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

Applying Möbius inversion to the identity with $g(n) = n$ and $f(d) = \phi(d)$:

$$\phi(n) = \sum_{d|n} \mu\left(\frac{n}{d}\right) \cdot d$$

Making the substitution $d' = \frac{n}{d}$ (so $d = \frac{n}{d'}$), we get:

$$\phi(n) = \sum_{d'|n} \mu(d') \cdot \frac{n}{d'} = \sum_{d|n} \mu(d) \cdot \frac{n}{d}$$

Factoring out n :

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

Since $\mu(d) = 0$ whenever d has a squared prime factor, the only non-zero terms in the sum come from square-free divisors of n . Each square-free divisor corresponds to choosing a subset $S \subseteq \{1, 2, \dots, k\}$ and forming $d = \prod_{i \in S} p_i$.

In other terms, we can write $\phi(n)$ as:

$$\phi(n) = n \sum_{|S|=0}^k \sum_{i \in S} \frac{\mu(\prod_{i \in S} p_i)}{\prod_{i \in S} p_i}$$

As μ is multiplicative:

$$\phi(n) = n \sum_{|S|=0}^k \sum_{i \in S} \prod_{i \in S} \frac{\mu(p_i)}{p_i}$$

This is equivalent to:

$$\phi(n) = \prod_{i=1}^k \left(1 + \frac{\mu(p_i)}{p_i} \right)$$

As $\mu(p_i) = -1$ for all values of $i \in \{1, 2, \dots, k\}$, we can write:

$$\phi(n) = \prod_{i=1}^k \left(1 + \frac{-1}{p_i} \right) \tag{7.4.1}$$

$$\phi(n) = \prod_{i=1}^k \left(1 - \frac{1}{p_i} \right) \tag{7.4.2}$$

Therefore:

$$\phi(n) = n \prod_{p|n} \left(1 - \frac{1}{p} \right)$$

This is the familiar product formula for Euler's totient function.

□

Let us use Möbius inversion to compute $\phi(12)$. We have $12 = 2^2 \cdot 3$, so the divisors of 12 are: 1, 2, 3, 4, 6, 12. Using the formula $\phi(12) = \sum_{d|12} \mu(d) \cdot \frac{12}{d}$:

$$\phi(12) = \mu(1) \cdot 12 + \mu(2) \cdot 6 + \mu(3) \cdot 4 + \mu(4) \cdot 3 + \mu(6) \cdot 2 + \mu(12) \cdot 1 \tag{7.4.3}$$

$$= 1 \cdot 12 + (-1) \cdot 6 + (-1) \cdot 4 + 0 \cdot 3 + 1 \cdot 2 + 0 \cdot 1 \tag{7.4.4}$$

$$= 4 \tag{7.4.5}$$

The natural numbers less than and co-prime to 12 are $\{1, 5, 7, 11\}$, so indeed $\phi(12) = 4$. We could have also used the product formula to calculate the number of co-primes.

Having worked through Euler's totient function, let us look at an another application of the inversion formula.

Example 7.4.1

How many fractions $\frac{a}{b}$ in lowest terms satisfy $1 \leq a \leq m$ and $1 \leq b \leq n$?

Solution: A fraction $\frac{a}{b}$ is in lowest terms if and only if $\gcd(a, b) = 1$.

Let $f(d)$ be the number of pairs (a, b) with $1 \leq a \leq m$, $1 \leq b \leq n$, and $\gcd(a, b) = d$.

Let $g(d)$ be the number of pairs (a, b) with $1 \leq a \leq m$, $1 \leq b \leq n$, and $d \mid \gcd(a, b)$.

Then:

$$g(d) = \left\lfloor \frac{m}{d} \right\rfloor \left\lfloor \frac{n}{d} \right\rfloor$$

since both a and b must be multiples of d .

We have:

$$g(d) = \sum_{k:d|k} f(k)$$

By Möbius inversion:

$$f(d) = \sum_{k:d|k} \mu\left(\frac{k}{d}\right) g(k) = \sum_{\ell=1}^{\min(\lfloor m/d \rfloor, \lfloor n/d \rfloor)} \mu(\ell) \left\lfloor \frac{m}{d\ell} \right\rfloor \left\lfloor \frac{n}{d\ell} \right\rfloor$$

For fractions in lowest terms, we want $f(1)$:

$$f(1) = \sum_{\ell=1}^{\min(m,n)} \mu(\ell) \left\lfloor \frac{m}{\ell} \right\rfloor \left\lfloor \frac{n}{\ell} \right\rfloor$$

For the special case $m = n$, this becomes:

$$f(1) = \sum_{\ell=1}^n \mu(\ell) \left\lfloor \frac{n}{\ell} \right\rfloor^2$$

7.5 Practice Exercises

Exercise 7.1

Let Id be the identity function where $Id(n) = n$, and μ be the Möbius function. Calculate the value of the Dirichlet convolution $(\mu * Id)(18)$.

Exercise 7.2

Let $\omega(n)$ denote the number of distinct prime factors of an integer $n > 1$, with $\omega(1) = 0$. Prove the identity:

$$\sum_{d|n} |\mu(d)| = 2^{\omega(n)}$$

where μ is the Möbius function.

Exercise 7.3

Let $\tau(n)$ be the number of divisors of n (the divisor function). We know that $\tau = u * u$, where u is the unit function ($u(n) = 1$ for all n). Use this fact and the properties of Dirichlet convolution to prove that $\sum_{d|n} \mu(d)\tau\left(\frac{n}{d}\right) = 1$ for all $n \geq 1$.

Exercise 7.4

The function $u(n) = 1$ for all n is the Dirichlet inverse of the Möbius function μ . What is the Dirichlet inverse of μ itself? That is, find the function g such that $\mu * g = \varepsilon$. Prove your answer.

Exercise 7.5

Let $\sigma(n)$ be the sum of the positive divisors of n . Given the identity $\sigma = Id * u$ (where $Id(n) = n$ and $u(n) = 1$), use the Möbius inversion formula to express $Id(n)$ as a sum involving σ and μ .

Exercise 7.6

An arithmetic function f is called completely multiplicative if $f(mn) = f(m)f(n)$ for all positive integers m, n . Prove that if f is completely multiplicative, then its Dirichlet inverse is given by $f^{-1}(n) = \mu(n)f(n)$.

Exercise 7.7

Let $\Lambda(n)$ be the von Mangoldt function, defined as $\log p$ if n is a power of a prime p , and 0 otherwise. Prove the identity $\sum_{d|n} \mu(d) \log(d) = -\Lambda(n)$.

Exercise 7.8

Prove that for any two multiplicative functions f and g , their Dirichlet convolution $f * g$ is also multiplicative.

Exercise 7.9

Let $q(n)$ be the characteristic function of square-free integers (i.e., $q(n) = 1$ if n is square-free, and $q(n) = 0$ otherwise). Prove that $q(n)$ is equal to $|\mu(n)|$ for all n .

Exercise 7.10

Let f be an arithmetic function and define its summatory function $F(n) = \sum_{d|n} f(d)$. Prove or disprove: if F is completely multiplicative, then f must be completely multiplicative as well.

Chapter 8

Primitive Root



Leonhard Euler (1707-1783)

Euler, one of history's most prolific mathematicians, made pioneering contributions to the theory of primitive roots and congruences. He introduced the totient function $\phi(n)$ which counts numbers coprime to n - a concept crucial for understanding primitive roots. Euler was the first to study primitive roots systematically, proving that if p is prime, then there exists at least one primitive root modulo p . He also established that for a prime p , the congruence $x^n \equiv a \pmod{p}$ has exactly $\gcd(n, p - 1)$ solutions when a is a primitive root modulo p . This groundwork became essential for modern cryptography.

Despite losing vision in his right eye at age 31 and becoming completely blind in his later years, Euler's mathematical output actually increased after his blindness. He developed remarkable mental calculation abilities and dictated his discoveries to assistants, producing nearly half of his 850+ publications while blind. His legendary remark, "Now I will have fewer distractions," after losing his sight exemplifies his extraordinary dedication to mathematics. When Frederick the Great once asked him to solve a complex problem that had baffled the court's other mathematicians, Euler immediately produced the answer in his head, remarking, "I calculate it as easily as other men breathe."

In the previous chapters, we explored Euler's totient function, studied multiplicative properties of number-theoretic functions, and discovered Wilson's theorem. These tools revealed deep structures in modular arithmetic, particularly how the quantity $\phi(n)$ captures the count of numbers coprime to n . Now we investigate elements that fully exploit this structure: primitive roots, which are numbers whose powers cycle through all $\phi(n)$ coprime residues. In this chapter, we explore when such maximal generators exist, prove their existence for certain moduli, and discover the surprising fact that they exist only for very specific types of numbers.

8.1 Introduction to Primitive Roots

When working with modular arithmetic, we often encounter fascinating cyclical patterns as we multiply a number by itself repeatedly. Consider what happens when we take the number 2 modulo 7 and keep multiplying by 2:

$$2^1 \equiv 2 \pmod{7} \quad (8.1.1)$$

$$2^2 \equiv 4 \pmod{7} \quad (8.1.2)$$

$$2^3 \equiv 1 \pmod{7} \quad (8.1.3)$$

After three steps, we return to 1, and the pattern repeats. As discussed earlier, this cyclic property of an integer g modulo n is represented by $\text{ord}_n(g)$. In this case, $\text{ord}_7(2) = 3$. However, the maximum possible order modulo 7 is $\phi(7) = 6$, so 2 does not achieve the longest possible cycle. Integers that do achieve the maximum order of $\phi(n)$ are called primitive roots modulo n .

Definition: Primitive Root

An element $g \in \{1, 2, \dots, n-1\}$ is called a primitive root modulo n if $\text{ord}_n(g) = \phi(n)$, where $\text{ord}_n(g)$ denotes the smallest positive integer k such that $g^k \equiv 1 \pmod{n}$. That is, g is a primitive root if the sequence g, g^2, g^3, \dots modulo n does not repeat until the $\phi(n)$ -th term, at which point $g^{\phi(n)} \equiv 1 \pmod{n}$.

For prime p , this becomes $\text{ord}_p(g) = p-1$ since $\phi(p) = p-1$.

8.2 Primitive Roots

Given a natural number n , it is interesting to understand whether there always exists a primitive root a modulo n . If not for all composite numbers, can we at least say that primitive roots exist for all prime numbers?

Before we dive into proving the existence of primitive roots, let's develop an intuitive understanding of the underlying concepts.

8.2.1 Exploring Order of an Integer in Modular Arithmetic

The concept of order of an integer modulo n is tightly coupled with the concept of primitive roots in modular arithmetic. For modulo n , it follows from the Euler's Theorem that the maximum possible order of an integer a can be $\phi(n)$. Can order of an integer a modulo n take any value between 1 and $\phi(n)$ or are there a few subtleties to it?

Theorem: The Order Dividing Theorem

For any number a coprime to n , the order of a modulo n must divide $\phi(n)$, where ϕ is Euler's totient function.

Proof

Let k be the order of a modulo n . We know from Euler's theorem that $a^{\phi(n)} \equiv 1 \pmod{n}$.

If we divide $\phi(n)$ by k , we get some quotient q and remainder r where $0 \leq r < k$:

$$\phi(n) = qk + r \quad (8.2.1)$$

Now consider $a^{\phi(n)}$:

$$1 \equiv a^{\phi(n)} \pmod{n} \quad (8.2.2)$$

$$\equiv a^{qk+r} \pmod{n} \quad (8.2.3)$$

$$\equiv (a^k)^q \cdot a^r \pmod{n} \quad (8.2.4)$$

$$\equiv 1^q \cdot a^r \pmod{n} \quad (8.2.5)$$

$$\equiv a^r \pmod{n} \quad (8.2.6)$$

Since k is the smallest positive integer such that $a^k \equiv 1 \pmod{n}$, and $r < k$, the only way $a^r \equiv 1 \pmod{n}$ is if $r = 0$. Therefore, k must divide $\phi(n)$.

□

This theorem reveals an elegant structure in modular arithmetic: the possible orders are restricted to divisors of $\phi(n)$. For example, when working modulo 19, where $\phi(19) = 18$, every order must divide 18, restricting the possible cycle lengths to 1, 2, 3, 6, 9, and 18.

Let us try to understand the cyclic nature of modular arithmetic in more detail by examining what happens when we repeatedly multiply an element with itself. We will work with modulo 19 and test whether 2 might be a primitive root by computing its powers systematically.

Example 8.2.1

Let's compute the powers of 2 modulo 19.

Solution:

$$\begin{array}{llll}
 2^1 \equiv 2 & \pmod{19} & 2^{10} \equiv 36 \equiv 17 & \pmod{19} \\
 2^2 \equiv 4 & \pmod{19} & 2^{11} \equiv 34 \equiv 15 & \pmod{19} \\
 2^3 \equiv 8 & \pmod{19} & 2^{12} \equiv 30 \equiv 11 & \pmod{19} \\
 2^4 \equiv 16 & \pmod{19} & 2^{13} \equiv 22 \equiv 3 & \pmod{19} \\
 2^5 \equiv 32 \equiv 13 & \pmod{19} & 2^{14} \equiv 6 & \pmod{19} \\
 2^6 \equiv 26 \equiv 7 & \pmod{19} & 2^{15} \equiv 12 & \pmod{19} \\
 2^7 \equiv 14 & \pmod{19} & 2^{16} \equiv 24 \equiv 5 & \pmod{19} \\
 2^8 \equiv 28 \equiv 9 & \pmod{19} & 2^{17} \equiv 10 & \pmod{19} \\
 2^9 \equiv 18 & \pmod{19} & 2^{18} \equiv 20 \equiv 1 & \pmod{19}
 \end{array}$$

We observe that $2^{18} \equiv 1 \pmod{19}$, and then the pattern would repeat. The element 2 has order 18 modulo 19. As $\text{ord}_{19}(2) = 18$, we have 2 as a primitive root modulo 19. Notice that the powers of 2 generate all 18 non-zero residues modulo 19 before returning to 1. This observation leads us to the following general lemma.

Lemma 8.2.1:

Let p be a prime number and a be an element of order k modulo p . Then the powers $\{a^1, a^2, \dots, a^k\}$ are all distinct modulo p .

Proof

We need to show that $a^i \not\equiv a^j \pmod{p}$ for any $1 \leq i < j \leq k$.

Suppose, for contradiction, that $a^i \equiv a^j \pmod{p}$ for some $1 \leq i < j \leq k$.

Since p is prime and a has order k modulo p , we know $\gcd(a, p) = 1$, which implies $\gcd(a^i, p) = 1$. Therefore, a^i has a multiplicative inverse modulo p . Multiplying both sides by $(a^i)^{-1}$, we get:

$$a^{j-i} \equiv 1 \pmod{p} \quad (8.2.7)$$

Since $1 \leq j - i < k$, we have found a positive integer $\ell = j - i < k$ such that $a^\ell \equiv 1 \pmod{p}$. But this contradicts our assumption that k is the order of a , which means k is the smallest positive integer such that $a^k \equiv 1 \pmod{p}$.

Therefore, the powers a^1, a^2, \dots, a^k must all be distinct modulo p .

□

It is important to realize that for any integer m , such that $1 \leq m \leq k$, $(a^m)^k \equiv 1 \pmod{p}$. This follows because $(a^m)^k = (a^k)^m \equiv 1 \pmod{p}$.

8.2.2 The Polynomial Connection: Roots of $x^k \equiv 1 \pmod{p}$

As we investigate integers a of order k modulo n , we realize that a is a solution to the polynomial congruence $x^k \equiv 1 \pmod{p}$. What are the other solutions to the polynomial congruence?

Lemma 8.2.2:

If k divides $p - 1$ and a has order k modulo p , then: The k powers $\{a^1, a^2, \dots, a^k\}$ are precisely all the solutions to $x^k \equiv 1 \pmod{p}$.

Proof

We've already shown that these k powers are distinct. Each satisfies $(a^m)^k = a^{km} = (a^k)^m \equiv 1^m \equiv 1 \pmod{p}$, so they are all solutions to $x^k \equiv 1 \pmod{p}$. Since a polynomial of degree k can have at most k solutions modulo p , these must be all the solutions.

