

# Mathematics by Experiment: Plausible Reasoning in the 21st Century

Jonathan M. Borwein

Centre for Experimental and Constructive Mathematics

Department of Mathematics

Simon Fraser University

David H. Bailey

Lawrence Berkeley National Laboratory

August 4, 2003

## Preface

[I]ntuition comes to us much earlier and with much less outside influence than formal arguments which we cannot really understand unless we have reached a relatively high level of logical experience and sophistication. . . . In the first place, the beginner must be convinced that proofs deserve to be studied, that they have a purpose, that they are interesting.

George Polya, *Mathematical Discovery: On Understanding, Learning and Teaching Problem Solving*, 1968

The authors first met in 1985, when Bailey used the Borwein quartic algorithm for  $\pi$  as part of a suite of tests on the new Cray-2 then being installed at the NASA Ames Research Center in California. As our collaboration has grown over the past 18 years, we have become more and more convinced of the power of experimental techniques in mathematics. When we started our collaboration, relatively few mathematicians employed computations in serious research work. In fact, there appeared to be a widespread view in the field that “real mathematicians don’t compute.” In the ensuing years, computer hardware has skyrocketed in power and plummeted in cost, thanks to the remarkable phenomenon of Moore’s Law. In addition, numerous powerful mathematical software products, both commercial and noncommercial, have become available. But just importantly, a new generation of mathematicians is eager to use these tools, and consequently numerous new results are being discovered.

The experimental methodology described in this book, as well as in the second volume of this work, *Experimentation in Mathematics: Computational Paths to Discovery* [48], provides a compelling way to generate understanding and insight; to generate and confirm or confront conjectures; and generally to make mathematics more tangible, lively and fun for both the professional researcher and the novice. Furthermore, the experimental approach helps broaden the interdisciplinary nature of mathematical research: a chemist, physicist, engineer, and a mathematician may not understand each others’ motivation or technical

language, but they often share an underlying computational approach, usually to the benefit of all parties involved.

Our views have been expressed well by Epstein and Levy in a 1995 article on experiment and proof [88].

The English word “prove”—as its Old French and Latin ancestors—has two basic meanings: to try or test, and to establish beyond doubt. The first meaning is largely archaic, though it survives in technical expressions (printer’s proofs) and adages (the exception proves the rule, the proof of the pudding). That these two meanings could have coexisted for so long may seem strange to us mathematicians today, accustomed as we are to thinking of “proof” as an unambiguous term. But it is in fact quite natural, because the most common way to establish something in everyday life is to examine it, test it, probe it, experiment with it.

As it turns out, much the same is true in mathematics as well. Most mathematicians spend a lot of time thinking about and analyzing particular examples. This motivates future development of theory and gives one a deeper understanding of existing theory. Gauss declared, and his notebooks attest to it, that his way of arriving at mathematical truths was “through systematic experimentation.” It is probably the case that most significant advances in mathematics have arisen from experimentation with examples. For instance, the theory of dynamical systems arose from observations made on the stars and planets and, more generally, from the study of physically motivated differential equations. A nice modern example is the discovery of the tree structure of certain Julia sets by Douady and Hubbard: this was first observed by looking at pictures produced by computers and was then proved by formal arguments.

Our goal in these books is to present a variety of *accessible* examples of modern mathematics where intelligent computing plays a significant role (along with a few examples showing the limitations of computing). We have concentrated primarily on examples from analysis and number theory, as this is where we have the most experience, but there are numerous excursions into other areas of mathematics as well (see the Table of Contents). For the most part, we have

contented ourselves with outlining reasons and exploring phenomena, leaving a more detailed investigation to the reader. There is, however, a substantial amount of new material, including numerous specific results that have not yet appeared in the mathematical literature, as far as we are aware.

This work is divided into two volumes, each of which can stand by itself. This volume, *Mathematics by Experiment: Plausible Reasoning in the 21st Century*, presents the rationale and historical context of experimental mathematics, and then presents a series of examples that exemplify the experimental methodology. We include in this volume a reprint of an article co-authored by one of us that complements this material. The second book, *Experimentation in Mathematics: Computational Paths to Discovery*, continues with several chapters of additional examples. Both volumes include a chapter on numerical techniques relevant to experimental mathematics.

Each volume is targeted to a fairly broad cross-section of mathematically trained readers. Most of this volume should be readable by anyone with solid undergraduate coursework in mathematics. Most of the second volume should be readable by persons with upper-division undergraduate or graduate-level coursework. None of this material involves highly abstract or esoteric mathematics.

The subtitle of this volume is taken from George Polya's well-known work, *Mathematics and Plausible Reasoning* [166]. This two-volume work has been enormously influential—if not uncontroversial—not only in the field of artificial intelligence, but also in the mathematical education and pedagogy community.

Some programming experience is valuable to address the material in this book. Readers with no computer programming experience are invited to try a few of our examples using commercial software such as *Mathematica* and *Maple*. Happily, much of the benefit of computational-experimental mathematics can be obtained on any modern laptop or desktop computer—a major investment in computing equipment and software is not required.

Each chapter concludes with a section of commentary and exercises. This permits us to include material that relates to the general topic of the chapter, but which does not fit nicely within the chapter exposition. This material is not necessarily sorted by topic nor graded by difficulty, although some hints, discussion and answers are given. This is because mathematics in the raw does not announce, “I am solved using such and such a technique.” In most cases, half the battle is to determine how to start and which tools to apply.

We should mention two recent books on mathematical experimentation: [106]

and [142]. In both cases, however, the focus and scope centers on the teaching of students and thus is quite different from ours.

We are grateful to our colleagues Victor Adamchik, Heinz Bauschke, Peter Borwein, David Bradley, Gregory Chaitin, David and Gregory Chudnovsky, Robert Corless, Richard Crandall, Richard Fateman, Greg Fee, Helaman Ferguson, Steven Finch, Ronald Graham, Andrew Granville, Christoph Haenel, David Jeffrey, Jeff Joyce, Adrian Lewis, Petr Lisonek, Russell Luke, Mathew Morin, David Mumford, Andrew Odlyzko, Hristo Sendov, Luis Serrano, Neil Sloane, Daniel Rudolph, Asia Weiss, and John Zucker who were kind enough to help us prepare and review material for this book; to Mason Macklem, who helped with material, indexing (note that in the index definitions are marked in bold, and quotes with a suffix “Q”), and more; to Jen Chang and Rob Scharein, who helped with graphics; to Janet Vertesi who helped with bibliographic research; to Will Galway, Xiaoye Li, and Yozo Hida, who helped with computer programming; and to numerous others who have assisted in one way or another in this work. We thank Roland Girgensohn in particular for contributing a significant amount of material and reviewing several drafts. We owe a special debt of gratitude to Klaus Peters for urging us to write this book and for helping us nurse it into existence. Finally, we wish to acknowledge the assistance and the patience exhibited by our spouses and family members during the course of this work.

Borwein’s work is supported by the Canada Research Chair Program and the Natural Sciences and Engineering Council of Canada. Bailey’s work is supported by the Director, Office of Computational and Technology Research, Division of Mathematical, Information, and Computational Sciences of the U.S. Department of Energy, under contract number DE-AC03-76SF00098.

Jonathan M. Borwein  
Simon Fraser University  
Burnaby, British Columbia V5A 1S6, Canada  
E-mail: [jborwein@cecm.sfu.ca](mailto:jborwein@cecm.sfu.ca)

David H. Bailey  
Lawrence Berkeley National Laboratory  
Berkeley, CA 94720, USA  
Email: [dhbailey@lbl.gov](mailto:dhbailey@lbl.gov)

August 2003



## Photo and Illustration Credits

We are grateful to the following for permission to reproduce material:

Béla Bollobás (Littlewood's Miscellany)  
David and Gregory Chudnovsky (Random Walk on Pi)  
George Paul Csicsery (Paul Erdős photo)  
Helaman Ferguson (Sculpture photos)  
Göttingen University Library (Riemann manuscript)  
Mathematical Association of America (Polya's coin graphic)  
Andrew Odlyzko (Data and graphs of Riemann zeta function)  
Brailey Sims (Pooh Math)  
The Smithsonian Institution (ENIAC computer photo)  
Nick Trefethen (Daisy pseudospectrum graphic)  
Asia Weiss (Coxeter's memorabilia)

Several of the nicer graphics that appear in black-and-white in various chapters of this volume, as well as some from the second volume, have been rendered in the Color Supplement (Chapter 8).

## Experimental Mathematics Web Site

The authors have established a web site containing an updated collection of links to many of the URLs mentioned in the two volumes, plus errata, software, tools, and other web useful information on experimental mathematics. This can be found at the following URL:

<http://www.expmath.info>

# List of Figures

1.1	Gauss on the lemniscate. . . . .	16
1.2	Gauss on the arithmetic-geometric mean. . . . .	17
1.3	Kelly’s 1948 “Proof from the Book.” . . . .	21
1.4	Paul Bert’s 1886 experimental geometry text, pages 66–67. . . . .	23
1.5	Paul Bert’s text, pages 68–69. . . . .	24
1.6	Polya’s illustration of the change solution. . . . .	25
1.7	The Sierpinski cube. . . . .	36
1.8	The torus and the two-sphere. . . . .	45
2.1	How not to perform experiments: guess and check. . . . .	54
2.2	Pascal triangles modulo two? . . . . .	56
2.3	Sierpinski triangle by random selections. . . . .	58
2.4	Bifurcation in the logistic iteration. . . . .	61
2.5	Ferguson’s “Eight-Fold Way” sculpture. . . . .	64
2.6	Ferguson’s “Figure-Eight Knot Complement” sculpture. . . . .	65
2.7	The ten tetrahedral configurations. . . . .	71
2.8	Page one of Riemann’s 1859 manuscript. . . . .	79
2.9	Page two of Riemann’s 1859 manuscript. . . . .	80
2.10	$\zeta$ function zeroes (dotted) vs pseudo-random data (dashed) . . . . .	82
2.11	Nearest neighbor spacing versus Montgomery’s distribution. . . . .	84
2.12	Root two is irrational. . . . .	86
2.13	Duijvestijn’s dissection. . . . .	95
2.14	Takeya’s conjectured answer—the three-cusp hypocycloid. . . . .	96
2.15	Donald Coxeter’s own kaleidoscope . . . . .	97
2.16	A projection of a four dimensional polytope. . . . .	99
2.17	Weeks’ two-bridge discovery. . . . .	107



2.18	7th iterate of Sierpinski triangle made by paper cutting and folding.	108
2.19	The modulus of the $\zeta$ function. . . . .	110
3.1	Newton's scheme to compute $\pi$ . . . . .	120
3.2	Ramanujan's 75th birthday stamp. . . . .	125
3.3	Fax from <i>The Simpsons</i> show. . . . .	134
3.4	The ENIAC computer. . . . .	153
4.1	A random walk based on one million digits of $\pi$ . . . . .	168
4.2	A 1993 photo of Paul Erdős. . . . .	196
5.1	Experimental $Q$ -minimizing function and matching Gaussian. . .	212
5.2	Difference between $Q$ -minimizing function and matching Gaussian.	212
6.1	Computational cost for the devil's convolution. . . . .	253
6.2	The first 16000 poles of $B(v)$ , approaching the Julia set of $v \rightarrow v+v^2$ .	267
6.3	Three images quantized at quality 50 (L), 48 (C) and 75 (R). . . .	269
7.1	Graphical statistics of our experiments . . . . .	297
8.1	Pascal triangles modulo two? . . . . .	305
8.2	A pictorial proof of Archimedes' inequality . . . . .	305
8.3	Root two is irrational. . . . .	305
8.4	Ferguson's "Eight-Fold Way" and "Figure-Eight Knot Comple- ment." . . . . .	306
8.5	Ferguson's snow sculpture of the Costa surface . . . . .	307
8.6	The "Firmament" by John Robinson. . . . .	307
8.7	Ramanujan's seventy-fifth birthday stamp. . . . .	307
8.8	Coxeter's own kaleidoscope model. . . . .	308
8.9	The modulus of the modular function $k$ . . . . .	308
8.10	A random walk based on one million digits of $\pi$ . . . . .	309
8.11	The ENIAC computer. . . . .	309
8.12	The knots $10_{161}$ and $10_{162}$ . . . . .	310
8.13	How does one identify two knots? . . . . .	310
8.14	Projection of 4-D polyhedron. . . . .	310
8.15	The Riemann surface for the Lambert function. . . . .	311
8.16	Julia set associated with Newton solutions of $x^3 - 1 = 0$ . . . . .	311
8.17	Roots of zero-one polynomials. . . . .	312

# List of Tables

2.1	Formulas found by PSLQ for the ten tetrahedral diagrams. . . . .	71
2.2	The prime-counting function $\pi(x)$ and Gauss' approximation. . . . .	77
3.1	Pre-calculus era $\pi$ calculations. . . . .	123
3.2	Calculus era $\pi$ calculations. . . . .	124
3.3	Digital era $\pi$ calculations. . . . .	130
3.4	Computed hexadecimal digits of $\pi$ . . . . .	143
3.5	Binary BBP-type formulas. . . . .	146
3.6	Nonbinary BBP-type formulas. . . . .	147
4.1	Decimal values of various mathematical constants. . . . .	165
4.2	Binary values of various mathematical constants. . . . .	166
4.3	Statistics for the first trillion decimal digits of $\pi$ . . . . .	166
4.4	Statistics for the first trillion hexadecimal digits of $\pi$ . . . . .	167
5.1	$Q$ values for various functions. . . . .	209
7.1	Serendipity and experimentation . . . . .	279
7.2	Some experimental results . . . . .	281
7.3	Shrinking or encapsulating mathematics . . . . .	282
7.4	An uninteresting(?) proof . . . . .	284
7.5	Background on normality . . . . .	293
7.6	Data and statistics . . . . .	295



# Contents

<b>Preface</b>	<b>i</b>
<b>List of Figures</b>	<b>vii</b>
<b>List of Tables</b>	<b>ix</b>
<b>1 What is Experimental Mathematics?</b>	<b>1</b>
1.1 Background . . . . .	1
1.2 Complexity Considerations . . . . .	4
1.3 Proof versus Truth . . . . .	7
1.4 Paradigm Shifts . . . . .	11
1.5 Gauss, the Experimental Mathematician . . . . .	14
1.6 Geometric Experiments . . . . .	15
1.6.1 On Picture-Writing . . . . .	22
1.7 Sample Problems of Experimental Math . . . . .	26
1.8 Internet-Based Mathematical Resources . . . . .	31
1.9 Commentary and Additional Examples . . . . .	37
<b>2 Experimental Mathematics in Action</b>	<b>53</b>
2.1 Pascal's Triangle . . . . .	53
2.2 A Curious Anomaly in the Gregory Series . . . . .	57
2.3 Bifurcation Points in the Logistic Iteration . . . . .	60
2.4 Experimental Mathematics and Sculpture . . . . .	63
2.5 Recognition of Euler Sums . . . . .	67
2.6 Quantum Field Theory . . . . .	69
2.7 Definite Integrals and Infinite Series . . . . .	70
2.8 Prime Numbers and the Zeta Function . . . . .	75

2.9	Two Observations about $\sqrt{2}$ . . . . .	85
2.9.1	Irrationality . . . . .	85
2.9.2	Rationality . . . . .	87
2.10	Commentary and Additional Examples . . . . .	88
<b>3</b>	<b>Pi and Its Friends</b>	<b>117</b>
3.1	A Short History of Pi . . . . .	117
3.2	Fascination with Pi . . . . .	131
3.3	Behind the Cubic and Quartic Iterations . . . . .	133
3.4	Computing Individual Digits of Pi . . . . .	136
3.5	Unpacking the BBP Formula for Pi . . . . .	143
3.6	Other BBP-Type Formulas . . . . .	145
3.7	Does Pi Have a Nonbinary BBP Formula? . . . . .	150
3.8	Commentary and Additional Examples . . . . .	152
<b>4</b>	<b>Normality of Numbers</b>	<b>163</b>
4.1	Normality: A Stubborn Question . . . . .	163
4.2	BBP Constants and Normality . . . . .	169
4.3	A Class of Provably Normal Constants . . . . .	174
4.4	Algebraic Irrationals . . . . .	178
4.5	Periodic Attractors and Normality . . . . .	181
4.6	Commentary and Additional Examples . . . . .	186
<b>5</b>	<b>The Power of Constructive Proofs I</b>	<b>197</b>
5.1	The Fundamental Theorem of Algebra . . . . .	197
5.1.1	Liouville's Theorem and the Sine Product . . . . .	200
5.1.2	Roots and Branches . . . . .	202
5.2	The Uncertainty Principle . . . . .	206
5.3	A Concrete Approach to Inequalities . . . . .	213
5.4	The Gamma Function . . . . .	216
5.4.1	Volumes of Balls . . . . .	219
5.5	Stirling's Formula . . . . .	221
5.6	Derivative Methods of Evaluation . . . . .	223
5.6.1	Legendre's Relation and Applications . . . . .	224
5.6.2	The AGM Iteration . . . . .	225
5.6.3	Other Mean Iterations . . . . .	227
5.6.4	Gregory and Euler . . . . .	228

5.7	Commentary and Additional Examples . . . . .	230
<b>6</b>	<b>Numerical Techniques I</b>	<b>241</b>
6.1	Convolutions and Fourier Transforms . . . . .	242
6.2	High-Precision Arithmetic . . . . .	245
6.2.1	Double-Double and Quad-Double Arithmetic . . . . .	245
6.2.2	Arbitrary Precision Arithmetic . . . . .	247
6.2.3	Karatsuba Multiplication . . . . .	249
6.2.4	FFT-Based Multiplication . . . . .	250
6.2.5	Algebraic and Transcendental Functions . . . . .	254
6.3	Constant Recognition . . . . .	257
6.3.1	PSLQ . . . . .	258
6.3.2	Multipair PSLQ . . . . .	260
6.3.3	Practical Constant Recognition . . . . .	262
6.4	Commentary and Additional Examples . . . . .	263
<b>7</b>	<b>Making Sense of Experimental Math</b>	<b>273</b>
7.1	Introduction . . . . .	273
7.1.1	Our Goals . . . . .	274
7.1.2	Unifying Themes . . . . .	275
7.2	Experimental Mathematics . . . . .	276
7.2.1	Journal of . . . . .	276
7.2.2	The Deductivist Style . . . . .	280
7.3	Zeilberger and the Encapsulation of Identity . . . . .	283
7.3.1	Putting a Price on Reliability . . . . .	283
7.3.2	It's All About Insight . . . . .	285
7.4	Experiment and "Theory" . . . . .	287
7.5	"Theoretical" Experimentation . . . . .	289
7.6	A Mathematical Experiment . . . . .	292
7.6.1	Experimentation . . . . .	292
7.6.2	Verification . . . . .	292
7.6.3	Interpretation . . . . .	294
7.7	Conclusion . . . . .	296
<b>8</b>	<b>A Color Supplement</b>	<b>301</b>
	<b>Bibliography</b>	<b>313</b>

xiv

*CONTENTS*

**Index**

**329**

# Chapter 1

## What is Experimental Mathematics?

The computer has in turn changed the very nature of mathematical experience, suggesting for the first time that mathematics, like physics, may yet become an empirical discipline, a place where things are discovered because they are seen.

David Berlinski, 1997 [35]

If mathematics describes an objective world just like physics, there is no reason why inductive methods should not be applied in mathematics just the same as in physics.

Kurt Gödel, 1951 [98]

### 1.1 Background

One of the greatest ironies of the information technology revolution is that while the computer was conceived and born in the field of pure mathematics, through the genius of giants such as John von Neumann and Alan Turing, until recently this marvelous technology had only a minor impact within the field that gave it birth.

This has not been the case in applied mathematics, as well as in most other scientific and engineering disciplines, which have aggressively integrated com-



puter technology into their methodology. For instance, physicists routinely utilize numerical simulations to study exotic phenomena ranging from supernova explosions to big bang cosmology—phenomena that in many cases are beyond the reach of conventional laboratory experimentation. Chemists, molecular biologists, and material scientists make use of sophisticated quantum-mechanical computations to unveil the world of atomic-scale phenomena. Aeronautical engineers employ large-scale fluid dynamics calculations to design wings and engines for jet aircraft. Geologists and environmental scientists utilize sophisticated signal processing computations to probe the earth’s natural resources. Biologists harness large computer systems to manage and analyze the exploding volume of genome data. And social scientists—economists, psychologists, and sociologists—make regular use of computers to spot trends and inferences in empirical data.

In the late 1980s, recognizing that its members were lagging behind in embracing computer technology, the American Mathematical Society began a regular “Computers and Mathematics” section in the monthly newsletter, *Notices of the American Mathematical Society*, edited at first by Jon Barwise and subsequently by Keith Devlin. This continued until the mid-1990s and helped to convince the mathematical community that the computer can be a useful research tool. In 1992, a new journal, *Experimental Mathematics*, was launched, founded on the belief “that theory and experiment feed on each other, and that the mathematical community stands to benefit from a more complete exposure to the experimental process.” It encouraged the submission of algorithms, results of experiments, and descriptions of computer programs, in addition to formal proofs of new results [87].

Perhaps the most important advancement along this line is the development of broad spectrum mathematical software products such as *Mathematica* and *Maple*. These days, many mathematicians are highly skilled with these tools and use them as part of their day-to-day research work. As a result, we are starting to see a wave of new mathematical results discovered partly or entirely with the aid of computer-based tools. Further developments in hardware (the gift of Moore’s Law of semiconductor technology), software tools, and the increasing availability of valuable Internet-based facilities, are all ensuring that mathematicians will have their day in the computational sun.

This new approach to mathematics—the utilization of advanced computing technology in mathematical research—is often called *experimental mathemat-*

*ics.* The computer provides the mathematician with a “laboratory” in which he or she can perform experiments: analyzing examples, testing out new ideas, or searching for patterns. Our book is about this new, and in some cases not so new, way of doing mathematics. To be precise, by experimental mathematics, we mean the methodology of doing mathematics that includes the use of computations for:

1. Gaining insight and intuition.
2. Discovering new patterns and relationships.
3. Using graphical displays to suggest underlying mathematical principles.
4. Testing and especially falsifying conjectures.
5. Exploring a possible result to see if it is worth formal proof.
6. Suggesting approaches for formal proof.
7. Replacing lengthy hand derivations with computer-based derivations.
8. Confirming analytically derived results.

Note that the above activities are, for the most part, quite similar to the role of laboratory experimentation in the physical and biological sciences. In particular, they are very much in the spirit of what is often termed “computational experimentation” in physical science and engineering, which is why we feel the qualifier “experimental” is particularly appropriate in the term experimental mathematics.

We should note that one of the more valuable benefits of the computer-based experimental approach in modern mathematics is its value in rejecting false conjectures (Item 4): A single computational example can save countless hours of human effort that would otherwise be spent attempting to prove false notions.

With regards to Item 5, we observe that mathematicians generally do not know during the course of research how it will pan out, but nonetheless must, in a conventional mathematical approach, prove all the pieces along the way as assurance that the project makes sense and remains on course. The methods of experimental mathematics allow mathematicians to maintain a reasonable level of assurance without nailing down all the lemmas the first time through. At the

end of the day, they can decide if the result merits proof. If it is not the answer that was sought, or if it is simply not interesting enough, much less time will have been spent coming to this conclusion.

Many mathematicians remain uncomfortable with the appearance in published articles of expressions such as “proof by *Mathematica*” or “established by *Maple*” (see Item 7 above). There is, however, a clear trend in this direction, and it seems to us to be both futile and counterproductive to resist it. In Chapter 7 we will further explore the nature of mathematical experimentation and proof.

## 1.2 Complexity Considerations

Gordon Moore, the co-founder of Intel Corporation, noted in a 1965 article

The complexity for minimum component costs has increased at a rate of roughly a factor of two per year. . . . Certainly over the short term this rate can be expected to continue, if not to increase. Over the longer term, the rate of increase is a bit more uncertain, although there is no reason to believe it will not remain nearly constant for at least 10 years. [155]

With these sentences, Moore stated what is now known as Moore’s Law, namely the observation that semiconductor technology approximately doubles in capacity and overall performance roughly every 18 to 24 months (not quite every year as Moore predicted above). This trend has continued unabated for nearly 40 years, and, according to Moore and other industry analysts, there is still no end in sight—at least another ten years is assured [2]. This astounding record of sustained exponential progress has no peer in the history of technology. What’s more, we will soon see mathematical computing tools implemented on parallel computer platforms, which will provide even greater power to the research mathematician.

However, we do not suggest that amassing huge amounts of processing power can solve all mathematical problems, even those that are amenable to computational analysis. There are doubtless some cases where a dramatic increase in computation could, by itself, result in significant breakthroughs, but it is easier to find examples where this is unlikely to happen.

For example, consider Clement Lam’s 1991 proof of the nonexistence of a finite projective plane of order ten [140]. This involved a search for a configuration

of  $n^2 + n + 1$  points and equally many lines. Lam's computer program required thousands of hours of run time on a Cray computer system. Lam estimates that the next case ( $n = 18$ ) susceptible to his methods would take millions of years on any conceivable architecture.

Along this line, although a certain class of computer-based mathematical analysis is amenable to "embarrassingly parallel" (the preferred term is now "naturally parallel") processing, these tend not to be problems of central interest in mathematics. A good example of this is the search for Mersenne primes, namely primes of the form  $2^n - 1$  for integer  $n$ . While such computations are interesting demonstrations of mathematical computation, they are not likely to result in fundamental breakthroughs. By contrast let us turn to perhaps the most fundamental of current algorithmic questions.

**The P versus NP problem.** (This discussion is taken from [45].) Of the seven million-dollar Millennium Prize problems, the one that is most germane to our present voyage is the so-called "P versus NP problem," also known as the " $P \neq NP$ " problem. We quote from the discussion on the Clay web site:

It is Saturday evening and you arrive at a big party. Feeling shy, you wonder whether you already know anyone in the room. Your host proposes that you must certainly know Rose, the lady in the corner next to the dessert tray. In a fraction of a second you are able to cast a glance and verify that your host is correct. However, in the absence of such a suggestion, you are obliged to make a tour of the whole room, checking out each person one by one, to see if there is anyone you recognize. This is an example of the general phenomenon that generating a solution to a problem often takes far longer than verifying that a given solution is correct. Similarly, if someone tells you that the number 13,717,421 can be written as the product of two smaller numbers, you might not know whether to believe him, but if he tells you that it can be factored as 3607 times 3803, then you can easily check that it is true using a hand calculator. One of the outstanding problems in logic and computer science is determining whether questions exist whose answer can be quickly checked (for example by computer), but which require a much longer time to solve from scratch (without knowing the answer). There certainly seem to be many such questions. But so far no one has proved that

any of them really does require a long time to solve; it may be that we simply have not yet discovered how to solve them quickly. Stephen Cook formulated the P versus NP problem in 1971.

Although in many instances one may question the practical distinction between polynomial and nonpolynomial algorithms, this problem really is central to our current understanding of computing. Roughly it conjectures that many of the problems we currently find computationally difficult must *per force* be that way. It is a question about methods, not about actual computations, but it underlies many of the challenging problems one can imagine posing. A question that requests one to “compute such and such a sized incidence of this or that phenomena” always risks having the answer, “It’s just not possible,” because  $P \neq NP$ .

With the “*NP*” caveat (though factoring is difficult it is not generally assumed to be in the class of *NP*-hard problems), let us offer two challenges that are far fetched, but not inconceivable, goals for the next few decades.

**First Challenge.** Design an algorithm that can reliably factor a random thousand digit integer.

Even with a huge effort, current algorithms get stuck at about 150 digits. Details can be found at <http://www.rsasecurity.com/rsalabs/challenges/factoring>, where the current factoring challenges are listed. One possible solution to the factorization problem may come through quantum computing, using an algorithm found by Peter Shor in 1994 [178]. However, it is still not clear whether quantum phenomenon can be harnessed on the scale required for this algorithm to be practical.

With regards to cash prizes, there is also \$100,000 offered for any honest 10,000,000 digit prime: <http://www.mersenne.org/prime.htm>.

Primality checking is currently easier than factoring, and there are some very fast and powerful *probabilistic* primality tests—much faster than those providing “certificates” of primality. There is also the recently discovered “AKS” deterministic polynomial time algorithm for primality, whose implementations, as we note in Section 7.2 of the second volume, keep improving.

Given that any computation has potential errors due to: (i) subtle (or even not-so-subtle) programming bugs, (ii) compiler errors, (iii) system software errors, and (iv) undetected hardware integrity errors, it seems increasingly pointless to distinguish between these two types of primality tests. Many would take

their chances with a  $(1 - 10^{-100})$  probability statistic over a “proof” any day (more on this topic is presented in Section 7.2 of the second volume).

The above questions are intimately related to the Riemann Hypothesis and its extensions, though not obviously so. They are also critical to issues of Internet security. If someone learns how to rapidly factor large numbers, then many current security systems are no longer secure.

Many old problems lend themselves to extensive exploration. One example that arose in signal processing is called the *Merit Factor problem*, and is due in large part to Marcel Golay with closely related versions due to Littlewood and to Erdős. It has a long pedigree though certainly not as elevated as the Riemann Hypothesis.

The problem can be formulated as follows. Suppose  $\mathcal{A}_n$  consists of all sequences  $(a_0 = 1, a_1, \dots, a_n)$  of length  $n + 1$  where each  $a_i$  is restricted to 1 or  $-1$ , for  $i > 0$ . If  $c_k = \sum_{j=0}^{n-k} a_j a_{j+k}$ , then the problem is to minimize  $\sum_{k=-n}^n c_k^2$  over  $\mathcal{A}_n$  for each fixed  $n$ . This is discussed at length in [56].