□

Let us recall the nature of a primitive root: a primitive root a modulo p generates all the elements in the set $\{1, \dots, p - 1\}$ as it continues to multiply with itself over $p - 1$ times. This happens because $\{a, a^2, \dots, a^{p-1}\}$ are all distinct modulo p .

It follows from the previous Lemma that given a primitive root a , its powers form the solution to $x^{p-1} \equiv 1 \pmod{p}$. It is important to realize that even though a is a primitive root, not all of a^m for $1 \leq m \leq p - 1$ are primitive roots. Why does this happen?

If we know the order of an integer a modulo p , is the order of a^m same as order of a for $1 \leq m \leq p - 1$?

Theorem: The Order of Powers

If $\text{ord}_n(a) = k$, then $\text{ord}_n(a^m) = \frac{k}{\gcd(k, m)}$.

Proof

Let $d = \gcd(k, m)$ and set $t = \frac{k}{d}$. We need to prove two things: (1) that $(a^m)^t \equiv 1 \pmod{n}$, and (2) that t is the smallest such positive integer.

For the first part:

$$(a^m)^t = a^{mt} = a^{m \cdot \frac{k}{d}} = a^{\frac{mk}{d}} = (a^k)^{\frac{m}{d}} \equiv 1^{\frac{m}{d}} \pmod{n} \equiv 1 \pmod{n}$$

For the second part, suppose there exists a smaller positive integer $s < t$ such that $(a^m)^s \equiv 1 \pmod{n}$. This means $a^{ms} \equiv 1 \pmod{n}$. Since k is the order of a modulo n , we know k must divide ms .

Since $d = \gcd(k, m)$, we can write $k = da'$ and $m = db'$ where $\gcd(a', b') = 1$. Then:

$$\begin{aligned} k \mid ms &\implies da' \mid db's \\ &\implies a' \mid b's \end{aligned}$$

Since $\gcd(a', b') = 1$, we have $a' \mid s$. But $a' = \frac{k}{d} = t$, so $t \mid s$. This contradicts our assumption that $s < t$.

Therefore, $t = \frac{k}{\gcd(k, m)}$ is indeed the order of a^m modulo n .

□

From the Order of Powers theorem, we can ask: for a primitive root a modulo prime p , when is a^m also a primitive root?

Since a is a primitive root modulo p , we have $\text{ord}_p(a) = p - 1$. By the Order of Powers theorem:

$$\text{ord}_p(a^m) = \frac{p - 1}{\gcd(p - 1, m)} \tag{8.2.8}$$

For a^m to be a primitive root, we need $\text{ord}_p(a^m) = p - 1$. This happens if and only if:

$$\frac{p - 1}{\gcd(p - 1, m)} = p - 1 \iff \gcd(p - 1, m) = 1 \tag{8.2.9}$$

Therefore, a^m is a primitive root modulo p if and only if $\gcd(p - 1, m) = 1$. As a corollary, this implies that for any divisor d of $p - 1$, a^d is not a primitive root modulo p .

Example 8.2.2

Let's return to our element $a = 2$ with order 18 modulo 19. What is the order of $a^6 = 2^6 \equiv 7 \pmod{19}$?

Solution: Using the Order of Powers theorem:

$$\gcd(6, 18) = 6 \quad (8.2.10)$$

$$\text{ord}_{19}(2^6) = \frac{18}{6} = 3 \quad (8.2.11)$$

We can verify this directly by computing powers of 7:

$$2^6 \equiv 7 \pmod{19} \quad (8.2.12)$$

$$2^{12} \equiv 11 \pmod{19} \quad (8.2.13)$$

$$2^{18} \equiv 1 \pmod{19} \quad (8.2.14)$$

Indeed, the order of 2^6 is 3, which equals $\frac{18}{\gcd(6, 18)}$, confirming our theorem.

This example reveals an important insight. We found that 2^6 has order 3 modulo 19, so it is not a primitive root. However, 2^6 is still a solution to $x^{18} \equiv 1 \pmod{19}$ (since all elements satisfy this by Fermat's Little Theorem).

More significantly, since $\text{ord}_{19}(2^6) = 3$, we know that 2^6 is also a solution to the smaller congruence $x^3 \equiv 1 \pmod{19}$. This illustrates a key principle: elements satisfy $x^k \equiv 1 \pmod{p}$ for any multiple k of their order.

This leads us to formalize the following general principle:

Theorem:

Let p be a prime and d a positive integer. The solutions to the congruence $x^d \equiv 1 \pmod{p}$ are precisely those elements whose orders divide d .

Proof

Let a be an element with order k modulo p . Then:

(\Rightarrow) If $a^d \equiv 1 \pmod{p}$, then k must divide d , since k is the smallest positive integer such that $a^k \equiv 1 \pmod{p}$.

(\Leftarrow) If k divides d , then $d = qk$ for some integer q , and $a^d = a^{qk} = (a^k)^q \equiv 1^q \equiv 1 \pmod{p}$.

□

This tells us that the solutions to $x^d \equiv 1 \pmod{p}$ consist of all elements whose orders divide d . In other words, if we want to find all solutions to $x^d \equiv 1 \pmod{p}$, we need to collect all elements of order 1, all elements of order 2 (if $2|d$), all elements of order 3 (if $3|d$), and so on for every divisor of d .

8.2.3 Counting Elements by Order

We've seen how elements of different orders contribute to the solutions of congruences like $x^d \equiv 1 \pmod{p}$. This naturally leads to a fundamental question: for a given prime p , how many elements have each possible order? Since every non-zero element modulo p has some order that divides $p - 1$, let's count how many elements have each specific order.

Theorem: Counting Formula

For any divisor k of $p - 1$, the number of elements with order exactly k modulo p is $\phi(k)$.

Proof

Let $\psi(k)$ denote the number of elements with order exactly k modulo p .

First, we show that $\psi(k)$ is either 0 or $\phi(k)$. If there exists at least one element a with order k , then by our earlier work on powers of elements, the elements $\{a^1, a^2, \dots, a^k\}$ are all distinct and form the complete set of solutions to $x^k \equiv 1 \pmod{p}$. Among these, exactly $\phi(k)$ elements have order exactly k (those a^m where $\gcd(m, k) = 1$). So either no elements have order k , or exactly $\phi(k)$ elements do.

Since every element in $\{1, 2, \dots, p - 1\}$ has some order that divides $p - 1$:

$$\sum_{k|p-1} \psi(k) = p - 1 \quad (8.2.15)$$

We also know that:

$$\sum_{k|p-1} \phi(k) = p - 1 \quad (8.2.16)$$

Since $\psi(k) \in \{0, \phi(k)\}$ for each k , we have $\psi(k) \leq \phi(k)$. If $\psi(k) = 0$ for any k , then $\sum \psi(k) < \sum \phi(k)$, contradicting the equality. Therefore, $\psi(k) = \phi(k)$ for all divisors k of $p - 1$.

□

This theorem guarantees that elements of every possible order (dividing $p - 1$) exist modulo p .

The intuitive idea behind this result is that when we have an element of maximum order (a primitive root), its various powers automatically create elements of all smaller possible orders in exactly the right quantities. The Order of Powers theorem shows us how these powers distribute among different orders, and the counts work out perfectly to give us exactly $\phi(k)$ elements of each order k .

Example 8.2.3

For our element $a = 2$ with order 18 modulo 19, what does the Counting Formula tell us about the distribution of orders among all non-zero elements modulo 19?

Solution: The theorem predicts the following distribution of orders among the 18 non-zero elements modulo 19:

- $\phi(18) = 6$ elements have order exactly 18 (the primitive roots modulo 19)
- $\phi(9) = 6$ elements have order exactly 9
- $\phi(6) = 2$ elements have order exactly 6
- $\phi(3) = 2$ elements have order exactly 3
- $\phi(2) = 1$ element has order exactly 2
- $\phi(1) = 1$ element has order exactly 1 (the element 1)

Total: $6 + 6 + 2 + 2 + 1 + 1 = 18$ elements, which accounts for all non-zero residues modulo 19.

8.2.4 The Existence of Primitive Roots for Primes

We have established that if primitive roots exist for a prime p , then there are exactly $\phi(p - 1)$ of them, and they create a beautiful distribution of orders among all non-zero elements. But we haven't yet proven that primitive roots actually exist! Let's now prove the fundamental theorem that guarantees their existence.

Theorem: Existence of Primitive Roots

Every prime number p has exactly $\phi(p - 1)$ primitive roots modulo p .

Proof

This follows directly from the Counting Formula. We proved that for any divisor k of $p - 1$, there are exactly $\phi(k)$ elements with order k modulo p .

In particular, taking $k = p - 1$, there are exactly $\phi(p - 1)$ elements with order $p - 1$ modulo p .

Since $p \geq 2$, we have $p - 1 \geq 1$. For any positive integer n , $\phi(n) \geq 1$ (since $\gcd(1, n) = 1$ always). Therefore $\phi(p - 1) \geq 1$.

This means there exists at least one element with order $p - 1$ modulo p , which is by definition a primitive root modulo p . In fact, there are exactly $\phi(p - 1)$ such primitive roots.

□

An important consequence of this result is that primitive roots are generators for the multiplicative structure modulo p . If g is a primitive root modulo p , then every non-zero element modulo p can be expressed uniquely as a power of g . That is, the powers $\{g^1, g^2, \dots, g^{p-1}\}$ produce exactly the complete set of non-zero residues $\{1, 2, 3, \dots, p - 1\}$ modulo p in some order.

Through this discussion, we have seen how the concept of order in modular arithmetic leads naturally to the existence of primitive roots. By examining the structure of solutions to polynomial congruences $x^k \equiv 1 \pmod{p}$, we discovered how elements of different orders distribute themselves among these solutions.

This approach not only proves that primitive roots exist but gives us a comprehensive understanding of the cyclic structure of the multiplicative group modulo a prime. This structure is fundamentally what makes primitive roots so powerful in number theory and its applications.

8.3 Primitive Roots for Composite Numbers

We have established that every prime has primitive roots that generate all non-zero residues. A natural question arises: what about composite numbers? Do they also have primitive roots?

For composite numbers, a primitive root a modulo n would generate all integers coprime to n . Since there are $\phi(n)$ such integers, we need $\text{ord}_n(a) = \phi(n)$. Let's explore whether such generators exist for composite moduli.

For instance, a primitive root a modulo 10 would be one of the numbers coprime to 10, namely $\{1, 3, 7, 9\}$, that can generate all the others through its powers. The required order must be exactly $\phi(10) = 4$.

Let's test $g = 3$:

$$3^1 \equiv 3 \pmod{10} \quad (8.3.1)$$

$$3^2 \equiv 9 \pmod{10} \quad (8.3.2)$$

$$3^3 \equiv 27 \equiv 7 \pmod{10} \quad (8.3.3)$$

$$3^4 \equiv 81 \equiv 1 \pmod{10} \quad (8.3.4)$$

So 3 is indeed a primitive root modulo 10, generating all numbers coprime to 10.

But do all composite numbers have primitive roots? Let's check $n = 8$, where $\phi(8) = 4$ and we need an element of order 4 among $\{1, 3, 5, 7\}$:

$$1^1 \equiv 1 \pmod{8} \quad (\text{ord}_8(1) = 1) \quad (8.3.5)$$

$$3^2 \equiv 9 \equiv 1 \pmod{8} \quad (\text{ord}_8(3) = 2) \quad (8.3.6)$$

$$5^2 \equiv 25 \equiv 1 \pmod{8} \quad (\text{ord}_8(5) = 2) \quad (8.3.7)$$

$$7^2 \equiv 49 \equiv 1 \pmod{8} \quad (\text{ord}_8(7) = 2) \quad (8.3.8)$$

No element has order 4, so there's no primitive root modulo 8! This raises the question: which composite numbers have primitive roots?

8.3.1 Primitive Roots for Prime Powers

We've seen that some composite numbers (such as 10) have primitive roots while others (such as 8) don't. To understand the pattern, let's start with the simplest case: prime powers p^k .

We know primitive roots exist for prime moduli p . The natural question is: if g is a primitive root modulo p , does it remain a primitive root modulo p^2, p^3 , and so on? If so, we could easily construct primitive roots for all prime powers. If not, we need to understand when and how to modify our approach.

The Lifting Lemma allows us to “lift” primitive roots from modulo p to modulo p^k . Let us first develop an intuitive understanding of why this works.

Why Lifting Can Fail

The process of extending a primitive root from modulo p to modulo p^k is called “lifting.” Unfortunately, this doesn't always work automatically. Let's see why with a concrete example.

Consider $p = 5$. We can verify that $g = 2$ is a primitive root modulo 5 because:

$$2^1 \equiv 2 \pmod{5} \quad (8.3.9)$$

$$2^2 \equiv 4 \pmod{5} \quad (8.3.10)$$

$$2^3 \equiv 8 \equiv 3 \pmod{5} \quad (8.3.11)$$

$$2^4 \equiv 16 \equiv 1 \pmod{5} \quad (8.3.12)$$

So 2 has order $4 = \phi(5)$ modulo 5. But when we check 2 modulo 25:

$$2^4 \equiv 16 \pmod{25}$$

$$2^8 \equiv 16^2 \equiv 256 \equiv 6 \pmod{25}$$

$$2^{12} \equiv 6 \cdot 16 \equiv 96 \equiv 21 \pmod{25}$$

$$2^{16} \equiv 21 \cdot 16 \equiv 336 \equiv 11 \pmod{25}$$

$$2^{20} \equiv 11 \cdot 16 \equiv 176 \equiv 1 \pmod{25}$$

So 2 has order $20 = \phi(25)$ modulo 25, meaning it lifts successfully to a primitive root modulo 25.

But not all primitive roots behave this way. Let's see what can go wrong with a different example. Consider $p = 7$, where $g = 3$ is a primitive root modulo 7 (since $3^6 \equiv 1 \pmod{7}$ and no smaller power works).

However, when we check modulo 49:

$$3^6 \equiv 1 \pmod{7} \quad (8.3.13)$$

$$3^6 \equiv 1 \pmod{49} \quad (8.3.14)$$

Since $3^6 \equiv 1 \pmod{49}$ already, the order of 3 modulo 49 is at most 6, not the required $42 = \phi(49)$. So 3 fails to lift to a primitive root modulo 49.

The Critical Condition

The key insight is that there's a simple test to determine whether a primitive root will lift successfully. A primitive root g modulo p will lift to a primitive root modulo p^k for all $k \geq 2$ if and only if:

$$g^{p-1} \not\equiv 1 \pmod{p^2}$$

This condition is critical because it determines whether the "extra structure" in higher powers p^k can be fully captured by the powers of g .

Let's test this condition with our earlier examples.

For $p = 5$, we have:

$$2^4 = 16 \equiv 1 \pmod{5} \quad (8.3.15)$$

$$2^4 = 16 \not\equiv 1 \pmod{25} \quad (8.3.16)$$

So $g = 2$ satisfies the condition $g^{p-1} \not\equiv 1 \pmod{p^2}$ and lifts successfully to a primitive root modulo 25.

For $p = 7$ and $g = 3$:

$$3^6 \equiv 1 \pmod{7} \quad (8.3.17)$$

$$3^6 \equiv 1 \pmod{49} \quad (8.3.18)$$

Here, $g = 3$ fails the condition and doesn't lift to a primitive root modulo 49.

The Deep Intuition Behind the Critical Condition

Why does this seemingly technical condition $g^{p-1} \not\equiv 1 \pmod{p^2}$ have such powerful consequences? Let's develop the intuition by thinking about what happens as we move from smaller to larger moduli.

Imagine you have a number g that works perfectly as a generator modulo p . This means its powers $g^1, g^2, g^3, \dots, g^{p-1}$ give you all the numbers from 1 to $p - 1$ (modulo p), and g^{p-1} is the first time you get back to 1 (mod p). Think of it like taking $p - 1$ steps around a circle with $p - 1$ meaningful spots, landing back at 1.

Now, we move to a bigger circle, modulo p^2 . This circle has $\phi(p^2) = p(p - 1)$ spots (numbers coprime to p^2). We want to know if g is still a generator here. Does it take exactly $p(p - 1)$ steps to get back to 1?