Minima have been found up to about  $n = 50$ . The search space of sequences of size 50 is  $2^{50} \approx 10^{15}$ , which approaches the limit of the very large-scale calculations feasible today. The records use a branch and bound algorithm which grows more or less like  $1.8^n$ . This is marginally better than the naive  $2^n$  growth of a completely exhaustive search but is still painfully exponential.

**Second Challenge.** Find the minima in the merit factor problem for sizes  $n \leq 100$ .

The best hope for a solution lies in development of better algorithms. The problem is widely acknowledged as a very hard problem in combinatorial optimization, but it isn’t known to be in one of the recognized hard classes like *NP*. The next best hope is a radically improved computer technology, perhaps quantum computing. And there is always a remote chance that analysis will lead to a mathematical solution.

## 1.3 Proof versus Truth

In any discussion of an experimental approach to mathematical research, the questions of reliability and standards of proof justifiably come to center stage. We certainly do not claim that computations utilized in an experimental approach to mathematics by themselves constitute rigorous proof of the claimed

results. Rather, we see the computer primarily as an exploratory tool to discover mathematical truths, and to suggest avenues for formal proof.

For starters, it must be acknowledged that no amount of straightforward case checking constitutes a proof. For example, the “proof” of the Four Color Theorem in the 1970s, namely that every planar map can be colored with four colors so adjoining countries are never the same color, was considered a proof because prior mathematical analysis had reduced the problem to showing that a large but finite number of bad configurations could be ruled out. The “proof” was viewed as somewhat flawed because the case analysis was inelegant, complicated and originally incomplete (this computation was recently redone after a more satisfactory analysis). Though many mathematicians still yearn for a simple proof, there is no particular reason to believe that all elegant true conjectures have elegant proofs. What’s more, given Gödel’s result, some may have no proofs at all.

Nonetheless, we feel that in many cases computations constitute very strong evidence, evidence that is at least as compelling as some of the more complex formal proofs in the literature. Prominent examples include: (1) the determination that the Fermat number  $F_{24} = 2^{2^{24}} + 1$  is composite, by Crandall, Mayer, and Papadopoulos [80]; (2) the recent computation of  $\pi$  to more than one trillion decimal digits by Yasumasa Kanada and his team; and (3) the Internet-based computation of binary digits of  $\pi$  beginning at position one quadrillion organized by Colin Percival. These are among the largest computations ever done, mathematical or otherwise (the  $\pi$  computations will be described in greater detail in Chapter 3). Given the numerous possible sources of error, including programming bugs, hardware bugs, software bugs, and even momentary cosmic-ray induced glitches (all of which are magnified by the sheer scale of these computations), one can very reasonably question the validity of these results.

But for exactly such reasons, computations such as these typically employ very strong validity checks. For example, the Crandall-Mayer-Papadopoulos computation employed a “wavefront” scheme. Here a faster computer system computed a chain of squares modulo  $F_{24}$ , such as  $3^{2^{1000000}} \bmod F_{24}$ ,  $3^{2^{2000000}} \bmod F_{24}$ ,  $3^{2^{3000000}} \bmod F_{24}$ ,  $\dots$ . Then each of a set of slower computers started with one of these intermediate values, squared it 1,000,000 times modulo  $F_{24}$ , and checked to see if the result (a 16-million-bit integer) precisely reproduced the next value in the chain. If it did, then this is very strong evidence that both computations were correct. If not, then the process was repeated [80, page 187].

In the case of computations of digits of  $\pi$ , it has been customary for many years to verify a result either by repeating the computation using a different algorithm, or by repeating with a slightly different index position. For example, if one computes hexadecimal digits of  $\pi$  beginning at position one trillion (we shall see how this can be done in Chapter 3), then this can be checked by repeating the computation at hexadecimal position one trillion minus one. It is easy to verify (see Algorithm 3 in Section 3.4) that these two calculations take almost completely different trajectories, and thus can be considered “independent.” If both computations generate 25 hexadecimal digits beginning at the respective positions, then 24 digits should perfectly overlap. If these 24 hexadecimal digits do agree, then we can argue that the probability that these digits are in error, in a very strong (albeit heuristic) sense, is roughly one part in  $16^{24} \approx 7.9 \times 10^{28}$ , a figure much larger even than Avogadro’s number ( $6.022 \times 10^{22}$ ). Percival’s actual computation of the quadrillionth binary digit (i.e., the 250 trillionth hexadecimal digit) of  $\pi$  was verified by a similar scheme, which for brevity we have simplified here.

Kanada and his team at the University of Tokyo, who just completed a computation of the first 1.24 trillion decimal digits of  $\pi$ , employed an even more impressive validity check (Kanada’s calculation will be discussed in greater detail in Section 3.1). They first computed more than one trillion hexadecimal digits, using two different formulas. The hexadecimal digit string produced by both of these formulas, beginning at hex digit position 1,000,000,000,001, was **B4466E8D21 5388C4E014**. Next, they employed the algorithm, mentioned in the previous paragraph and described in more detail in Chapter 3, which permits one to directly compute hexadecimal digits beginning at a given position (in this case 1,000,000,000,001). This result was **B4466E8D21 5388C4E014**. Needless to say, these two sets of results, obtained by utterly different computational approaches, are in complete agreement. After this step, they converted the hexadecimal expansion to decimal, then back to hexadecimal as a check. When this final check succeeded, they felt safe to announce their results.

As a rather different example, a computation jointly performed by one of the present authors and David Broadhurst, a British physicist, discovered a previously unknown integer relation involving a set of 125 real constants associated with the largest real root of Lehmer’s polynomial [22]. This computation was performed using 50,000 decimal digit arithmetic and required 44 hours on 32 processors of a Cray T3E parallel supercomputer. The 125 integer coefficients



discovered by the program ranged in size up to  $10^{292}$ . The certification of this relation to 50,000 digit precision was thus at least 13,500 decimal digits beyond the level ( $292 \times 125 = 36,500$ ) that could reasonably be ascribed to numerical roundoff error. This result was separately affirmed by a computation on a different computer system, using 59,000-digit arithmetic, or roughly 22,500 decimal digits beyond the level of plausible roundoff error.

Independent checks and extremely high numerical confidence levels still do not constitute formal proofs of correctness. What's more, we shall see in Section 1.4 of the second volume some examples of "high-precision frauds," namely "identities" that hold to high precision, yet are not precisely true. Even so, one can argue that many computational results are as reliable, if not more so, than a highly complicated piece of human mathematics. For example, perhaps only 50 or 100 people alive can, given enough time, digest *all* of Andrew Wiles' extraordinarily sophisticated proof of Fermat's Last Theorem. If there is even a one percent chance that each has overlooked the same subtle error (and they may be psychologically predisposed to do so, given the numerous earlier results that Wiles' result relies on), then we must conclude that computational results are in many cases actually *more* secure than the proof of Fermat's Last Theorem.

Richard Dedekind's marvelous book, *Two Essays on Number Theory* [85], originally published in 1887, provides a striking example of how the nature of what is a satisfactory proof changes over time. In this work, Dedekind introduces Dedekind cuts and a modern presentation of the construction of the reals (see Item 2 at the end of this chapter). In the second essay, "The Nature and Meaning of Numbers," an equally striking discussion of finite and infinite sets takes place. Therein, one is presented with Theorem 66:

**Theorem.** There exist infinite systems.

**Proof.** My own realm of thoughts, i.e., the totality  $S$  of all things, which can be objects of my thought, is infinite. For if  $s$  signifies an element of  $S$ , then is the thought  $s'$ , that  $s$  can be object of my thought, itself an element of  $S$ . If we regard this as transform  $\phi(s)$  of the element  $s$  then has the transformation  $\phi$  of  $S$ , thus determined, the property that the transform  $S'$  is part of  $S$ ; and  $S'$  is certainly proper part of  $S$ , because there are elements in  $S$  (e.g., my own ego) which are different from such thought  $s'$  and therefore are not contained in  $S'$ . Finally, it is clear that if  $a, b$  are different elements of

$S$ , their transformation  $\phi$  is a distinct (similar) transformation (26). Hence  $S$  is infinite, which was to be proved.

A similar presentation is found in §13 of the *Paradoxes des Unendlichen*, by Bolzano (Leipzig, 1851). Needless to say, such a “proof” would not be acceptable today. In our modern formulation of mathematics there is an *Axiom of Infinity*, but recall that this essay predates the publications of Frege and Russell and the various paradoxes of modern set theory.

Some other examples of this sort are given by Judith Grabiner, who for instance compares Abel’s comments on the lack of rigor in 18th-century arguments with the standards of Cauchy’s 19th-century *Cours d’analyse* [100].

## 1.4 Paradigm Shifts

We acknowledge that the experimental approach to mathematics that we propose will be difficult for some in the field to swallow. Many may still insist that mathematics is all about formal proof, and from their viewpoint, computations have no place in mathematics. But in our view, mathematics is not ultimately about formal proof; it is instead about secure mathematical knowledge. We are hardly alone in this regard—many prominent mathematicians throughout history have either exemplified or explicitly espoused such a view.

Gauss expressed an experimental philosophy, and utilized an experimental approach on numerous occasions. In the next section, we shall present one significant example. Examples from de Morgan, Klein, and others will be given in subsequent sections.

Georg Friedrich Bernhard Riemann (1826–1866) was one of the most influential scientific thinkers of the past 200 years. However, he proved very few theorems, and many of the proofs that he did supply were flawed. But his conceptual contributions, such as Riemannian geometry and the Riemann zeta function, as well as his contributions to elliptic and Abelian function theory, were epochal.

Jacques Hadamard (1865–1963) was perhaps the greatest mathematician to think deeply and seriously about cognition in mathematics. He is quoted as saying “. . . in arithmetic, until the seventh grade, I was last or nearly last,” which should give encouragement to many young students. Hadamard was both the author of *The Psychology of Invention in the Mathematical Field* [107], a

1945 book that still rewards close inspection, and co-prover of the Prime Number Theorem in 1896, which stands as one of the premier results of 19th century mathematics and an excellent example of a result whose discovery and eventual proof involved detailed computation and experimentation. He nicely declared:

The object of mathematical rigor is to sanction and legitimize the conquests of intuition, and there was never any other object for it. (J. Hadamard, from E. Borel, “Lecons sur la theorie des fonctions,” 1928, quoted in [165])

G. H. Hardy was another of the 20th century’s towering figures in mathematics. In addition to his own mathematical achievements in number theory, he is well known as the mentor of Ramanujan. In his Rouse Ball lecture in 1928, Hardy emphasized the intuitive and constructive components of mathematical discovery:

I have myself always thought of a mathematician as in the first instance an observer, a man who gazes at a distant range of mountains and notes down his observations. . . . The analogy is a rough one, but I am sure that it is not altogether misleading. If we were to push it to its extreme we should be led to a rather paradoxical conclusion; that we can, in the last analysis, do nothing but point; that proofs are what Littlewood and I call gas, rhetorical flourishes designed to affect psychology, pictures on the board in the lecture, devices to stimulate the imagination of pupils. This is plainly not the whole truth, but there is a good deal in it. The image gives us a genuine approximation to the processes of mathematical pedagogy on the one hand and of mathematical discovery on the other; it is only the very unsophisticated outsider who imagines that mathematicians make discoveries by turning the handle of some miraculous machine. Finally the image gives us at any rate a crude picture of Hilbert’s metamathematical proof, the sort of proof which is a ground for its conclusion and whose object is to convince. [59, Preface]

As one final example, in the modern age of computers, we quote John Milnor, a contemporary Fields medalist:

If I can give an abstract proof of something, I'm reasonably happy. But if I can get a concrete, computational proof and actually produce numbers I'm much happier. I'm rather an addict of doing things on computer, because that gives you an explicit criterion of what's going on. I have a visual way of thinking, and I'm happy if I can see a picture of what I'm working with. [169, page 78]

We should point out that paradigm shifts of this sort in scientific research have always been difficult to accept. For example, in the original 1859 edition of *Origin of the Species*, Charles Darwin wrote,

Although I am fully convinced of the truth of the views given in this volume . . . , I by no means expect to convince experienced naturalists whose minds are stocked with a multitude of facts all viewed, during a long course of years, from a point of view directly opposite to mine. . . . [B]ut I look with confidence to the future—to young and rising naturalists, who will be able to view both sides of the question with impartiality. [81, page 453]

In the 20th century, a very similar sentiment was expressed by Max Planck regarding quantum physics:

[A] new scientific truth does not triumph by convincing its opponents and making them see the light, but rather because its opponents die and a new generation grows up that is familiar with it. [163, page 33–34]

Thomas Kuhn observed in his epochal work *The Structure of Scientific Revolutions*,

I would argue, rather, that in these matters neither proof nor error is at issue. The transfer of allegiance from paradigm to paradigm is a conversion experience that cannot be forced. [134, page 151]

Two final quotations deal with the dangers of overreliance on tradition and “authority” in scientific research. The first is an admonition by the early English scholar-scientist Robert Record, in his 1556 cosmology textbook *The Castle of Knowledge*:

No man can worthily praise Ptolemye . . . yet muste ye and all men take heed, that both in him and in all mennes workes, you be not abused by their autoritye, but evermore attend to their reasons, and examine them well, ever regarding more what is saide, and how it is proved, than who saith it, for autorite often times deceaveth many menne. [90, page 47]

The following is taken from an intriguing, recently published account of why one of the most influential articles of modern mathematical economics, which in fact later led to a Nobel Prize for its authors (John R. Hicks and Kenneth J. Arrow), was almost not accepted for publication:

[T]o suggest that the normal processes of scholarship work well on the whole and in the long run is in no way contradictory to the view that the processes of selection and sifting which are essential to the scholarly process are filled with error and sometimes prejudice. (Kenneth Arrow [190])

## 1.5 Gauss, the Experimental Mathematician

Carl Friedrich Gauss once confessed

I have the result, but I do not yet know how to get it. [8, page 115]

Gauss was particularly good at seeing meaningful patterns in numerical data. When just 14 or 15 years old, he conjectured that  $\pi(n)$ , the number of primes less than  $n$ , is asymptotically approximated by  $n/\log n$ . This conjecture is, of course, the Prime Number Theorem, eventually proved by Hadamard and de la Vallée Poussin in 1896. This will be discussed in greater detail in Section 2.8.

Here is another example of Gauss's prowess at "mental experimental mathematics." One day in 1799, while examining tables of integrals provided originally by James Stirling, he noticed that the reciprocal of the integral

$$\frac{2}{\pi} \int_0^1 \frac{dt}{\sqrt{1-t^4}},$$

agreed numerically with the limit of the rapidly convergent arithmetic-geometric mean iteration:  $a_0 = 1$ ,  $b_0 = \sqrt{2}$ ;

$$a_{n+1} = \frac{a_n + b_n}{2}, \quad b_{n+1} = \sqrt{a_n b_n}. \quad (1.5.1)$$

The sequences  $(a_n)$  and  $(b_n)$  have the common limit 1.1981402347355922074... Based on this purely computational observation, Gauss was able to conjecture and subsequently prove that the integral is indeed equal to this common limit. It was a remarkable result, of which he wrote in his diary (see [50, pg. 5] and below) “[the result] will surely open up a whole new field of analysis.” He was right. It led to the entire vista of 19th century elliptic and modular function theory.

We reproduce the relevant pages from his diary as Figures 1.1 and 1.2. The first shows the now familiar hypergeometric series which, along with the arithmetic-geometric mean iteration, we discuss in some detail in Section 5.6.2.

In Figure 1.2, an excited Gauss writes:

*Novus in analysi campus se nobis aperuit, scilicet investigatio functionem etc.* (October 1798) [A new field of analysis has appeared to us, evidently in the study of functions etc.]

And in May 1799 (a little further down the page), he writes:

*Terminum medium arithmetico-geometricum inter 1 et (root 2) esse pi/omega usque ad figuram undecimam comprobavimus, qua re demonstrata prorsus novus campus in analysi certo aperietur.* [We have shown the limit of the arithmetical-geometric mean between 1 and root 2 to be pi/omega up to eleven figures, which on having been demonstrated, a whole new field in analysis is certain to be opened up.]

## 1.6 Geometric Experiments

Augustus de Morgan (1806–71), the first President of the London Mathematical Society, was equally influential as an educator and a researcher [170]. As the following two quotes from De Morgan show, neither a pride in numerical skill nor a desire for better geometric tools is new. De Morgan like many others saw profound differences between two and three dimensions:

Problemu eliminationis ita solutum ut nihil  
 amplius desiderari possit. Gott. J. G.  
 Vera elegantiuscula circa attractionem  
 sphaerae

$$2 + \frac{1}{3} \frac{1 \cdot 3}{4 \cdot 4} + \frac{1}{5!} \frac{1 \cdot 3 \cdot 5 \cdot 7}{4 \cdot 4 \cdot 8 \cdot 8} + \frac{1}{729} \frac{1 \cdot 3 \cdot 5 \cdot 7 \cdot 9 \cdot 11}{4 \cdot 4 \cdot 8 \cdot 8 \cdot 16 \cdot 16} \dots = \sqrt{5}$$

$$1,02220\dots = \frac{1,9110\dots}{3,1415\dots}$$

arc. sin. lemn.  $\frac{1}{2} \varphi$  - ar. sin. lemn.  $\frac{1}{2} \varphi$

$$= \frac{\frac{1}{2} \varphi}{\pi} - \frac{\frac{1}{2} \varphi}{\pi}$$

sin. lemn.  $\frac{1}{2} \varphi$  = 0,95500898 sin.  $\frac{1}{2} \varphi$   
 = 0,0430493 sin.  $\frac{3}{2} \varphi$   
 + 0,0018605 sin.  $\frac{5}{2} \varphi$   
 - 0,0008803 sin.  $\frac{7}{2} \varphi$   
 sin<sup>2</sup> lemn.  $\frac{1}{2} \varphi$  = 0,4569472 =  $\frac{1}{2} \cos \varphi$

Arc. sin. lemn.  $\frac{1}{2} \varphi$  =  $\frac{\varphi}{\pi}$   
 +  $\left(\frac{\pi}{4} - \frac{2}{\pi}\right) \sin 2\varphi$   
 +  $\left(\frac{7\pi}{8} - \frac{17}{8}\right) \sin 4\varphi$

$$\sin^2 = 0,475221\dots \sin - + 0,03$$

Figure 1.1: Gauss on the lemniscate.

Sic laudate, et quae summe omnes expectati-  
 ones superantia acquisivimus et quae  
 per methodos suae campum propus-  
 itum nobis aperuit. Gott. Jul.

# Solutio problematis ballistici Gott. Jul.

# Cometarum theoriam perfectiorem reddidi Gott. Jul.

Novus in analysi campus de nobis aperuit,  
 scilicet investigatio functionum etc.

# Formas superiores considerare coepimus  
 Mr. Febr. 1777

Formulas exactas pro parallaxi  
 eruimus ————— Mr. Apr. 8.

# Terminum medium arithmetico-geometricum  
 inter 1 et  $\sqrt{2}$  esse =  $\frac{\pi}{10}$  usque  
 ad figuram undecimam comprobavimus, quare  
 demonstrata proinde novus campus in analysi  
 certo aperietur Mr. Mai 30.

# In principis Geometriae egregios progressus  
 fecimus Mr. Sept.

# Circa terminos medios arithmetico-geometricos  
 multae nova deteximus Mr. Novemb.

Figure 1.2: Gauss on the arithmetic-geometric mean.



Considerable obstacles generally present themselves to the beginner, in studying the elements of Solid Geometry, from the practice which has hitherto uniformly prevailed in this country, of never submitting to the eye of the student, the figures on whose properties he is reasoning, but of drawing perspective representations of them upon a plane. . . . I hope that I shall never be obliged to have recourse to a perspective drawing of any figure whose parts are not in the same plane.

There is considerable evidence that young children see more naturally in three than two dimensions.

Elsewhere, de Morgan celebrates:

In 1831, Fourier's posthumous work on equations showed 33 figures of solution, got with enormous labor. Thinking this is a good opportunity to illustrate the superiority of the method of W. G. Horner, not yet known in France, and not much known in England, I proposed to one of my classes, in 1841, to beat Fourier on this point, as a Christmas exercise. I received several answers, agreeing with each other, to 50 places of decimals. In 1848, I repeated the proposal, requesting that 50 places might be exceeded: I obtained answers of 75, 65, 63, 58, 57, and 52 places.

Angela Vierling's web page <http://math.bu.edu/people/angelav/projects/models> well describes 19th-century desire for aids to visualization:

During many of these investigations, models were built to illustrate properties of these surfaces. The construction and study of plaster models was especially popular in Germany (particularly in Göttingen under the influence of Felix Klein). Many of the models were mass produced by publishing houses and sold to mathematicians and mathematics departments all over the world. Models were built of many other types of surfaces as well, including surfaces arising from the study of differential geometry and calculus. Such models enjoyed a wonderful reception for a while, but after the 1920's production and interest waned.

Felix Klein, like De Morgan, was equally influential as researcher and as educator. These striking and very expensive models still exist in many university departments and can be viewed as a precursor to modern electronic visualization tools such as Rob Scharein's KnotPlot site <http://www.colab.sfu.ca/KnotPlot>, which now has a three dimensional “immersive reality” version, and the remarkable Electronic Geometry Models site <http://www.eg-models.de>.

While we will not do a great deal of geometry in this book, this arena has great potential for visualization and experimentation. Three beautiful theorems come to mind where visualization, in particular, plays a key role:

1. *Pick's theorem* on the area of a simple lattice polygon,  $P$ :

$$A(P) = I(P) + \frac{1}{2}B(P) - 1, \quad (1.6.2)$$

where  $I(P)$  is the number of lattice points inside  $P$  and  $B(P)$  is the number of lattice points on the boundary of  $P$  including the vertices.

2. Minkowski's seminal result in the geometry of numbers that *a symmetric convex planar body must contain a nonzero lattice point in its interior if its area exceeds four.*
3. Sylvester's theorem: *Given a noncollinear finite set in the plane, one can always draw a line through exactly two points of the set.*

In the case of Pick's theorem it is easy to think of a useful experiment (one of the present authors has invited students to do this experiment). It is reasonable to first hunt for a formula for acute-angled triangles. One can then hope to piece together the more general result by triangulating the polygon (even if it is nonconvex), and then clearly for right-angled triangles. Now place the vertices at  $(n, 0)$ ,  $(0, m)$ , and  $(0, 0)$  and write a few lines of code that separately totals the number of times  $(j, k)$  lies on the boundary lines or inside the triangle as  $j$  ranges between 0 and  $n$  and  $k$  between 0 and  $m$ . A table of results for small  $m$  and  $n$  will expose the result. For example, if we consider all right-angled

triangles of height  $h$  and width  $w$  with area 30, we obtain:

$$\begin{bmatrix} h & w & A & I & B \\ 10 & 6 & 30 & 22 & 18 \\ 12 & 5 & 30 & 22 & 18 \\ 15 & 4 & 30 & 21 & 20 \\ 20 & 3 & 30 & 19 & 24 \\ 30 & 2 & 30 & 14 & 34 \\ 60 & 1 & 30 & 0 & 62 \end{bmatrix}. \quad (1.6.3)$$

It is more of a challenge to think of a useful experiment to determine that “4” is the right constant in Minkowski’s theorem. Both of these results are very accessibly described in [160].

James Sylvester, mentioned in Item 3 above, was president of the London Mathematical Society in the late 19th century. He once wrote, “The early study of Euclid made me a hater of geometry” [146]. Discrete geometry (now much in fashion as “computational geometry” and another example of very useful pure mathematics) was clearly more appealing to Sylvester. For Sylvester’s theorem (posed but not solved by Sylvester), one can imagine various Java applets but scattering a fair number of points on a sheet of white paper and using a ruler seems more than ample to get a sense of the truth of the result.

Sylvester’s conjecture was largely forgotten for 50 years. It was first established—“badly” in the sense that the proof is much more complicated than it needed to be—by Gallai (1943) and also by Paul Erdős, who named “the Book” in which God keeps elegant and aesthetically perfect proofs. Kelly’s proof, which was declared by Erdős to be “in the Book,” was actually published by Donald Coxeter in the MAA Monthly in 1948 (this is a good example of how easily the archival record is often obscured). A marvellous eponymous book is [3]. It is chock full of proofs that are or should be in the book and, for example, gives six proofs of the infinitude of primes.

**Proof Sketch.** Let  $S$  be the given set of points. Consider the collection  $\mathcal{C}$  of pairs  $(L, p)$ , where  $L$  is a line through (at least two distinct) points in  $S$  and where  $p$  is a point in  $S$  not on  $L$ . Then  $\mathcal{C}$  is nonempty and contains only finitely many such pairs. Among those, pick  $(L, p)$  such that the distance from  $p$  to  $L$  is *minimal*. We claim that  $L$  harbors exactly two points from  $S$ .

Assume not, then,  $L$  contains 3 or more points. In Figure 1.3,  $L$  is represented as the horizontal line. Let  $q$  be the projection of  $p$  onto  $L$ . In Figure 1.3, we

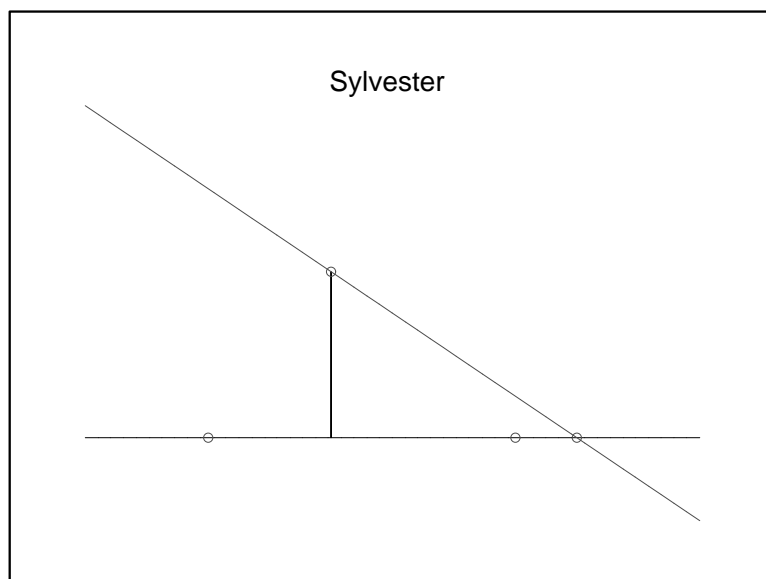


Figure 1.3: Kelly's 1948 "Proof from the Book."

drew 3 points of  $S$  on  $L$ . Label these points  $a, b, c$ . (Two must be on one side of  $q$ .) Consider  $b$  and draw the line  $L'$  through  $p$  and either  $a$  or  $c$ , whichever line is closer to  $b$ . In Figure 1.3,  $L'$  is the slanted line. Then  $(L', b)$  belongs to  $\mathcal{C}$  and the distance from  $b$  to  $L'$  is strictly smaller than the distance from  $p$  to  $L$ . But this contradicts the choice of  $(L, p)$ .  $\square$

As with the visual proof of the irrationality of  $\sqrt{2}$ , we will give in Section 2.9, we see forcibly the power of the right *minimal configuration*.

Dirac conjectured that every sufficiently large set of  $n$  noncollinear points contains at least  $n/2$  proper (or elementary) lines through exactly two points.

By contrast,

The Desmic configuration, discovered by Stephanos in 1890, is  $[\dots]$  a configuration spanning 3-space, consisting of three tetrads of points, each two of the tetrads being in perspective from the four points of the third tetrad. This means that any line intersecting two of the tetrads also intersects the third. [55]

It is conjectured that in many senses this configuration (built from the corners of the cube and a point at infinity) is unique [55]. One can view the Desmic configuration as showing that the Sylvester-Gallai theorem “fails in three dimensions.”

Sylvester had a most colorful and somewhat difficult life which included a seminal role in the founding of Johns Hopkins University, and ended as the first Jewish Chair in Oxford. Educated in Cambridge, he could not graduate until 1871, when theological tests were finally abolished. This is engagingly described in *Oxford Figures* [90].

Along this line, readers may be interested in Figures 1.4 and 1.5, which are taken from Part VII of a 19th century experimental geometry book by the French educator Paul Bert. The intention was to make school geometry more intuitive and empirical, quite far from Euclid’s *Elements*.

### 1.6.1 On Picture-Writing

George Polya, in an engaging eponymous *American Mathematical Monthly* article, provides three compelling examples of converting pictorial representations of problems into generating function solutions [164]:

1. *In how many ways can you make change for a dollar?*

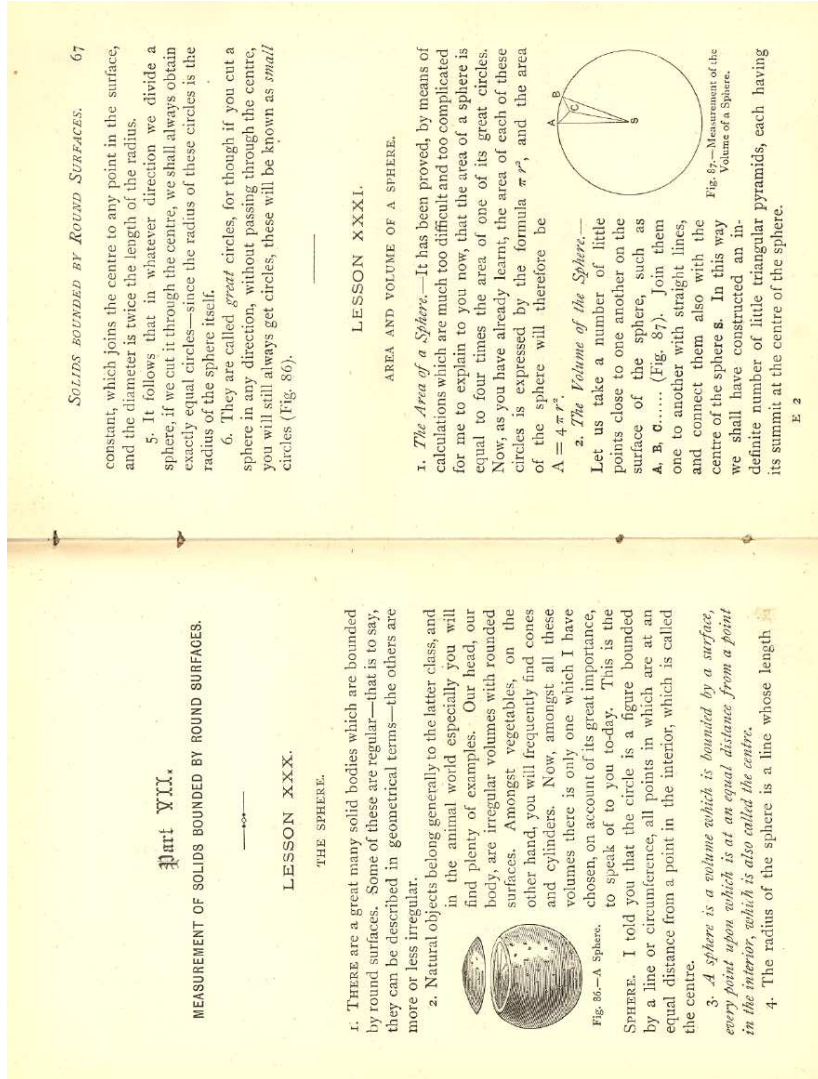
This leads to the (US currency) generating function

$$\sum_{k=1}^{\infty} P_k x^k = \frac{1}{(1-x)(1-x^5)(1-x^{10})(1-x^{25})(1-x^{50})},$$

which one can easily expand using a *Mathematica* command,

```
Series[1/((1-x)*(1-x^5)*(1-x^10)*(1-x^25)*(1-x^50)), {x, 0, 100}]
```

to obtain  $P_{100} = 292$  (242 for Canadian currency, which lacks a 50 cent piece). Polya’s illustration is shown in Figure 1.6.



constant, which joins the centre to any point in the surface, and the diameter is twice the length of the radius.

5. It follows that in whatever direction we divide a sphere, if we cut it through the centre, we shall always obtain exactly equal circles—since the radius of these circles is the radius of the sphere itself.

6. They are called *great circles*, for though if you cut a sphere in any direction, without passing through the centre, you will still always get circles, these will be known as *small circles* (Fig. 86).

Part VII.

MEASUREMENT OF SOLIDS BOUNDED BY ROUND SURFACES.

LESSON XXX.

THE SPHERE.

1. There are a great many solid bodies which are bounded by round surfaces. Some of these are regular—that is to say, they can be described in geometrical terms—the others are more or less irregular.

2. Natural objects belong generally to the latter class, and in the animal world especially you will find plenty of examples. Our head, our body, are irregular volumes with rounded surfaces. Amongst vegetables, on the other hand, you will frequently find cones and cylinders. Now, amongst all these volumes there is only one which I have chosen, on account of its great importance, to speak of to you to-day. This is the SPHERE. I told you that the circle is a figure bounded by a line or circumference, all points in which are at an equal distance from a point in the interior, which is called the centre.

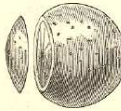


Fig. 86.—A Sphere.

3. A sphere is a volume which is bounded by a surface, every point upon which is at an equal distance from a point in the interior, which is also called the centre.

4. The radius of the sphere is a line whose length

LESSON XXXI.

AREA AND VOLUME OF A SPHERE.

1. *The Area of a Sphere.*—It has been proved, by means of calculations which are much too difficult and too complicated for me to explain to you now, that the area of a sphere is equal to four times the area of one of its great circles. Now, as you have already learnt, the area of each of these circles is expressed by the formula  $\pi r^2$ , and the area of the sphere will therefore be

$$A = 4\pi r^2.$$

2. *The Volume of the Sphere.*—Let us take a number of little points close to one another on the surface of the sphere, such as A, B, C,.... (Fig. 87). Join them one to another with straight lines, and connect them also with the centre of the sphere s. In this way we shall have constructed an indefinite number of little triangular pyramids, each having its summit at the centre of the sphere.

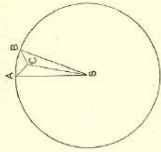


Fig. 87.—Measurement of the Volume of a Sphere.

Figure 1.4: Paul Bert's 1886 experimental geometry text, pages 66–67.

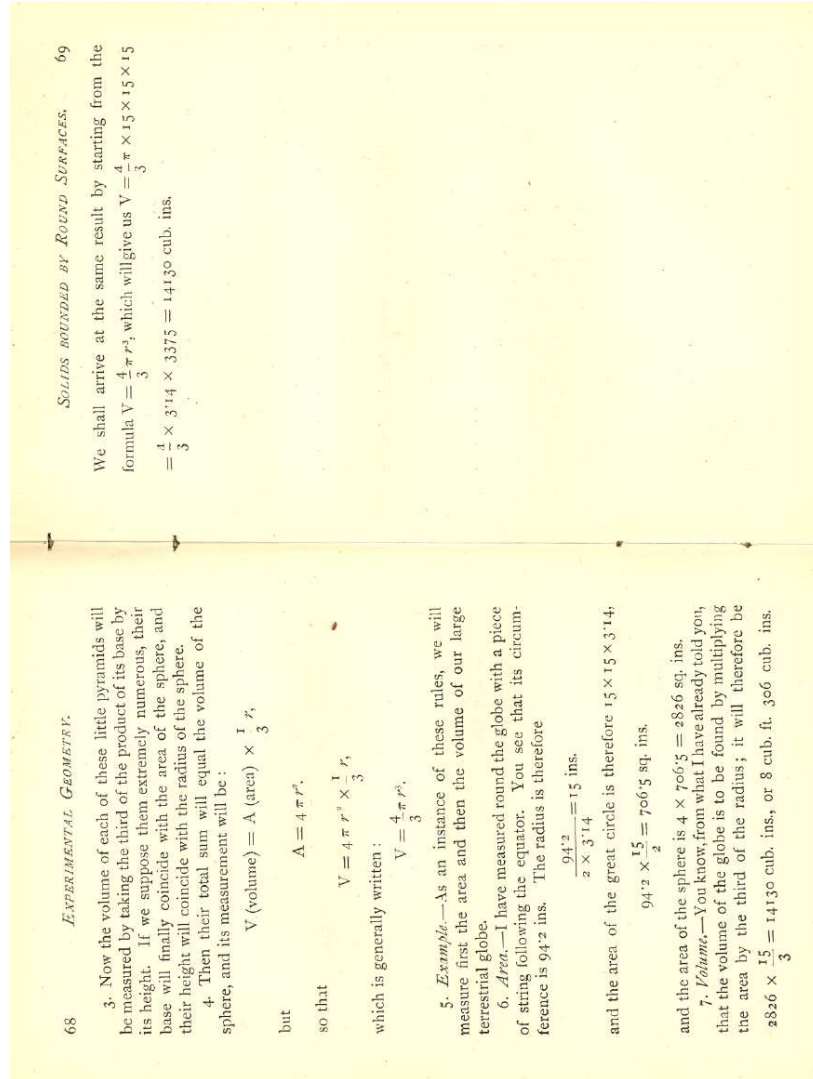


Figure 1.5: Paul Bert's text, pages 68–69.

$$\begin{aligned}
 & (\square + \textcircled{1} + \textcircled{1}\textcircled{1} + \textcircled{1}\textcircled{1}\textcircled{1} + \dots) \cdot \\
 & (\square + \textcircled{5} + \textcircled{5}\textcircled{5} + \textcircled{5}\textcircled{5}\textcircled{5} + \dots) \cdot \\
 & (\square + \textcircled{10} + \textcircled{10}\textcircled{10} + \textcircled{10}\textcircled{10}\textcircled{10} + \dots) \cdot \\
 & (\square + \textcircled{25} + \textcircled{25}\textcircled{25} + \textcircled{25}\textcircled{25}\textcircled{25} + \dots) \cdot \\
 & (\square + \textcircled{50} + \textcircled{50}\textcircled{50} + \textcircled{50}\textcircled{50}\textcircled{50} + \dots) \cdot \\
 & = \dots + \square \cdot \textcircled{5}\textcircled{5}\textcircled{5} \cdot \textcircled{10} \cdot \textcircled{25} \cdot \textcircled{50} + \dots
 \end{aligned}$$

Figure 1.6: Polya’s illustration of the change solution.

2. *Dissect a polygon with  $n$  sides into  $n - 2$  triangles by  $n - 3$  diagonals and compute  $D_n$ , the number of different dissections of this kind.*

This leads to the fact that the generating function for  $D_3 = 1, D_4 = 2, D_5 = 5, D_6 = 14, D_7 = 42, \dots$

$$D(x) = \sum_{k=1}^{\infty} D_k x^k$$

satisfies

$$D(x) = x [1 + D(x)]^2,$$

whose solution is therefore

$$D(x) = \frac{1 - 2x - \sqrt{1 - 4x}}{x},$$

and  $D_{n+2}$  turns out to be the  $n$ -th Catalan number  $\binom{2n}{n}/(n + 1)$ .

3. *Compute  $T_n$ , the number of different (rooted) trees with  $n$  knots.*

The generating function of the  $T_n$  becomes a remarkable result due to Cayley:

$$T(x) = \sum_{k=1}^{\infty} T_k x^k = x \prod_{k=1}^{\infty} (1 - x^k)^{-T_k}, \tag{1.6.4}$$



where remarkably the product and the sum share their coefficients. This produces a recursion for  $T_n$  in terms of  $T_1, T_2, \dots, T_{n-1}$ , which starts:  $T_1 = 1, T_2 = 1, T_3 = 2, T_4 = 4, T_5 = 9, T_6 = 20, \dots$  We shall revisit such products in Section 4.2 of the second volume of this work.

In each case, Polya's main message is that one can usefully draw pictures of the component elements—(a) in pennies, nickels dimes and quarters (plus loonies in Canada and half dollars in the US), (b) in triangles and (c) in the simplest trees (with the fewest knots).

## 1.7 Sample Problems of Experimental Math

In the January 2002 issue of *SIAM News*, Nick Trefethen of Oxford University presented ten diverse problems used in teaching graduate numerical analysis students at Oxford University, the answer to each being a certain real number. Readers were challenged to compute ten digits of each answer, with a \$100 prize to the best entrant. Trefethen wrote, "If anyone gets 50 digits in total, I will be impressed."

Success in solving these problems requires a broad knowledge of mathematics and numerical analysis, together with significant computational effort to obtain solutions and ensure correctness of the results. The strengths and limitations of *Maple*, *Mathematica*, *MATLAB* and other software tools are strikingly revealed in these exercises.

A total of 94 teams, representing 25 different nations, submitted results. Twenty of these teams received a full 100 points (10 correct digits for each problem). Since these results were much better than expected, an anonymous donor provided funds for a \$100 award to each team. The present authors and Greg Fee entered, but failed to qualify for an award. The ten problems are:

1. What is  $\lim_{\epsilon \rightarrow 0} \int_{\epsilon}^1 x^{-1} \cos(x^{-1} \log x) dx$ ?
2. A photon moving at speed 1 in the  $x$ - $y$  plane starts at  $t = 0$  at  $(x, y) = (1/2, 1/10)$  heading due east. Around every integer lattice point  $(i, j)$  in the plane, a circular mirror of radius  $1/3$  has been erected. How far from the origin is the photon at  $t = 10$ ?

3. The infinite matrix  $A$  with entries  $a_{11} = 1$ ,  $a_{12} = 1/2$ ,  $a_{21} = 1/3$ ,  $a_{13} = 1/4$ ,  $a_{22} = 1/5$ ,  $a_{31} = 1/6$ , etc., is a bounded operator on  $\ell^2$ . What is  $\|A\|$ ?
4. What is the global minimum of the function  $\exp(\sin(50x)) + \sin(60e^y) + \sin(70 \sin x) + \sin(\sin(80y)) - \sin(10(x + y)) + (x^2 + y^2)/4$ ?
5. Let  $f(z) = 1/\Gamma(z)$ , where  $\Gamma(z)$  is the gamma function, and let  $p(z)$  be the cubic polynomial that best approximates  $f(z)$  on the unit disk in the supremum norm  $\|\cdot\|_\infty$ . What is  $\|f - p\|_\infty$ ?
6. A flea starts at  $(0, 0)$  on the infinite 2-D integer lattice and executes a biased random walk: At each step it hops north or south with probability  $1/4$ , east with probability  $1/4 + \epsilon$ , and west with probability  $1/4 - \epsilon$ . The probability that the flea returns to  $(0, 0)$  sometime during its wanderings is  $1/2$ . What is  $\epsilon$ ?
7. Let  $A$  be the  $20000 \times 20000$  matrix whose entries are zero everywhere except for the primes  $2, 3, 5, 7, \dots, 224737$  along the main diagonal and the number 1 in all the positions  $a_{ij}$  with  $|i - j| = 1, 2, 4, 8, \dots, 16384$ . What is the  $(1, 1)$  entry of  $A^{-1}$ .
8. A square plate  $[-1, 1] \times [-1, 1]$  is at temperature  $u = 0$ . At time  $t = 0$  the temperature is increased to  $u = 5$  along one of the four sides while being held at  $u = 0$  along the other three sides, and heat then flows into the plate according to  $u_t = \Delta u$ . When does the temperature reach  $u = 1$  at the center of the plate?
9. The integral  $I(a) = \int_0^2 [2 + \sin(10\alpha)] x^\alpha \sin(\alpha/(2 - x)) dx$  depends on the parameter  $\alpha$ . What is the value  $\alpha \in [0, 5]$  at which  $I(\alpha)$  achieves its maximum?
10. A particle at the center of a  $10 \times 1$  rectangle undergoes Brownian motion (i.e., 2-D random walk with infinitesimal step lengths) till it hits the boundary. What is the probability that it hits at one of the ends rather than at one of the sides?

These problems and their solutions are described in detail in a forthcoming book [41]. Answers correct to 40 digits are available at the URL <http://web.comlab.ox.ac.uk/oucl/work/nick.trefethen/hundred.html>.

Inspired by this set of problems, the present authors have assembled a similar set of problems, similar in style to the SIAM/Oxford 100 Digit Challenge, but emphasizing the flavor of experimental mathematics. As in the above problem set, a real constant is defined in each case. The objective here is to produce at least 100 correct digits, so that a total of 1000 points can be earned. Several of these can be solved by fairly direct application of numerical computation, but others require mathematical analysis and reduction before computation can be done in reasonable time. Each problem provides an “extra credit” question, for which an additional 100 points may be earned. The maximum total score is thus 2000 points.

In each case, these problems can be solved with techniques presented either in this volume or in the companion volume. Answers, with references, can be found at the URL <http://www.expmath.info>.

1. Compute the value of  $r$  for which the chaotic iteration  $x_{n+1} = rx_n(1 - x_n)$ , starting with some  $x_0 \in (0, 1)$ , exhibits a bifurcation between 4-way periodicity and 8-way periodicity.

Extra credit: This constant is an algebraic number of degree not exceeding 20. Find the minimal polynomial with integer coefficients that it satisfies.

2. Evaluate

$$\sum_{(m,n,p) \neq 0} \frac{(-1)^{m+n+p}}{\sqrt{m^2 + n^2 + p^2}}, \quad (1.7.5)$$

where convergence means the limit of sums over the integer lattice points enclosed in increasingly large cubes surrounding the origin.

Extra credit: Identify this constant.

3. Evaluate the sum

$$\sum_{k=1}^{\infty} \left( 1 - \frac{1}{2} + \cdots + (-1)^{k+1} \frac{1}{k} \right)^2 (k+1)^{-3}.$$

Extra credit: Evaluate this constant as a multiterm expression involving well-known mathematical constants. This expression has seven terms, and

involves  $\pi$ ,  $\log 2$ ,  $\zeta(3)$ , and  $\text{Li}_5(1/2)$ , where  $\text{Li}_n(x) = \sum_{k>0} x^n/n^k$ . Hint: The expression is “homogenous,” in the sense that each term has the same total “degree.” The degrees of  $\pi$  and  $\log 2$  are each 1, the degree of  $\zeta(3)$  is 3, the degree of  $\text{Li}_5(1/2)$  is 5, and the degree of  $\alpha^n$  is  $n$  times the degree of  $\alpha$ .

4. Evaluate

$$\prod_{k=1}^{\infty} \left[ 1 + \frac{1}{k(k+2)} \right]^{\log_2 k} = \prod_{k=1}^{\infty} k^{\left[ \log_2 \left( 1 + \frac{1}{k(k+2)} \right) \right]}.$$

Extra credit: Evaluate this constant in terms of a less-well-known mathematical constant.

5. Given  $a, b, \eta > 0$ , define

$$R_{\eta}(a, b) = \frac{a}{\eta + \frac{b^2}{\eta + \frac{4a^2}{\eta + \frac{9b^2}{\eta + \dots}}}}.$$

Calculate  $R_1(2, 2)$ .

Extra credit: Evaluate this constant as a two-term expression involving a well-known mathematical constant.

6. Calculate the expected distance between two random points on different sides of the unit square.

Hint: This can be expressed in terms of integrals as

$$\frac{2}{3} \int_0^1 \int_0^1 \sqrt{x^2 + y^2} dx dy + \frac{1}{3} \int_0^1 \int_0^1 \sqrt{1 + (y - u)^2} du dy.$$

Extra credit: Express this constant as a three-term expression involving algebraic constants and an evaluation of the natural logarithm with an algebraic argument.

7. Calculate the expected distance between two random points on different faces of the unit cube.

Hint: This can be expressed in terms of integrals as

$$\frac{4}{5} \int_0^1 \int_0^1 \int_0^1 \int_0^1 \sqrt{x^2 + y^2 + (z - w)^2} dw dx dy dz$$

$$+ \frac{1}{5} \int_0^1 \int_0^1 \int_0^1 \int_0^1 \sqrt{1 + (y - u)^2 + (z - w)^2} du dw dy dz.$$

Extra credit: Express this constant as a six-term expression involving algebraic constants and two evaluations of the natural logarithm with algebraic arguments.

8. Calculate

$$\int_0^\infty \cos(2x) \prod_{n=1}^\infty \cos\left(\frac{x}{n}\right) dx.$$

Extra credit: Express this constant as an analytic expression. Hint: It is *not* what it first appears to be.

9. Calculate

$$\sum_{i>j>k>l>0} \frac{1}{i^3 j k^3 l}.$$

Extra credit: Express this constant as a single-term expression involving a well-known mathematical constant.

10. Evaluate

$$W_1 = \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} \frac{1}{3 - \cos(x) - \cos(y) - \cos(z)} dx dy dz.$$

Extra credit: Express this constant in terms of the Gamma function.

## 1.8 Internet-Based Mathematical Resources

We list here some Internet-based resources that we have found useful in our research. We have selected those that we believe will be fairly permanent, but some may become defunct with the passage of time. An updated list of these URLs, with links, can be found at <http://www.expmath.info>.

### Institutional Sites

1. American Institute of Mathematics:  
<http://www.aimath.org>
2. Canadian Mathematical Society's KaBoL site:  
<http://camel.math.ca/Kabol>
3. Clay Mathematics Institute:  
<http://www.claymath.org>
4. Experimental Mathematics Journal:  
<http://www.expmath.org>
5. European Mathematical Society:  
<http://elib.uni-osnabrueck.de>
6. Mathematical Association of America Online:  
<http://www.maa.org>
7. Math-Net (International Mathematical Union):  
<http://www.Math-Net.org>
8. MathSciNet:  
<http://e-math.ams.org/mathscinet>
9. Society for Industrial and Applied Mathematics:  
<http://www.siam.org>

### Commercial Sites

1. Apple Computer's Advanced Computation Group research site:  
<http://developer.apple.com/hardware/ve/acgresearch.html>

2. The Cinderella interactive geometry site:  
<http://www.cinderella.de>
3. Integrals.com (operated by Wolfram Research):  
<http://www.integrals.com>
4. Maplesoft (*Maple*):  
<http://www.maplesoft.com>  
<http://www.mapleapps.com>
5. MathResources interactive dictionaries:  
<http://www.mathresources.com>
6. Mathworks (MATLAB):  
<http://www.mathworks.com>
7. NEC's CiteSeer database:  
<http://citeseer.nj.nec.com>
8. Eric Weisstein's World of Mathematics site:  
<http://mathworld.wolfram.com>
9. Wolfram Research, Inc. (*Mathematica*):  
<http://www.wolfram.com>

### Noncommercial Software and Tools

1. Alf-Christian Achilles' computer science bibliography:  
<http://liinwww.ira.uka.de/bibliography>
2. The Algorithm Project software library:  
<http://algo.inria.fr/libraries/software.html>
3. The ArXiv mathematics article database:  
<http://front.math.ucdavis.edu>
4. The Boyer-Moore theorem prover:  
<http://www.cs.utexas.edu/users/moore/best-ideas/nqthm>
5. The CECM Euler-zeta computation tool:  
<http://www.cecm.sfu.ca/projects/ezface+>

6. The CECM integer relation tool (see also Item 7):  
<http://www.cecm.sfu.ca/projects/IntegerRelations>
7. The CECM inverse symbolic calculator:  
<http://www.cecm.sfu.ca/projects/ISC>
8. The CECM RevEng tool (see also Item 7):  
<http://psg.cecm.sfu.ca/projects/reveng/client/RevEngClient.html>
9. Richard Crandall's integer computation software:  
<http://www.perfsci.com>
10. Richard Fateman's online integration tool:  
<http://torte.cs.berkeley.edu:8010/tilu>
11. The FFTW site (FFT software):  
<http://www.fftw.org>
12. The FIZ-Karlsruhe journal and abstract database:  
<http://www.zblmath.fiz-karlsruhe.de>
13. The GNU high-precision arithmetic library:  
<http://www.gnu.org/software/gmp/gmp.html>
14. The *Magma* computational algebra system:  
<http://magma.maths.usyd.edu.au/magma>
15. The NERSC double-double, quad-double, and arbitrary precision computation software:  
<http://www.expmath.info>
16. The NERSC Experimental Mathematician's Toolkit:  
<http://www.expmath.info>
17. The NERSC PiSearch facility (searches for names or hex digit sequences in the first several billion binary digits of  $\pi$ ):  
<http://pi.nerisc.gov>
18. The Netlib software repository (linear algebra and other math software):  
<http://www.netlib.org>



19. Neil Sloane's online dictionary of integer sequences:  
<http://www.research.att.com/~njas/sequences>

### Other Resources

1. Lee Borrell's "Absolute Certainty?" site:  
<http://www.fortunecity.com/emachines/e11/86/certain.html>
2. Jonathan Borwein's  $\pi$  pages:  
[http://www.cecm.sfu.ca/personal/jborwein/pi\\_cover.html](http://www.cecm.sfu.ca/personal/jborwein/pi_cover.html)
3. The CECM  $\pi$  recital site (recites  $\pi$  in numerous languages):  
<http://www.cecm.sfu.ca/pi/yapPing.html>
4. Gregory Chaitin's site on algorithmic information theory:  
<http://www.cs.umaine.edu/~chaitin>.
5. The de Smit-Lenstra site on the mathematics of Escher's "Print Gallery":  
<http://escherdroste.math.leidenuniv.nl>
6. Stewart Dickson's math art site:  
<http://emsh.calarts.edu/~mathart>
7. The electronic geometry site:  
<http://www.eg-models.de>
8. The Embree-Trefethen-Wright pseudospectra and eigenproblem site:  
<http://web.comlab.ox.ac.uk/projects/pseudospectra>
9. Helaman Ferguson's mathematical sculpture site:  
<http://www.helasculpt.com>
10. Steven Finch's mathematical constant site:  
<http://pauillac.inria.fr/algo/bsolve/constant/constant.html>
11. The Geometry Analysis Numerics Graphics (GANG) site:  
<http://www.gang.umass.edu>
12. Xavier Gourdon and Pascal Sebah's site for famous math constants:  
<http://numbers.computation.free.fr/Constants/constants.html>

13. Andrew Granville's Pascal triangle site:  
<http://www.cecm.sfu.ca/organics/papers/granville/support/pascalform.html>
14. David Griffeath's cellular automata site:  
<http://psoup.math.wisc.edu>
15. Jerry Grossman's Erdős number site:  
<http://www.acs.oakland.edu/~grossman/erdoshp.html>
16. Thomas Hales' Kepler problem site:  
<http://www.math.pitt.edu/~thales/kepler98>
17. David Joyce's site (Java implementation of Euclid's *Elements*, and the Mandelbrot and Julia set explorer):  
<http://aleph0.clarku.edu/~djoyce>
18. Yasumasa Kanada's  $\pi$  site:  
<http://www.super-computing.org>
19. Donald Knuth's site on primes in  $\pi$ :  
<http://mathpages.com/home/kmath184/kmath184.htm>
20. The Mersenne prime site:  
<http://www.mersenne.org>
21. National Institute of Standards and Technology's *Digital Library of Mathematical Functions*:  
<http://dlmf.nist.gov>
22. The Organic Mathematics Project site:  
<http://www.cecm.sfu.ca/organics>
23. The Piworld site (features artwork based on  $\pi$ ):  
<http://www.piworld.de>
24. George Reese's Buffon needle site:  
<http://www.mste.uiuc.edu/reese/buffon/buffon.html>
25. RSA Security's factorization challenge site:  
<http://www.rsasecurity.com/rsalabs/challenges/factoring>

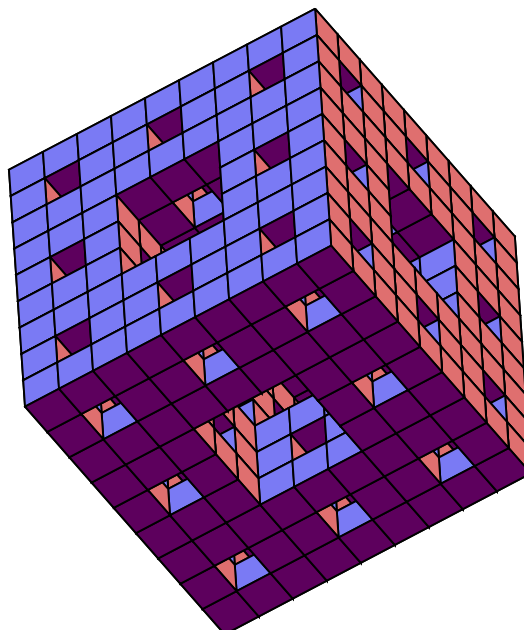


Figure 1.7: The Sierpinski cube.

26. Saint Andrews College's history of mathematics site (and curve resource):  
<http://www-gap.dcs.st-and.ac.uk/history>
27. Rob Scharein's KnotPlot research and development site:  
<http://www.colab.sfu.ca/KnotPlot>
28. Angela Vierling's mathematical models site:  
<http://math.bu.edu/people/angelav/projects/models>
29. Jeff Weeks' topology and geometry site:  
<http://www.geometrygames.org>
30. Figure 1.7 shows a Sierpinski cube plotted in *JavaView*  
<http://www.cecm.sfu.ca/news/coolstuff/JVL/htm/gallery.htm>

## 1.9 Commentary and Additional Examples

1. **The old and the new.** David Joyce's Java implementation of Euclid's *Elements* and his *Mandelbrot and Julia Set Explorer* show what technology can offer, for old and new material, when used appropriately: <http://aleph0.clarku.edu/~djoyce/java/elements/elements.html> and <http://aleph0.clarku.edu/~djoyce/julia/explorer.html>.
2. **Dedekind's preference for mental constructs** is also apparent in his development of the real numbers [85]:

Many authors who adopted Dedekind's basic ideas preferred not to follow him in defining the real numbers as creations of the mind corresponding to cuts in the system of rational numbers. . . . Bertrand Russell emphasised the advantage of defining the real numbers simply as . . . segments of the rationals. . . . But Dedekind had his reasons . . . for defining the real numbers as he did. When Heinrich Weber expressed his opinion in a letter to Dedekind that an irrational number should be taken to be the cut, instead of something new which is created in the mind and supposed to correspond to the cut, Dedekind replied "We have the right to grant ourselves such a creative power, and besides it is much more appropriate to proceed thus because of the similarity of all numbers. The rational numbers surely also produce cuts, but I will certainly not give out the rational number as identical with the cut generated by it; and also by introduction of the irrational numbers, one will often speak of cut-phenomena with such expressions, granting them such attributes, which applied to the numbers themselves would sound quite strange." [104, page 224]

3. **Hardy's "apology."** G. H. Hardy (1877–1947), the leading British analyst of the first half of the 20th century, wrote compellingly in defense of pure mathematics. In his essay, *A Mathematician's Apology* [108], he noted that

All physicists and a good many quite respectable mathematicians are contemptuous about proof.

Hardy's *Apology* is also a spirited defense of beauty over utility:

Beauty is the first test. There is no permanent place in the world for ugly mathematics.