We know from working modulo p that g^{p-1} is *something* like 1. More precisely, $g^{p-1} = 1 + (a \cdot p)$. Let's write this as:

$$g^{p-1} = 1 + a \cdot p \quad (8.3.19)$$

Now, the crucial condition comes in: $g^{p-1} \not\equiv 1 \pmod{p^2}$. This tells us that the 'multiple of p ' ($a \cdot p$) is *not* a multiple of p^2 . This means the number a itself is *not* divisible by p . This a is special!

Why is this special a important?

Getting to p^2 : Since g^{p-1} is *not* 1 (mod p^2), the cycle length (order) of g modulo p^2 cannot be just $p - 1$. It has to be longer. It is important to note that $\text{ord}_p(g) \mid \text{ord}_{p^2}(g)$. This holds because $p^2 \mid (g^k - 1) \implies p \mid (g^k - 1)$.

Since the order must divide $\phi(p^2) = p(p - 1)$ and be a multiple of the order mod p (which is $p - 1$), the only remaining option is that the order modulo p^2 is exactly $p(p - 1)$. So, the condition ensures g works for p^2 .

Continuing to p^3, p^4, \dots : This is where the special a really does its job. Let's see what happens when we calculate powers relevant to the *next* modulus, p^3 . We need to know the order mod p^3 , and the options are $p(p - 1)$ or $p^2(p - 1)$. To decide, we need to check the power $g^{p(p-1)}$. Is it $1 \pmod{p^3}$?

Let's calculate it using what we know:

$$g^{p(p-1)} = (g^{p-1})^p \quad (8.3.20)$$

$$= (1 + ap)^p \quad (8.3.21)$$

If you expand $(1 + ap)^p$ using the binomial theorem, you get:

$$(1 + ap)^p = 1 + p(ap) + (\text{terms with } p^3 \text{ or higher powers of } p) \quad (8.3.22)$$

$$= 1 + ap^2 + (\text{terms divisible by } p^3) \quad (8.3.23)$$

So, $g^{p(p-1)} \equiv 1 + ap^2 \pmod{p^3}$.

Now, look at that ap^2 term. Since a was *not* divisible by p , ap^2 is *not* divisible by p^3 . This means:

$$g^{p(p-1)} \not\equiv 1 \pmod{p^3} \quad (8.3.24)$$

Because g raised to the power $\phi(p^2)$ is not 1 modulo p^3 , the order modulo p^3 cannot be $\phi(p^2)$. It must be the larger option, $\phi(p^3) = p^2(p - 1)$.

The Pattern:

The condition $g^{p-1} = 1 + ap$ (with $p \nmid a$) sets up a chain reaction:

- $g^{\phi(p)} = g^{p-1} \equiv 1 + ap \pmod{p^2}$ (Not 1 mod p^2) \Rightarrow Order mod p^2 is $\phi(p^2)$.
- $g^{\phi(p^2)} = (g^{\phi(p)})^p \equiv (1 + ap)^p \equiv 1 + ap^2 \pmod{p^3}$ (Not 1 mod p^3) \Rightarrow Order mod p^3 is $\phi(p^3)$.
- $g^{\phi(p^3)} = (g^{\phi(p^2)})^p \equiv (1 + ap^2)^p \equiv 1 + ap^3 \pmod{p^4}$ (Not 1 mod p^4) \Rightarrow Order mod p^4 is $\phi(p^4)$.
- ... and so on.

Each time you calculate g raised to the power $\phi(p^k)$, you find it's equal to $1 + ap^k \pmod{p^{k+1}}$. Because a wasn't divisible by p initially, ap^k is never divisible by p^{k+1} . This non-zero "error term" prevents the order from being $\phi(p^k)$ when working modulo p^{k+1} , forcing it to be the maximum possible value, $\phi(p^{k+1})$.

Why This Matters

This insight gives us a practical method to find primitive roots for prime powers:

1. Find a primitive root g modulo p
2. Test if $g^{p-1} \not\equiv 1 \pmod{p^2}$
3. If it passes, g is a primitive root for all p^k
4. If it fails, try $g + p$ instead

The beauty of this result is that we only need to check the condition once (modulo p^2), and we're guaranteed that the primitive root works for all higher powers p^k .

Lemma 8.3.3: Lifting Lemma

Let p be an odd prime. There exists a primitive root g modulo p such that $g^{p-1} \not\equiv 1 \pmod{p^2}$. Any such g is a primitive root modulo p^k for all $k \geq 1$.

Proof

Let g_0 be a primitive root modulo p .

Step 1: We first show there exists a primitive root g modulo p such that $g^{p-1} \not\equiv 1 \pmod{p^2}$.

If $g_0^{p-1} \not\equiv 1 \pmod{p^2}$, we simply take $g = g_0$.

If $g_0^{p-1} \equiv 1 \pmod{p^2}$, we set $g = g_0 + p$ and observe that $g \equiv g_0 \pmod{p}$, so g is also a primitive root modulo p . Using the binomial expansion:

$$g^{p-1} = (g_0 + p)^{p-1} \tag{8.3.25}$$

$$\equiv g_0^{p-1} + (p-1)g_0^{p-2}p \pmod{p^2} \tag{8.3.26}$$

$$\equiv 1 + (p-1)pg_0^{p-2} \pmod{p^2} \tag{8.3.27}$$

$$\equiv 1 - pg_0^{p-2} \pmod{p^2} \tag{8.3.28}$$

Since $p \nmid g_0$, we have $pg_0^{p-2} \not\equiv 0 \pmod{p^2}$. Thus $g^{p-1} \not\equiv 1 \pmod{p^2}$.

Step 2: Now, let g be a primitive root modulo p with $g^{p-1} \not\equiv 1 \pmod{p^2}$. We can write $g^{p-1} = 1 + ap$ where $p \nmid a$.

We claim that for all $k \geq 2$:

$$g^{p^{k-2}(p-1)} \equiv 1 + a_k p^{k-1} \pmod{p^k} \quad \text{where } p \nmid a_k \tag{8.3.29}$$

We prove this by induction on k :

- *Base case ($k = 2$):* Already established with $a_2 = a$.

- *Inductive step: Assume the claim holds for $j = k \geq 2$. We show it for $k + 1$.*

Raising to power p :

$$g^{p^{k-1}(p-1)} = (1 + a_k p^{k-1})^p \quad (8.3.30)$$

$$= 1 + \binom{p}{1} a_k p^k + \binom{p}{2} a_k^2 p^{2k-2} + \dots \quad (8.3.31)$$

$$\equiv 1 + a_k p^k \pmod{p^{k+1}} \quad (8.3.32)$$

The higher-order terms vanish modulo p^{k+1} since p is odd and $k \geq 2$, making $2k - 2 \geq k + 1$.

Step 3: Finally, we determine the order of g modulo p^k .

Let $d = \text{ord}_{p^k}(g)$. Since g is a primitive root modulo p with order $p - 1$, we know $(p - 1)|d$. Also, $d|\phi(p^k) = p^{k-1}(p - 1)$.

Therefore, $d = p^j(p - 1)$ for some $0 \leq j \leq k - 1$.

From our induction result, we have $g^{p^{k-2}(p-1)} \not\equiv 1 \pmod{p^k}$, which means d cannot be $p^{k-2}(p - 1)$ or smaller.

The only possibility is $d = p^{k-1}(p - 1) = \phi(p^k)$, proving that g is a primitive root modulo p^k .

□

Having established primitive roots for prime powers p^k , we naturally ask: what about $2p^k$ where p is an odd prime?

The key insight is that since $\gcd(2, p^k) = 1$, we have $\phi(2p^k) = \phi(2)\phi(p^k) = \phi(p^k)$. This means we need the same cycle length as for p^k alone.

Lemma 8.3.4: Primitive Roots for $2p^k$

Primitive roots exist for $n = 2p^k$, where p is an odd prime and $k \geq 1$.

Proof

Since $\gcd(2, p^k) = 1$, Euler's totient function gives us $\phi(2p^k) = \phi(2)\phi(p^k) = 1 \cdot \phi(p^k) = \phi(p^k)$.

Let h be a primitive root modulo p^k (which exists by the Lifting Lemma). We construct a primitive root modulo $2p^k$ by considering two cases:

Case 1: If h is odd, let $g = h$. Then $\gcd(g, 2p^k) = 1$ since g is odd and coprime to p^k .

Let $d = \text{ord}_{2p^k}(g)$. Since $g^d \equiv 1 \pmod{2p^k}$, we have $g^d \equiv 1 \pmod{p^k}$. Therefore $\text{ord}_{p^k}(g) \mid d$, which means $\phi(p^k) \mid d$.

But also $d \mid \phi(2p^k) = \phi(p^k)$. Combining these divisibility conditions, we get $d = \phi(p^k) = \phi(2p^k)$.

Therefore g is a primitive root modulo $2p^k$.

Case 2: If h is even, let $g = h + p^k$. Since p^k is odd, g is odd, so $\gcd(g, 2p^k) = 1$.

Since $g \equiv h \pmod{p^k}$, we have $\text{ord}_{p^k}(g) = \text{ord}_{p^k}(h) = \phi(p^k)$.

By the same argument as Case 1, $\text{ord}_{2p^k}(g) = \phi(p^k) = \phi(2p^k)$.

Therefore g is a primitive root modulo $2p^k$.

In both cases, we have constructed a primitive root modulo $2p^k$.

□

To complete our classification, we need to verify that primitive roots exist for the remaining small cases: $n = 1, 2, 4$. These are special cases that don't fit the patterns p^k or $2p^k$.

- **For $n = 1$:** We have $\phi(1) = 1$, and the only element to consider is 1. Since $1^1 \equiv 1 \pmod{1}$, we have $\text{ord}_1(1) = 1 = \phi(1)$. So 1 is a primitive root modulo 1.
- **For $n = 2$:** We have $\phi(2) = 1$, and the only element coprime to 2 is 1. Since $1^1 \equiv 1 \pmod{2}$, we have $\text{ord}_2(1) = 1 = \phi(2)$. So 1 is a primitive root modulo 2.
- **For $n = 4$:** We have $\phi(4) = 2$, and the elements coprime to 4 are $\{1, 3\}$. Let's check $g = 3$:

$$\begin{aligned} 3^1 &\equiv 3 \pmod{4} \\ 3^2 &\equiv 9 \equiv 1 \pmod{4} \end{aligned}$$

So $\text{ord}_4(3) = 2 = \phi(4)$, making 3 a primitive root modulo 4.

This confirms that primitive roots exist for $n = 1, 2, 4$.

8.4 Non-Existence of Primitive Roots

We've established that primitive roots exist for $n = 1, 2, 4, p^k$, and $2p^k$ where p is an odd prime and $k \geq 1$. But what about all other composite numbers?

For example, we saw earlier that $n = 8$ has no primitive roots, and we might wonder about numbers like $15 = 3 \times 5$ or $12 = 2^2 \times 3$. In this section, we'll prove that primitive roots do NOT exist for any composite number outside our special cases.

The key insight is that composite numbers with “too much” prime structure have elements whose orders are too small to generate all coprime residues.

Lemma 8.4.5: Orders modulo 2^k for $k \geq 3$

For $k \geq 3$, if a is an odd integer, then $a^{2^{k-2}} \equiv 1 \pmod{2^k}$.

Proof

We prove this by induction on k .

Base case ($k = 3$): We need to show $a^{2^{3-2}} = a^2 \equiv 1 \pmod{2^3}$, i.e., $a^2 \equiv 1 \pmod{8}$ for any odd integer a .

Since a is odd, we can write $a = 2m + 1$ for some integer m . Then:

$$a^2 = (2m + 1)^2 = 4m^2 + 4m + 1 = 4m(m + 1) + 1 \quad (8.4.1)$$

Since consecutive integers m and $m + 1$ have opposite parity, one is even, making $m(m + 1)$ even. Therefore $4m(m + 1) \equiv 0 \pmod{8}$, and $a^2 \equiv 1 \pmod{8}$.

Inductive step: Assume $a^{2^{k-2}} \equiv 1 \pmod{2^k}$ for some $k \geq 3$. We can write $a^{2^{k-2}} = 1 + c \cdot 2^k$ for some integer c .

Squaring both sides:

$$a^{2^{k-1}} = (a^{2^{k-2}})^2 = (1 + c \cdot 2^k)^2 \quad (8.4.2)$$

$$= 1 + 2c \cdot 2^k + c^2 \cdot 2^{2k} \quad (8.4.3)$$

$$= 1 + c \cdot 2^{k+1} + c^2 \cdot 2^{2k} \quad (8.4.4)$$

Since $k \geq 3$, we have $2k \geq k + 1$, so $c^2 \cdot 2^{2k} \equiv 0 \pmod{2^{k+1}}$. Therefore:

$$a^{2^{k-1}} \equiv 1 + c \cdot 2^{k+1} \equiv 1 \pmod{2^{k+1}} \quad (8.4.5)$$

This completes the induction, proving the result for all $k \geq 3$.

□

This shows that every odd integer modulo 2^k (for $k \geq 3$) has order dividing 2^{k-2} , which is strictly less than $\phi(2^k) = 2^{k-1}$. Therefore, no primitive roots exist for 2^k when $k \geq 3$.

Theorem: Complete Classification of Primitive Roots

An integer $n > 1$ has a primitive root if and only if n is of the form 2 , 4 , p^k , or $2p^k$ where p is an odd prime and $k \geq 1$.

Proof

We've already shown that primitive roots exist for $n = 2, 4, p^k$, and $2p^k$. We've also shown that primitive roots do not exist for 2^k when $k \geq 3$. It remains to show they don't exist for composite numbers with at least two distinct prime factors.

Let $n = p_1^{k_1} p_2^{k_2} \cdots p_r^{k_r}$ be the prime factorization of n , where at least two of the primes are distinct.

Let a be any integer with $\gcd(a, n) = 1$. By Euler's theorem applied to each prime power:

$$a^{\phi(p_i^{k_i})} \equiv 1 \pmod{p_i^{k_i}} \text{ for each } i \quad (8.4.6)$$

Define $L = \text{lcm}(\phi(p_1^{k_1}), \phi(p_2^{k_2}), \dots, \phi(p_r^{k_r}))$. By properties of congruences and the Chinese Remainder Theorem, we have:

$$a^L \equiv 1 \pmod{n} \quad (8.4.7)$$

This means the order of a modulo n must divide L . We now show $L < \phi(n)$ by considering the possible combinations of distinct prime factors:

Case 1: n is divisible by two distinct odd primes p_i and p_j

Since p_i and p_j are odd primes, both $\phi(p_i^{k_i}) = p_i^{k_i-1}(p_i - 1)$ and $\phi(p_j^{k_j}) = p_j^{k_j-1}(p_j - 1)$ are even (as $p - 1$ is even when p is odd).

Since at least two terms in the LCM are even, their LCM will be at most half their product. This means the overall LCM of all terms will be strictly less than the product of all terms. More precisely, for any two even numbers x and y :

$$\text{lcm}(x, y) = \frac{xy}{\gcd(x, y)} \leq \frac{xy}{2} < xy \quad (8.4.8)$$

since $\gcd(x, y) \geq 2$ when both are even.

Therefore:

$$L = \text{lcm}(\phi(p_i^{k_i}), \phi(p_j^{k_j}), \dots) \quad (8.4.9)$$

$$< \phi(p_i^{k_i}) \cdot \phi(p_j^{k_j}) \cdots \quad (8.4.10)$$

$$= \phi(n) \quad (8.4.11)$$

Case 2: n is divisible by 2^k (where $k \geq 2$) and an odd prime p

If $k = 2$: We have $L = \text{lcm}(\phi(2^2), \phi(p^j), \dots) = \text{lcm}(2, \phi(p^j), \dots)$. Since $\phi(p^j)$ is even, $\text{lcm}(2, \phi(p^j)) = \phi(p^j)$. Therefore, the LCM calculation is effectively performed on a set of numbers whose product is $\phi(n)/\phi(2^2) = \phi(n)/2$. Since the LCM is always less than or equal to the product of the numbers:

$$L \leq \phi(p^j)\phi(\text{other factors}) = \frac{\phi(n)}{\phi(2^2)} = \frac{\phi(n)}{2} < \phi(n) \quad (8.4.12)$$

If $k \geq 3$: From our earlier lemma, for any odd integer a , we have $a^{2^{k-2}} \equiv 1 \pmod{2^k}$ where $2^{k-2} = \phi(2^k)/2$. Also, $\phi(p^j)$ is even. Therefore:

$$L = \text{lcm}(2^{k-2}, \phi(p^j), \dots) \leq \frac{2^{k-2}\phi(p^j)\dots}{2} = \frac{\phi(n)}{4} < \phi(n) \quad (8.4.13)$$

In all cases, we find $L < \phi(n)$. Since the order of any element a must divide L , we have $\text{ord}_n(a) < \phi(n)$ for all a coprime to n . This means no primitive root exists for composite numbers with at least two distinct prime factors.