Along this line, many have noted his quote,

Real mathematics . . . is almost wholly “useless.”

This has been overplayed and is now very dated, given the importance of cryptography, data compression and other applications of algebra and number theory that have arisen in recent years. But Hardy does acknowledge that if number theory could be employed for any “practical and obviously honourable purpose,” then “neither Gauss nor any other mathematician would have been so foolish as to decry or regret such applications.”

4. **Ramsey's theorem.** Ramsey's theorem asserts that given positive integers  $k$  and  $l$  there is an integer  $R(k, l)$  so that any graph with  $R(k, l)$  vertices either possesses a clique (complete subgraph) with  $k$  vertices or an independent set with  $l$  vertices.  $R(3) = R(3, 3) = 6$  and  $R(4) = R(4, 4) = 18$ . This is often described as saying that at a six-person dinner party either there are three friends or there are three strangers. Such numbers  $R(k, 1)$  are very hard to compute.

Indeed, Paul Erdős suggested that if an alien demanded we give it the value of  $R(5)$  in order to save the Earth, we should set all mathematicians and computers to calculating the value. But if the alien demanded  $R(6)$ , then humanity should attempt to destroy it before it destroyed us.

The following is from the site: <http://www.math.uiuc.edu/~miletic/Museum/ramsey.html>:

One version of Ramsey's Theorem states that no matter which number  $k$  you choose, you can find a number  $n$  such that given any arrangement of  $n$  pegs, [connected by red or green string] there must exist a monochromatic collection of  $k$  pegs. We will denote the smallest such  $n$  that works for a given  $k$  by  $R(k)$ . The above results can be stated more succinctly by saying that  $R(3) = 6$  and  $R(4) = 18$ . It is easy to see that  $R(1) = 1$

(given only 1 peg, there are no pieces of string, so it forms a monochromatic collection for vacuous reasons) and  $R(2) = 2$  (If you only have 2 pegs, there is only piece of string, so the 2 pegs form a monochromatic collection).

Perhaps surprisingly, nobody knows the value of  $R(k)$  for any  $k$  larger than 4. The best results currently known state that  $R(5)$  is somewhere between 43 and 49 (inclusive) and  $R(6)$  is somewhere between 102 and 165. You might wonder why, given the incredible computing power at our disposal, we can not simply search through all arrangements of string for 43 pegs through 49 pegs to find the actual value of  $R(5)$ . However, one can calculate that there are  $2^{903}$  (a number that has 272 decimal digits!) ways to arrange red and green string among 43 pegs, which is a number beyond ordinary comprehension (scientists estimate that there are about 80 digits in the number of electrons in the universe). By using symmetry, one can drastically lower the number of such arrangements a computer would have to look at, but even if we only had to examine 1 out of every 1 trillion configurations, we would still be left with over  $2^{864}$  arrangements (a number that has 261 digits!). Estimating the values of  $R(k)$  requires mathematical ingenuity in addition to brute force calculations.

5. **Gödel's theorem and complexity.** While there are still no “ordinary” Gödel statements (that is, true but unprovable), the *Paris-Harrington theorem* comes close. It is “[a]n arithmetically expressible true statement from finitary combinatorics . . . that is not provable in Peano arithmetic. The statement  $S$  in question is the strengthening of the finite Ramsey theorem by requiring the homogeneous set  $H$  to be ‘relatively large,’ i.e.,  $\text{card } H \geq \min H$ .” (Math Reviews). See also <http://www.cs.utexas.edu/users/moore/best-ideas/nqthm/> and <http://www.fortunecity.com/emachines/e11/86/certain.html>.
6. **Goodstein's theorem.** Another very striking example is that of *Goodstein sequences* [180]. Consider writing a number base  $b$ . Doing the same for each of the exponents in the resulting representations, until the process stops, yields the *hereditary base  $b$  representation of  $n$* . For example, the

hereditary base 2 representation of 266 is

$$266 = 2^{2^{2+1}} + 2^{2+1} + 2^1.$$

**Base change.** Let  $B_b(n)$  be the natural number obtained on replacing each  $b$  by  $b + 1$  in the hereditary base  $b$  representation of  $n$ . For example, “bumping the base” from 2 to 3 above gives

$$B_2(266) = 3^{3^{3+1}} + 3^{3+1} + 3.$$

Consider a sequence of integers obtained by repeatedly applying the operation: *Bump the base and subtract one from the result*. Iteratively performing this procedure for 266 yields

$$\begin{aligned} 266_0 &= 266 \\ 266_1 &= 3^{3^{3+1}} + 3^{3+1} + 2 \\ 266_2 &= 4^{4^{4+1}} + 4^{4+1} + 1 \\ 266_3 &= 5^{5^{5+1}} + 5^{5+1} \\ 266_4 &= 6^{6^{6+1}} + 6^{6+1} - 1 \\ &= 6^{6^{6+1}} + 5 \cdot 6^6 + 5 \cdot 6^5 + \cdots + 5 \cdot 6 + 5 \\ 266_5 &= 7^{7^{7+1}} + 5 \cdot 7^7 + 5 \cdot 7^5 + \cdots + 5 \cdot 7 + 4 \\ &\dots \end{aligned}$$

Done generally, this determines the *Goodstein sequence* starting at  $n$ . That is, we recursively define nonnegative integers  $n_0 = n, n_1, \dots, n_k, \dots$  by

$$n_{k+1} = B_{k+2}(n_k) - 1,$$

if  $n_k > 0$  and  $n_{k+1} = 0$  otherwise. We initially obtain very rapid growth:  $266_8 \approx 10^{10^{11}}$  and  $4_k$  first reaches 0 for  $k = 3(2^{402653211} - 1) \approx 10^{121210695}$ .

However, Goodstein in 1944 proved that *every Goodstein sequence converges to 0*.

Remarkably, in 1982 Paris and Kirby showed that Goodstein’s theorem is not provable in ordinary Peano arithmetic, despite being a fairly ordinary sounding number-theoretic fact.

There are quite natural, closely related games on finite trees such as “Hercules and the Hydra” in which the fact that Hercules always has a winning strategy to defeat a many-headed hydra is independent of arithmetic. Such results show that even in a seemingly computational framework, we may bump into Gödel’s theorem [132].

7. **Hales’ computer-assisted proof of Kepler’s conjecture.** In 1611, Kepler described the stacking of equal-sized spheres into the familiar arrangement we see for oranges in the grocery store. He asserted that this packing is the tightest possible. This assertion is now known as the Kepler conjecture, and has persisted for centuries without rigorous proof. Hilbert included the Kepler conjecture in his famous list of unsolved problems in 1900. In 1994, Thomas Hales, now at the University of Pittsburgh, proposed a five-step program that would result in a proof: (a) treat maps that only have triangular faces; (b) show that the face-centered cubic and hexagonal-close packings are local maxima in the strong sense that they have a higher score than any Delaunay star with the same graph; (c) treat maps that contain only triangular and quadrilateral faces (except the pentagonal prism); (d) treat maps that contain something other than a triangle or quadrilateral face; (e) treat pentagonal prisms.

In 1998, Hales announced that the program was now complete, with Samuel Ferguson (son of Helaman Ferguson) completing the crucial fifth step. This project involved extensive computation, using an interval arithmetic package, a graph generator, and *Mathematica*. The computer files containing the source code and computational results occupy more than three Gbytes of disk space. Additional details, including papers, are available at the URL <http://www.math.pitt.edu/~thales/kepler98>.

As this book was going to press, the *Annals of Mathematics* has decided to publish Hales’ paper, but with a cautionary note, because although a team of referees is “99% certain” that the computer-assisted proof is sound, they have not been able to verify every detail [185]. One wonders if every other article in this journal has implicitly been certified to be correct with more than 99% certainty.

8. **“That’s Mathematics”** in song.



1. Counting sheep  
 When you're trying to sleep,  
 Being fair  
 When there's something to share,  
 Being neat  
 When you're folding a sheet,  
 That's mathematics!

2. When a ball  
 Bounces off of a wall,  
 When you cook  
 From a recipe book,  
 When you know  
 How much money you owe,  
 That's mathematics!

3. How much gold can you hold in  
 an elephant's ear?  
 When it's noon on the moon,  
 then what time is it here?  
 If you could count for a year,  
 would you get to infinity,  
 Or somewhere in that vicinity?

4. When you choose  
 How much postage to use,  
 When you know  
 What's the chance it will snow,  
 When you bet  
 And you end up in debt,  
 Oh try as you may,  
 You just can't get away  
 From mathematics!

5. Andrew Wiles gently smiles,  
 Does his thing, and voila!  
 Q.E.D., we agree,  
 And we all shout hurrah!  
 As he confirms what Fermat  
 Jotted down in that margin,  
 Which could've used some enlargin'.

6. Tap your feet,  
 Keepin' time to a beat  
 Of a song  
 While you're singing along,  
 Harmonize  
 With the rest of the guys,  
 Yes, try as you may,  
 You just can't get away  
 From mathematics!

© Tom Lehrer 1995 (used by permission).

9. **Mathematics of Escher's "Print Galley."** In Maurits C. Escher's 1956 painting "Prententoonstelling," a young man is viewing a painting in an exhibition gallery. As his eyes follow the waterfront buildings shown in this painting around in a circle, he discovers among these buildings the very gallery he is standing in. Bart de Smit and Hendrik Lenstra have shown that the painting can be viewed as drawn on an elliptic curve over

the complex plane, and if continued would repeat itself, with each iteration reduced in size by a factor of 22.5836845286 . . . and rotated clockwise by 157.6255960832 . . . degrees [83]. This research received feature coverage in the *New York Times* [172]. Details are available at the URL <http://escherdroste.math.leidenuniv.nl>.

#### 10. Techniques for putting the Internet to work.

As a teenager in early 19th Century Britain, Michael Faraday struggled to overcome a lack of formal education by reading (among other things) self-help books that were popular at the time. It was from one such book (*Improvement of the Mind*, authored by Isaac Watts) that Faraday learned four ways to become smarter: (1) attend lectures, (2) take notes, (3) correspond with people of similar interests, (4) join a discussion group. (Scott Butner, ChemAlliance Staff)

#### 11. Gravitational boosting. “The Voyager Neptune Planetary Guide” (JPL Publication 89–24) has an excellent description of Michael Minovitch’s computational and unexpected discovery of gravitational boosting (otherwise known as slingshot magic) at the Jet Propulsion Laboratory in 1961.

The article starts by quoting Arthur C. Clarke: “Any sufficiently advanced technology is indistinguishable from magic.” Until Minovitch discovered that the so-called *Hohmann transfer ellipses* were not the minimum energy way of getting to the outer planets, “most planetary mission designers considered the gravity field of a target planet to be somewhat of a nuisance, to be cancelled out, usually by onboard Rocket thrust.” For example, without a gravitational boost from the orbits of Saturn, Jupiter and Uranus, the Earth-to-Neptune Voyager mission (achieved in 1989 in little more than a decade) would have taken more than 30 years!

#### 12. John Maynard Keynes. Two excerpts follow from *Keynes the Man*, written on the 50th anniversary of the great economist’s death, by Sir Alec Cairncross, in the *Economist*, April 20, 1996:

Keynes distrusted intellectual rigour of the Ricardian type as likely to get in the way of original thinking and saw that it

was not uncommon to hit on a valid conclusion before finding a logical path to it. . . . “I don’t really start,” he said, “until I get my proofs back from the printer. Then I can begin serious writing.”

Keynes’ undergraduate training was in mathematics at Cambridge, where he excelled as at most things he tried. He was an avid collector of rare books and manuscripts, Newton’s included. Keynes and Hardy were virtually the only scientists who intersected with the Bloomsbury group.

13. **A world of doughnuts and spheres.** As this book was going to press, the Russian mathematician Grigori Perelman was lecturing on a proof of the *Poincaré Conjecture*. His potentially ground-breaking work, if found to be valid, may earn him a share of a \$1 million prize for solving one of the Clay Mathematics Institute’s “Millennium Prize Problems.” The Clay Institute’s web site describes the Poincaré Conjecture in these terms:

If we stretch a rubber band around the surface of an apple, then we can shrink it down to a point by moving it slowly, without tearing it and without allowing it to leave the surface. On the other hand, if we imagine that the same rubber band has somehow been stretched in the appropriate direction around a doughnut, then there is no way of shrinking it to a point without breaking either the rubber band or the doughnut [Figure 1.8]. We say the surface of the apple is “simply connected,” but that the surface of the doughnut is not. Poincaré, almost a hundred years ago, knew that a two dimensional sphere is essentially characterized by this property of simple connectivity, and asked the corresponding question for the three dimensional sphere (the set of points in four dimensional space at unit distance from the origin). This question turned out to be extraordinarily difficult, and mathematicians have been struggling with it ever since.

Peter Sarnak, a well-known Princeton University mathematician, described Perelman’s work in these words: “He’s not facing Poincaré directly, he’s just trying to do [a] grander scheme.” The Poincaré result is merely “a million dollar afterthought.” Prof. Sun-Yung Alice Chang observed that

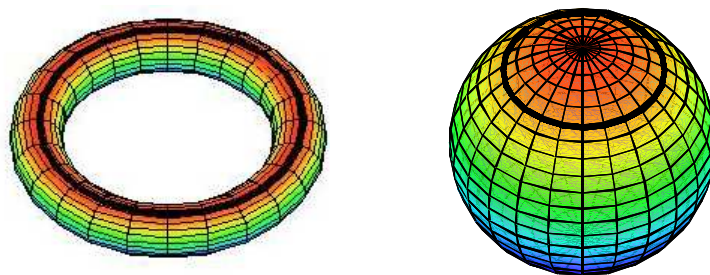


Figure 1.8: The torus and the two-sphere.

the Poincaré Conjecture is “in the same scale as Fermat’s Last Theorem. [Proving] it puts you in the history of mathematics; the dream of every mathematician” [65]. Along this line, a *New York Times* report [123] observed, “That grown men and women can make a living pondering such matters is a sign that civilization, as fragile as it may sometimes seem, remains intact.”

14. **The number partitioning problem.** Given a set of  $n$  nonnegative integers  $a_1, a_2, \dots, a_n$ , the number partitioning problem is to divide this set into two subsets such that the sums of the numbers in each subset are as nearly equal as possible. Brian Hayes calls this “the easiest hard problem” [112]. It is well known to be  $NP$ -complete. Nonetheless, some reasonably effective heuristic algorithms are known for solution. Hayes provides the following analogy for one of these:

One of the cherished customs of childhood is choosing up sides for a ball game. Where I grew up, we did it this way: The two chief bullies of the neighborhood would appoint themselves captains of the opposing teams, and then they would take turns picking other players. On each round, a captain would choose the most capable (or, toward the end, the least inept) player from the pool of remaining candidates, until everyone present had been assigned to one side or the other. The aim of this ritual was to produce two evenly matched teams and, along the way, to remind each of us of our precise ranking in the neighborhood pecking order. It usually worked.

None of us in those days—not the hopefuls waiting for our name to be called, and certainly not the two thick-necked team leaders—recognized that our scheme for choosing sides implements a greedy heuristic for the balanced number partitioning problem. And we had no idea that this problem is NP-complete—that finding the optimum team rosters is certifiably hard. We just wanted to get on with the game.

Solutions to this problem exhibit a curious experimental phenomenon: When  $n$  is smaller than the number of bits needed to encode the  $a_i$ , the computational cost grows exponentially with  $n$ . When  $n$  is larger than this value, the computational cost actually decreases, then levels off, growing only linearly with  $n$  [154].

15. **Hersh’s humanist philosophy of mathematics.** However hard mathematical paradigm shifts are to accept and whatever the result of these tides, mathematics is and will remain a uniquely human undertaking. Indeed, Reuben Hersh’s arguments for a humanist philosophy of mathematics [114, 115], as paraphrased below, become even more convincing in our experimental setting:

1. *Mathematics is human.* It is part of and fits into human culture. It does not match Frege’s concept of an abstract, timeless, tenseless, objective reality.

2. *Mathematical knowledge is fallible.* As in science, mathematics can advance by making mistakes and then correcting or even re-correcting them. The “fallibilism” of mathematics is brilliantly argued in Lakatos’ *Proofs and Refutations*.

3. *There are different versions of proof or rigor.* Standards of rigor can vary depending on time, place, and other things. The use of computers in formal proofs, exemplified by the computer-assisted proof of the four color theorem in 1977, is just one example of an emerging nontraditional standard of rigor.

4. *Empirical evidence, numerical experimentation and probabilistic proof all can help us decide what to believe in mathematics.* Aristotelian logic isn’t necessarily always the best way of deciding.

5. *Mathematical objects are a special variety of a social-cultural-historical object.* Contrary to the assertions of certain post-modern detractors, mathematics cannot be dismissed as merely a new form of literature or religion. Nevertheless, many mathematical objects can be seen as shared ideas, like Moby Dick in literature, or the Immaculate Conception in religion.

16. **Quantum Computer. n.** a radical, still largely theoretical, alternative to the VON NEUMANN ARCHITECTURE of present DIGITAL COMPUTERS. First posited by Richard Feynman in 1982, and detailed by Deutsch (1985) it relies on exploiting quantum mechanical properties of: quantization, entanglement, and especially quantum interference. Interference means *outcomes of quantum processes depend on all possible histories of a process*, and makes quantum computers exponentially more powerful than classical ones. Entanglement means *spatially separated, non-interacting, systems with prior interaction may still have locally inaccessible information in common*. This is what makes quantum encryption possible. The field exploded after 1994 when Peter Shor discovered a quantum algorithm for efficient FACTORIZATION of very large numbers. Such methods would render current CODING THEORY and encryption techniques, such as RSA, ineffective.

**qubit, n.** the QUANTUM COMPUTER analogue of a computer BIT. It models the fact that atoms can be prepared in a SUPERPOSITION of two different electronic states. Analogously, a quantum two-state system—a quantum bit or qubit—can be prepared in a superposition of its two logical states 0 and 1. Thus a qubit can simultaneously be both 0 and 1. (The above two definitions are taken from [42].)

17. **Mathematics, history and philosophy.**

The history of mathematics, lacking the guidance of philosophy, has become blind, while the philosophy of mathematics, turning its back on the most intriguing phenomena in the history of mathematics, has become empty. (Imre Lakatos, in *Proofs and Refutations* [138, 125])

18. **Nature of mathematical genius.** Mathematical genius appears to be a rare but remarkable phenomenon. Is it a genetic? An emergent accident of DNA? The result of good parenting? A stimulating intellectual environment? The product of inspiring teachers?

- (a) Recent interviews (Peter Liljedahl, SFU, unpublished PhD work) with top-level mathematicians suggests that most do not read very much, preferring to have results described to them, and put a large emphasis on the role of chance—which always favors the prepared mind.
- (b) Mathematical genius is almost always noted early in life and blossoms rather soon. As G. H. Hardy observed in his *A Mathematician's Apology*, “Mathematics, more than any other art or science, is a young man’s game. . . . I do not know an instance of a major mathematical advance initiated by a man past fifty. . . . [Newton’s] “greatest ideas of all, fluxions and the law of gravitation, came to him about 1666, when he was twenty-four.” [108, pg. 78]. Needless to say, this does not bode well for the present authors!
- (c) Archimedes, Newton, Euler, Gauss, Ramanujan, and others all seem to have had extraordinary facilities for numerical and algebraic computation. Most of us know the story of Gauss who, when his teacher asked the class to sum the integers from 1 to 100, quickly noted that this was  $50 \times 101 = 5050$ , and was the only student to obtain the correct answer! Ramanujan’s genius was his incredible skill at algebraic manipulation, a skill that flowered largely in a vacuum of advanced training in modern mathematics. The following is told of John von Neumann, who made fundamental contributions to computer science, mathematical economics, meteorology, probability theory, and quantum mechanics in the early 20th century [149, pg. 10]:

Two bicyclists are 20 miles apart and head toward each other at 10 miles per hour each. At the same time a fly traveling at a steady 15 miles per hour starts from the front wheel of the northbound bicycle. It lands on the front wheel of the southbound bicycle, and then instantly turns around and flies back, and after next landing instantly flies north again. Question: What total distance did the fly cover before it was crushed between the two front wheels?

The slow way of answering is to calculate the distance that the fly travels on its first trip to the southbound front wheel, then the distance it travels on its next trip to the northbound wheel, and finally to sum the infinite series so obtained. . . . The short way is to note that the bicycles will meet exactly an hour after starting, by which time the 15-miles-per-hour fly must have covered 15 miles. When the question was put to [John von Neumann], he danced and answered immediately, “15 miles.” “Oh, you’ve heard the trick before,” said the disappointed questioner. “What trick?” asked the puzzled Johnny. “I simply summed the infinite series.”

- (d) Others are known for their prodigious skill and boundless energy in ranging over a wide variety of very difficult topics with ease. The following is told of Nobel-prize-winning mathematical physicist Richard Feynman, during a lecture at CalTech [97, 145]:

Feynman immediately rose, astonishingly, to say that such objects would be gravitationally unstable. Furthermore, he said that the instability followed from general relativity. The claim required a calculation of the subtle countervailing effects of stellar forces and relativistic gravity. Fowler thought he was talking through his hat. A colleague later discovered that Feynman had done a hundred pages of work on the problem years before. The Chicago astrophysicist Subrahmanyan Chandrasekhar independently produced Feynman’s result—it was part of the work for which he won a Nobel Prize twenty years later. Feynman himself never bothered to publish. Someone with a new idea always risked finding, as one colleague said, “that Feynman had signed the guest book and already left.”

John Maynard Keynes, who studied the original writings of Isaac Newton while riding taxicabs between British treasury board meetings, wrote this of Newton on the tricentenary of his birth [129, 145]:

His peculiar gift was the power of holding continuously in his mind a purely mental problem until he had seen straight through it. I fancy his pre-eminence is due to his muscles of



intuition being the strongest and most enduring with which a man has ever been gifted. Anyone who has ever attempted pure scientific or philosophical thought knows how one can hold a problem momentarily in one's mind and apply all one's powers of concentration to piercing through it, and how it will dissolve and escape and you find that what you are surveying is a blank. I believe that Newton could hold a problem in his mind for hours and days and weeks until it surrendered to him its secret. Then being a supreme mathematical technician he could dress it up, how you will, for purposes of exposition, but it was his intuition which was pre-eminently extraordinary—"so happy in his conjectures," said de Morgan, "as to seem to know more than he could possibly have any means of proving."

- (e) In a recent provocative essay, David Lykken explores the emergence of genius. Here is his conclusion [145]:

Ericsson and Charness [89] are willing to acknowledge that genetic differences in temperament and "preferred activity level" may determine which of us go for the gold but, curiously, they cling to the assumption that individual genetic differences in both physical and mental capacities are not important, perhaps nonexistent. This would require us to believe that ... little Gauss's ability to correct his father's arithmetic at three and confound his school master at ten resulted, not from extraordinary mental hardware, but from mental software acquired through self-directed practice in an intellectually unstimulating environment.

Those of us who have studied MZ [monozygote, i.e., identical] twins reared apart from one another find these assumptions ... incredible. We cannot believe that MZA twins [monozygote twins reared apart] correlate .75 in IQ merely because, in their separate environments, their similarities in temperament led them to indulge in very similar amounts of practice on very similar topics. ...

I think we must agree with Ericsson, however, that works of

genius tend to be the product of minds enriched by years of concentrated effort. Isaac Newton often became so caught up in cerebration that he would forget to eat or sleep. Edwin Land, inventor of the instant Polaroid camera and of a sophisticated computational theory of color vision, sometimes worked at his desk for 36 hours or more, unaware of the passage of time until he felt faint on standing up. Similar stories were told of Edison. It does not follow, however, that these were ordinary minds to begin with.

Edison, Feynman, Land, and Newton all from their boyhood had intense curiosity, an enthusiasm or zeal for discovery and understanding. Each of them was able to take seriously hypotheses that others thought to be implausible (or had not thought about at all). All four possessed a kind of intellectual arrogance that permitted them to essay prodigious tasks, to undertake to solve problems that most of their contemporaries believed to be impossible. And each of them had quite extraordinary powers of concentration. . . .

I think what lies at the heart of these mysteries is genetic, probably emergenic. The configuration of traits of intellect, mental energy, and temperament with which, during the plague years of 1665–6, Isaac Newton revolutionized the world of science were, I believe, the consequence of a genetic lottery that occurred about nine months prior to his birth, on Christmas day, in 1642.



## Chapter 2

# Experimental Mathematics in Action

The purpose of computing is insight, not numbers.

Richard Hamming, *Numerical Methods for Scientists and Engineers*, 1962

In this chapter, we will present a few particularly engaging examples of modern experimental mathematics in action. We invite those readers with access to some of the computational tools we mention below to personally try some of these examples—with some sense of realism and with a caution not to emulate Winnie the Pooh’s heuristics in Figure 2.1 too closely.

### 2.1 Pascal’s Triangle

Although Blaise Pascal (1623–1662) is commonly credited with what is now known as Pascal’s triangle (namely the array of integers such that row  $m$  and column  $n$  has the value  $\binom{m}{n}$ , typically shown centered on the page), this well-known numerical scheme is actually much older. The earliest written account may be in a 1303 Chinese book by Chu Shih-Chieh, entitled *Precious Mirror of the Four Elements*, but even here, Pascal’s triangle is called “The Old Method.” Thus it is at least 700 years old, and probably centuries older still. Given its simplicity and antiquity, it is natural to expect that every conceivable aspect of



Figure 2.1: How not to perform experiments: guess and check.

this venerable scheme has been discovered and studied at length. But even in such arenas, diligent researchers can sometimes glean new insights.

Start with the familiar binomial coefficient,

$$\binom{m}{n} = \frac{m!}{n!(m-n)!}, \quad (2.1.1)$$

and consider the remainders on division by 2. That is, set

$$F_2(m, n) = \binom{m}{n} \bmod 2. \quad (2.1.2)$$

Let  $N_2(m)$  be the number of 1s in the set  $(F_2(m, n), 0 \leq n \leq m)$ .

By means of an algebraic argument, it can be shown that  $N_2(m)$  is given by

$$N_2(m) = 2^{B_1(m)}, \quad (2.1.3)$$

where  $B_1(m)$  is the number of 1s in the binary representation of  $m$  [102].

But there is another proof of this fact that demonstrates the power of visualization in experimental mathematics.

Pascal's triangle has long been used as a tool to visualize the behavior of the binomial coefficient function. For example, the combinatorial identity  $\binom{m}{n} = \binom{m-1}{n-1} + \binom{m-1}{n}$  is equivalent to the Pascal triangle rule that the value of entry  $(m, n)$  is merely the sum of the entries to the left and right of it, on the previous row. As another example, the fact that binomial coefficients are greatest when  $n = \lfloor m/2 \rfloor$  is evident from visual inspection—centrally located cells have had the greatest benefit of the combining add operation in the rows above, and this visual observation can indeed be fashioned into a rigorous proof.

Consider now a modified Pascal triangle where entries are reduced modulo 2, as shown in Figure 2.2. Such a figure was easy to generate using the Internet-based tool available at <http://www.cecm.sfu.ca/organics/papers/granville/support/pascalform.html>. It is evident from the left-hand image in Figure 2.2, which comes from the interface, that the entries of this Pascal triangle form a Sierpinski triangle, which in the limit is a self-similar figure akin to those studied by Mandelbrot [151]. In fact, it can be seen that this figure exhibits the fractional dimension  $\log_2 3 = 1.58496\dots$  [192].

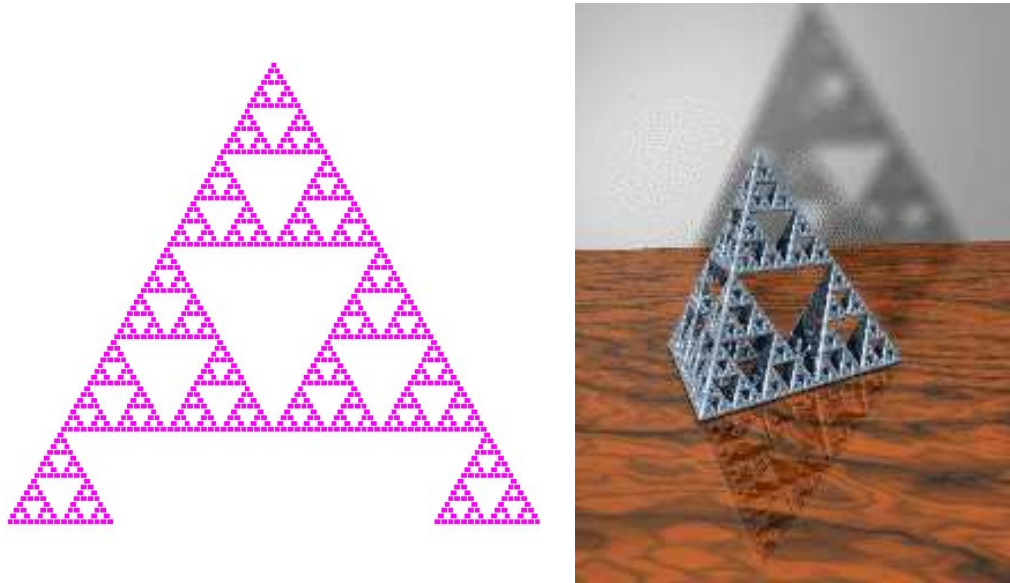


Figure 2.2: Pascal triangles modulo two?

The right-hand figure shows a computer realization of this process on a tetrahedron, due to Darrell Hepting. Now let's take a look at a proof of identity (2.1.3) in the style of experimental mathematics, a proof that we argue is significantly easier than the usual algebraic proof of this result.

Given a positive integer  $m$ , construct a Pascal triangle modulo 2 of size  $2^{\lceil \log_2 m \rceil}$ . In this way the  $m$ -th row corresponds to a line that crosses the lower half of the largest upright triangle. Note that successive binary digits of  $m$  specify whether the line crosses the upper (0) or the lower (1) half of successively smaller upright triangles. The upper halves always contain one half-sized upright triangle, whereas the lower halves contain two such triangles. The total number of triangles crossed by the line is thus multiplied by a factor of two whenever the lower half is crossed. The total number of ones in the  $m$ -th row is thus the product of the factors of two associated with each one in the binary expansion of  $m$ . This proves (2.1.3). See Stephen Wolfram's article [192] and the Crandall-Pomerance book [78] (research problem 8.22) for additional details and some generalizations.

We should add here that Wolfram has recently published his long-awaited tome *A New Kind of Science*, an exhaustive study of cellular automata with

connections to numerous other fields of science. Wolfram’s approach in this book is almost exclusively experimental, employing nearly 1000 high-resolution graphical examples to highlight his observations, building up to his “Principle of Computational Equivalence.” In this regard, one could view Pascal’s triangle modulo 2 as merely a particular instance of a one-dimensional cellular automata [193].

We might note in passing that the Sierpinski triangle has a habit of appearing in quite a few unexpected places in mathematics. One remarkable example, in quite a different context, is the following. Construct an “arbitrary” triangle, with vertices  $(x_1, y_1)$ ,  $(x_2, y_2)$  and  $(x_3, y_3)$ . Specify a starting point  $(x, y)$  within the triangle. Then iterate the following construction: First select a random integer  $r$  in the set  $(1, 2, 3)$ , and then construct a new point  $(x', y')$  as follows:

$$(x', y') = \left( \frac{1}{2}(x + x_r), \frac{1}{2}(y + y_r) \right). \quad (2.1.4)$$

This procedure is then iterated indefinitely.

If you carry out this procedure on a computer with a high-resolution display, it is immediately evident that, no matter what triangle is specified, and no matter what initial point is selected, the resulting figure (after possibly the first few points) is a Sierpinski triangle contained in the original triangle—see, for example, Figure 2.3. In other words, the Sierpinski triangle is the “attractor” set for this iteration (by the way, it is easier to “see” this fact than to prove it). The study of these fractal objects and the schemes used to generate them is by itself a fascinating example of experimental mathematics in action. See, for example, Michael Barnsley’s interesting book, *Fractals Everywhere* [30].

## 2.2 A Curious Anomaly in the Gregory Series

In 1988, Joseph Roy North of Colorado Springs observed that Gregory’s series for  $\pi$ ,

$$\pi = 4 \sum_{k=1}^{\infty} \frac{(-1)^{k+1}}{2k-1} = 4(1 - 1/3 + 1/5 - 1/7 + \cdots), \quad (2.2.5)$$

when truncated to 5,000,000 terms, gives a value that differs strangely from the true value of  $\pi$ . Here is the truncated Gregory value and the true value of  $\pi$ :



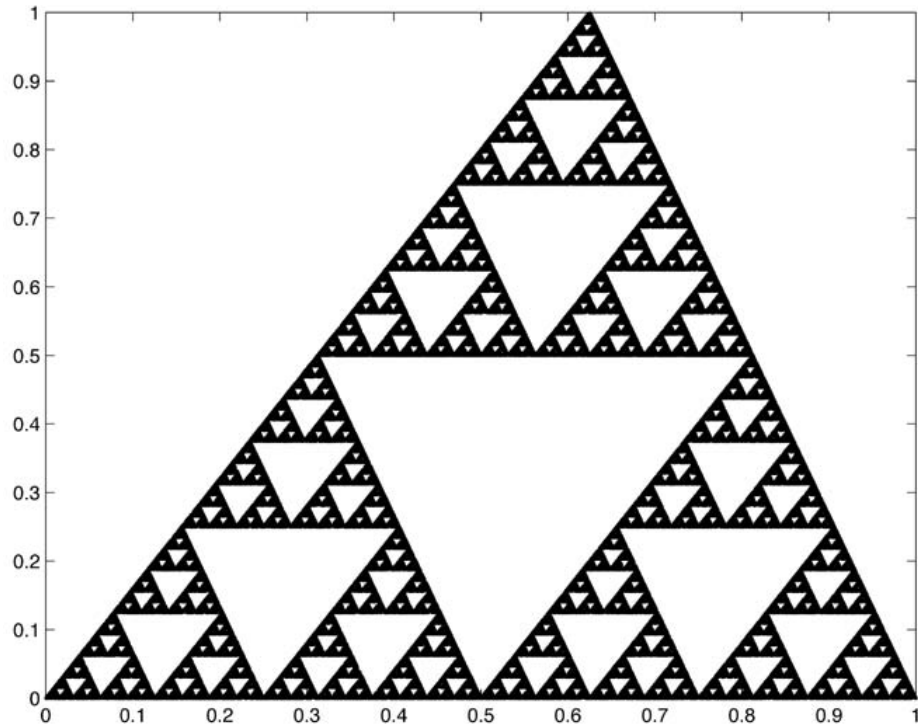


Figure 2.3: Sierpinski triangle by random selections.

```

3.14159245358979323846464338327950278419716939938730582097494182230781640...
3.14159265358979323846264338327950288419716939937510582097494459230781640...
      2             -2             10             -122             2770

```

The series value differs, as one might expect from a series truncated to 5,000,000 terms, in the seventh decimal place—a “4” where there should be a “6.” But the next 13 digits are correct! Then, following another erroneous digit, the sequence is once again correct for an additional 12 digits. In fact, of the first 46 digits, only four differ from the corresponding decimal digits of  $\pi$ . Further, the “error” digits appear to occur in positions that have a period of 14, as shown above. Such anomalous behavior begs explanation.

Once observed, it is natural (and easy given a modern computer algebra system) to ask if something similar happens with the logarithm. Indeed it does, as the following value obtained by truncating the series  $\log 2 = 1 - 1/2 + 1/3 - 1/4 + \dots$  shows:

0.69314708055995530941723212125817656807551613436025525140068000949418722...  
 0.69314718055994530941723212145817656807550013436025525412068000949339362...  
           1      -1                  2                  -16                  272                  -7936

Here again, the “erroneous” digits appear in locations with a period of 14. In the first case, the differences from the “correct” values are (2, -2, 10, -122, 2770), while in the second case the differences are (1, -1, 2, -16, 272, -7936). We note that each integer in the first set is even; dividing by two, we obtain (1, -1, 5, -122, 1385).

How can we find out exactly what is going on here? A great place to start is by enlisting the help of an excellent resource for the computational mathematician: Neil Sloane and Simon Plouffe’s Internet-based integer sequence recognition tool, available at <http://www.research.att.com/~njas/sequences>. This tool has no difficulty recognizing the first sequence as “Euler numbers” and the second as “tangent numbers.” Euler numbers and tangent numbers are defined in terms of the Taylor’s series for  $\sec x$  and  $\tan x$ , respectively:

$$\begin{aligned}\sec x &= \sum_{k=0}^{\infty} \frac{(-1)^k E_{2k} x^{2k}}{(2k)!} \\ \tan x &= \sum_{k=0}^{\infty} \frac{(-1)^{k+1} T_{2k+1} x^{2k+1}}{(2k+1)!}.\end{aligned}\tag{2.2.6}$$

Indeed, this discovery, made originally through the print version of the sequence recognition tool available more than a decade ago, led to a formal proof that these sequences are indeed the source of the “errors” in these sequences. The precise result is that the following asymptotic expansions hold:

$$\frac{\pi}{2} - 2 \sum_{k=1}^{N/2} \frac{(-1)^{k+1}}{2k-1} \approx \sum_{m=0}^{\infty} \frac{E_{2m}}{N^{2m+1}}\tag{2.2.7}$$

$$\log 2 - \sum_{k=1}^{N/2} \frac{(-1)^{k+1}}{k} \approx \frac{1}{N} + \sum_{m=1}^{\infty} \frac{T_{2m-1}}{N^{2m}}.\tag{2.2.8}$$

Now the genesis of the anomaly mentioned above is clear: North, in computing  $\pi$  by Gregory’s series, had by chance truncated the series at 5,000,000 terms, which is exactly one-half of a fairly large power of ten. Indeed, setting  $N = 10,000,000$

in Equation (2.2.7) shows that the first hundred or so digits of the truncated series value are small perturbations of the correct decimal expansion for  $\pi$ . And the asymptotic expansions show up on the computer screen, as we observed above. Similar phenomena occur for other constants. (See [47] for proofs of (2.2.7) and (2.2.8), together with some additional details.)

## 2.3 Bifurcation Points in the Logistic Iteration

Of course, our earlier remarks about pure mathematicians being slow to take advantage of the computer do have some exceptions. In particular, the entire area of dynamical systems (sometimes referred to as chaos theory) owes its very existence to modern information technology. And one of the classic examples of a chaotic iteration is known as the *logistic* iteration: Fix a real number  $r > 0$ , select  $x_0$  in the unit interval  $(0, 1)$ , and then iterate

$$x_{k+1} = rx_k(1 - x_k). \quad (2.3.9)$$

This is termed the “logistic” iteration because of its roots in computational ecology: It mimics the behavior of a biological population, which, if it becomes too numerous, exhausts its available food supply and then falls back to a smaller population, possibly oscillating in an irregular manner over many generations.

For values of  $r < 1$ , the iterates  $(x_k)$  quickly converge to zero. For  $1 < r < 3$ , the iterates converge to a single nonzero limit point. At  $r = 3$ , a bifurcation occurs: For  $3 < r < 3.449489\dots = 1 + \sqrt{6}$ , the iterates oscillate between two distinct limit points. A second bifurcation occurs at  $r = 1 + \sqrt{6}$ . In particular, for  $1 + \sqrt{6} < r < 3.544090359\dots$ , the iterates oscillate in a periodic fashion between four distinct limit points. This pattern of limit point bifurcation and period doubling occurs at successively shorter intervals, until  $r > 3.5699457\dots$ , when iterates behave in a completely chaotic manner. This behavior is shown in Figure 2.4.

Until recently, the identity of the third bifurcation point, namely the constant  $b_3 = 3.544090359\dots$ , was not known. It is fairly straightforward, by means of recursive substitutions of Equation (2.3.9), to demonstrate that this constant must be algebraic, but the bound on the degree of the integer polynomial that  $b_3$  satisfies is quite large and thus not very useful.

A tool that can be used in such situations is an *integer relation algorithm*. This is an algorithm which, when given  $n$  real numbers  $(x_1, x_2, \dots, x_n)$ , returns

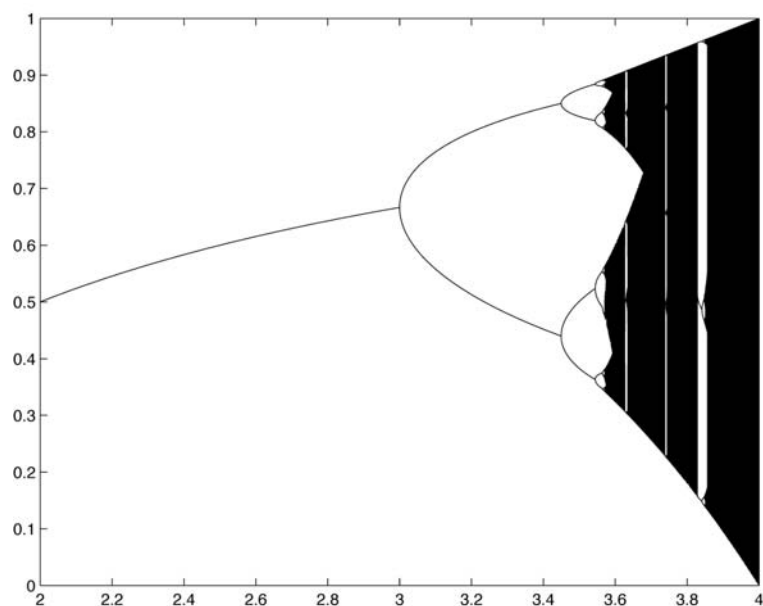


Figure 2.4: Bifurcation in the logistic iteration.

integers  $(a_1, a_2, \dots, a_n)$ , not all zero, such that  $a_1x_1 + a_2x_2 + \dots + a_nx_n = 0$  (if such a solution exists). Such computations must be done using very high precision arithmetic, or else the results are not numerically significant. At present the best algorithm for integer relation detection appears to be the “PSLQ” algorithm of mathematician-sculptor Helaman Ferguson [92, 15, 23], although the “LLL” algorithm is also often used. We will discuss integer relation detection in greater depth in Section 6.3.1. For the time being we mention the Internet-based integer relation tool at <http://www.cecm.sfu.ca/projects/IntegerRelations> and the Experimental Mathematician’s Toolkit at <http://www.expmath.info>.

One straightforward application of an integer relation tool is to recover the polynomial satisfied by an algebraic number. If you suspect that a constant  $\alpha$ , whose numerical value can be calculated to high precision, is algebraic of degree  $n$ , then you can test this conjecture by computing the  $(n + 1)$ -long vector  $(1, \alpha, \alpha^2, \dots, \alpha^n)$ , and then using this vector as input to an integer relation calculation. If it finds a solution vector  $(a_0, a_1, a_2, \dots, a_n)$  with a sufficiently high degree of numerical accuracy, then you can be fairly confident that these integers are precisely the coefficients of the polynomial satisfied by  $\alpha$ .

In the present example, where  $\alpha = b_3$ , a predecessor algorithm to PSLQ recovered the polynomial

$$0 = 4913 + 2108t^2 - 604t^3 - 977t^4 + 8t^5 + 44t^6 + 392t^7 - 193t^8 - 40t^9 + 48t^{10} - 12t^{11} + t^{12}. \quad (2.3.10)$$

You might like to try to rediscover this polynomial by using the Internet-based tool mentioned above. To do this requires a high-precision value of  $b_3$ . Its value correct to 120 decimal digits is:

3.5440903595 5192285361 5965986604 8045405830 9984544457 3675457812  
2530305842 9428588630 1225625856 6424891799 9626089927 7589974545

If you do not wish to type this number in, you may find it by using *Mathematica*:

```
FindRoot[4913 + 2108*t^2 - 604*t^3 - 977*t^4 + 8*t^5 +
44*t^6 + 392*t^7 - 193*t^8 - 40*t^9 + 48*t^10 - 12*t^11 +
t^12 == 0, {t, 3.544}, WorkingPrecision -> 125]
```

or by using a similar command with the Experimental Mathematician's Toolkit.

Recently, the fourth bifurcation point  $b_4 = 3.564407266095\dots$  was identified by a similar, but much more challenging, integer relation calculation. In particular, it was found that  $\alpha = -b_4(b_4 - 2)$  satisfies a certain integer polynomial of degree 120. The recovered coefficients descend monotonically from  $257^{30} \approx 1.986 \times 10^{72}$  down to 1. This calculation required 10,000 decimal digit precision arithmetic, and more than one hour on 48 processors of a parallel computer system. Full details can be found in [23].

We add in passing that the limiting ratio of the intervals between successive bifurcation points in the logistic iteration is  $4.669201609\dots$ . This is known as Feigenbaum's constant, named for Mitchell Feigenbaum who first observed, as a result of computational experiments, that this limiting ratio holds for a wide class of chaotic iterations, not just the logistic iteration. A fascinating account of this discovery is given in [96]. Researchers have subsequently made numerous attempts to recognize Feigenbaum's constant as an algebraic number, or as a simple formula involving other well-known constants of mathematics, but these efforts have failed so far. An algorithm for calculating Feigenbaum's constant to high precision is given in [60].

## 2.4 Experimental Mathematics and Sculpture

In the previous section, we mentioned the PSLQ algorithm, which was discovered in 1993 by Helaman Ferguson. This is certainly a signal accomplishment—for example, the PSLQ algorithm (with associated lattice reduction algorithms) was recently named one of ten “algorithms of the century” by *Computing in Science and Engineering* [15]. Nonetheless Ferguson is even more well-known for his numerous mathematics-inspired sculptures, which grace numerous research institutes in the United States. Photos and highly readable explanations of these sculptures can be seen in a lovely book written by his wife, Claire [91]. Together, the Fergusons recently won the 2002 Communications Award, bestowed by the Joint Policy Board of Mathematics. The citation for this award declares that the Fergusons “have dazzled the mathematical community and a far wider public with exquisite sculptures embodying mathematical ideas, along with artful and accessible essays and lectures elucidating the mathematical concepts.”

Ferguson notes that the PSLQ algorithm can be thought of as a  $n$ -dimension extension of the Euclidean algorithm, and is, like the Euclidean scheme, fundamentally a “subtractive” algorithm. As Ferguson explains, “It is also true that my sculptural form of expression is subtractive: I get my mathematical forms by direct carving of stone.” [167]

There is a remarkable and unanticipated connection between Ferguson’s PSLQ algorithm and at least one of Ferguson’s sculptures. It is known that the volumes of complements of certain knot figures (which volumes in  $\mathbb{R}^3$  are infinite) are finite in hyperbolic space, and sometimes are given by certain explicit formulas. This is not true of all knots. Many of these hyperbolic complements of knots correspond to certain discrete quotient subgroups of matrix groups.

One of Ferguson’s sculptures, known as the “Eight-Fold Way,” is housed at the Mathematical Sciences Research Institute in Berkeley, California (see Figure 2.5 and item 33 at the end of this chapter).

Another of Ferguson’s well-known sculptures is the “Figure-Eight Complement II” (see Figure 2.6 and Item 33 at the end of this chapter). It has been known for some time that the hyperbolic volume  $V$  of the figure-eight knot



Figure 2.5: Ferguson's "Eight-Fold Way" sculpture.



Figure 2.6: Ferguson's "Figure-Eight Knot Complement" sculpture.



complement is given by the formula

$$V = 2\sqrt{3} \sum_{n=1}^{\infty} \frac{1}{n \binom{2n}{n}} \sum_{k=n}^{2n-1} \frac{1}{k} \quad (2.4.11)$$

$$= 2.029883212819307250042405108549 \dots \quad (2.4.12)$$

In 1998, British physicist David Broadhurst conjectured that  $V/\sqrt{3}$  is a rational linear combination of

$$C_j = \sum_{n=0}^{\infty} \frac{(-1)^n}{27^n (6n+j)^2}. \quad (2.4.13)$$

Indeed, it is, as Broadhurst [61] found using a PSLQ program:

$$V = \frac{\sqrt{3}}{9} \sum_{n=0}^{\infty} \frac{(-1)^n}{27^n} \left( \frac{18}{(6n+1)^2} - \frac{18}{(6n+2)^2} - \frac{24}{(6n+3)^2} - \frac{6}{(6n+4)^2} + \frac{2}{(6n+5)^2} \right). \quad (2.4.14)$$

You can verify this yourself, using for example the Mathematician's Toolkit, available at <http://www.expmath.info>. Just type the following lines of code:

```
v = 2 * sqrt[3] * sum[1/(n * binomial[2*n,n]) * sum[1/k, \
  {k, n, 2*n-1}], {n, 1, infinity}]
pslq[v/sqrt[3], table[sum[(-1)^n/(27^n*(6*n+j)^2), \
  {n, 0, infinity}], {j, 1, 6}]]
```

When this is done you will recover the solution vector  $(9, -18, 18, 24, 6, -2, 0)$ . A proof that formula (2.4.14) holds, together with a number of other identities for  $V$ , is given at the end of this chapter in Items 34 and 35. As you will see, this proof is a classic example of a proof from experimental methodology, in that it relies on “knowing” ahead of time that the formula holds.

As we shall see in Section 3.4, constants given by a formula of the general type given in (2.4.14), namely a “BBP-type” formula, possess some remarkable properties, among them the fact that you can calculate the  $n$ -th digit (base-3 digit in this case) of such constants by means of a simple algorithm, without having to compute any of the first  $n - 1$  digits.

Ferguson comments that the discovery of this BBP-type expression for  $V$  is a “major advance toward understanding the figure-eight knot complement volume.” Accordingly, he has carved Broadhurst’s formula on the figure-eight knot complement sculptures commissioned by the Clay Mathematics Institute, both the Inner Mongolian black granite piece and the smaller bronzes (the Clay Math Award pieces). As he explains, “Finally the subtractive sculpture and the subtractive algorithm have come together.”

## 2.5 Recognition of Euler Sums

In April 1993, Enrico Au-Yeung, an undergraduate at the University of Waterloo, brought to the attention of one of us (Borwein) the curious result [43]

$$\begin{aligned} \sum_{k=1}^{\infty} \left(1 + \frac{1}{2} + \cdots + \frac{1}{k}\right)^2 k^{-2} &= 4.59987\dots \\ &\approx \frac{17}{4}\zeta(4) = \frac{17\pi^4}{360}. \end{aligned} \quad (2.5.15)$$

The function  $\zeta(s)$  in (2.5.15) is the classical *Riemann zeta function*,

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

Bernoulli showed that for even integers,  $\zeta(2n)$  is a rational multiple of  $\pi^{2n}$  [50]. (Bernoulli’s result is proved in Section 3.2 of the second volume of this work.)

Au-Yeung had computed the sum in (2.5.15) to 500,000 terms, giving an accuracy of 5 or 6 decimal digits. Suspecting that his discovery was merely a numerical coincidence, Borwein sought to compute the sum to a higher level of precision. Using Fourier analysis and Parseval’s equation, he obtained

$$\frac{1}{2\pi} \int_0^{\pi} (\pi - t)^2 \log^2\left(2 \sin \frac{t}{2}\right) dt = \sum_{n=1}^{\infty} \frac{(\sum_{k=1}^n \frac{1}{k})^2}{(n+1)^2}. \quad (2.5.16)$$

The idea here is that the series on the right of (2.5.16) permits one to evaluate (2.5.15), while the integral on the left can be computed using the numerical quadrature facility of *Mathematica* or *Maple*. When he did this, he was surprised

to find that the conjectured identity holds to more than 30 digits. We should add here that by good fortune,  $17/360 = 0.047222\dots$  has period one and thus can plausibly be recognized from its first six digits, so that Au-Yeung's numerical discovery was not entirely far-fetched.

What Borwein did not know at the time was that Au-Yeung's suspected identity follows directly from a related result proved by De Doelder in 1991 [86]. In fact, it had cropped up even earlier as a problem in the *American Mathematical Monthly*, but the story goes back further still. Some historical research showed that Euler considered these summations. In response to a letter from Goldbach, he examined sums that are equivalent to

$$\sum_{k=1}^{\infty} \left( 1 + \frac{1}{2^m} + \dots + \frac{1}{k^m} \right) (k+1)^{-n}. \quad (2.5.17)$$

The great Swiss mathematician was able to give explicit values for certain of these sums in terms of the Riemann zeta function. For example, he found an explicit formula for the case  $m = 1$ ,  $n \geq 2$ .

In retrospect, perhaps it was for the better that Borwein had not known of De Doelder's and Euler's results, because Au-Yeung's intriguing numerical discovery launched a fruitful line of research by a number of researchers that continued until nearly the present day. Sums of this general form are nowadays known as "Euler sums" or "Euler-Zagier sums."

In order to explore them more rigorously, it has been necessary to develop an efficient means to calculate their value to high precision—namely the 200 or more digit accuracy needed to obtain numerically significant results using integer relation calculations. Along this line, one of us (Bailey [20]) found a satisfactory scheme involving the Euler-Maclaurin summation formula. Another, more powerful scheme is described in [45]. These techniques will be discussed in Section 7.5 of the second volume of this work.

High precision calculations of many of these sums, together with considerable investigations involving heavy use of *Maple's* symbolic manipulation facilities, eventually yielded numerous new results.

Below are just a few of the interesting results that were first discovered numerically and have since been established analytically [44]. Since these results were first obtained in 1994, many more specific identities have been discovered, and a growing body of general formulas and other results have been proven.

These results, together with the underlying numerical and symbolic techniques used in their derivation, are discussed further in Chapter 3 of the second volume.

$$\begin{aligned}
\sum_{k=1}^{\infty} \left(1 + \frac{1}{2} + \cdots + \frac{1}{k}\right)^2 (k+1)^{-4} &= \frac{37}{22680} \pi^6 - \zeta^2(3) \\
\sum_{k=1}^{\infty} \left(1 + \frac{1}{2} + \cdots + \frac{1}{k}\right)^3 (k+1)^{-6} &= \\
&\zeta^3(3) + \frac{197}{24} \zeta(9) + \frac{1}{2} \pi^2 \zeta(7) - \frac{11}{120} \pi^4 \zeta(5) - \frac{37}{7560} \pi^6 \zeta(3) \\
\sum_{k=1}^{\infty} \left(1 - \frac{1}{2} + \cdots + (-1)^{k+1} \frac{1}{k}\right)^2 (k+1)^{-3} &= \\
&4 \operatorname{Li}_5\left(\frac{1}{2}\right) - \frac{1}{30} \log^5(2) - \frac{17}{32} \zeta(5) - \frac{11}{720} \pi^4 \log(2) + \frac{7}{4} \zeta(3) \log^2(2) \\
&+ \frac{1}{18} \pi^2 \log^3(2) - \frac{1}{8} \pi^2 \zeta(3), \tag{2.5.18}
\end{aligned}$$

where  $\operatorname{Li}_n(x) = \sum_{k>0} x^k/k^n$  denotes the polylogarithm function.

## 2.6 Quantum Field Theory

In another recent development, David Broadhurst (who discovered the identity (2.4.14) for Ferguson’s Clay Math Award sculpture) has found, using similar methods, that there is an intimate connection between Euler sums and constants resulting from evaluation of Feynman diagrams in quantum field theory [63, 64]. In particular, the renormalization procedure (which removes infinities from the perturbation expansion) involves multiple zeta values, which we will discuss in detail in Chapter 3 of the second volume.

Broadhurst’s recent results are even more remarkable. He has shown [61], using PSLQ computations, that in each of ten cases with unit or zero mass, the finite part of the scalar 3-loop tetrahedral vacuum Feynman diagram reduces to four-letter “words” that represent iterated integrals in an alphabet of seven “letters” comprising the single 1-form  $\Omega = dx/x$  and the six 1-forms  $\omega_k = dx/(\lambda^{-k} - x)$ , where  $\lambda = (1 + \sqrt{-3})/2$  is the primitive sixth root of unity, and  $k$

runs from 0 to 5. A four-letter word here is a four-dimensional iterated integral, such as

$$U = \zeta(\Omega^2\omega_3\omega_0) = \int_0^1 \frac{dx_1}{x_1} \int_0^{x_1} \frac{dx_2}{x_2} \int_0^{x_2} \frac{dx_3}{(-1-x_3)} \int_0^{x_3} \frac{dx_4}{(1-x_4)} = \sum_{j>k>0} \frac{(-1)^{j+k}}{j^3k}.$$

There are  $7^4$  such four-letter words. Only two of these are primitive terms occurring in the 3-loop Feynman diagrams:  $U$ , above, and

$$V = \operatorname{Re}[\zeta(\Omega^2\omega_3\omega_1)] = \sum_{j>k>0} \frac{(-1)^j \cos(2\pi k/3)}{j^3k}.$$

The remaining terms in the diagrams reduce to products of constants found in Feynman diagrams with fewer loops. These ten cases are shown in Figure 2.7. In these diagrams, dots indicate particles with nonzero rest mass. The formulas that have been found, using PSLQ, for the corresponding constants are given in Table 2.1. In the Table the constant  $C = \sum_{k>0} \sin(\pi k/3)/k^2$ .

## 2.7 Definite Integrals and Infinite Series

We mention here one particularly useful application of experimental mathematics methodology: evaluating definite integrals and sums of infinite series by means of numerical calculations. In one sense, there is nothing new here, since mathematicians have utilized computers to compute the approximate numerical value of definite integrals and infinite series since the dawn of computing. What we suggest here, however, is a slightly different approach: Use advanced numerical quadrature techniques and series summations methods, extended to the realm of high-precision arithmetic, and then use the computed values (typically accurate to tens or even hundreds of decimal digits) as input to a computer-based constant recognition tool, which hopefully can recognize the constant as a simple expression involving known mathematical constants.

We will discuss techniques for computing definite integrals and sums of series to high precision in Section 7.4 of the second volume of this work. For the time being, we simply note that both *Mathematica* and *Maple* have incorporated some

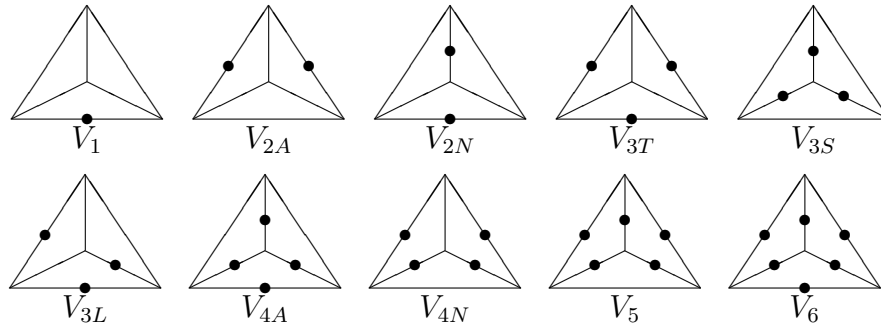


Figure 2.7: The ten tetrahedral configurations.