This completes the proof of the complete classification.

□

8.5 Practice Exercises

Exercise 8.1

Find all primitive roots modulo 11.

Exercise 8.2

Determine whether 3 is a primitive root modulo 19.

Exercise 8.3

If p is a prime number and g is a primitive root modulo p , prove that $g^{\frac{p-1}{2}} \equiv -1 \pmod{p}$.

Exercise 8.4

Determine all values of n for which 2 is a primitive root modulo n .

Exercise 8.5

Let p be an odd prime and g be a primitive root modulo p . Determine the number of solutions to the congruence $x^2 \equiv 1 \pmod{p}$.

Exercise 8.6

Prove that if $p \equiv 3 \pmod{4}$ is a prime, then -1 is not a quadratic residue modulo p .

Exercise 8.7

Find the order of 5 modulo 24.

Exercise 8.8

Prove that for any odd prime p and integer a with $\gcd(a, p) = 1$, we have $a^{p-1} \equiv 1 \pmod{p}$ (Fermat's Little Theorem).

Exercise 8.9

If p is a prime and g is a primitive root modulo p , how many solutions does the congruence $x^3 \equiv 1 \pmod{p}$ have?

Exercise 8.10

If p is an odd prime and g is a primitive root modulo p , prove that $g + p \cdot \mathbb{Z}$ is a generator of the multiplicative group $(\mathbb{Z}/p\mathbb{Z})^*$.

Chapter 9

Quadratic Residues and Reciprocity

Adrien-Marie Legendre (1752-1833)

Adrien-Marie Legendre made fundamental contributions to number theory, analysis, and mathematical physics during the revolutionary period of French mathematics. His “*Essai sur la théorie des nombres*” (1798) was one of the first systematic treatments of number theory, where he laid the foundations of quadratic reciprocity and introduced the Legendre symbol, a notation that remains standard today.



Legendre formulated an early version of the law of quadratic reciprocity, though he was unable to provide a complete proof. His work on elliptic integrals led to the classification of elliptic functions and the development of what are now called Legendre polynomials. Legendre also made the first rigorous proof that π is irrational and conjectured the prime number theorem, which states that the number of primes less than or equal to a positive real number x , denoted as $\pi(x)$, is approximately $\frac{x}{\ln x}$.

Despite living through the turbulent French Revolution and Napoleonic era, Legendre maintained his mathematical productivity well into his seventies. His “*Éléments de géométrie*” became the standard geometry textbook in France and America for over a century, demonstrating his gift for clear mathematical exposition alongside his research achievements.

Building on our previous work with modular arithmetic, this chapter explores when a number can be expressed as a perfect square modulo a prime, leading us to the celebrated Law of Quadratic Reciprocity. Quadratic residues represent one of the most elegant areas of number theory, where algebraic and arithmetic properties intertwine to reveal deep structural patterns in the integers.

The Legendre symbol provides a powerful notation that encapsulates whether a number is a quadratic residue modulo a prime, and its multiplicative properties lead us naturally to quadratic reciprocity, which Gauss called the “golden theorem” of number theory.

9.1 Quadratic Residues

Definition: Quadratic Residue

*Let p be an odd prime and a be an integer not divisible by p . We say that a is a **quadratic residue modulo p** if there exists an integer x such that:*

$$x^2 \equiv a \pmod{p}$$

*If no such x exists, then a is called a **quadratic non-residue modulo p** .*

For example, consider $p = 7$. Let's determine which numbers are quadratic residues modulo 7:

$$1^2 \equiv 1 \pmod{7} \tag{9.1.1}$$

$$2^2 \equiv 4 \pmod{7} \tag{9.1.2}$$

$$3^2 \equiv 9 \equiv 2 \pmod{7} \tag{9.1.3}$$

$$4^2 \equiv 16 \equiv 2 \pmod{7} \tag{9.1.4}$$

$$5^2 \equiv 25 \equiv 4 \pmod{7} \tag{9.1.5}$$

$$6^2 \equiv 36 \equiv 1 \pmod{7} \tag{9.1.6}$$

Thus, the quadratic residues modulo 7 are $\{1, 2, 4\}$, and the quadratic non-residues are $\{3, 5, 6\}$.

Theorem: Number of Quadratic Residues

For an odd prime p , there are exactly $\frac{p-1}{2}$ quadratic residues and $\frac{p-1}{2}$ quadratic non-residues among the integers $1, 2, \dots, p-1$.

Proof

Consider the squares $1^2, 2^2, \dots, \left(\frac{p-1}{2}\right)^2$ modulo p . We will first show using contradiction that these are all distinct values.

Let us assume $i^2 \equiv j^2 \pmod{p}$ where $1 \leq i < j \leq \frac{p-1}{2}$. Then $p \mid (j^2 - i^2) = (j-i)(j+i)$. Since p is prime, either $p \mid (j-i)$ or $p \mid (j+i)$.

As $0 < j-i < \frac{p-1}{2} < p$, we have $p \nmid (j-i)$. Since $2 \leq j+i \leq p-1 < p$, we have $p \nmid (j+i)$. Thus we have a contradiction, which shows that the squares $1^2, 2^2, \dots, \left(\frac{p-1}{2}\right)^2$ are distinct modulo p .

Next, note that if $x^2 \equiv a \pmod{p}$, then $(-x)^2 \equiv a \pmod{p}$ as well. Since $-x \equiv p-x \pmod{p}$ and $x \neq p-x$ for $1 \leq x \leq \frac{p-1}{2}$ (as this would imply $2x \equiv 0 \pmod{p}$, which is impossible since $\gcd(2, p) = 1$ and $x \not\equiv 0 \pmod{p}$), each quadratic residue arises from exactly two squares.

Therefore, there are exactly $\frac{p-1}{2}$ distinct quadratic residues among $\{1, 2, \dots, p-1\}$, and hence $\frac{p-1}{2}$ quadratic non-residues.

□

Let us refer back to the example previously discussed. Notice the symmetry pattern: x and $7-x$ give the same quadratic residue:

$$1^2 \equiv 6^2 \equiv 1 \pmod{7} \quad (9.1.7)$$

$$2^2 \equiv 5^2 \equiv 4 \pmod{7} \quad (9.1.8)$$

$$3^2 \equiv 4^2 \equiv 2 \pmod{7} \quad (9.1.9)$$

Since each pair $(x, 7-x)$ produces the same residue, we only need to check half the values: $x = 1, 2, 3$ (that is, $x = 1, 2, \dots, \frac{7-1}{2}$).

From our calculations, these give distinct quadratic residues:

$$1^2 \equiv 1 \pmod{7} \quad (9.1.10)$$

$$2^2 \equiv 4 \pmod{7} \quad (9.1.11)$$

$$3^2 \equiv 2 \pmod{7} \quad (9.1.12)$$

$$(9.1.13)$$

Since 1, 4, and 2 are all different, we get exactly $\frac{7-1}{2} = 3$ distinct quadratic residues modulo 7.

9.2 The Legendre Symbol

The Legendre symbol provides an elegant notation for quadratic residues.

Definition: Legendre Symbol

Let p be an odd prime and a be an integer. The **Legendre symbol** $\left(\frac{a}{p}\right)$ is defined as:

$$\left(\frac{a}{p}\right) = \begin{cases} 0 & \text{if } p \mid a \\ 1 & \text{if } a \text{ is a quadratic residue modulo } p \\ -1 & \text{if } a \text{ is a quadratic non-residue modulo } p \end{cases}$$

The Legendre symbol gives us a clean way to encode information about quadratic residues. For instance, with $p = 7$:

$$\left(\frac{1}{7}\right) = 1, \quad \left(\frac{2}{7}\right) = 1, \quad \left(\frac{3}{7}\right) = -1, \quad \left(\frac{4}{7}\right) = 1$$

Theorem: Euler's Criterion

Let p be an odd prime and a be an integer not divisible by p . Then:

$$\left(\frac{a}{p}\right) \equiv a^{\frac{p-1}{2}} \pmod{p}$$

Proof

Fermat's Little Theorem tells us that when $\gcd(a, p) = 1$ for prime p , we have $a^{p-1} \equiv 1 \pmod{p}$. This means $(a^{\frac{p-1}{2}})^2 \equiv 1 \pmod{p}$, so $a^{\frac{p-1}{2}} \equiv \pm 1 \pmod{p}$.

Case 1: Suppose a is a quadratic residue, so $a \equiv x^2 \pmod{p}$ for some x .

Then:

$$a^{\frac{p-1}{2}} \equiv (x^2)^{\frac{p-1}{2}} \equiv x^{p-1} \equiv 1 \pmod{p}$$

where the last congruence follows from Fermat's Little Theorem.

Case 2: Suppose a is a quadratic non-residue.

Consider the polynomial $f(x) = x^{\frac{p-1}{2}} - 1$. This has degree $\frac{p-1}{2}$, so it has at most $\frac{p-1}{2}$ roots modulo p .

From Case 1, we know that all $\frac{p-1}{2}$ quadratic residues are roots of this polynomial.

Since there are $\frac{p-1}{2}$ quadratic non-residues, and each must satisfy $a^{\frac{p-1}{2}} \equiv \pm 1 \pmod{p}$, we must have $a^{\frac{p-1}{2}} \equiv -1 \pmod{p}$ for all quadratic non-residues a .

Therefore, $\left(\frac{a}{p}\right) \equiv a^{\frac{p-1}{2}} \pmod{p}$.

□

Euler's criterion gives us a computational method for determining quadratic residues without having to check all possible squares.

9.2.1 Properties of the Legendre Symbol

Theorem: Properties of the Legendre Symbol

Let p be an odd prime and a, b be integers. Then:

1. $\left(\frac{ab}{p}\right) = \left(\frac{a}{p}\right) \left(\frac{b}{p}\right)$ (multiplicativity)
2. $\left(\frac{a^2}{p}\right) = 1$ if $\gcd(a, p) = 1$
3. $\left(\frac{-1}{p}\right) = (-1)^{\frac{p-1}{2}}$
4. $\left(\frac{2}{p}\right) = (-1)^{\frac{p^2-1}{8}}$

Proof

We will prove properties 1, 3, and 4; property 2 follows immediately from the definition.

Property 1: Using Euler's criterion:

$$\left(\frac{ab}{p}\right) \equiv (ab)^{\frac{p-1}{2}} \pmod{p} \quad (9.2.1)$$

$$\equiv a^{\frac{p-1}{2}} b^{\frac{p-1}{2}} \pmod{p} \quad (9.2.2)$$

$$\equiv \left(\frac{a}{p}\right) \left(\frac{b}{p}\right) \pmod{p} \quad (9.2.3)$$

Since both sides are in $\{-1, 0, 1\}$, equality holds.

Property 3: Using Euler's criterion:

$$\left(\frac{-1}{p}\right) \equiv (-1)^{\frac{p-1}{2}} \pmod{p}$$

Since $(-1)^{\frac{p-1}{2}} \in \{-1, 1\}$ and $p > 2$, we have equality.

Note that $(-1)^{\frac{p-1}{2}} = 1$ if and only if $\frac{p-1}{2}$ is even, which occurs when $p \equiv 1 \pmod{4}$.

Property 4: Part 1: Proof of the Congruence

We aim to prove that for any odd prime p :

$$2^{\frac{p-1}{2}} \equiv (-1)^{\frac{p^2-1}{8}} \pmod{p}$$

Let us define $M = \frac{p-1}{2}$. Consider the product of the first M even integers:

$$P = 2 \cdot 4 \cdot 6 \cdots (p-1) \tag{9.2.4}$$

$$= 2^M (1 \cdot 2 \cdot 3 \cdots M) \tag{9.2.5}$$

$$= 2^M M! \tag{9.2.6}$$

So, we have our first expression for the product:

$$P \equiv 2^{\frac{p-1}{2}} \left(\frac{p-1}{2} \right)! \pmod{p}$$

Now, let's examine the terms $\{2, 4, 6, \dots, p-1\}$ individually modulo p . We'll rearrange them into a set of residues with the smallest possible absolute values. For each term $2k$ (where $k = 1, \dots, M$):

- If $2k \leq \frac{p-1}{2}$, the term remains as it is.
- If $2k > \frac{p-1}{2}$, we can write $2k \equiv 2k - p \pmod{p}$. Notice that $-(2k - p) = p - 2k$, which is a positive integer.

Let's count how many of our terms fall into the second category. A term $2k$ is greater than $p/2$ if $k > p/4$. Let's call the number of such terms m .

$$m = (\text{total terms}) - \left(\text{terms with } k \leq \frac{p}{4} \right) = \frac{p-1}{2} - \left\lfloor \frac{p}{4} \right\rfloor$$

For these m terms, we replace $2k$ with the congruent value $-(p - 2k)$. This introduces a factor of $(-1)^m$ into our product. The new set of values consists of:

- The original terms $2k$ where $k \leq \lfloor p/4 \rfloor$.
- The new positive values $p - 2k$ where $k > p/4$.

It can be shown that this new collection of numbers $\{2, 4, \dots, 2\lfloor p/4 \rfloor\}$ and $\{p - 2(\lfloor p/4 \rfloor + 1), \dots, p - (p - 1) = 1\}$ is simply a permutation of the set $\{1, 2, \dots, M\}$.

Therefore, the product P can also be written as:

$$P \equiv (-1)^m (1 \cdot 2 \cdot 3 \cdots M) \equiv (-1)^m M! \pmod{p}$$

We now have two expressions for $P \pmod{p}$. Equating them gives:

$$2^M M! \equiv (-1)^m M! \pmod{p}$$

Since p is a prime and $M = \frac{p-1}{2} < p$, $M!$ is not divisible by p . Thus, we can cancel $M!$ from both sides:

$$2^M \equiv (-1)^m \pmod{p}$$

Substituting $M = \frac{p-1}{2}$ and $m = \frac{p-1}{2} - \lfloor p/4 \rfloor$, we have:

$$2^{\frac{p-1}{2}} \equiv (-1)^{\frac{p-1}{2} - \lfloor p/4 \rfloor} \pmod{p}$$

By analyzing the possible values of p modulo 8, we can show that the exponent m has the same parity as $\frac{p^2-1}{8}$.

- If $p = 8k \pm 1$, m is even and $\frac{p^2-1}{8}$ is even.
- If $p = 8k \pm 3$, m is odd and $\frac{p^2-1}{8}$ is odd.

Therefore, we can replace the exponent m with $\frac{p^2-1}{8}$, completing the first part of the proof:

$$2^{\frac{p-1}{2}} \equiv (-1)^{\frac{p^2-1}{8}} \pmod{p}$$

Now we use the congruence to prove the final property. Euler's Criterion suggests:

$$\left(\frac{2}{p}\right) \equiv 2^{\frac{p-1}{2}} \pmod{p} \quad (9.2.7)$$

$$\equiv (-1)^{\frac{p^2-1}{8}} \pmod{p} \quad (9.2.8)$$

Since both sides of this congruence can only be 1 or -1, and p is an odd prime (so $p > 2$), the congruence must be an equality:

$$\left(\frac{2}{p}\right) = (-1)^{\frac{p^2-1}{8}}$$

□

Let us compute $\left(\frac{5}{13}\right)$ using Euler's criterion to see the direct computational approach. We need to compute $5^{\frac{13-1}{2}} = 5^6 \pmod{13}$.

$$\begin{aligned} 5^2 &\equiv 25 \equiv 12 \equiv -1 \pmod{13} \\ 5^4 &\equiv (-1)^2 \equiv 1 \pmod{13} \\ 5^6 &\equiv 5^4 \cdot 5^2 \equiv 1 \cdot (-1) \equiv -1 \pmod{13} \end{aligned}$$

Therefore, $\left(\frac{5}{13}\right) = -1$, so 5 is a quadratic non-residue modulo 13. We will later verify this result using quadratic reciprocity, which provides a more efficient method for larger numbers.