$V_1$	$= 6\zeta(3) + 3\zeta(4)$
$V_{2A}$	$= 6\zeta(3) - 5\zeta(4)$
$V_{2N}$	$= 6\zeta(3) - \frac{13}{2}\zeta(4) - 8U$
$V_{3T}$	$= 6\zeta(3) - 9\zeta(4)$
$V_{3S}$	$= 6\zeta(3) - \frac{11}{2}\zeta(4) - 4C^2$
$V_{3L}$	$= 6\zeta(3) - \frac{15}{4}\zeta(4) - 6C^2$
$V_{4A}$	$= 6\zeta(3) - \frac{77}{12}\zeta(4) - 6C^2$
$V_{4N}$	$= 6\zeta(3) - 14\zeta(4) - 16U$
$V_5$	$= 6\zeta(3) - \frac{469}{27}\zeta(4) + \frac{8}{3}C^2 - 16V$
$V_6$	$= 6\zeta(3) - 13\zeta(4) - 8U - 4C^2$

Table 2.1: Formulas found by PSLQ for the ten tetrahedral diagrams.

reasonably good numerical facilities for this purpose, and it is often sufficient to rely on these packages when numerical values are needed.

For our first example, we use *Maple* or *Mathematica* to compute the following three integrals to over 100 decimal digit accuracy:

$$\int_0^1 \frac{t^2 \log(t) dt}{(t^2 - 1)(t^4 + 1)} =$$

0.180671262590654942792308128981671615337114571018296766266  
240794293758566224133001770898254150483799707740 ...

$$\int_0^{\pi/4} \frac{t^2 dt}{\sin^2(t)} =$$

0.843511841685034634002620051999528151651689086421444293697  
112596906587355669239938399327915596371348023976 ...

$$\int_0^\pi \frac{x \sin x dx}{1 + \cos^2 x} =$$

2.467401100272339654708622749969037783828424851810197656603  
337344055011205604801310750443350929638057956006 ... (2.7.19)

(the third of these is from [106]). Both *Maple* and *Mathematica* attempt to evaluate these definite integrals analytically. In each case, however, while the results appear to be technically correct, they are not very useful, in that they are either rather lengthy, or involve advanced functions and complex entities. We suspect that there are considerably simpler closed-form versions.

Indeed, using the Inverse Symbolic Calculator (ISC) tool (a constant recognition facility) at <http://www.cecm.sfu.ca/projects/ISC>, we obtain the following, based solely on the numerical values above:

$$\begin{aligned} \int_0^1 \frac{t^2 \log(t) dt}{(t^2 - 1)(t^4 + 1)} &= \frac{\pi^2(2 - \sqrt{2})}{32} \\ \int_0^{\pi/4} \frac{t^2 dt}{\sin^2(t)} &= -\frac{\pi^2}{16} + \frac{\pi \log(2)}{4} + G \\ \int_0^\pi \frac{x \sin x dx}{1 + \cos^2 x} &= \frac{\pi^2}{4}, \end{aligned} \quad (2.7.20)$$

where  $G$  denotes Catalan's constant

$$G = \sum_{n=0}^{\infty} \frac{(-1)^n}{(2n+1)^2}.$$

We might add that Catalan's constant is widely believed to be irrational, but this has never been proved. Using an integer relation tool, together with a high-precision numerical value of Catalan (which can easily be found by typing `N[Catalan, 100]` in *Mathematica*), you can see that it is not a root of an integer polynomial with reasonable degree and reasonable-sized coefficients.

As an exercise, we challenge you to evaluate

$$\int_0^{\pi/4} \frac{t^3 dt}{\sin^2(t)}.$$

Other examples of numerical identification of definite integrals will be presented in Section 7.4 of the second volume of this work.

As a second example, as we shall see in Section 3.1, rational linear combinations of arctangent formulas can be devised to produce relatively efficient algorithms for computing  $\pi$ . In fact, such formulas have been used for centuries. One approach to finding such formulas is to analytically derive them. Another approach is to simply explore for them using the numerical values of individual arctan formulas. For instance, by computing values of the individual arctans below to moderately high precision, and applying an integer relation tool, one can easily deduce the relations

$$\begin{aligned} \pi &= 48 \arctan \frac{1}{49} + 128 \arctan \frac{1}{57} - 20 \arctan \frac{1}{239} + 48 \arctan \frac{1}{110443} \\ \pi &= 176 \arctan \frac{1}{57} + 28 \arctan \frac{1}{239} - 48 \arctan \frac{1}{682} + 96 \arctan \frac{1}{12943}. \end{aligned} \tag{2.7.21}$$

One way to do this is by using the "Experimental Mathematician's Toolkit," which is available at <http://www.expmath.info>. Type the string

```
pslq[pi, arctan[1/49], arctan[1/57], arctan[1/239],
arctan[1/110443]]
```



and you will obtain the vector  $(1, -48, -128, 20, -48)$ , which gives the first formula of (2.7.21). These particular formulas were used by Kanada in his latest computation of  $\pi$  to more than one trillion decimal digits.

A third example, one of the most remarkable examples of constant recognition in the literature, is due to Ronald Graham and Henry Pollak [101]. They were studying the sequences  $(a_n)$  and  $(b_n)$ , defined by  $a_0 = m$ , then

$$a_{n+1} = \lfloor \sqrt{2a_n(a_n + 1)} \rfloor \quad (2.7.22)$$

followed by

$$b_n = a_{2n+1} - 2a_{2n-1}. \quad (2.7.23)$$

Can the sequence  $(a_n)$  and/or  $(b_n)$  be identified? Graham and Pollak found that if you define the constants  $\alpha(m)$ , for integers  $m \geq 1$ ,

$$\alpha(m) = 0.b_1b_2b_3 \cdots_2, \quad (2.7.24)$$

meaning that the sequence  $(b_n)$  is to be interpreted as the binary expansion of the constant  $\alpha(m)$ , then the resulting constants are simple algebraic numbers. In particular,

$$\begin{aligned} \alpha(1) &= \sqrt{2} - 1 \\ \alpha(2) &= \sqrt{2} - 1 \\ \alpha(3) &= 2\sqrt{2} - 2 \\ \alpha(4) &= 2\sqrt{2} - 2 \\ \alpha(5) &= 3\sqrt{2} - 4 \\ \alpha(6) &= 4\sqrt{2} - 5 \\ \alpha(7) &= 3\sqrt{2} - 4 \\ \alpha(8) &= 5\sqrt{2} - 7 \\ \alpha(9) &= 4\sqrt{2} - 5 \\ \alpha(10) &= 6\sqrt{2} - 8. \end{aligned}$$

This recognition led to an explicit formula for the sequence  $(a_n)$  as

$$a_n = \lfloor \tau(2^{(n-1)/2} + 2^{(n-1)/2}) \rfloor, \quad (2.7.25)$$

where  $\tau$  is the  $m$ -th smallest real number in the set  $\{1, 2, 3, \dots\} \cup \{\sqrt{2}, 2\sqrt{2}, 3\sqrt{2}, \dots\}$ . See [101] for complete details. It is not known if generalized sequences such as

$$a_{n+1} = \lfloor \sqrt{3a_n(a_n + 1)} \rfloor \quad (2.7.26)$$

$$a_{n+1} = \lfloor \sqrt[3]{2a_n(a_n + 1)(a_n + 2)} \rfloor \quad (2.7.27)$$

have analogous properties.

We will discuss constant recognition techniques more in Section 6.3.

## 2.8 Prime Numbers and the Zeta Function

The oldest systematic treatment of prime numbers is in Euclid's *Elements*, written about 300 BCE. Euclid's results on prime numbers, like his results on geometry, are presented entirely as theorems and proofs, with scarcely any motivating examples or mention of how these results were first discovered. Thus it is not possible to learn to what extent experimental methods were used in the original discovery of these principles. In more recent times, we have written accounts of the discoveries of various prime number results, and we can find several clear examples of experimental mathematics in action. As a canonical example, in 1849 Gauss recounted his analysis of the density of prime numbers:

I pondered this problem as a boy, in 1792 or 1793, and found that the density of primes around  $t$  is  $1/\log t$ , so that the number of primes up to a given bound  $x$  is approximately  $\int_2^x dt/\log t$ . [103, 99]

Formal proof that the above approximation is asymptotically correct, which is now known as the Prime Number Theorem, did not come until 1896, more than 100 years after Gauss' experimental discovery [78]. In his writings on the subject, Gauss expressed a fascination with prime numbers that is reminiscent of the fascination other mathematicians have expressed with the digits of  $\pi$ :

The problem of distinguishing prime numbers from composite numbers and of resolving the latter into their prime factors is known to be one of the most important and useful in arithmetic. It has engaged the industry and wisdom of ancient and modern geometers to

such an extent that it would be superfluous to discuss the problem at length. . . . Further, the dignity of the science itself seems to require that every possible means be explored for the solution of a problem so elegant and so celebrated. [95, Article 329 (1801)]

This fascination with prime numbers has not diminished in the 201 years since this was written. Compare for example the comments of Don Zagier, in a recent article analyzing the first 50 million prime numbers:

[T]here is no apparent reason why one number is prime and another not. To the contrary, upon looking at these numbers one has the feeling of being in the presence of one of the inexplicable secrets of creation. [194]

In recent years, with the advent of computers, the prime-counting function  $\pi(x)$  has been tabulated precisely for rather large values of  $x$ . This has been done by using some rather advanced techniques that we will not describe here. Interested readers are referred to the new book, *Prime Numbers: A Computational Perspective*, by Richard Crandall and Carl Pomerance [78], where two different advanced techniques are presented. Table 2.2 gives values of  $\pi(x)$  for powers of ten up to  $10^{22}$ , together with the corresponding values of the approximate formula given by the Prime Number Theorem, and the differences  $\int_2^x dt/\log t - \pi(x)$ . The last-listed and most recent calculation is due to P. Demichel and X. Gourdon (February 2001). This data was obtained from Eric Weisstein's very useful "World of Mathematics" web site <http://mathworld.wolfram.com>.

One of the principal motivations of computing values of the prime-counting function  $\pi(x)$  is to explore a fascinating and highly promising connection to the Riemann zeta function, a discovery which is a classic example of an experimental approach to mathematical research. The following material is adapted in part from an excellent and very readable article by Andrew Granville [103].

In examining the data in Table 2.2, we observe that the overcount values in the last column appear to be roughly  $\sqrt{x}$ . Thus one might think that by subtracting  $\sqrt{x}$ , one would get an asymptotically better approximation. This is not the case—in 1914, Littlewood proved that the differences between  $\pi(x)$  and Gauss' approximation change sign infinitely often. In 1933, Skewes showed that the first sign change must occur at least by  $10^{10^{34}}$ , a number known as the *Skewes number* because of its prodigious size (see [78], Exercise 1.35). This

$x$	$\pi(x)$	$\int_2^x dt/\log t$	Difference
$10^1$	4	5	1
$10^2$	25	29	4
$10^3$	168	177	9
$10^4$	1229	1245	16
$10^5$	9592	9629	37
$10^6$	78498	78627	129
$10^7$	6 64579	6 64917	338
$10^8$	57 61455	57 62208	753
$10^9$	508 47534	508 49234	1700
$10^{10}$	4550 52511	4550 55614	3103
$10^{11}$	41180 54813	41180 66400	11587
$10^{12}$	3 76079 12018	3 76079 50280	38262
$10^{13}$	34 60655 36839	34 60656 45809	1 08970
$10^{14}$	320 49417 50802	320 49420 65691	3 14889
$10^{15}$	2984 45704 22669	2984 45714 75287	10 52618
$10^{16}$	27923 83410 33925	27923 83442 48556	32 14631
$10^{17}$	2 62355 71576 54233	2 62355 71656 10821	79 56588
$10^{18}$	24 73995 42877 40860	24 73995 43096 90414	219 49554
$10^{19}$	234 05766 72763 44607	234 05766 73762 22381	998 77774
$10^{20}$	2220 81960 25609 18840	2220 81960 27836 63483	2227 44643
$10^{21}$	21127 26948 60187 31928	21127 26948 66161 26181	5973 94253
$10^{22}$	2 01467 28668 93159 06290	2 01467 28669 12482 61497	19323 55207

Table 2.2: The prime-counting function  $\pi(x)$  and Gauss' approximation.

bound has recently been reduced to a “tiny” number, a mere  $10^{316}$  [103]. Littlewood [143, pg. 110–112] gives a fascinating discussion of his interaction with Skewes.

A key development in this area stems from Riemann’s discovery in 1859 that

$$\frac{\pi(x) - \int_2^x dt/\log t}{\sqrt{x}/\log x} \approx -1 - 2 \sum_{\gamma \in S} \frac{\sin(\gamma \log x)}{\gamma}, \quad (2.8.28)$$

where  $S$  is the set of  $\gamma$  such that  $\gamma > 0$  and  $1/2 + i\gamma$  is a zero of the Riemann zeta function  $\zeta(s) = \sum_{k>0} 1/k^s$ . The left-hand side of (2.8.28) is simply the overcount noted in Table 2.2, divided by what we earlier observed to be the apparent size of the error term, namely  $\sqrt{x}/\log x$  (a more accurate version of the approximation  $\sqrt{x}$ ). The right-hand side of (2.8.28) is suggestive of a Fourier series—it is as if the prime numbers have “music” in them.

In Riemann’s 1859 memoir, in which he presented this approximation formula (2.8.28), he noted that the result was contingent upon (actually equivalent to) the hypothesis that the only zeroes of  $\zeta(s)$  for complex numbers  $s = \sigma + i\gamma$  with  $0 \leq \sigma \leq 1$  are for  $\sigma = 1/2$  [171]. This is the famous Riemann Hypothesis, whose proof has eluded the best mathematicians for nearly 150 years. Photocopies of the first two pages of Riemann’s memoir on the zeta function, stored in Göttingen, are shown in Figures 2.8 and 2.9.

Riemann’s 1859 memoir did not contain any clues as to how he was led to make this conjecture. For many years mathematicians believed that Riemann had come to this conjecture on the basis of some profound intuition. Indeed, the Riemann Hypothesis has been held up as a premier example of the heights one could attain by sheer intellect alone. In 1929, long after Riemann’s death, the renowned number theorist Carl Ludwig Siegel (1896–1981) learned that Riemann’s widow had donated his working papers to the Göttingen University library. Among these papers Siegel found several pages of heavy-duty numerical calculations, with a number of the lowest-order zeroes of the zeta function calculated to several decimal places each. As Andrew Granville has observed, “So much for pure thought alone” [103]. The Riemann-Siegel formula is developed further in Item 2 of Chapter 3 in the second volume.

In 1936 Cramer proposed a probabilistic model of the primes. In particular, he defined a sequence  $X_3, X_4, \dots$  of independent random variables on  $\{0, 1\}$  with  $P(X_n = 1) = 1/\log n$ , and  $P(X_n = 0) = 1 - 1/\log n$  (here,  $P$  denotes

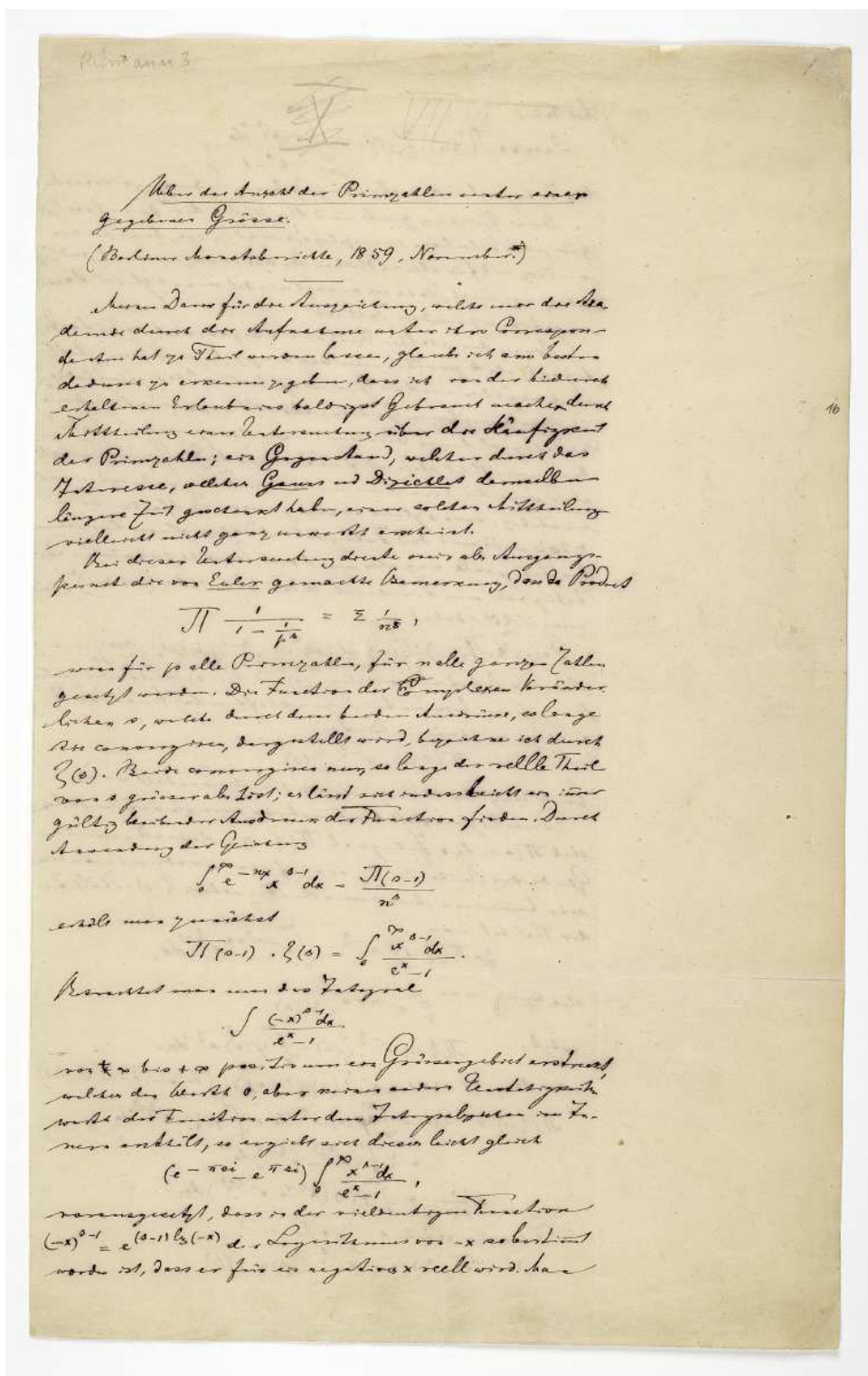


Figure 2.8: Page one of Riemann's 1859 manuscript.

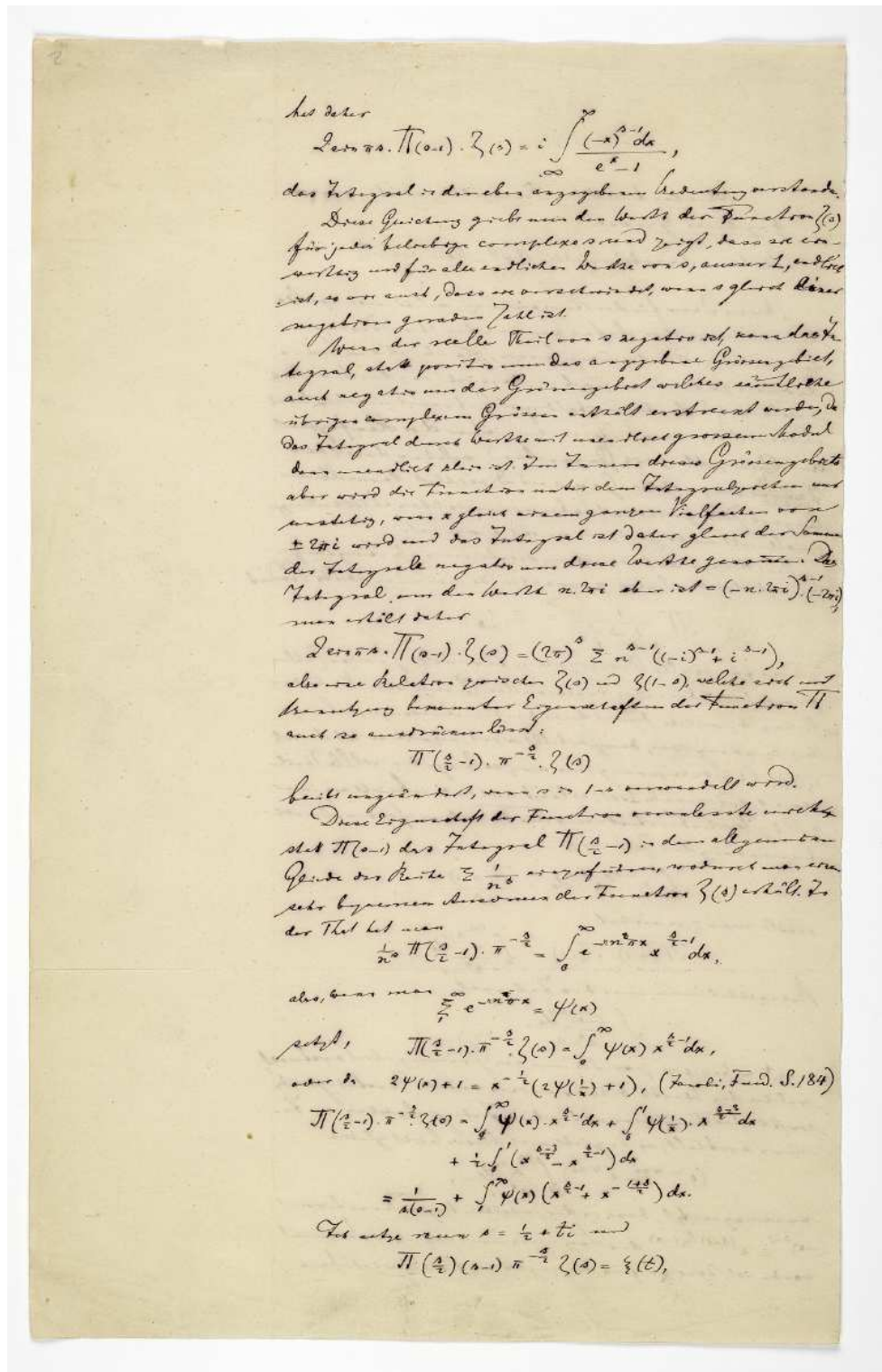


Figure 2.9: Page two of Riemann's 1859 manuscript.

probability). This turns out to be a fairly good model for occurrence of prime numbers. In fact, it can be shown that

$$\lim_{n \rightarrow \infty} \frac{\sum_{x \leq n \leq x+h} X_n}{\int_x^{x+h} dt / \log t} = 1 \quad (2.8.29)$$

with probability 1. In such comparisons of statistical distributions, it is natural to consider the variance,

$$V = E \left( \sum_{x \leq n \leq x+h} X_n - \int_x^{x+h} \frac{dt}{\log t} \right)^2. \quad (2.8.30)$$

(here,  $E$  denotes expected value). Unfortunately, it can be shown that this variance is *not* the same as the corresponding statistical variance of the prime counting function  $\pi(x)$ . In Figure 2.10, we show 100 zeroes of the zeta function, beginning with the  $(10^{21} + 1)$ -th zero, displayed as the dotted line, with pseudorandom numbers (added and scaled to the endpoints of the graph) displayed as the dashed line for comparison. The ranks of the zeroes, minus  $(10^{21} + 1)$ , are plotted on the  $x$ -axis, while the imaginary parts of the zeroes, minus the imaginary part of the  $(10^{21} + 1)$ -th zero (namely  $144176897509546973538.498 \dots$ ), are plotted on the  $y$  axis. It is clear from this plot that the zeroes of the Riemann zeta function are hardly “randomly” spaced—they adhere to a straight line much more closely than the pseudorandom data. It is as if the zeroes of the Riemann zeta function consciously “repel” each other, so as to stay much closer to constant spacing.

Given how well Gauss’ model has worked in the past, this breakdown was both unexpected and disconcerting. As Paul Erdős wrote, “God may not play dice with the universe, but something strange is going on with the primes.” [147]

In 1976, Julia Mueller revisited the question of the variance of the prime counts in her PhD dissertation. Building on her work, Daniel Goldston and Hugh Montgomery soon made the remarkable discovery that a good understanding of the variance is actually equivalent to the spacing between pairs of zeroes of the zeta function. Three years earlier, while investigating a different question of number theory, Montgomery had been led to the following conjecture: The expected number of zeroes of the zeta function in a gap of length  $T$  times the



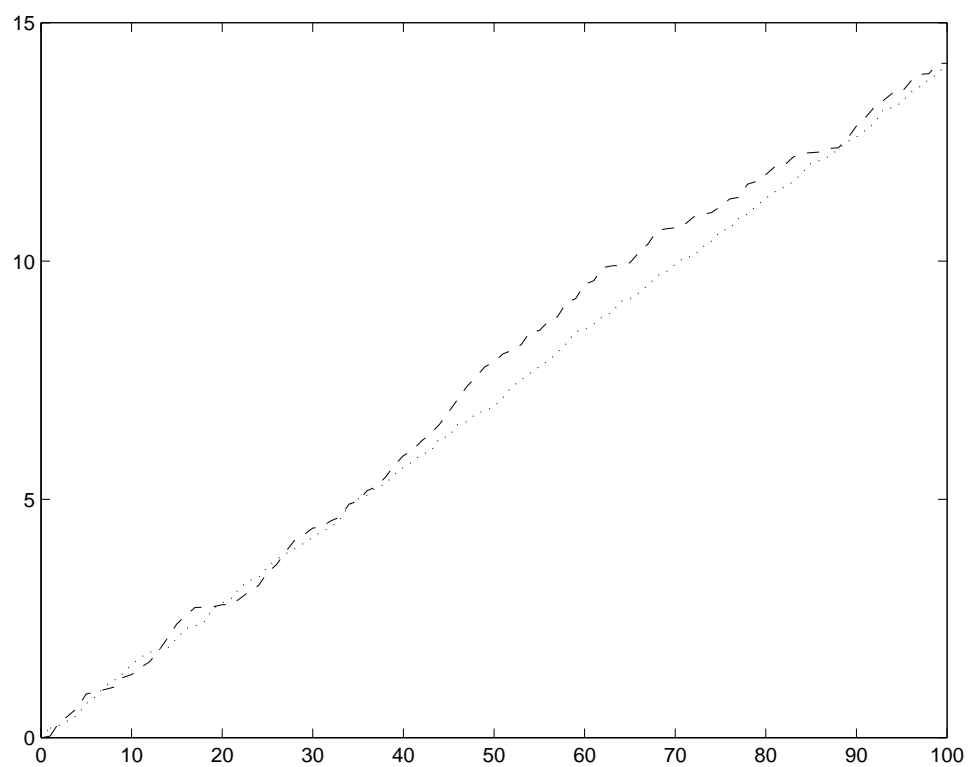


Figure 2.10:  $\zeta$  function zeroes (dotted) vs pseudo-random data (dashed)

average gap, following a zero, is

$$\int_0^T \left( 1 - \left( \frac{\sin \pi u}{\pi u} \right)^2 \right) du. \quad (2.8.31)$$

If the zeroes of the zeta function were randomly spaced, this value would simply be  $T$ . This conjecture agrees with the observed repulsion of zeroes, as observed in Figure 2.10. For example, we would expect to find another zero within 0.01 times the average spacing roughly one time in 100 trials for random spacing, whereas this formula predicts that such close spacing would occur only one time in nearly 1,000,000 instances. We can compare how closely this data matches with actual nearest-neighbor spacing for the zeta function, by examining calculated data by Andrew Odlyzko [158]. Figure 2.11 shows this spacing data based on a billion zeroes near the  $1.3 \times 10^{16}$ -th zero.

As you can see from this graph, this function agrees extremely well with the experimentally measured spacings. Montgomery has now partially proven that his conjecture is correct—he has shown that the Fourier transform of his distribution function is correct in a small range, assuming the Riemann Hypothesis.

The next step in this saga is even more remarkable. Soon after his experimental discovery on the prime counts and zero spacing, Montgomery visited the Institute for Advanced Study in Princeton, hoping to discuss his new ideas with Selberg and Bombieri, who were both at the Institute at the time. In a communal afternoon tea, Montgomery happened to chat with Freeman Dyson, the renowned mathematical physicist (although he originally was a number theorist). When Montgomery told Dyson of his new ideas on the zeroes of the Riemann zeta function, Dyson responded that he was very familiar with this problem, and in fact was even aware of the issue of the spacing of the zeroes. But it wasn't from number theory that Dyson was aware of these phenomena; instead, Dyson had encountered these questions through quantum mechanics. In fact, Montgomery's distribution was precisely the function that Dyson had found ten years earlier when modeling energy levels in quantum chaos. These researchers immediately realized that this serendipitous coincidence would likely lead to more progress on both the number theory and quantum physics fronts.