9.3 Quadratic Reciprocity

Quadratic reciprocity theorem tells us that the question “is p a quadratic residue modulo q?” is intimately connected to the question “is q a quadratic residue modulo p?”

Euler first empirically observed quadratic reciprocity in the mid-18th century. Legendre was the first one to clearly formulated the reciprocity law in 1785. Even though he provided a proof, it had crucial gaps. Gauss provided the first complete proof of the theorem at the age of 19 in his *Disquisitiones Arithmeticae* (1801). He was so captivated with the proof that he discovered eight different proofs for the theorem throughout his lifetime, each revealing new insights between number theory, algebra, and geometry.

Theorem: Gauss's Lemma

Let p be an odd prime and a be an odd integer with $\gcd(a, p) = 1$. Consider the integers $a, 2a, 3a, \dots, \frac{p-1}{2}a$ and their least positive residues modulo p. Let n be the number of these residues that exceed $\frac{p}{2}$. Then:

$$\left(\frac{a}{p}\right) = (-1)^n$$

Proof

Consider the set of the first $\frac{p-1}{2}$ multiples of a:

$$S_{\text{mult}} = \left\{ a, 2a, 3a, \dots, \frac{p-1}{2}a \right\}$$

Let their least positive residues modulo p form the set $R = \{r_1, r_2, \dots, r_{\frac{p-1}{2}}\}$, where $ka \equiv r_k \pmod{p}$ and $1 \leq r_k \leq p-1$.

All these residues r_k are distinct. If not, we would have $ia \equiv ja \pmod{p}$ for some $1 \leq i, j \leq \frac{p-1}{2}$. Since $\gcd(a, p) = 1$, we can cancel a to get $i \equiv j \pmod{p}$. This implies $i = j$. Furthermore, no residue can be 0 since p does not divide a or k .

Now, we partition the set of residues R into two disjoint subsets:

$$S_1 = \left\{ r \in R \mid r < \frac{p}{2} \right\} \text{ and } S_2 = \left\{ r \in R \mid r > \frac{p}{2} \right\}$$

Let $n = |S_2|$. Note that no residue can be exactly $\frac{p}{2}$ since p is odd.

Claim: $S_1 \cup \{p - r \mid r \in S_2\} = \{1, 2, \dots, \frac{p-1}{2}\}$

Let us define $T = S_1 \cup \{p - r \mid r \in S_2\}$ and $U = \{1, 2, \dots, \frac{p-1}{2}\}$. To prove $T = U$, we will show that all elements of T are distinct and lie in the range $[1, \frac{p-1}{2}]$. Since $|T| = |S_1| + |S_2| = \frac{p-1}{2} = |U|$, this will be sufficient.

We know that if $x \in S_1$, then by definition $1 \leq x < \frac{p}{2}$. If $y = p - r$ with $r \in S_2$, then by definition $\frac{p}{2} < r \leq p - 1$. This implies $0 < p - r < \frac{p}{2}$. Thus, every element of T is an integer in the set $\{1, 2, \dots, \frac{p-1}{2}\}$.

Next we need to show the distinctness of elements in T . We already know that all the elements of S_1 and $\{p - r \mid r \in S_2\}$ are distinct. We must now show that an element from S_1 cannot be equal to an element from $\{p - r \mid r \in S_2\}$.

Let us assume for contradiction that there exists $x \in S_1$ and $y = p - r$ (with $r \in S_2$) such that $x = y$. This means $x = p - r$, or $x + r = p$.

By definition, $x \equiv ia \pmod{p}$ and $r \equiv ja \pmod{p}$ for some distinct $i, j \in \{1, 2, \dots, \frac{p-1}{2}\}$. Substituting these into the congruence $x + r \equiv 0 \pmod{p}$:

$$ia + ja \equiv 0 \pmod{p}$$

$$a(i + j) \equiv 0 \pmod{p}$$

Since $\gcd(a, p) = 1$, we can cancel a to get $i + j \equiv 0 \pmod{p}$. However, $1 \leq i \leq \frac{p-1}{2}$ and $1 \leq j \leq \frac{p-1}{2}$. This gives the strict bounds $2 \leq i + j \leq p - 1$. The sum $i + j$ cannot be a multiple of p in this range. This is a contradiction.

Therefore, the sets S_1 and $\{p - r \mid r \in S_2\}$ are disjoint.

Since all elements of T are distinct and lie in U , and $|T| = |U|$, the sets must be identical. This proves the claim.

Let us take a look at the product of all the residues in $R = \{r_1, \dots, r_{\frac{p-1}{2}}\}$:

$$\prod_{k=1}^{\frac{p-1}{2}} r_k \equiv \prod_{k=1}^{\frac{p-1}{2}} (ka) \pmod{p} \quad (9.3.1)$$

$$\equiv \left(\prod_{k=1}^{\frac{p-1}{2}} k \right) \left(\prod_{k=1}^{\frac{p-1}{2}} a \right) \pmod{p} \quad (9.3.2)$$

$$\equiv \left(\frac{p-1}{2} \right)! \cdot a^{\frac{p-1}{2}} \pmod{p} \quad (9.3.3)$$

The product of the residues also equals to the product of elements from S_1 and S_2 . For each $r \in S_2$, we can write $r \equiv -(p-r) \pmod{p}$. There are $n = |S_2|$ such elements.

$$\prod_{k=1}^{\frac{p-1}{2}} r_k = \left(\prod_{r \in S_1} r \right) \cdot \left(\prod_{r \in S_2} r \right) \quad (9.3.4)$$

$$\equiv \left(\prod_{r \in S_1} r \right) \cdot \left(\prod_{r \in S_2} (-(p-r)) \right) \pmod{p} \quad (9.3.5)$$

$$\equiv (-1)^n \left(\prod_{r \in S_1} r \right) \cdot \left(\prod_{r \in S_2} (p-r) \right) \pmod{p} \quad (9.3.6)$$

The set of numbers inside the products is $S_1 \cup \{p-r \mid r \in S_2\}$. Following the claim, this set is exactly $\{1, 2, \dots, \frac{p-1}{2}\}$. Therefore, the product of these numbers is simply $\left(\frac{p-1}{2}\right)!$. In other words:

$$\prod_{k=1}^{\frac{p-1}{2}} r_k \equiv (-1)^n \cdot \left(\frac{p-1}{2} \right)! \pmod{p}$$

Equating the two derivations, we have:

$$a^{\frac{p-1}{2}} \cdot \left(\frac{p-1}{2} \right)! \equiv (-1)^n \cdot \left(\frac{p-1}{2} \right)! \pmod{p}$$

Since p is prime, it does not divide any integer from 1 to $\frac{p-1}{2}$, so $\gcd\left(\left(\frac{p-1}{2}\right)!, p\right) = 1$. We can safely cancel the factorial from both sides of the congruence:

$$a^{\frac{p-1}{2}} \equiv (-1)^n \pmod{p}$$

By Euler's criterion, the left side is congruent to the Legendre symbol $\left(\frac{a}{p}\right)$. Since both sides are either 1 or -1 , the congruence implies equality:

$$\left(\frac{a}{p}\right) = (-1)^n$$

□

The Law of Quadratic Reciprocity states that for two distinct odd primes p and q :

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

Before we discuss a formal proof of the theorem, let us first develop a geometric intuition about the statement. The key insight lies in visualizing the problem through lattice points in the coordinate plane. This elegant approach, developed by Gotthold Eisenstein, transforms the abstract question of quadratic residues into a concrete counting problem.

For the distinct odd primes p and q , we work within the rectangle defined by:

$$1 \leq x \leq \frac{p-1}{2} \quad \text{and} \quad 1 \leq y \leq \frac{q-1}{2}$$

Since p and q are odd, both $\frac{p-1}{2}$ and $\frac{q-1}{2}$ are integers, so our rectangle contains exactly $\frac{p-1}{2} \cdot \frac{q-1}{2}$ lattice points (points with integer coordinates).

The crucial element is the diagonal line from the origin, with equation:

$$y = \frac{q}{p}x$$

This line divides our rectangle into two regions: lattice points below the diagonal and lattice points above it. It is important to realize that there are no lattice points on the line itself.

Claim: *There are no lattice points on the diagonal line $y = \frac{q}{p}x$ within our rectangle.*

Proof: If (a, b) were a lattice point on the line within our rectangle, then $b = \frac{qa}{p}$. Since a and b are integers, this would require p to divide qa . Since $\gcd(p, q) = 1$, we would need p to divide a . But $1 \leq a \leq \frac{p-1}{2} < p$, so a cannot be divisible by p . This contradiction shows no such lattice points exist.

For the Legendre symbol $\left(\frac{q}{p}\right)$, Gauss's Lemma tells us to examine the multiples $q, 2q, 3q, \dots, \frac{p-1}{2}q$

modulo p . Let μ be the number of these multiples whose least positive residues exceed $\frac{p}{2}$. Then:

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

The key insight is that this count μ equals the number of lattice points (x, y) with integer coordinates satisfying:

- $1 \leq x \leq \frac{p-1}{2}$ and $1 \leq y \leq \frac{q-1}{2}$
- The point lies strictly below the diagonal line $y = \frac{q}{p}x$

To establish this connection, consider a multiple kq where $1 \leq k \leq \frac{p-1}{2}$. We can write:

$$kq = \left\lfloor \frac{kq}{p} \right\rfloor \cdot p + r$$

where r is the remainder when kq is divided by p .

The remainder $r > \frac{p}{2}$ if and only if:

$$\frac{p}{2} < kq - \left\lfloor \frac{kq}{p} \right\rfloor \cdot p \iff \left\lfloor \frac{kq}{p} \right\rfloor + \frac{1}{2} < \frac{kq}{p}$$

This is equivalent to saying that the lattice point $\left(k, \left\lfloor \frac{kq}{p} \right\rfloor\right)$ lies strictly below the diagonal line $y = \frac{q}{p}x$, because:

$$\left\lfloor \frac{kq}{p} \right\rfloor < \frac{kq}{p} = \frac{q}{p} \cdot k$$

Moreover, since $1 \leq k \leq \frac{p-1}{2}$ and $\left\lfloor \frac{kq}{p} \right\rfloor \leq \frac{kq}{p} < \frac{q(p-1)}{2p} < \frac{q}{2}$, we have $1 \leq \left\lfloor \frac{kq}{p} \right\rfloor \leq \frac{q-1}{2}$, so this lattice point lies within our rectangle. Therefore, each multiple kq with remainder $> \frac{p}{2}$ corresponds to exactly one lattice point below the diagonal in our rectangle.

Similarly, for the Legendre symbol $\left(\frac{p}{q}\right)$, we examine multiples $p, 2p, 3p, \dots, \frac{q-1}{2}p$ modulo q . Let ν be the number of these whose remainders exceed $\frac{q}{2}$. Then:

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

This count ν equals the number of lattice points (x, y) in the same rectangle that lie strictly above the diagonal line $y = \frac{q}{p}x$.

Since every lattice point is either strictly above or strictly below the diagonal:

$$\mu + \nu = \frac{p-1}{2} \cdot \frac{q-1}{2}$$

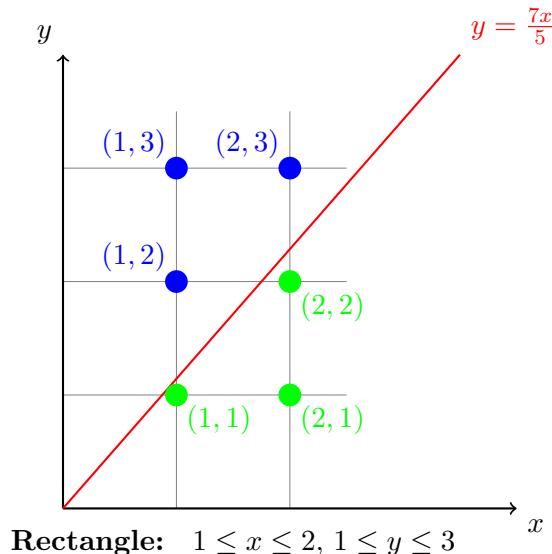
Quadratic Reciprocity follows immediately:

$$\left(\frac{q}{p}\right) \left(\frac{p}{q}\right) = (-1)^\mu \cdot (-1)^\nu = (-1)^{\mu+\nu} = (-1)^{\frac{p-1}{2} \cdot \frac{q-1}{2}}$$

This geometric argument reveals the profound symmetry underlying quadratic reciprocity. The abstract algebraic relationship between two Legendre symbols becomes a concrete statement about counting lattice points on opposite sides of a diagonal line.

A Visual Example

Let us look at quadratic reciprocity for $p = 5$ and $q = 7$. The rectangle contains lattice points (x, y) where $1 \leq x \leq 2$ and $1 \leq y \leq 3$. The diagonal line is $y = \frac{7x}{5} = 1.4x$.



Count:

- μ = points below diagonal = $\{(1,1), (2,1), (2,2)\} \Rightarrow \mu = 3$
- ν = points above diagonal = $\{(1,2), (1,3), (2,3)\} \Rightarrow \nu = 3$

Quadratic Reciprocity:

$$\left(\frac{7}{5}\right) \left(\frac{5}{7}\right) = (-1)^\mu \cdot (-1)^\nu = (-1)^3 \cdot (-1)^3 = (-1) \cdot (-1) = 1$$

This matches the formula: $(-1)^{\frac{p-1}{2} \cdot \frac{q-1}{2}} = (-1)^{2 \cdot 3} = (-1)^6 = 1$.

This geometric visualization transforms the mysterious patterns of quadratic residues into an elegant counting problem, making the reciprocity law both intuitive and beautiful.

Theorem: Law of Quadratic Reciprocity

Let p and q be distinct odd primes. Then:

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

Equivalently:

$$\left(\frac{p}{q}\right) = \left(\frac{q}{p}\right) \quad \text{if } p \equiv 1 \pmod{4} \text{ or } q \equiv 1 \pmod{4} \quad (9.3.7)$$

$$\left(\frac{p}{q}\right) = -\left(\frac{q}{p}\right) \quad \text{if } p \equiv q \equiv 3 \pmod{4} \quad (9.3.8)$$

Proof

For $\left(\frac{q}{p}\right)$, let n be the number of residues among $\{q, 2q, 3q, \dots, \frac{p-1}{2}q\}$ that exceed $\frac{p}{2}$ when reduced modulo p . Similarly, for $\left(\frac{p}{q}\right)$, let m be the number of residues among $\{p, 2p, 3p, \dots, \frac{q-1}{2}p\}$ that exceed $\frac{q}{2}$ when reduced modulo q .

By Gauss's lemma:

$$\left(\frac{q}{p}\right) = (-1)^n \quad \text{and} \quad \left(\frac{p}{q}\right) = (-1)^m$$

Through geometric arguments involving lattice points, one can show that:

$$n + m = \frac{p-1}{2} \cdot \frac{q-1}{2}$$

Therefore:

$$\left(\frac{p}{q}\right) \left(\frac{q}{p}\right) = (-1)^m \cdot (-1)^n = (-1)^{n+m} = (-1)^{\frac{p-1}{2} \cdot \frac{q-1}{2}}$$

□

Example 9.3.1

Compute $\left(\frac{5}{13}\right)$ using quadratic reciprocity.

Solution: Using quadratic reciprocity with $p = 13$ and $q = 5$:

$$\left(\frac{5}{13}\right) \left(\frac{13}{5}\right) = (-1)^{\frac{13-1}{2} \cdot \frac{5-1}{2}} = (-1)^{6 \cdot 2} = (-1)^{12} = 1$$

Since the exponent is even, we have $\left(\frac{5}{13}\right) = \left(\frac{13}{5}\right)$.

Now, $13 \equiv 3 \pmod{5}$, so:

$$\left(\frac{13}{5}\right) = \left(\frac{3}{5}\right)$$

To find $\left(\frac{3}{5}\right)$, we use quadratic reciprocity again. Since both 3 and 5 are odd primes with $3 \equiv 5 \equiv 3 \pmod{4}$:

$$\left(\frac{3}{5}\right) \left(\frac{5}{3}\right) = (-1)^{\frac{3-1}{2} \cdot \frac{5-1}{2}} = (-1)^{1 \cdot 2} = (-1)^2 = 1$$

Since the exponent is even, we have $\left(\frac{3}{5}\right) = \left(\frac{5}{3}\right)$.