The quantum chaos equations are comparatively simpler than those of prime number theory, and so they were better understood. As a result, there was a flurry of new activity in the field. Researchers focused not only on close pairs of

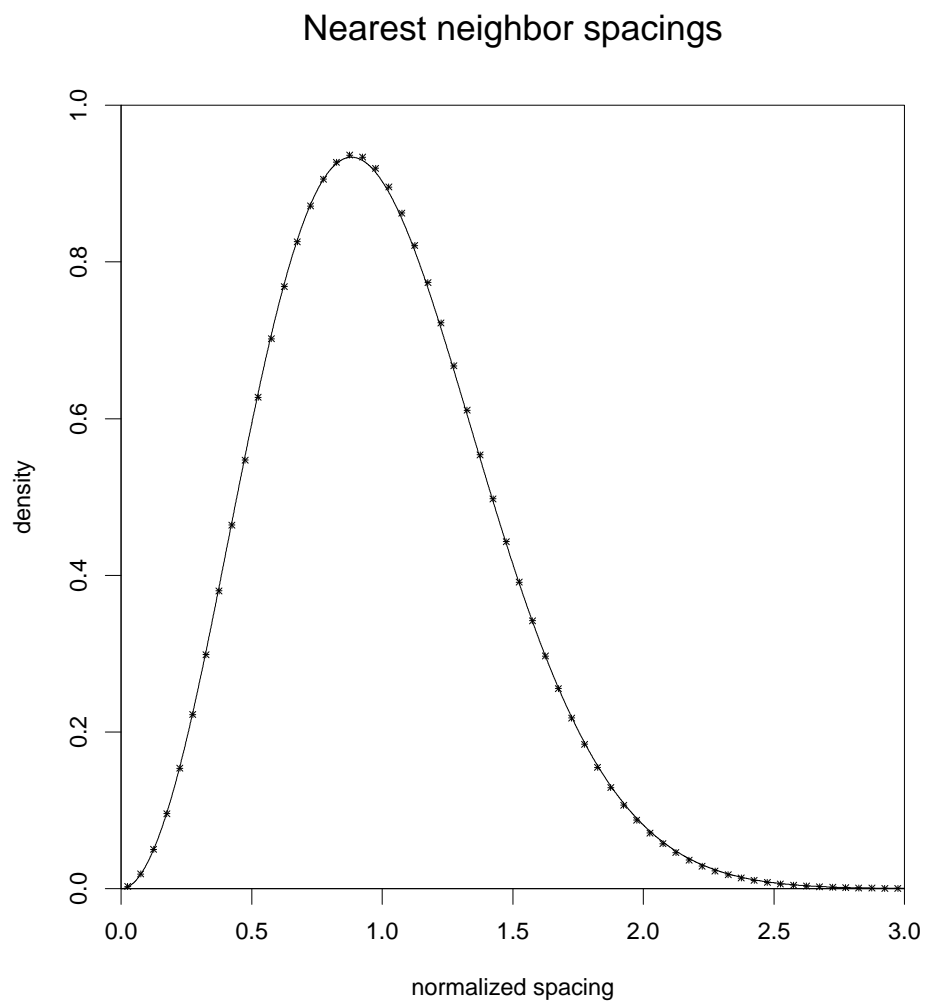


Figure 2.11: Nearest neighbor spacing versus Montgomery's distribution.

zeroes, but also three-at-a-time correlations and  $n$ -at-a-time correlations. However, it was 20 years later (1994) before Rudnick and Sarnak proved the quantum mechanics analogue of Montgomery's result (that the Fourier transform of the distribution function is correct in a small range, assuming the Riemann Hypothesis) [174]. Then in 1996, two mathematical physicists (Bogolmony and Keating) rederived the Montgomery-Dyson prediction for  $n$ -at-a-time correlations from a new angle. They took a classic conjecture of number theory, namely the Hardy-Littlewood version of the prime  $k$ -tuplets conjecture, and showed that this also led to a similar conclusion [40]. Since then there has been a flurry of new research in this area, both on the number theory and mathematical physics fronts. These developments have led some in the field to optimistically predict that the long-sought proof of the Riemann Hypothesis may be at hand. As Michael Berry said in 2000,

I have a feeling that the Riemann Hypothesis will be cracked in the next few years. I see the strands coming together. Someone will soon get the million dollars. [130]

The million dollars Berry mentions here refers to a prize that has been offered for the proof of the Riemann Hypothesis (and for several other outstanding mathematical problems) by the Clay Mathematics Institute. Details of this prize are available on the Clay Institute web site <http://www.claymath.org>. A very readable recent survey is to be found in [73].

## 2.9 Two Observations about $\sqrt{2}$

As we observed earlier in the discussion of Pascal's triangle, it is still possible to find new insights even in the oldest areas of mathematics. We illustrate this principle again with two intriguing observations on that oldest of irrational constants, the square root of two.

### 2.9.1 Irrationality

We first present Tom Apostol's lovely new geometric proof of the irrationality of  $\sqrt{2}$  [6].

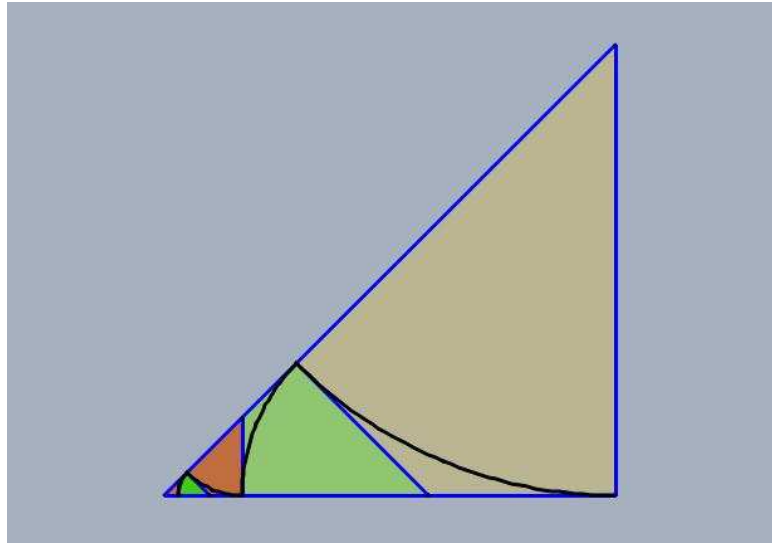


Figure 2.12: Root two is irrational.

**Proof.** To say  $\sqrt{2}$  is rational is to draw a right-angled isosceles triangle with integer sides. Consider the *smallest* right-angled isosceles triangle with integer sides—that is with shortest hypotenuse.

Suppose ABC is that triangle with apex A and hypotenuse AC. Draw a circle centered at A of length AB and mark the point D where it cuts the hypotenuse. Now draw the tangent to the circle at D and mark the point E where it cuts the base of the triangle. Then DEC is a *smaller* isosceles right-angled triangle and again has integer sides (see Figure 2.12). This is a contradiction.  $\square$

As Figure 2.12 suggests, this can be beautifully illustrated in a dynamic geometry package such as *Geometer's Sketchpad* or *Cinderella*. We can continue to draw smaller and smaller integer-sided similar triangles until the area drops below  $\frac{1}{2}$ . But we give it here to emphasize the ineffably human component of the best proofs, and to suggest the role of the visual.

Algebraically this leads to the following proof that nonperfect squares have irrational roots: If  $p/q = \sqrt{n}$ , then

$$\frac{p'}{q'} = \frac{nq - \lfloor \sqrt{n} \rfloor p}{p - \lfloor \sqrt{n} \rfloor q} = \frac{p}{q} \quad (2.9.32)$$

and  $0 < p' < p$ ,  $0 < q' < q$ . It is left to the reader to convert this into a picture along the lines of Figure 2.12.

### 2.9.2 Rationality

Ironically,  $\sqrt{2}$  can sometimes make things rational:

$$\left(\sqrt{2}^{\sqrt{2}}\right)^{\sqrt{2}} = \sqrt{2}^{(\sqrt{2}\cdot\sqrt{2})} = \sqrt{2}^2 = 2. \quad (2.9.33)$$

Hence, by the *principle of the excluded middle*,

$$\text{either } \sqrt{2}^{\sqrt{2}} \in \mathbf{Q} \text{ or } \sqrt{2}^{\sqrt{2}} \notin \mathbf{Q}.$$

In either case, we can deduce that there are irrational numbers  $\alpha$  and  $\beta$  with  $\alpha^\beta$  rational. But how do we know which ones? This is not an adequate proof for an Intuitionist nor a Constructivist. It is entirely ineffective, and we may build a whole mathematical philosophy project around such issues. Actually,  $\sqrt{2}^{\sqrt{2}}$  is transcendental by the Gelfond-Schneider theorem (Exercise 28), but proofs of this are hard and usually suffer from the same flaws.

It is instructive to compare this result with the assertion that  $\alpha = \sqrt{2}$  and  $\beta = 2 \log_2(3)$  yield  $\alpha^\beta = 3$  as *Mathematica* confirms. This illustrates nicely that verification is often easier than discovery. Similarly, the fact that multiplication is easier than factorization is at the base of secure encryption schemes for e-commerce.

Indeed, there are eight possible rational/irrational triples:  $\alpha^\beta = \gamma$ ; finding examples of all cases is now a good exercise (Exercise 29). Note how much can be taught about computation with rational numbers, approximation to irrationals, rates of convergence, etc. from these simple pieces.

We close the section with a description of the meeting between the inventor of logarithms (John Napier) and the scientist who made them into technology (Henry Briggs) [188]:

[W]here almost one quarter hour was spent, each beholding the other with admiration before one word was spoken: at last Mr. Briggs began “My Lord, I have undertaken this long journey purposely to see your person, and to know by what wit or ingenuity you first

came to think of this most excellent help unto Astronomy, viz. the Logarithms: but my Lord, being by you found out, I wonder nobody else found it out before, when now being known it appears so easy.”

## 2.10 Commentary and Additional Examples

1. **The hardest possible proof?** Use Fermat’s last theorem to prove  $2^{1/n}$  is irrational for integer  $n > 2$ . Generalize.
2. **The final digit of a sum.** Problem: Determine the final digit  $\ell_n$  of  $\sigma_n = \sum_{k=1}^n k$ . (Taken from [106]).  
Solution: Computational experimentation shows the pattern repeats modulo 20;  $\ell_{n+20} = \ell_n \pmod{20}$  is easily proven from  $\sigma_{n+20} = \sigma_n + \sum_{k=1}^{n+20} k$ .
3. **The  $3x+1$  problem.** This is a classic example of an innocent looking, but highly intractable problem:

For integer  $x$ , let  $T(x) = (3x + 1)/2$  for  $x$  odd and  $x/2$  for  $x$  even. The  $3x + 1$  conjecture is that starting from any positive integer  $n$ , repeated iteration of  $T$  eventually returns to 1.

This problem is best described in the interactive article by Jeff Lagarias at <http://www.cecm.sfu.ca/organics/papers/lagarias> (see also [137]), with records stored at <http://www.ieeta.pt/~tos/3x+1.html>. This conjecture has been “checked” to at least  $100 \cdot 2^{50}$ .

4. **Limit of a simple iteration.** Establish the limit of the iteration that starts with  $a_0 = 0, a_1 = 1/2$  and iterates  $a_{n+1} = (1 + a_n + a_{n-1}^3)/3$ , for  $n > 1$ . Determine what happens as  $a_1 = a$  is allowed to vary.
5. **Putnam problem 1985–B5.** Evaluate

$$\mathcal{K} = \int_0^\infty t^{-1/2} e^{-1985(t+t^{-1})} dt.$$

Answer:  $\mathcal{K} = \sqrt{\pi} e^{-3970} / \sqrt{1985}$ . The Putnam problems listed here and in subsequent chapters are taken from [128]. Hint: This is problematic to

evaluate numerically as stated, since its value is tiny. So consider instead  $\mathcal{K}(\alpha) = \int_0^\infty t^{-1/2} e^{-\alpha(t+t^{-1})} dt$  for some other specific (or general) constant  $\alpha$ .

6. **Putnam problem 1987–A6.** Let  $n$  be a positive integer and let  $a_3(n)$  be the number of zeroes in the ternary expansion of  $n$ . Determine for which positive  $x$  the series  $\sum_{n=1}^\infty x^{a_3(n)}/n^3$  converges. Answer: For  $x < 25$ . In the  $b$ -ary analogue,  $\sum_{n=1}^\infty x^{a_b(n)}/n^b$  converges if and only if  $x < b^b - b + 1$ .

7. **Putnam problem 1987–B1.** Evaluate

$$\int_2^4 \frac{\sqrt{\log(9-x)}}{\sqrt{\log(9-x)} + \sqrt{\log(3+x)}} dx \quad (= 1).$$

8. **Putnam problem 1991–A5.** Find the supremum of

$$\int_0^y \sqrt{x^4 + (y - y^2)^2} dx,$$

for  $0 \leq y \leq 1$ . Hint: Plot the function.

9. **Putnam problem 1992–A2.** Evaluate

$$\int_0^1 C(-y-1) \sum_{k=1}^{1992} \frac{1}{y+k} dy,$$

where  $C(\alpha)$  is the coefficient of  $x^{1992}$  in the Maclaurin expansion of  $(1+x)^\alpha$ . Answer: 1992.

10. **Putnam problem 1992–B3.** Consider the dynamical system generated by  $a_0 = x$  and

$$a_{n+1} = \frac{y^2 + a_n^2}{2},$$

for  $n \geq 0$ . Determine the region in the plane for which the iteration converges. What is its area? Hint: Try computing some values and plot the results. Assuming without loss of generality  $x, y > 0$ , the limit must satisfy  $2\ell - \ell^2 = y^2$ . Thus the region defined is the convex hull of unit circles centered at  $(\pm 1, 0)$ . Answer:  $\pi + 4$ .



11. **Random projections.** Consider a arbitrary point inside a triangle. Determine what happens asymptotically when the point is projected to successive sides of the triangle, where the side is selected either in cyclical order or pseudo-randomly. Hint: Consider first what happens in an obtuse triangle.
12. **Putnam problem 1995–B4.** Determine a simple expression for

$$\sigma = \sqrt[8]{\frac{2207 - \frac{1}{2207 - \frac{1}{2207 - \frac{1}{2207 - \dots}}}}{2207 - \frac{1}{2207 - \frac{1}{2207 - \dots}}}}. \quad (2.10.34)$$

Hint: Calculate this limit to 15 decimal place accuracy, using ordinary double-precision arithmetic. Then use the ISC tool, with the “integer relation algorithm” option, to recognize the constant as a simple algebraic number. The result can be proved by noting that  $\sigma^8 = 2207 - 1/\sigma^8$ , so that  $\sigma^4 + \sigma^{-4} = 47$ . Answer:  $(3 + \sqrt{5})/2$ .

13. **Berkeley problem 1.1.35.** Find the derivative at  $x = 0$  of

$$\int_{\sin(x)}^{\cos(x)} e^{t^2+xt} dt.$$

Hint: Plot it. Answer =  $(e - 3)/2$ . The Berkeley problems listed here and in subsequent chapters are taken from [84].

14. **Berkeley problem 7.6.6.** Compute  $A^{10^6}$  and  $A^{-7}$  for  $A = \begin{bmatrix} 3/2 & 1/2 \\ -1/2 & 1/2 \end{bmatrix}$ .  
Hint: Use *Mathematica* or *Maple* to evaluate  $A^n$  for various integers  $n$ .

15. **Two radical expressions.** (From [113, pg. 81, 84]). Express

$$\sqrt[3]{\cos\left(\frac{2}{7}\pi\right)} + \sqrt[3]{\cos\left(\frac{4}{7}\pi\right)} + \sqrt[3]{\cos\left(\frac{6}{7}\pi\right)}$$

$$\sqrt[3]{\cos\left(\frac{2}{9}\pi\right)} + \sqrt[3]{\cos\left(\frac{4}{9}\pi\right)} + \sqrt[3]{\cos\left(\frac{8}{9}\pi\right)}$$

as radicals. Hint: Calculate to high precision, then use the ISC tool to find the polynomial they satisfy.

Answers:  $\sqrt[3]{\frac{1}{2}(5 - 3\sqrt[3]{7})}$  and  $\sqrt[3]{\frac{3}{2}\sqrt[3]{9} - 3}$ .

16. **Some simple continued fractions.** Compute the simple continued fraction for  $e$ ,  $(e - 1)/2$ ,  $e^2$ ,  $\log(2)$ ,  $\log(10)$ ,  $3^{1/2}$ ,  $2^{1/3}$ ,  $\pi$ ,  $\pi/2$ ,  $e^\pi$ , and  $\pi^e$ .
17. **Crandall's continued fraction.** Compute, then guess and prove, the continued fraction for

$$\sqrt{2} \frac{e^{\sqrt{2}} + 1}{e^{\sqrt{2}} - 1},$$

which fraction is manifestly *not* periodic, proving in this way that  $e^{\sqrt{2}}$  is irrational. Richard Crandall informs us that this (in 1968) was the first and perhaps only interesting thing he ever proved all by himself.

18. **Putnam problem 1988–B2.** Prove or disprove: If  $x$  and  $y$  are real numbers with  $y \geq 0$  and  $y(y + 1) \leq (x + 1)^2$ , then  $y(y - 1) \leq x^2$ . Hint: Plot it.
19. **Putnam problem 1989–A3.** Show that all roots of

$$11z^{10} + 10iz^9 + 10iz - 11 = 0$$

lie on the unit circle. Hint: This can be solved explicitly using *Maple*, *Mathematica*, or a custom-written root-finding program that employs Newton iterations. A detailed discussion of polynomial root-finding techniques can be found in Section 7.3 of the second volume.

20. **Putnam problem 1992–A3.** For a given positive integer  $m$  find all triples  $(n, x, y)$  of positive integers with  $(n, m) = 1$  solving  $(x^2 + y^2)^m = (xy)^n$ . Hint: Using a symbolic math program, try finding solutions for various integer pairs  $x, y$ . Answer: The only solution is  $(m + 1, 2^{m/2}, 2^{m/2})$  for  $m$  even.
21. **Berkeley problem 6.11.5.** Prove that  $\sqrt{2} + \sqrt[3]{3}$  is irrational. Hint: Use *Maple* or *Mathematica* to find the minimum polynomial of this constant.

22. **Fateman's web integrator.** Richard Fateman maintains an on line integration tool, <http://torte.cs.berkeley.edu:8010/tilu>. In its early days, as much as two of every three requests were ill formed; to address this, the interface performs a significant amount of heuristic work to decide what users intend by expressions such as

$$\int \frac{ax + bx}{cx + dx} dx \quad \text{or} \quad \int \sin x dx.$$

The tool also runs directly on the Macintosh graphing calculator.

23. **The happy end problem.** The happy end problem is to find, for  $n \geq 3$ , the smallest positive integer  $N(n)$  such that any set of  $N(n)$  points, no three of which are collinear, must contain  $n$  points that are the vertices of a convex  $n$ -gon. It is so called because Ester Klein, who posed the problem, married George Szekeres shortly after he and Paul Erdős proved the first bounds on the problem [144]. It is still open [144].
24. **The Mann iteration.** For any continuous function  $f : [0, 1] \mapsto [0, 1]$ , the iteration  $x_0 = x \in [0, 1]$  and

$$x_n = \frac{1}{n} \sum_{k=0}^{n-1} f(x_k)$$

(the Césaro average) always converges to a fixed point of  $f$ . One can study many other summability methods similarly. Since the function can be highly obstreperous, this is largely a theoretical real-variable iteration, albeit a beautiful one. Easy examples will convince one of how painfully slow or unstable convergence can be. This is especially understandable in light of

**Theorem 2.10.1 (Sharkovsky).** *If a continuous self-mapping of the reals,  $f$ , possesses a periodic point of order  $m$ , then  $f$  will possess a periodic point of order  $n$ , precisely when  $n$  follows  $m$  in the following ordering of the natural numbers:*

$$3, 5, 7, \dots, 2^3, 2^5, 2^7, \dots, 2^2 3, 2^2 5, 2^2 7, \dots, 2^3 3, 2^3 5, 2^3 7, \dots, \dots, 2^3, 2^2 2, 2, 1.$$

In particular this includes the famous result of Li and Yorke to the effect that “period three implies chaos.”

25. **Finding coefficients of an integer polynomial.** The numerical identification methods described here work especially well when you know *a priori* that a given integral, sum or other constant is a linear combination of given quantities, and you wish only to obtain the precise integer coefficients. For example, if you are told that, for any integer  $N > 0$ ,

$$\sum_{n=0}^{\infty} \frac{n^n e^{-n}}{(n+N)!}$$

evaluates to a polynomial of degree  $N$  in  $e$ , then it is an easy matter, for any given small  $N$ , to pick off the integers. Indeed, for  $N < 10$ , say, you can discover that

$$Q_N = \sum_{n=0}^{\infty} \frac{n^n e^{-n}}{(n+N)!} - \sum_{k=1}^N \frac{(-1)^{k-1} e^k}{(N-k)! k^k} \quad (2.10.35)$$

is a rational number. The first four are  $-1, -1/4, -7/108, -97/6912$ . Continuing in this manner, you can ultimately discover that

$$Q_N = \sum_{k=1}^N \frac{(-1)^k}{(N-k)! k^k} \sum_{n=0}^{k-1} \frac{k^n}{n!}. \quad (2.10.36)$$

This relies on replacing  $\exp(-n)$  by its series, exchanging order of summation, and then discovering and deriving the identity

$$\sum_{k=0}^n (-1)^k k^n \binom{N+n}{N+k} = (-1)^n \sum_{k=1}^N k^n (-1)^{k-1} \binom{N+n}{N-k} \quad (2.10.37)$$

or equivalently

$$\sum_{k=0}^{M+N} (-1)^k (M-k)^M \binom{M+N}{k} = 0 \quad (2.10.38)$$

for all  $M, N > 0$ . This in turn follows (using the binomial theorem) from

$$\sum_{k=0}^P (-1)^k k^n \binom{P}{k} = 0, \quad (2.10.39)$$

for all  $0 \leq n < P$ . The final identity (2.10.39) is established by setting  $Df(x) = xf'(x)$  and observing that  $D^k(1-x)^P$  has a zero at 1 for  $k < P$ .

26. **A taste of J. E. Littlewood's Miscellany.** We record a variety of observations, mathematical, stylistic and personal from Littlewood's wonderful *Miscellany*, a collection originally published in 1953, but republished with an excellent introduction by Béla Bollobás in 1988 [143]. More of Littlewood is to be found at the end of Chapter 4 of the second volume.

(a) (Entry (10) page 28) *Dissection of squares and cubes into squares and cubes, finite in number and all unequal.* "The square dissection is possible in an infinity of different ways (the simplest, due to Duijvestijn, is shown in [Figure 2.13]). A cube dissection is not possible." The proof of the second is short and elegant.

(b) (Entry (19) page 32) *An isoperimetric problem: an area of (greatest) diameter not greater than 1 is at most  $\frac{1}{4}\pi$ .*

Outline: "It is easy to see that we may suppose the area convex and on one side of one of its 'tangents'." With polar coordinates

$$\text{area} = \frac{1}{2} \int_0^{\frac{\pi}{2}} \left( r^2(\theta) + r^2\left(\theta - \frac{\pi}{2}\right) \right) d\theta,$$

and the integrand never exceeds one.

(c) (Page 38) *Keakeya's problem.* "Find the region of least area in which a segment of unit length can turn continuously through  $360^\circ$  (minimize the area swept over). It was long thought that the answer was as in [Figure 2.14] and the area  $\frac{1}{8}\pi$ . In 1930 [sic], however, A. S. Besicovitch, (*Math. Zeit.* **27** (1928), 312–320.) showed that the answer is zero area (unattained): given an arbitrarily small  $\epsilon$  the area swept can be less than  $\epsilon$ ."

(d) (§2, Misprints, page 56)

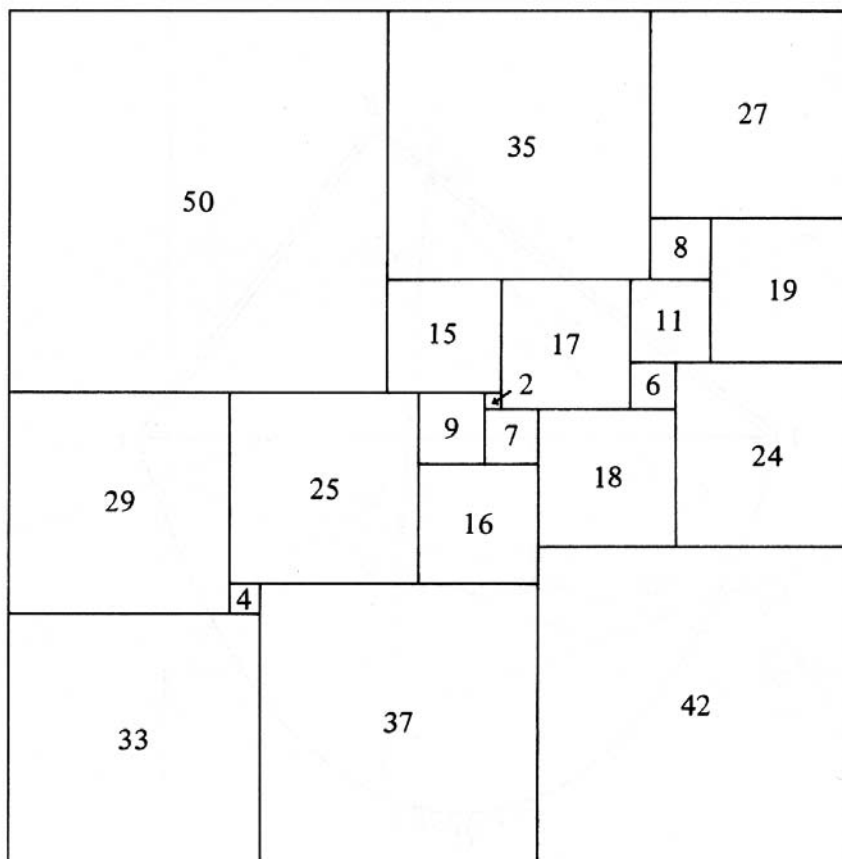


Figure 3

Figure 2.13: Duijvestijn's dissection.

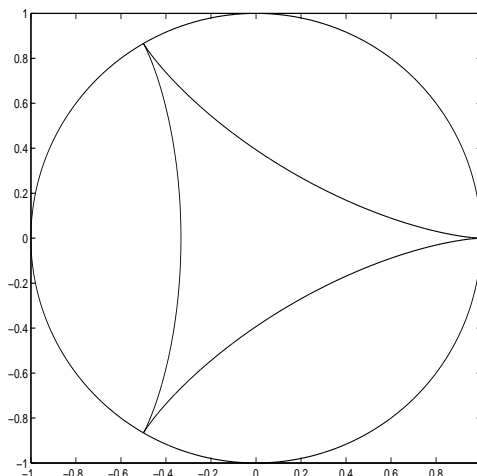


Figure 2.14: Kakeya’s conjectured answer—the three-cusp hypocycloid.

“I once challenged Hardy to find a misprint on a certain page of a joint paper. He failed: it was in his own name ‘G, H. Hardy’.”

“A minute I wrote (about 1917) for the Ballistics office ended with the sentence ‘Thus  $\sigma$  should be made as small as possible.’ This did not appear in the printed minute. But P. J. Brigg said ‘What is that?’ A speck in the blank space turned out to be the tiniest  $\sigma$  I have ever seen (the printers must have scoured London for it).”

27. **The hypocycloid.** Cartesian coordinates are given by

$$x = (a - b) \cos(t) + b \cos((a/b - 1)t), y = (a - b) \sin(t) - b \sin((a/b - 1)t)$$

for  $a > b > 0$ . This is plotted for  $a = 1, b = 2/3$  in Figure 2.14.

An excellent source for curves is <http://www-gap.dcs.st-and.ac.uk/history/Curves/Curves.html>.

28. **The Gelfond-Schneider theorem.** This is the result that for complex algebraic numbers  $\alpha \neq 0, 1$  and  $\beta$  irrational, the quantity  $\alpha^\beta$  is transcendental. It follows that



Figure 2.15: Donald Coxeter's own kaleidoscope

- (a)  $\sqrt{2}^{\sqrt{2}}$  is transcendental; and  
 (b)  $e^{\pi} = (-1)^{-i}$  is transcendental.
29. **Eight solutions.** Find examples of all eight rational and irrational possibilities of  $\alpha^{\beta} = \gamma$ .
30. **Quantum connections.** Another connection between quantum theory and the Riemann zeta is the observation, by Richard Crandall in 1991, that there exists an initial wave function  $\psi(x, 0)$ —having no zero-crossings whatever (it looks like a “bell curve,” for example)—which, under standard Schroedinger propagation, has evolved at a certain time  $T$  into a  $\psi(x, T)$  with *infinitely many* zero-crossings, each such zero corresponding to a critical zero of  $\zeta(1/2 + ix)$ . It is not yet known whether there are additional consequences of this observation. Some details are given in [78].
31. **H. S. M. (Donald) Coxeter (1907–2003).** The renowned Canadian geometer H. S. M. Coxeter passed away in late March 2003. Coxeter was known for making extensive use of physical models in his research. A portion of his collection is on display at the University of Toronto, where he worked for 67 years. The model shown in Figure 2.15 now resides at York University in Toronto.

Among his numerous published books, *Regular Complex Polytopes*, for example, is lavishly illustrated with beautiful and often intricate figures. He



was a friend of Maurits C. Escher, the graphic artist. In a 1997 paper, Coxeter showed that Escher, despite knowing no mathematics, had achieved “mathematical perfection” in his etching “Circle Limit III.” “Escher did it by instinct,” Donald Coxeter noted, “I did it by trigonometry.”