Since $5 \equiv 2 \pmod{3}$:

$$\left(\frac{5}{3}\right) = \left(\frac{2}{3}\right) = (-1)^{\frac{3^2-1}{8}} = (-1)^{\frac{9-1}{8}} = (-1)^1 = -1$$

Therefore:

$$\left(\frac{3}{5}\right) = -1$$

Thus, $\left(\frac{5}{13}\right) = -1$.

We can verify this matches our earlier computation using Euler's criterion: $5^{\frac{13-1}{2}} = 5^6 \equiv (-1) \pmod{13}$, confirming our result.

9.4 Computing Legendre Symbols Efficiently

Quadratic reciprocity, combined with the properties of the Legendre symbol, gives us an efficient algorithm for computing $\left(\frac{a}{p}\right)$ without using Euler's criterion.

Algorithm: Computing Legendre Symbols

To compute $\left(\frac{a}{p}\right)$ where p is an odd prime:

1. Do a prime factorization of a , where $a = 2^j \cdot p_1^{e_1} \cdot p_2^{e_2} \cdots p_r^{e_r}$ where each p_i is an

- odd prime
2. Use $\left(\frac{2^j}{p}\right) = \left(\frac{2}{p}\right)^j$ and $\left(\frac{2}{p}\right) = (-1)^{\frac{p^2-1}{8}}$
 3. For each odd prime factor p_i of a , use quadratic reciprocity to compute $\left(\frac{p_i}{p}\right)$
 4. Get the final result using multiplicativity: $\left(\frac{a}{p}\right) = \left(\frac{2^j}{p}\right) \prod_{i=1}^r \left(\frac{p_i}{p}\right)^{e_i}$

Example 9.4.1

Compute $\left(\frac{45}{13}\right)$ efficiently.

Solution: Using multiplicativity:

$$\left(\frac{45}{13}\right) = \left(\frac{3^2 \cdot 5}{13}\right) = \left(\frac{3^2}{13}\right) \left(\frac{5}{13}\right) = 1 \cdot \left(\frac{5}{13}\right)$$

We already computed $\left(\frac{5}{13}\right) = -1$. Therefore, $\left(\frac{45}{13}\right) = -1$.

9.5 The Jacobi Symbol

The Legendre symbol can be generalized to composite numbers through the Jacobi symbol.

Definition: Jacobi Symbol

Let $n = p_1^{a_1} p_2^{a_2} \cdots p_k^{a_k}$ be the prime factorization of an odd positive integer $n > 1$, and let a be an integer coprime to n . The **Jacobi symbol** $\left(\frac{a}{n}\right)$ is defined as:

$$\left(\frac{a}{n}\right) = \prod_{i=1}^k \left(\frac{a}{p_i}\right)^{a_i}$$

where each $\left(\frac{a}{p_i}\right)$ is a Legendre symbol. Note that $\left(\frac{a}{p_i}\right)^{a_i}$ means the Legendre symbol raised to the power a_i .

For example, if $n = 3^2 \cdot 5 = 45$, then:

$$\left(\frac{a}{45}\right) = \left(\frac{a}{3}\right)^2 \left(\frac{a}{5}\right) = \left(\frac{a}{3}\right)^2 \left(\frac{a}{5}\right)$$

Since $\left(\frac{a}{3}\right) \in \{-1, 1\}$, we have $\left(\frac{a}{3}\right)^2 = 1$ always, so:

$$\left(\frac{a}{45}\right) = \left(\frac{a}{5}\right)$$

The Jacobi symbol retains many properties of the Legendre symbol, including multiplicativity and a generalized form of quadratic reciprocity. However, $(\frac{a}{n}) = 1$ does not necessarily mean that a is a quadratic residue modulo n when n is composite.

Example 9.5.1

Compute $(\frac{2}{15})$ using the Jacobi symbol.

Solution: Since $15 = 3 \cdot 5$:

$$\left(\frac{2}{15}\right) = \left(\frac{2}{3}\right) \left(\frac{2}{5}\right)$$

Using the formula for $\left(\frac{2}{p}\right)$:

$$\left(\frac{2}{3}\right) = (-1)^{\frac{3^2-1}{8}} = (-1)^{\frac{8}{8}} = -1 \quad (9.5.1)$$

$$\left(\frac{2}{5}\right) = (-1)^{\frac{5^2-1}{8}} = (-1)^{\frac{24}{8}} = (-1)^3 = -1 \quad (9.5.2)$$

Therefore:

$$\left(\frac{2}{15}\right) = (-1)(-1) = 1$$

Note that this doesn't mean 2 is a quadratic residue modulo 15, since $x^2 \equiv 2 \pmod{15}$ has no integer solutions.

9.6 Quadratic Forms and the Legendre Symbol

The study of quadratic forms, particularly those of the type $x^2 + ky^2$ where k is a positive integer, reveals deep connections to quadratic reciprocity and the Legendre symbol. This section explores how the representability of integers by such forms relates to quadratic residues modulo primes, and extends these results to composite numbers where possible.

9.6.1 Representation of Primes

Our fundamental question is: for which primes p do there exist integers x, y such that $x^2 + ky^2 = p$? Such a representation is called *primitive* if $\gcd(x, y) = 1$.

When representing a prime p , any solution is necessarily primitive. To see this, suppose a representation $x^2 + ky^2 = p$ exists and let $d = \gcd(x, y)$. We can write $x = da$ and $y = db$ for some integers a and b . Substituting these into the equation gives $d^2(a^2 + kb^2) = p$. This

implies that d^2 must divide p . Since the only positive square divisor of a prime number is 1, we must have $d^2 = 1$, and thus $d = 1$. Therefore, any representation of a prime by this form is automatically primitive.

Theorem: The Quadratic Form Residue Theorem

Let p be an odd prime and k be a positive integer with $\gcd(k, p) = 1$. If $x^2 + ky^2 = p$ has a solution in integers, then $\left(\frac{-k}{p}\right) = 1$.

Let us recall that the Legendre symbol $\left(\frac{a}{p}\right)$ for an odd prime p and integer a with $\gcd(a, p) = 1$ is defined as:

$$\left(\frac{a}{p}\right) = \begin{cases} 1 & \text{if } a \text{ is a quadratic residue modulo } p \\ -1 & \text{if } a \text{ is a quadratic non-residue modulo } p \end{cases}$$

Our theorem shows that a necessary condition for p to be represented by $x^2 + ky^2$ is $\left(\frac{-k}{p}\right) = 1$.

Examples:

- For $k = 1$: We need $\left(\frac{-1}{p}\right) = 1$, which holds if and only if $p \equiv 1 \pmod{4}$.
- For $k = 2$: We need $\left(\frac{-2}{p}\right) = \left(\frac{-1}{p}\right)\left(\frac{2}{p}\right) = 1$. This holds if and only if $p \equiv 1, 3 \pmod{8}$.

Proof

Suppose $x^2 + ky^2 = p$ for some integers x, y . This implies: $x^2 \equiv -ky^2 \pmod{p}$

We first show that p cannot divide y . If $p \mid y$, then $x^2 \equiv 0 \pmod{p}$, which implies $p \mid x$. In this case, we can write $x = px'$ and $y = py'$ for some integers x', y' . Substituting into the original equation gives $(px')^2 + k(py')^2 = p$, or $p^2(x'^2 + ky'^2) = p$. This implies $p(x'^2 + ky'^2) = 1$, which is impossible for integers. Thus, $p \nmid y$.

Since $p \nmid y$, y has a multiplicative inverse modulo p . We can therefore write: $(xy^{-1})^2 \equiv -k \pmod{p}$

This shows that $-k$ is a quadratic residue modulo p , which by definition means $\left(\frac{-k}{p}\right) = 1$.

□

The converse statement: if $\left(\frac{-k}{p}\right) = 1$, then p is representable by the form is a much deeper question. It is true for certain values of k (e.g., $k = 1, 2, 3$) but fails for many others.

Theorem: Sum of Two Squares for Primes

Let p be an odd prime. Then p can be represented as $x^2 + y^2$ iff $p \equiv 1 \pmod{4}$.

This is a case where the converse holds. The condition $p \equiv 1 \pmod{4}$ is precisely the condition for which $\left(\frac{-1}{p}\right) = 1$.

Proof

Necessity: Suppose $p = x^2 + y^2$ where p is an odd prime. Since p is odd, x and y cannot both be even. Also, they cannot both be odd (since then $x^2 + y^2 \equiv 1 + 1 = 2 \pmod{4}$, but no prime except 2 is congruent to 2 modulo 4). Therefore, one of x, y is even and the other is odd.

Without loss of generality, let x be odd and y be even. Then:

$$x^2 \equiv 1 \pmod{4}, \quad y^2 \equiv 0 \pmod{4}$$

So $p = x^2 + y^2 \equiv 1 + 0 = 1 \pmod{4}$.

Sufficiency: This direction requires more advanced techniques, which is beyond the scope of this book. However, the key idea is to use the fact that if $p \equiv 1 \pmod{4}$, then -1 is a quadratic residue modulo p . This means there exists an integer a such that $a^2 \equiv -1 \pmod{p}$, or equivalently, $a^2 + 1^2 \equiv 0 \pmod{p}$. Using properties of Gaussian integers (which is beyond our current scope), one can show this leads to an actual representation $p = x^2 + y^2$.

□

The condition $p \equiv 1 \pmod{4}$ is equivalent to saying that -1 is a quadratic residue modulo p . Recall from our study of Legendre symbols:

$$\left(\frac{-1}{p}\right) = (-1)^{\frac{p-1}{2}}$$

When $p \equiv 1 \pmod{4}$, we have $p = 4k + 1$ for some integer k , so:

$$\frac{p-1}{2} = \frac{4k+1-1}{2} = 2k$$

which is even. Therefore $\left(\frac{-1}{p}\right) = (-1)^{2k} = 1$.

When $p \equiv 3 \pmod{4}$, we have $p = 4k + 3$, so:

$$\frac{p-1}{2} = \frac{4k+3-1}{2} = 2k+1$$

which is odd. Therefore $\left(\frac{-1}{p}\right) = (-1)^{2k+1} = -1$.

Thus an odd prime p can be written as sum of two squares if and only if $\left(\frac{-1}{p}\right) = 1$. We can verify this result via manual checking. For instance, 17 can be represented as $17 = 1^2 + 4^2$, while 7 cannot be represented as a sum of two squares.

9.6.2 Extension to Composite Numbers

The extension to composite numbers is more delicate and depends heavily on the specific form and the arithmetic properties of k .

Necessary Conditions from Prime Factors

If a number n is representable by $x^2 + ky^2$, we can still gain information by reducing modulo its prime factors.

Theorem:

Let p be an odd prime dividing a positive integer n . If $n = x^2 + ky^2$ for some integers x, y with $\gcd(n, y) = 1$, then $\left(\frac{-k}{p}\right) = 1$.

Proof

Reducing $x^2 + ky^2 = n$ modulo p , we get $x^2 + ky^2 \equiv 0 \pmod{p}$. Since $\gcd(n, y) = 1$, we know $p \nmid y$, so y is invertible modulo p . As before, this leads to $(xy^{-1})^2 \equiv -k \pmod{p}$, proving that $\left(\frac{-k}{p}\right) = 1$.

□

This shows that primes for which $-k$ is a quadratic non-residue can only appear in the factorization of a representable number n in specific ways. For specific case where $k = 1$, a complete characterization of the previous theorem is as follows.

Theorem: Fermat-Euler Theorem on Sum of Two Squares

A positive integer n can be written as sum of two squares if and only if every prime $p \equiv 3 \pmod{4}$ appears to an even power in the prime factorization of n .

A complete proof of this theorem is again outside of the scope of the book. However, we can look at a couple of examples to develop an intuition. For instance, when $n = 12 = 2^2 \cdot 3$, the

prime 3 satisfies $3 \equiv 3 \pmod{4}$ and appears to power 1 (odd) in the factorization. So 12 cannot be written as sum of two squares. But for $n = 18 = 2 \cdot 3^2$, the prime 3 satisfies $3 \equiv 3 \pmod{4}$ but appears to power 2 (even). So 18 can be written as the sum of two squares in the form of $18 = 3^2 + 3^2$.

Let us discuss this more formally. Let us assume that $n = p_1 p_2$, where $p_1 \equiv p_2 \equiv 1 \pmod{4}$. Let us assume that, for a given natural number k , both the primes are represented as: $p_1 = a^2 + kb^2$ and $p_2 = c^2 + kd^2$. n can be represented as $(ac - kbd)^2 + k(ad + bc)^2$, which stems from the following algebraic identity: $(a^2 + kb^2)(c^2 + kd^2) = (ac - kbd)^2 + k(ad + bc)^2$.

The algebraic identity can be proved by expanding the right-hand side directly:

$$(ac - kbd)^2 + k(ad + bc)^2 = (ac)^2 - 2(ac)(kbd) + (kbd)^2 + k[(ad)^2 + 2(ad)(bc) + (bc)^2] \quad (9.6.1)$$

$$= a^2c^2 - 2abcdk + k^2b^2d^2 + ka^2d^2 + 2abckd + kb^2c^2 \quad (9.6.2)$$

$$= a^2c^2 + k^2b^2d^2 + ka^2d^2 + kb^2c^2 \quad (9.6.3)$$

$$= a^2(c^2 + kd^2) + kb^2(kd^2 + c^2) \quad (9.6.4)$$

$$= (a^2 + kb^2)(c^2 + kd^2) \quad (9.6.5)$$

This identity holds for all k . However, the identity is useful when n is a product of primes that can be represented as $a^2 + kb^2$ for some natural number k .

Conditions for Composite Representability

The complete characterization depends on k :

1. **For $k = 1$ (sums of two squares):** A positive integer n is representable by $x^2 + y^2$ iff every prime $p \equiv 3 \pmod{4}$ appears to an even power in the prime factorization of n .
2. **For $k = 2$:** A positive integer n is representable by $x^2 + 2y^2$ if and only if every prime $p \equiv 5, 7 \pmod{8}$ appears to an even power in the prime factorization of n .
3. **For $k = 3$:** A positive integer n is representable by $x^2 + 3y^2$ if and only if every prime $p \equiv 2 \pmod{3}$ appears to an even power in the prime factorization of n .
4. **For general k :** The conditions become more complex and the results are not that straightforward.

9.7 Practice Exercises

Exercise 9.1

Determine all prime pairs (p, q) such that $\left(\frac{q}{p}\right) = \left(\frac{p}{q}\right)$, where $p + q = 80$.

Exercise 9.2

Consider the system of congruences: $x^2 \equiv a \pmod{p}$ and $x^2 \equiv b \pmod{q}$ where p, q are distinct odd primes and $\gcd(a, pq) = \gcd(b, pq) = 1$. Determine the number of solutions modulo pq .

Exercise 9.3

Let $p = 2^n + 1$ be a Fermat prime. Prove that every quadratic non-residue modulo p is also a primitive root modulo p .

Exercise 9.4

Prove that there are infinitely many primes of the form $4k + 1$.

Exercise 9.5

Let $p \equiv 1 \pmod{4}$ be prime, written as $p = a^2 + b^2$ for integers a, b . Prove that $\left(\frac{a^2 - b^2}{p}\right) = \left(\frac{2}{p}\right)$.

Exercise 9.6

For which odd primes p does the congruence $x^4 \equiv -4 \pmod{p}$ have a solution?

Exercise 9.7

Let p be an odd prime and let a be an integer not divisible by p . Prove that the sum $\sum_{k=0}^{p-1} \left(\frac{k^2-a}{p} \right) = -1$.

Exercise 9.8

Let n be an odd prime. The Gaussian Sum is defined as $S_n = \sum_{k=1}^{n-1} \left(\frac{k}{n} \right) e^{2\pi i k/n}$. Prove that $|S_n|^2 = n$.

Exercise 9.9

Let p be a prime greater than 3. Prove that the congruence $x^2 + y^2 + z^2 \equiv 1 \pmod{p}$ always has a solution.

Exercise 9.10

Let p be an odd prime. For an integer a with $\gcd(a, p) = 1$, consider the permutation π_a of the set $\{1, 2, \dots, p-1\}$ defined by $\pi_a(k) \equiv ak \pmod{p}$. Prove Zolotarev's Lemma: the sign of this permutation, $\text{sgn}(\pi_a)$, is equal to the Legendre symbol $\left(\frac{a}{p} \right)$.

Part III

Past ISI Exam Questions

Chapter 10

Questions from Past UGA Papers

10.1 2025

ISI 2025, Question 1

- In the xy -plane, the curve $3x^3y + 6xy + 2xy^3 = 0$ represents
- (A) a pair of straight lines
 - (B) an ellipse
 - (C) a pair of straight lines and an ellipse
 - (D) a hyperbola

ISI 2025, Question 3

- The coefficient of x^8 in $(1 - 3x)^6(1 + 9x^2)^6(1 + 3x)^6$ is
- (A) $-3^9 \times 5$
 - (B) $3^9 \times 5$
 - (C) $-3^8 \times 5$
 - (D) $3^8 \times 5$

ISI 2025, Question 9

- The number of ordered pairs (a, b) of positive integers with $a < b$ satisfying $a^2 + b^2 = 2025$ is
- (A) 0

- (B) 1
- (C) 2
- (D) 6

ISI 2025, Question 19

Let a, b, c, d be positive integers such that the product $abcd = 999$. Then the number of different ordered 4-tuples (a, b, c, d) is

- (A) 20
- (B) 48
- (C) 80
- (D) 84

10.2 2024

ISI 2024, Problem 2

Let j be a number selected at random from $\{1, 2, \dots, 2024\}$. What is the probability that j is divisible by 9 and 15?

- (A) $\frac{1}{23}$ (B) $\frac{1}{46}$ (C) $\frac{1}{44}$ (D) $\frac{1}{253}$

ISI 2024, Problem 3

Let S_n be the set of all n -digit numbers whose digits are all 1 or 2 and there are no consecutive 2's. (Example: 112 is in S_3 but 221 is not in S_3). Then the number of elements in S_{10} is

- (A) 512 (B) 256 (C) 144 (D) 89

ISI 2024, Problem 11

Let $n \geq 1$. The maximum possible number of primes in the set $\{n + 6, n + 7, \dots, n + 34, n + 35\}$ is

- (A) 7 (B) 8 (C) 12 (D) 13

ISI 2024, Problem 16

Let $n > 1$ be the smallest composite integer that is coprime to $\frac{10000!}{9900!}$. Then

- (A) $n \leq 100$ (B) $100 < n \leq 9900$ (C) $9900 < n \leq 10000$ (D) $n > 10000$

ISI 2024, Problem 24

Let $p < q$ be prime numbers such that $p^2 + q^2 + 7pq$ is a perfect square. Then, the largest possible value of q is:

- (A) 7 (B) 11 (C) 23 (D) 29

ISI 2024, Problem 25

The set of all real numbers x for which $3^{2^{1-x^2}}$ is an integer has

- (A) 3 elements (B) 15 elements (C) 24 elements (D) infinitely many elements

10.3 2023

ISI 2023, Problem 4

The number of consecutive zeroes adjacent to the digit in the unit's place of 401^{50} is
(A) 3 (B) 4 (C) 49 (D) 50

ISI 2023, Problem 8

How many numbers formed by rearranging the digits of 234578 are divisible by 55?
(A) 0 (B) 12 (C) 36 (D) 72

ISI 2023, Problem 11

Suppose x and y are positive integers. If $4x + 3y$ and $2x + 4y$ are divided by 7, then the respective remainders are 2 and 5. If $11x + 5y$ is divided by 7, then the remainder equals
(A) 0. (B) 1. (C) 2. (D) 3.

ISI 2023, Problem 15

Let n be a positive integer having 27 divisors including 1 and n , which are denoted by d_1, \dots, d_{27} . Then the product of d_1, d_2, \dots, d_{27} equals

- (A) n^{13} . (B) n^{14} . (C) $n^{\frac{27}{2}}$. (D) $27n$.

ISI 2023, Problem 17

Suppose $z \in \mathbb{C}$ is such that the imaginary part of z is non-zero and $z^{25} = 1$. Then $\sum_{k=0}^{2023} z^k$ equals

- (A) 0. (B) 1. (C) $-1 - z^{24}$. (D) $-z^{24}$.

ISI 2023, Problem 20

If $[x]$ denotes the largest integer less than or equal to x , then $[(9 + \sqrt{80})^{20}]$ equals

- (A) $(9 + \sqrt{80})^{20} - (9 - \sqrt{80})^{20}$. (B) $(9 + \sqrt{80})^{20} + (9 - \sqrt{80})^{20} - 20$. (C)
 $(9 + \sqrt{80})^{20} + (9 - \sqrt{80})^{20} - 1$. (D) $(9 - \sqrt{80})^{20}$.

ISI 2023, Problem 24

The polynomial $x^{10} + x^5 + 1$ is divisible by

- (A) $x^2 + x + 1$. (B) $x^2 - x + 1$. (C) $x^2 + 1$. (D) $x^5 - 1$.

ISI 2023, Problem 25

Suppose $a, b, c \in \mathbb{R}$ and $f(x) = ax^2 + bx + c$, $x \in \mathbb{R}$. If $0 \leq f(x) \leq (x - 1)^2$ for all x , and $f(3) = 2$, then

- (A) $a = \frac{1}{2}$, $b = -1$, $c = \frac{1}{2}$. (B) $a = \frac{1}{3}$, $b = -\frac{1}{3}$, $c = 0$. (C) $a = \frac{2}{3}$, $b = -\frac{5}{3}$, $c = 1$. (D) $a = \frac{3}{4}$, $b = -2$, $c = \frac{5}{4}$.

ISI 2023, Problem 29

Suppose $f : \mathbb{Z} \rightarrow \mathbb{Z}$ is a non-decreasing function. Consider the following two cases:

Case 1. $f(0) = 2$, $f(10) = 8$, Case 2. $f(0) = -2$, $f(10) = 12$.

In which of the above cases it is necessarily true that there exists an n with $f(n) = n$?

- (A) In both cases. (B) In neither case. (C) In Case 1 but not necessarily in Case 2. (D) In Case 2 but not necessarily in Case 1.

ISI 2023, Problem 30

How many functions $f : \{1, 2, \dots, 10\} \rightarrow \{1, \dots, 2000\}$, which satisfy $f(i+1) - f(i) \geq 20$, for all $1 \leq i \leq 9$, are there?

- (A) $10! \binom{1829}{10}$ (B) $11! \binom{1830}{11}$ (C) $\binom{1829}{10}$ (D) $\binom{1830}{11}$

10.4 2022

ISI 2022, Problem 2

Any positive real number x can be expanded as $x = a_n \cdot 2^n + a_{n-1} \cdot 2^{n-1} + \cdots + a_1 \cdot 2^1 + a_0 \cdot 2^0 + a_{-1} \cdot 2^{-1} + a_{-2} \cdot 2^{-2} + \cdots$, for some $n \geq 0$, where each $a_i \in \{0, 1\}$. In the above-described expansion of 21.1875, the smallest positive integer k such that $a_{-k} \neq 0$ is:

- (A) 3 (B) 2 (C) 1 (D) 4

ISI 2022, Problem 6

Let \mathbb{Z} denote the set of integers. Let $f : \mathbb{Z} \rightarrow \mathbb{Z}$ be such that $f(x)f(y) = f(x+y) + f(x-y)$ for all $x, y \in \mathbb{Z}$. If $f(1) = 3$, then $f(7)$ equals

- (A) 840 (B) 844 (C) 843 (D) 842

ISI 2022, Problem 9

Suppose the numbers 71, 104 and 159 leave the same remainder r when divided by a certain number $N > 1$. Then, the value of $3N + 4r$ must equal:

- (A) 53 (B) 48 (C) 37 (D) 23

ISI 2022, Problem 10

In how many ways can we choose $a_1 < a_2 < a_3 < a_4$ from the set $\{1, 2, \dots, 30\}$ such that a_1, a_2, a_3, a_4 are in arithmetic progression?

- (A) 135 (B) 145 (C) 155 (D) 165

ISI 2022, Problem 19

The number of positive integers n less than or equal to 22 such that 7 divides $n^5 + 4n^4 + 3n^3 + 2022$ is

- (A) 7 (B) 8 (C) 9 (D) 10

ISI 2022, Problem 21

Let $1, \omega, \omega^2$ be the cube roots of unity. Then the product $(1 - \omega + \omega^2)(1 - \omega^2 + \omega^4)(1 - \omega^4 + \omega^8) \cdots (1 - \omega^{2^9} + \omega^{2^{10}})$ is equal to:

- (A) 2^{10} (B) 3^{10} (C) $2^{10}\omega$ (D) $3^{10}\omega^2$

ISI 2022, Problem 23

The number of triples (a, b, c) of positive integers satisfying the equation $\frac{1}{a} + \frac{1}{b} + \frac{1}{c} = 1 + \frac{2}{abc}$ and such that $a < b < c$, equals:

- (A) 3 (B) 2 (C) 1 (D) 0

10.5 2021

ISI 2021, Problem 3

The number of ways one can express $2^23^35^57^7$ as a product of two numbers a and b , where $\gcd(a, b) = 1$, and $1 < a < b$, is

- (A) 5. (B) 6. (C) 7. (D) 8.

ISI 2021, Problem 17

Define $a = p^3 + p^2 + p + 11$ and $b = p^2 + 1$, where p is any prime number. Let $d = \gcd(a, b)$. Then the set of possible values of d is

- (A) $\{1, 2, 5\}$ (B) $\{2, 5, 10\}$ (C) $\{1, 5, 10\}$ (D) $\{1, 2, 10\}$

ISI 2021, Problem 20

The number of all integer solutions of the equation $x^2 + y^2 + x - y = 2021$ is

- (A) 5. (B) 7. (C) 1. (D) 0.

ISI 2021, Problem 22

For a positive integer n , the equation $x^2 = n + y^2$, x, y integers, does not have a solution if and only if

- (A) $n = 2$. (B) n is a prime number. (C) n is an odd number. (D) n is an even number not divisible by 4.

10.6 2020

ISI 2020, Problem 1

The number of subsets of $\{1, 2, 3, \dots, 10\}$ having an odd number of elements is

- (A) 1024 (B) 512 (C) 256 (D) 50.

ISI 2020, Problem 8

Let a_n be the number of subsets of $\{1, 2, \dots, n\}$ that do not contain any two consecutive numbers. Then

- (A) $a_n = a_{n-1} + a_{n-2}$ (B) $a_n = 2a_{n-1}$ (C) $a_n = a_{n-1} - a_{n-2}$ (D) $a_n = a_{n-1} + 2a_{n-2}$.

Problem 19 (Revised Solution)

If a, b, c are distinct odd natural numbers, then the number of rational roots of the polynomial $ax^2 + bx + c = 0$ is

- (A) must be 0. (B) must be 1. (C) must be 2. (D) cannot be determined from the given data.

Exercise 10.1

Let $S = \{1, 2, \dots, n\}$. For any non-empty subset A of S , let $l(A)$ denote the largest number in A . If $f(n) = \sum_{A \subseteq S, A \neq \emptyset} l(A)$, that is, $f(n)$ is the sum of the numbers $l(A)$ while A ranges over all the nonempty subsets of S , then $f(n)$ is

- (A) $2^n(n + 1)$ (B) $2^n(n + 1) - 1$ (C) $2^n(n - 1)$ (D) $2^n(n - 1) + 1$

10.7 2019**ISI 2019, Problem 16**

A school allowed the students of a class to go to swim during the days March 11th to March 15, 2019. The minimum number of students the class should have had that ensures that at least two of them went to swim on the same set of dates is:

- (A) 6 (B) 32 (C) 33 (D) 121.

ISI 2019, Problem 17

Let $a_1 < a_2 < a_3 < a_4$ be positive integers such that $\sum_{i=1}^4 \frac{1}{a_i} = \frac{11}{6}$. Then, $a_4 - a_2$ equals

- (A) 11 (B) 10 (C) 9 (D) 8.

ISI 2019, Problem 22

Let the integers a_i for $0 \leq i \leq 54$ be defined by the equation $(1 + X + X^2)^{27} = a_0 + a_1X + a_2X^2 + \dots + a_{54}X^{54}$. Then, $a_0 + a_3 + a_6 + a_9 + \dots + a_{54}$ equals

- (A) 3^{26} (B) 3^{27} (C) 3^{28} (D) 3^{29} .

ISI 2019, Problem 26

The number of integers $n \geq 10$ such that the product $\binom{n}{10} \cdot \binom{n+1}{10}$ is a perfect square is:

- (A) 0 (B) 1 (C) 2 (D) 3

ISI 2019, Problem 27

Let $a \geq b \geq c \geq 0$ be integers such that $2^a + 2^b - 2^c = 144$. Then, $a + b - c$ equals:

- (A) 7 (B) 8 (C) 9 (D) 10.

ISI 2019, Problem 28

The number of integers n for which the cubic equation $X^3 - X + n = 0$ has 3 distinct integer solutions is:

- (A) 0 (B) 1 (C) 2 (D) infinite.

10.8 2018**ISI 2018, Problem 6**

A number is called a palindrome if it reads the same backward or forward. For example, 112211 is a palindrome. How many 6-digit palindromes are divisible by 495?

- (A) 10 (B) 11 (C) 30 (D) 45

ISI 2018, Problem 17

The number of pairs of integers (x, y) satisfying the equation $xy(x + y + 1) = 5^{2018} + 1$ is:

- (A) 0 (B) 2 (C) 1009 (D) 2018.

ISI 2018, Problem 25

The sum of all natural numbers a such that $a^2 - 16a + 67$ is a perfect square is:

- (A) 10 (B) 12 (C) 16 (D) 22.

10.9 2017**ISI 2017, Problem 6**

In the Mathematics department of a college, there are 60 first year students, 84 second year students, and 108 third year students. All of these students are to be divided into project groups such that each group has the same number of first year students, the same number of second year students, and the same number of third year students. What is the smallest possible size of each group?

- (A) 9 (B) 12 (C) 19 (D) 21.

Chapter 11

Questions from Past UGB Papers

11.1 2025

ISI 2025, UGB Problem 4

Let $S_1 = \{z \in \mathbb{C} : |z| = 1\}$ be the unit circle in the complex plane. Let $f : S_1 \rightarrow S_1$ be the map given by $f(z) = z^2$. We define $f^{(1)} := f$ and $f^{(k+1)} := f \circ f^{(k)}$ for $k \geq 1$. The smallest positive integer n such that $f^{(n)}(z) = z$ is called the period of z . Determine the total number of points in S_1 of period 2025.

(Hint: $2025 = 3^4 \times 5^2$)

ISI 2025, UGB Problem 5

Let a, b, c be nonzero real numbers such that $a + b + c \neq 0$. Assume that $\frac{1}{a} + \frac{1}{b} + \frac{1}{c} = \frac{1}{a+b+c}$. Show that for any odd integer k , $\frac{1}{a^k} + \frac{1}{b^k} + \frac{1}{c^k} = \frac{1}{a^k+b^k+c^k}$

ISI 2025, UGB Problem 6

Let \mathbb{N} denote the set of natural numbers, and let (a_i, b_i) , $1 \leq i \leq 9$, be nine distinct tuples in $\mathbb{N} \times \mathbb{N}$. Show that there are three distinct elements in the set $\{2^{a_i}3^{b_i} : 1 \leq i \leq 9\}$ whose product is a perfect cube.

ISI 2025, UGB Problem 8

Let $n \geq 2$ and let $a_1 \leq a_2 \leq \dots \leq a_n$ be positive integers such that $\sum_{i=1}^n a_i = \prod_{i=1}^n a_i$. Prove that $\sum_{i=1}^n a_i \leq 2n$ and determine when equality holds.

11.2 2024

ISI 2024, Question 8

In a sports tournament involving N teams, each team plays every other team exactly once. At the end of every match, the winning team gets 1 point and the losing team gets 0 points. At the end of the tournament, the total points received by the individual teams are arranged in decreasing order as follows:

$$x_1 \geq x_2 \geq \cdots \geq x_N.$$

Prove that for any $1 \leq k \leq N$,

$$\frac{N-k}{2} \leq x_k \leq N - \frac{k+1}{2}$$

11.3 2023

ISI 2023, Question 1

Determine all integers $n > 1$ such that every power of n has an odd number of digits.

ISI 2023, Question 4

Let n_1, n_2, \dots, n_{51} be distinct natural numbers each of which has exactly 2023 positive integer factors. For instance, 2^{2022} has exactly 2023 positive integer factors $1, 2, 2^2, \dots, 2^{2021}, 2^{2022}$. Assume that no prime larger than 11 divides any of the n_i 's. Show that there must be some perfect cube among the n_i 's. You may use the fact that $2023 = 7 \times 17 \times 17$.

11.4 2022

ISI 2022, Question 1

Consider a board having 2 rows and n columns. Thus there are $2n$ cells in the board. Each cell is to be filled in by 0 or 1.

- (a) In how many ways can this be done such that each row sum and each column sum is even?
- (b) In how many ways can this be done such that each row sum and each column sum is odd?

ISI 2022, Question 5

For any positive integer n , and $i = 1, 2$, let $f_i(n)$ denote the number of divisors of n of the form $3k + i$ (including 1 and n). Define, for any positive integer n ,

$$f(n) = f_1(n) - f_2(n).$$

Find the values of $f(5^{2022})$ and $f(21^{2022})$.

11.5 2021

ISI 2021, Question 2

Let $f : \mathbb{Z} \rightarrow \mathbb{Z}$ be a function satisfying $f(0) \neq 0 = f(1)$. Assume also that f satisfies equations (A) and (B) below.

$$f(xy) = f(x) + f(y) - f(x)f(y) \quad (\text{A})$$

$$f(x-y)f(x)f(y) = f(0)f(x)f(y) \quad (\text{B})$$

for all integers x, y .

- (i) Determine explicitly the set $\{f(a) : a \in \mathbb{Z}\}$.
- (ii) Assuming that there is a non-zero integer a such that $f(a) \neq 0$, prove that the set $\{b : f(b) \neq 0\}$ is infinite.

Exercise 11.1

Show that every positive rational number r can be uniquely expressed as a finite sum of the form

$$r = a_1 + \frac{a_2}{2!} + \frac{a_3}{3!} + \cdots + \frac{a_n}{n!}$$

where the a_k are integers satisfying $a_1 \geq 0$ and $0 \leq a_k \leq k-1$ for $k > 1$.

11.6 2020

ISI 2020, Question 1

Let i be a root of the equation $x^2 + 1 = 0$ and let ω be a root of the equation $x^2 + x + 1 = 0$. Construct a polynomial

$$f(x) = a_0 + a_1x + \dots + a_nx^n$$

where a_0, a_1, \dots, a_n are all integers such that $f(i + \omega) = 0$.

ISI 2020, Question 7

Consider a right-angled triangle with integer-valued sides $a < b < c$ where a, b, c are pairwise co-prime. Let $d = c - b$. Suppose d divides a . Then

- Prove that $d \leq 2$.
- Find all such triangles (i.e. all possible triplets a, b, c) with perimeter less than 100.

11.7 2019**ISI 2019, Question 1**

Prove that the positive integers n that cannot be written as a sum of r consecutive positive integers, with $r > 1$, are of the form $n = 2^l$ for some $l \geq 0$.

11.8 2018**ISI 2018, Question 7**

Let $a, b, c \in \mathbb{N}$ be such that

$$a^2 + b^2 = c^2 \text{ and } c - b = 1. \quad (11.8.1)$$

Prove that

- (i) a is odd,
- (ii) b is divisible by 4,
- (iii) $a^b + b^a$ is divisible by c .

11.9 2017

ISI 2017, Problem 1

Let the sequence $\{a_n\}_{n \geq 1}$ be defined by

$$a_n = \tan(n\theta),$$

where $\tan(\theta) = 2$. Show that for all n , a_n is a rational number which can be written with an odd denominator.

ISI 2017, Problem 5

Let $g : \mathbb{N} \rightarrow \mathbb{N}$ with $g(n)$ being the product of the digits of n .

- (a) Prove that $g(n) \leq n$ for all $n \in \mathbb{N}$.
- (b) Find all $n \in \mathbb{N}$, for which $n^2 - 12n + 36 = g(n)$.

ISI 2017, Problem 6

Let p_1, p_2, p_3 be primes with $p_2 \neq p_3$, such that $4 + p_1p_2$ and $4 + p_1p_3$ are perfect squares. Find all possible values of p_1, p_2, p_3 .

ISI 2017, Problem 8

Let k, n and r be positive integers.

- (a) Let $Q(x) = x^k + a_1x^{k+1} + \cdots + a_nx^{k+n}$ be a polynomial with real coefficients. Show that the function $\frac{Q(x)}{x^k}$ is strictly positive for all real x satisfying

$$0 < |x| < \frac{1}{1 + \sum_{i=1}^n |a_i|}.$$

- (b) Let $P(x) = b_0 + b_1x + \cdots + b_rx^r$ be a non-zero polynomial with real coefficients. Let m be the smallest number such that $b_m \neq 0$. Prove that the graph of $y = P(x)$ cuts the x -axis at the origin (i.e., P changes sign at $x = 0$) if and only if m is an odd integer.

Part IV

Mock Test

INDIAN STATISTICAL INSTITUTE
MOCK TEST: UGA

Time: 2 hours

Maximum Marks: 120

Instructions:

- UGA is a multiple choice examination. In each of the following questions, exactly one of the choices is correct.
 - You get four marks for each correct answer, one mark for each unanswered question, and zero marks for each incorrect answer.
 - Use of calculators is not permitted

- The number of primes p such that $p^2 + 11$ is also prime is
 - 0
 - 1
 - 2
 - infinitely many
 - Let n be a positive integer. The number of solutions to the congruence $x^2 \equiv 1 \pmod{15}$ is
 - 2
 - 4
 - 6
 - 8
 - For how many positive integers $n \leq 100$ does there exist a primitive root modulo n ?
 - 25
 - 26
 - 49
 - 50
 - The smallest positive integer k such that $2^k \equiv 1 \pmod{17}$ is
 - 4
 - 8
 - 16
 - 17
 - Let p be an odd prime. By Wilson's theorem, $(p - 1)! \equiv -1 \pmod{p}$. The value of $\frac{(p-1)!}{2} \pmod{p}$ when $p = 13$ is
 - 6
 - 7
 - 12
 - does not exist
 - The number of integer solutions to the Diophantine equation $3x + 5y = 100$ with $x, y \geq 0$ is

- (A) 6
(C) 8

- (B) 7
(D) 9

7. Let $p(x) = x^3 - 6x^2 + 11x - 6$. If $p(x)$ has integer roots, then the sum of all positive divisors of the product of these roots is

- (A) 12
(C) 28

- (B) 18
(D) 32

8. The Legendre symbol $\left(\frac{5}{p}\right) = 1$ for which of the following primes p ?

- (A) $p = 3$
(C) $p = 11$

- (B) $p = 7$
(D) $p = 13$

9. The number of primitive roots modulo 13 is

- (A) 4
(C) 8

- (B) 6
(D) 12

10. Let μ be the Möbius function. The value of $\sum_{d|30} \mu(d)$ is

- (A) -1
(C) 1

- (B) 0
(D) 2

11. The largest prime factor of $2^{12} - 1$ is

- (A) 7
(C) 17

- (B) 13
(D) 31

12. For which value of n does the polynomial $f(x) = x^n + x + 1$ have a rational root?

- (A) $n = 1$
(C) $n = 3$

- (B) $n = 2$
(D) No such n exists

13. The number of solutions to $x^2 - y^2 \equiv 0 \pmod{7}$ with $0 \leq x, y < 7$ is

- (A) 7
(C) 15

- (B) 13
(D) 21

14. If p is a prime and $p \equiv 1 \pmod{4}$, then the congruence $x^2 \equiv -1 \pmod{p}$ has

- (A) no solutions
(C) exactly two solutions

- (B) exactly one solution

- (D) more than two solutions

INDIAN STATISTICAL INSTITUTE
MOCK TEST: UGB

Time: 2 hours

Maximum Marks: 100

Q 1. Determine all positive integers n such that $2^n - 1$ divides $3^n - 1$.

Q 2. Let p be an odd prime and let a be an integer not divisible by p .

- (a) Prove that if $a^{(p-1)/2} \equiv 1 \pmod{p}$, then the congruence $x^2 \equiv a \pmod{p}$ has exactly two solutions modulo p .
- (b) Show that there are exactly $(p-1)/2$ quadratic residues modulo p among the integers $1, 2, \dots, p-1$.

Q 3. Let μ be the Möbius function. The **von Mangoldt function**, denoted $\Lambda(n)$, is defined for all positive integers n as:

$$\Lambda(n) = \begin{cases} \log p & \text{if } n = p^k \text{ for some prime } p \text{ and integer } k \geq 1 \\ 0 & \text{otherwise} \end{cases}$$

Prove that for any integer $n > 1$, $\sum_{d|n} \mu(d) \log d = -\Lambda(n)$ where the sum is over all positive divisors d of n .

Q 4. Find the number of ordered pairs (a, b) of positive integers such that $\gcd(a, b) = 1$, $a \leq 100$, $b \leq 100$, and $a^2 + b^2$ is divisible by 5.

Q 5. Let $n \geq 3$ be a positive integer. Consider the polynomial

$$P_n(x) = x^n + x^{n-1} + \cdots + x + 1$$

- (a) Show that if p is a prime divisor of $P_n(a)$ for some integer $a > 1$, then either p divides $a^{n+1} - 1$ or $p = n + 1$.
- (b) Prove that $P_n(a)$ has at least one prime divisor greater than n for any integer $a \geq 2$.

Q 6. Let $S = \{1, 2, 3, \dots, 2024\}$. Find the largest possible size of a subset $T \subseteq S$ such that no two distinct elements $x, y \in T$ satisfy $x + y = 2025$.

Q 7. Let p be an odd prime.

- (a) Prove that the congruence $x^2 \equiv -1 \pmod{p}$ has a solution if and only if $p \equiv 1 \pmod{4}$.
- (b) Using part (a) or otherwise, prove that there are infinitely many primes of the form $4k + 1$.

Q 8. Let $f : \mathbb{N} \rightarrow \mathbb{N}$ be a function such that for all positive integers m, n :

- (i) $f(mn) = f(m)f(n)$ if $\gcd(m, n) = 1$
- (ii) $f(p^k) = p^{k-1}(p - 1)$ for all primes p and positive integers k

Prove that $f(n) = \varphi(n)$ for all positive integers n , where φ is Euler's totient function.

Afterword

As we conclude this journey through Number Theory, I want to share what I believe is the most important insight for ISI aspirants. The entrance examination is designed to test your understanding of mathematical theory. Hence, the examination features limited questions of the purely numerical type.

Over the years, I have observed that those who succeed in the entrance examination are not necessarily the ones who memorize the most formulas or solve problems the fastest. They are the ones who develop an intuitive understanding of mathematical theory and learn to derive proofs.

The ISI entrance examination focuses on the questions of the theoretical kind, because the undergraduate program is designed to teach theoretical mathematics. Hence, it would be wise to develop a genuine interest for the theoretical mathematics prior to joining the institute. One must be ready to think about problems that are abstract in nature, because that is what mathematics is all about.

Consider the Goldbach Conjecture: every even number greater than 2 can be written as the sum of two primes. This statement is elementary and yet for the past 300 years, this statement has neither been proved nor disproved.

Consider another story for the quest to solve Fermat's Last Theorem. In 1637, Pierre de Fermat claimed that no three positive integers can satisfy $x^n + y^n = z^n$ for any integer n greater than 2. What followed was a 358 year wait that ended only in 1995 when Andrew Wiles finally succeeded in solving the theorem.

These stories capture, though extreme, the examples of abstractness that one must be ready to embrace to pursue a degree at ISI, Kolkata. When you struggle with a problem in this book, as you prepare for the ISI entrance examination, you are experiencing the same intellectual challenge that has driven mathematical discovery for millennia.

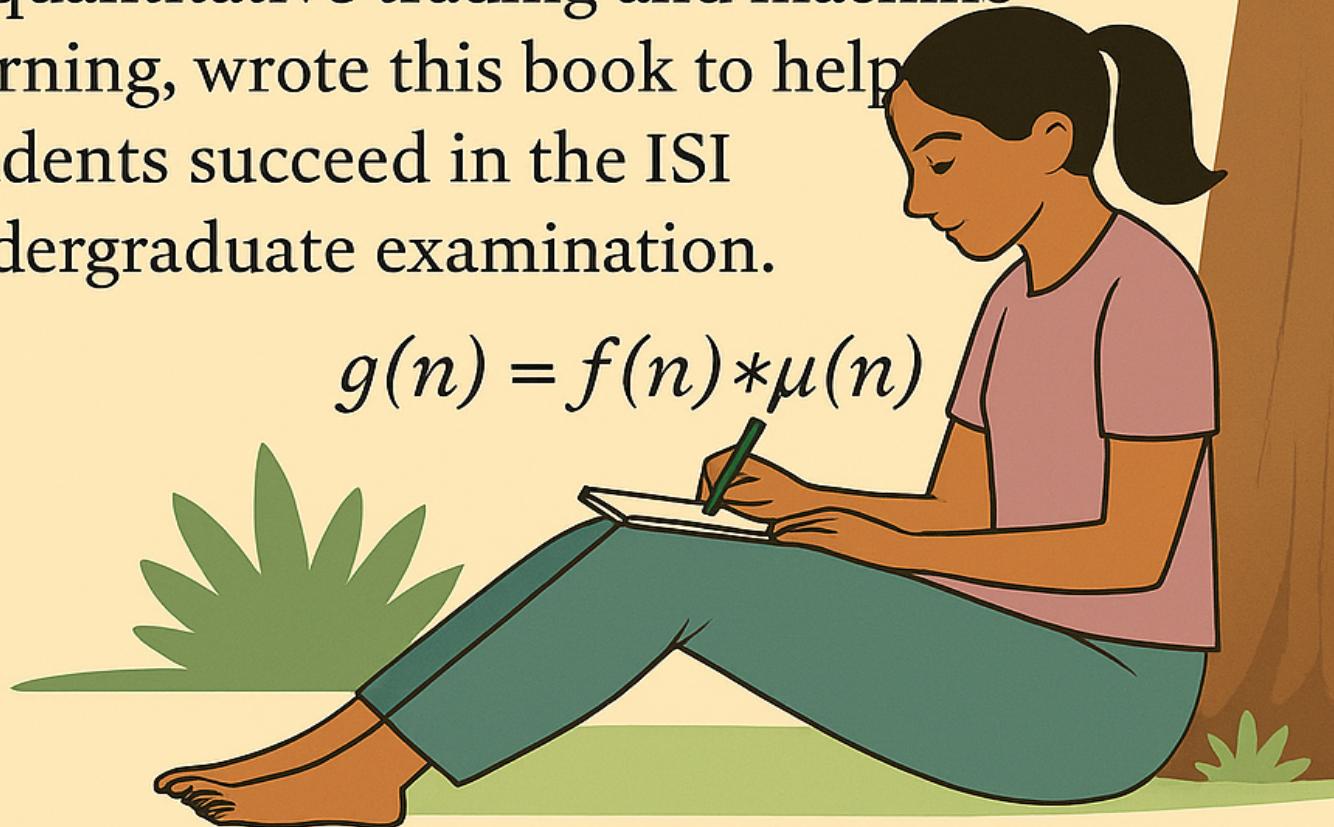
I have written this book because of the immense love for the subject and a realization about the lack of resources to prepare for the examination. If this book has helped you, I would consider my efforts in writing this one to be successful. Hopefully, apart from helping you prepare for the examination, this book also helped you grow and deepen the roots of your mathematical understanding, similar to the roots of a banyan tree!

Under the Banyan Tree: *Decoding Numbers*

transforms complex number theory into an accessible learning journey, featuring 65+ theorem proofs, 57 worked examples, 187 practice problems, and 2 full-length mock tests.

Author Sumit Gupta, an alumnus of UC Berkeley and ISI who specializes in quantitative trading and machine learning, wrote this book to help students succeed in the ISI undergraduate examination.

$$g(n) = f(n) * \mu(n)$$



Visit www.vatvriksha.com for additional resources and support.