His *Introduction to Geometry* was widely used as a text for university courses in geometry. The spirit of his work is perhaps best captured by a quote from Bertrand Russell that Coxeter includes in the preface to *Introduction to Geometry*:

Mathematics, rightly viewed, possesses not only truth, but supreme beauty—a beauty cold and austere, without appeal to any part of our weaker nature, without the gorgeous trappings of painting or music, yet sublimely pure, and capable of a stern perfection such as only the greatest art can show. [175, pg. 73]

This quote is doubly interesting given the fact that Russell may have been responsible for Coxeter pursuing a mathematical career, according to Arthur Sherk, the executor for Coxeter’s mathematics. Russell, who was a family friend, read a prize-winning essay on dimensionality that Coxeter wrote at age 16. He then told Coxeter’s father that his son was unusually gifted mathematically, and urged him to change the direction of Coxeter’s education.

Along this line, Robert Moody wrote the following, in a letter supporting Coxeter’s nomination to receive an honorary doctorate at York University in Toronto, Canada:

Modern science is often driven by fads and fashion, and mathematics is no exception. Coxeter’s style, I would say, is singularly unfashionable. He is guided, I think, almost completely by a profound sense of what is beautiful.

Two sculptures based on Coxeter’s work decorate the Fields Institute in Toronto. One, hanging from the ceiling, is a three-dimensional projection of a four-dimensional regular polytope whose 120 faces are dodecahedrons as shown in Figure 2.16.

Coxeter’s fascination with geometry extended beyond the purely intellectual. For example, he was intrigued by the suggestion that Einstein’s brain

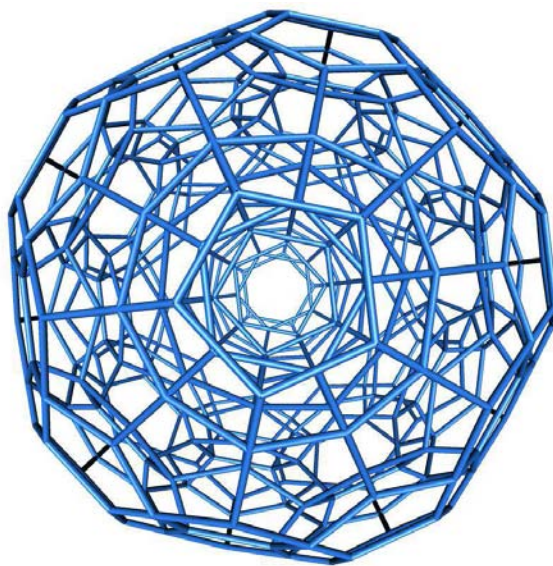


Figure 2.16: A projection of a four dimensional polytope.

exhibited unusual bilateral symmetry, which in part accounted for his genius, and has made provision for his brain to be similarly examined.

David Mumford recently commented that Coxeter placed great value on working out the details of complicated explicit examples:

In my book, Coxeter has been one of the most important 20th century mathematicians—not because he started a new perspective, but because he deepened and extended so beautifully an older esthetic. The classical goal of geometry is the exploration and enumeration of geometric configurations of all kinds, their symmetries and the constructions relating them to each other. The goal is not especially to prove theorems but to discover these perfect objects and, in doing this, theorems are only a tool that imperfect humans need to reassure themselves that they have seen them correctly. This is a flower garden whose beauty has almost been forgotten in the 20th century rush to abstraction and generality. I share Coxeter's love of this perspective,

which has deep roots in algebraic as well as Euclidean geometry. I always found the algebraic roots-and-weights approach to Lie groups arid and unsatisfying until I found Coxeter's work fleshing this out with a rich tapestry of examples. My own interest in explicit compactifications of the moduli space of Abelian varieties led me to wonderful parts of his work on explicit fundamental domains for quadratic forms modulo  $GL_n(\mathbb{Z})$ . In the 21st century, I think we see signs of a renewed appreciation of the significance of the classical geometry mode in mathematics, perhaps linked with having easily usable computational tools. I trust that this perspective will never die. [156]

A color illustration of the octahedral kaleidoscope Coxeter had built is shown in Chapter 8. Asia Weiss writes

The kaleidoscope was not built by Coxeter, but rather designed by him and made in a workshop in London, England I believe in the late 1920s. . . . Two kaleidoscopes were made: octahedral and icosahedral. (As the tetrahedral group is a subgroup in the octahedral group, the tetrahedral kaleidoscope was not made.) We also have the icosahedral one, but this one is damaged and I did not have it repaired as yet. [191]

32. **Math meets ice.** Figure 8.5 in the Color Supplement (Chapter 8) shows Helaman Ferguson's Costa surface sculpture, which he entered in the International Snow Sculpture Championship held at Breckenridge, Colorado, January 19-24, 1999 [70]. This piece was entitled "Invisible Handshake," because it is the space between two hands just before they clasp. Ferguson carved this piece with the assistance of Stan Wagon, Dan Schwalbe, Tomas Nemeth, and (posthumously) Alfred Gray. Here is Ferguson's account:

I choose this negative Gaussian curvature geometric form specifically because of the material properties of snow. Snow with its fair compressive strength and poor tensile strength is a caricature of stone. But negative Gaussian curvature, even in snow, presents a fabric of saddle points everywhere. Each point of the surface is the keystone of a pair of principal arches. There

were 14 other snow sculptures completed at the Breckenridge affair; a week later they all completely imploded in the heat wave. Our negative Gaussian curvature snow carving stood, retained its structure, sublimed, thinned, gracile. (Taken from <http://www.helasculpt.com>.)

33. **Mathematics of Ferguson’s sculptures.** The quotient group corresponding to Ferguson’s “Eight-Fold Way” sculpture is a certain subgroup of  $SL_2(Z)$ , i.e., the set of invertible, unit-determinant  $2 \times 2$  matrices with integer entries. One particular subgroup of  $SL_2(Z)$  corresponds to a tessellation of the hyperbolic circle by 7-gons. This design is used as the tiled base of Ferguson’s “Eight-Fold Way,” which is at the Mathematical Sciences Research Institute in Berkeley, CA. Another subgroup has a fundamental domain of 24 seven-gons, which has identifications corresponding to the surface of the marble sculpture above the base.

Ferguson’s “Figure-Eight Knot Complement II” sculpture corresponds to a certain subgroup of index 12 in the Bianchi group, which is  $SL_2(Z[\omega_3])$ , where  $\omega_3 = (-1 + i\sqrt{3})/2$  is a third root of unity. Let  $\Gamma$  denote the multiplicative group of  $2 \times 2$  matrices generated by the matrices

$$\begin{bmatrix} 1 & 0 \\ 0 & \omega_3 \end{bmatrix} \text{ and } \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}. \quad (2.10.40)$$

This  $\Gamma$  is the fundamental group of the figure-eight knot complement. The complement is the double quotient  $U_2 \backslash SL_2(C) / \Gamma$ , where  $U_2$  is the group of unitary matrices (i.e., inverse is conjugate transpose) in  $C$ . The first quotient  $U_2 \backslash SL_2(C)$  is geometrically the three-dimensional hyperbolic space. This group  $\Gamma$  is an “arithmetic” subgroup of  $SL_2(C)$ . The group  $SL_2(Z[\omega_3])$  is known as the Eisenstein integers, a ring of algebraic integers. The figure-eight knot is absolutely unique in this respect—there is no other knot with this double quotient possessing the discrete group arithmetic property. For further details of the group-theoretic structure of these objects, see [186].

34. **Clausen’s function and the figure-eight knot complement volume.** As we saw in Section 2.4, the volume of the figure-eight knot complement

is given by Equation (2.4.14). That is,

$$V = \frac{2}{\sqrt{27}} \sum_{k=0}^{\infty} \left( \frac{-1}{27} \right)^k \left( \frac{9}{(6k+1)^2} - \frac{9}{(6k+2)^2} - \frac{12}{(6k+3)^2} - \frac{3}{(6k+4)^2} + \frac{1}{(6k+5)^2} \right). \quad (2.10.41)$$

The reason is sketched below. Note that this proof crucially requires knowing the answer in advance, and moving back from the discovered result to something provable.

(a) We take the *log sine integral*

$$V = -2 \int_0^{\pi/3} \log \left( 2 \sin \left( \frac{t}{2} \right) \right) dt \quad (2.10.42)$$

as the defining formula for the requisite volume. Recall that the *Clausen function* is given by

$$\text{Cl}_2(\theta) = \sum_{n>0} \frac{\sin(n\theta)}{n^2}, \quad (2.10.43)$$

which satisfies  $\text{Cl}_2(\frac{\pi}{2}) = G$ . Then integration of (2.10.42) shows

$$\begin{aligned} V &= i \{ \text{Li}_2(e^{-i\pi/3}) - \text{Li}_2(e^{i\pi/3}) \} \\ &= 2 \text{ImLi}_2 \left( \frac{1+i\sqrt{3}}{2} \right) = 2 \text{Cl}_2 \left( \frac{\pi}{3} \right) = 3 \text{Cl}_2 \left( \frac{2\pi}{3} \right). \end{aligned}$$

All the requisite details about Clausen's function are to be found in Lewin [141].

(b) A hypergeometric equivalent formulation of (2.10.41) is

$$\begin{aligned} \frac{V}{\sqrt{3}} &\stackrel{?}{=} 2F \left( \frac{1}{6}, \frac{1}{6}, 1; \frac{7}{6}, \frac{7}{6}; \frac{-1}{27} \right) - \frac{1}{2} \left( \frac{1}{3}, \frac{1}{3}, 1; \frac{4}{3}, \frac{4}{3}; \frac{-1}{27} \right) \\ &\quad - \frac{8}{27} F \left( \frac{1}{2}, \frac{1}{2}, 1; \frac{3}{2}, \frac{3}{2}; \frac{-1}{27} \right) - \frac{1}{24} F \left( \frac{2}{3}, \frac{2}{3}, 1; \frac{5}{3}, \frac{5}{3}; \frac{-1}{27} \right) \\ &\quad + \frac{2}{225} F \left( \frac{5}{6}, \frac{5}{6}, 1; \frac{11}{6}, \frac{11}{6}; \frac{-1}{27} \right) \end{aligned}$$

which with some effort is expressible in terms of the dilogarithm. This leads to the following:

$$V \stackrel{?}{=} \operatorname{Im} \left\{ 4 \operatorname{Li}_2 \left( \frac{i\sqrt{3}}{3} \right) - \frac{8}{3} \operatorname{Li}_2 \left( \frac{i\sqrt{3}}{9} \right) + \operatorname{Li}_2 \left( \frac{1}{2} - \frac{i\sqrt{3}}{6} \right) + 8 \operatorname{Li}_2 \left( -\frac{1}{2} + \frac{i\sqrt{3}}{6} \right) \right\}.$$

(c) Now, Lewin in equation (5.5) of [141] gives

$$\operatorname{Im} \operatorname{Li}_2 (re^{i\theta}) = \omega \log(r) + \frac{1}{2} \operatorname{Cl}_2 (2\omega) - \frac{1}{2} \operatorname{Cl}_2 (2\omega + 2\theta) + \frac{1}{2} \operatorname{Cl}_2 (2\theta),$$

where  $\omega = \arctan(r \sin \theta / (1 - r \cos \theta))$ . Using this, a proof that (2.10.41) holds is reduced to showing that, with  $\alpha = \arctan(\sqrt{3}/9)$ ,

$$4 \operatorname{Cl}_2 \left( \frac{\pi}{3} \right) = 2 \operatorname{Cl}_2 (2\alpha) + \operatorname{Cl}_2 (\pi + 2\alpha) - 3 \operatorname{Cl}_2 \left( \frac{5}{3}\pi + 2\alpha \right),$$

which is true by applying the two variable identities for Clausen's function given in Equations (4.61) and (4.63) of [141], with  $\theta = \frac{\pi}{3}$ .

(d) Many other identities lie in (4.63) such as, with  $\alpha = \arctan \left( \frac{1}{3} \right)$ ,

$$6G = 2 \operatorname{Cl}_2 (2\alpha) - 3 \operatorname{Cl}_2 \left( 2\alpha - \frac{\pi}{2} \right) + \operatorname{Cl}_2 \left( 2\alpha + \frac{\pi}{2} \right).$$

A much harder identity is

$$\frac{7\sqrt{7}}{4} L_{-7}(2) = 3 \operatorname{Cl}_2 (\alpha) - 3 \operatorname{Cl}_2 (2\alpha) + \operatorname{Cl}_2 (3\alpha)$$

with  $\alpha = 2 \arctan (\sqrt{7})$ .

35. **Further identities for the knot complement volume** are explored below.

(a) On expanding the Clausen function of (2.10.43), we obtain

$$\begin{aligned} V &= \sum_{n \geq 1} \frac{2 \sin(n\pi/3)}{n^2} \\ &= \sqrt{3} \left( \sum_{n=0}^{\infty} \frac{1}{(6n+1)^2} + \sum_{n=0}^{\infty} \frac{1}{(6n+2)^2} \right. \\ &\quad \left. - \sum_{n=0}^{\infty} \frac{1}{(6n+4)^2} - \sum_{n=0}^{\infty} \frac{1}{(6n+5)^2} \right), \end{aligned}$$

or equivalently,

$$V = \frac{\sqrt{3}}{9} \sum_{n \geq 1} \frac{n - \frac{1}{2}}{(n - \frac{1}{3})^2 (n - \frac{2}{3})^2}.$$

(b) Changing variables in (2.10.42) and using the binomial theorem, we derive

$$V = -2 \int_0^1 \frac{\log(y)}{\sqrt{1 - (\frac{y}{2})^2}} dy = 2 \sum_{n=0}^{\infty} \frac{\binom{2n}{n}}{(2n+1)^2 16^n}.$$

The next two equivalent identities are derivable using (1.76) in [141] which gives the series for  $\text{Li}_2(e^{i\pi/3})$ :

(c)

$$V = \frac{2\pi}{3} \left( 1 - \log\left(\frac{\pi}{3}\right) + \sum_{m \geq 1} \frac{\zeta(2m)}{m(2m+1)6^{2m}} \right),$$

and

$$V = \frac{2\pi}{3} - \frac{2\pi}{3} \log\left(\frac{\pi}{3}\right) + 2 \sum_{n \geq 1} \frac{B_n}{2n} \frac{(\frac{\pi}{3})^{2n+1}}{(2n+1)!}.$$

(d) Show that

$$V = 2\sqrt{3} \text{Im} L_2 \left( \frac{i\sqrt{3}+1}{2} \right) = -\sqrt{3} \int_0^1 \frac{\log y}{1-y+y^2} dy.$$

(e) Hence expanding the last denominator and differentiating the  $\beta$ -function (see Section 5.4), deduce that

$$\begin{aligned} V &= 2\sqrt{3} \sum_{n \geq 1} \{\Psi(n) - \Psi(2n)\} \beta(n, n) = 2\sqrt{3} \sum_{n \geq 1} \frac{\sum_{k=n}^{2n-1} \frac{1}{k}}{n \binom{2n}{n}} \\ &= 2\sqrt{3} \sum_{n \geq 1} \frac{\sum_{k=1}^{2n-1} \frac{1}{k}}{n \binom{2n}{n}} - 2\sqrt{3} \sum_{n \geq 1} \frac{\sum_{k=1}^{n-1} \frac{1}{k}}{n \binom{2n}{n}}. \end{aligned}$$

(f) This is equivalent to

$$V = 2\sqrt{3} \int_0^{1/2} \frac{\log\left(\frac{1+s}{1-s}\right)}{(1-s^2)\sqrt{1-4s^2}} ds - 2\sqrt{3} \int_0^{1/2} \frac{-s \log(1-s^2)}{(1-s^2)\sqrt{1-4s^2}} ds,$$

where the first integral corresponds to  $\sum_{k=1}^{2n-1} 1/k$  and the second to  $\sum_{k=1}^{n-1} 1/k$ . These are equivalent to

$$\begin{aligned} V &= \sqrt{3} \int_0^{\pi/2} \frac{\log\left(\frac{1+(\sin t)/2}{1-(\sin t)/2}\right)}{1 - (\sin^2 t)/4} ds \\ &\quad - \sqrt{3} \int_0^{\pi/2} \left(\frac{-\sin t}{2}\right) \frac{\log(1 - (\sin^2 t)/4)}{1 - (\sin^2 t)/4} dt, \end{aligned}$$

where  $\int_0^{\pi/2} \sin^{2N+1} t dt = \Gamma^2(N+1)4^N/\Gamma(2N+2)$ , and  $\int_0^{\pi/2} \sin^{2N} t dt = \Gamma^2(N+1/2)2^{2N-1}/\Gamma(2N+1)$ . A symmetric rendition of the integral is:

$$V = 2\sqrt{3} \int_0^{1/2} \frac{(1+s) \log(1+s) - (1-s) \log(1-s)}{(1-s^2)\sqrt{1-4s^2}} ds.$$

(g) Deduce that

$$V = \sqrt{3} \sum_{n=0}^{\infty} \frac{\sum_{k=1}^{2n+1} (-1)^{k+1} k^{-1}}{\binom{2n}{n} (2n+1)}.$$

(h) Finally,

$$-\sqrt{3} \sum_{n=2}^{\infty} \frac{\sum_{k=1}^{n-1} k^{-1}}{n \binom{2n}{n}} = \frac{\sqrt{3}}{2} \int_0^1 \frac{\log(1-t(1-t))}{1-t+t^2} dt.$$



Let

$$H_1 = \sum_{n=1}^{\infty} \frac{\sum_{k=1}^{n-1} k^{-1}}{n \binom{2n}{n}}, \quad H_2 = \sum_{n=1}^{\infty} \frac{\sum_{k=1}^{2n-1} k^{-1}}{n \binom{2n}{n}}.$$

Then  $V = 2\sqrt{3} \{H_2 - H_1\}$  and, following the logic of the previous exercise,

$$H_1 = \sum_{k=0}^{\infty} \left(\frac{-1}{27}\right)^k \left\{ -\frac{1/3}{(6k+1)^2} + \frac{1}{(6k+2)^2} + \frac{10/9}{(6k+3)^2} + \frac{1/3}{(6k+4)^2} - \frac{1/27}{(6k+5)^2} \right\},$$

while

$$H_2 = \sum_{k=0}^{\infty} \left(\frac{-1}{27}\right)^k \left\{ \frac{2/3}{(6k+1)^2} - \frac{2/9}{(6k+3)^2} - \frac{2/27}{(6k+5)^2} \right\}.$$

See [141, 61] for additional details.

36. **Census of hyperbolic knots.** The URL [http://newweb.cecm.sfu.ca/cgi-bin/KnotPlot/object/hyperbolic\\_census?knot=k4.1](http://newweb.cecm.sfu.ca/cgi-bin/KnotPlot/object/hyperbolic_census?knot=k4.1) contains a census of hyperbolic knots (most, but not all, knots are hyperbolic). The figure-eight knot, not the trefoil, is now simplest when viewed by the number of perfect tetrahedra in its complement.
37. **Jeff Weeks on experimental topology.** Jeff Weeks recently wrote:

There are plenty of such [experimentally discovered topology] results, but most would require a fair amount of explanation to be significant to non-specialists. For example, one of my favorites is the (computationally discovered!) pattern in the canonical triangulation for two-bridge knot and link complements. The canonical triangulation is geometrically defined, yet it exactly encodes the twists in the (topological) tangle decomposition of the knot or link. The correspondence is remarkably simple, with exactly two tetrahedra realizing each twist. This result has since been proved, but the proof is long and difficult. . . . I could try to dig up a figure illustrating how the pairs of tetrahedra realize



Figure 2.17: Weeks' two-bridge discovery.

the twists, ... but then it's not clear that you could explain it easily so that a non-specialist would understand why it's such an amazingly beautiful thing (as opposed to the reader struggling just to get a dim grasp of what the result is saying, without really appreciating what makes it cool).

Figure 2.17 attempts to illustrate Jeff Weeks' beautiful discovery.

38. **Griffeath's "psoup" cellular automata web site.** David Griffeath well describes his site <http://psoup.math.wisc.edu> as follows:

Hi there. I'm a math prof at the University of Wisconsin-Madison whose research these days focuses on self-organization of random cellular automata. For the past ten years or so I have been producing colorful computer graphics and animations that illustrate the ability of local parallel update rules to generate spatial structure from disordered initial states. My work has appeared in numerous books and periodicals, including *Nonlinear Science Today* and *Scientific American*. Some of my cellular automaton (CA) rules have been featured in the simulation packages CALAB and James Gleick's *Chaos: The Software*. This home page is intended to provide a gallery of my images and movies, as well

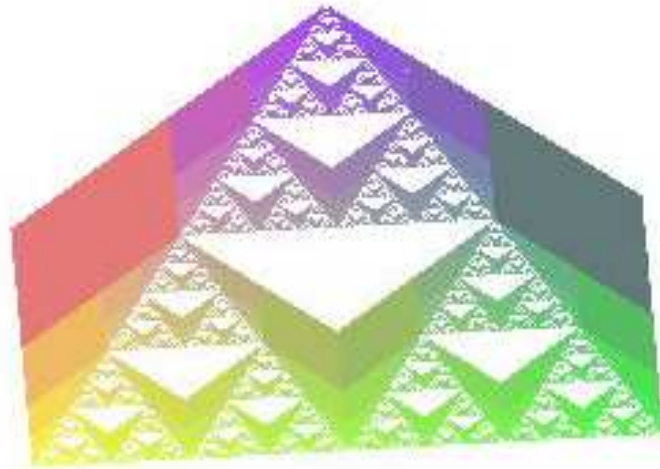


Figure 2.18: 7th iterate of Sierpinski triangle made by paper cutting and folding.

as links to software that allows you to design your own.

It's tempting to think of self-organization as evolution from primordial soup. Machines that orchestrate such dynamics are sometimes called mixmasters. I use both garden-variety PCs and dedicated Cellular Automaton Machines (CAMs) as mixmasters. Since my other great fixation in life is gastronomy (witness the important 3-letter word hidden in my Welsh last name), this site is laid out as a kitchen.

39. **Fractal cards.** Not all impressive representations require a computer. Elaine Simmt and Brent Davis [179] describe lovely constructions made by repeated paper folding and cutting that result in beautiful fractal “pop-up” cards. Nonetheless, in Figure 2.18, we show the seventh iterate of a Sierpinski triangle built in software by following those paper cutting and folding rules in software. Note the similarity to the tetrahedron in Figure 2.2. Recursive *Maple* code is given below:

```

sierpinski:= proc( n::nonnegint )
    local p1, p2, q1, q2, r1, r2, plotout;
    p1:= [1.,0.,0.]; q1:= [-1.,0.,0.]; r1:= [0.,0.,1.];
    p2:= [1.,1.,0.]; q2:= [-1.,1.,0.]; r2:= [0.,1.,1.];
    plotout:= polys(n, p1, p2, r1, r2, q1, q2);
    return PLOT3D( plotout, SCALING(CONSTRAINED),
        AXESSTYLE(NONE), STYLE(PATCHNOGRID), ORIENTATION(90,45) );
end:

polys:= proc( n::nonnegint, p1, p3, r1, r3, q1, q3 )
    local p2, q2, r2, s1, s2, s3, t1, t2, t3, u2, u3;
    if n=0 then return POLYGONS([p1,p3,r3,r1], [q1,q3,r3,r1]) fi;
    p2:= (p1+p3)/2; q2:= (q1+q3)/2; r2:= (r1+r3)/2;
    s1:= (p1+r1)/2; s2:= (p2+r2)/2; s3:= (p3+r3)/2;
    t1:= (q1+r1)/2; t2:= (q2+r2)/2; t3:= (q3+r3)/2;
    u2:= (p2+q2)/2; u3:= (p3+q3)/2;
    return polys(n-1, p2, p3, s2, s3, u2, u3),
        polys(n-1, s1, s2, r1, r2, t1, t2),
        polys(n-1, u2, u3, t2, t3, q2, q3),
        POLYGONS([p1,p2,s2,s1], [q1,q2,t2,t1]); end:

```

40. **Gauss on learning.** In Boris Stoicheff's often enthralling biography of Gerhard Herzberg (1903-1999), Gauss is recorded as writing

It is not knowledge, but the act of learning, not possession but the act of getting there which generates the greatest satisfaction.  
[182, page 42]

Herzberg, who fled Germany for Saskatchewan in 1935, won the 1971 Nobel Prize in Chemistry for his pioneering work in molecular spectroscopy. His impact on Canadian—and international—science, signally in chemistry, physics and astronomy, has been enormous.

41. **Horizontal Riemann curves.** Figure 2.19 plots the modulus of the function  $Z$  defined by  $(x, y) \mapsto |\zeta(x + iy)|$  for  $1/5 < x < 4/5$  and  $1 < y < 100$ . Much related material and many illustrations are to be found in Brian Conrey's survey [73].

Plot  $Z(x, y)$  as a function of  $x$  on the critical strip  $[0, 1]$  for various values of  $y$ . (Compare what happens with the roles of  $x$  and  $y$  reversed.) As

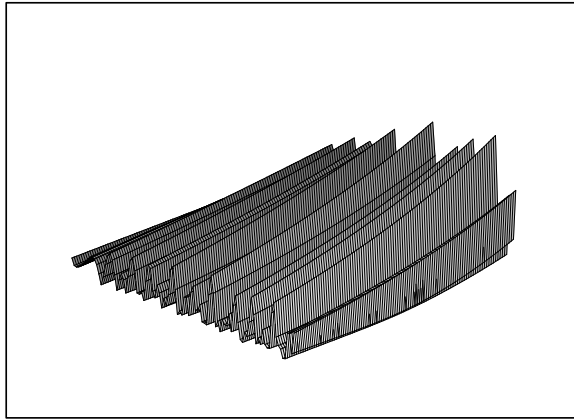


Figure 2.19: The modulus of the  $\zeta$  function.

it turns out, the monotonicity exhibited in the graph of level curves of  $Z(\cdot, y)$  implies the Riemann Hypothesis. This conclusion is given in [197], and was discovered while the first author of that paper was teaching an undergraduate complex variable class.

42. **Hardy and Littlewood's Four Axioms for Collaboration.** Harald Bohr, quoting from the preface of Béla Bollobás' 1988 edition of Littlewood's *Miscellany*, said:

The first [axiom] said that when one wrote to the other (they often preferred to exchange thoughts in writing instead of orally), it was completely indifferent whether what they said was right or wrong. As Hardy put it, otherwise they could not write completely as they pleased, but would have to feel a certain responsibility thereby. The second axiom was to the effect that, when one received a letter from the other, he was under no obligation whatsoever to read it, let alone answer it—because, as they said, it might be that the recipient of the letter would prefer not to

work at that particular time, or perhaps that he was just then interested in other problems. . . . The third axiom was to the effect that, although it did not really matter if they both thought about the same detail, still, it was preferable that they should not do so. And, finally, the fourth, and perhaps most important axiom, stated that it was quite indifferent if one of them had not contributed the least bit to the contents of a paper under their common name; otherwise there would constantly arise quarrels and difficulties in that now one, and now the other, would oppose being named co-author.

These rules work very well in the present age.

43. **What is a tool?** James O'Donnell wrote a mere five years ago:

I am a student of the works of St. Augustine and shall begin accordingly with confession. The single most transforming feature of cyberspace as we inhabit it in 1997 for my own scholarship can be found in a warehouse on the edges of downtown Seattle. I mean the nerve center of [www.amazon.com](http://www.amazon.com). [159]

Google, MathSciNet—the online version of Mathematical Reviews at <http://e-math.ams.org/mathscinet>, and Amazon.com are just as much tools of discovery as are PSLQ, the Inverse Calculator and *Mathematica*.

44. **Tea with John-Paul II.** James Glanz wrote in “Web Archive Opens a New Realm of Research” (*New York Times*, April 30, 2001):

Three years ago, said Dr. Zanelli in Chile, one of his own students posted a paper and the next day received an e-mail message from Dr. Witten. The student was at first so shocked that he accused Dr. Zanelli of playing an elaborate practical joke.

“We learned that great physicists do read the archives daily and they browse through all the preprints,” Dr. Zanelli said, “even if they come from an obscure place in the end of the world.”

Dr. Witten's instant response, Dr. Zanelli said, “was like having the pope drop by for tea.”

Such is the promise of the wired world, deluges of spam notwithstanding.

45. **False proofs of big theorems.** David Einstein has commented that in the last generation, there has always been at least one somewhat serious purported proof of the Poincaré conjecture under scrutiny. Similar comments hold for Fermat’s Last Theorem, the Riemann Hypothesis, the Four Color Theorem, and so on. Many famous mathematicians have survived premature publicity for proofs found later to be unavoidably flawed. Hans Rademacher’s putative proof of the *falsehood* of the Riemann Hypothesis made the pages of *Time* in the late forties, even after the flaw had been discovered by Siegel.[73]

Of course, even the existence of generally accepted proofs does not discourage many amateurs from hunting for counterexamples nor from trying to square the circle or trisect angles.

46. **Research, teaching and communication.** The Fields medalist Bill Thurston, in his 1995 article, “On proof and progress,” makes the observation about researchers’ habits that

Mathematicians have developed habits of communication that are often dysfunctional. Organizers of colloquium talks everywhere exhort speakers to explain things in elementary terms. Nonetheless, most of the audience at an average colloquium talk gets little of value from it. Perhaps they are lost within the first 5 minutes, yet sit silently through the remaining 55 minutes. Or perhaps they quickly lose interest because the speaker plunges into technical details without presenting any reason to investigate them. At the end of the talk, the few mathematicians who are close to the field of the speaker ask a question or two to avoid embarrassment.

This pattern is similar to what often holds in classrooms, where we go through the motions of saying for the record what we think the students “ought” to learn, while the students are trying to grapple with the more fundamental issues of learning our language and guessing at our mental models. Books compensate by giving samples of how to solve every type of homework problem.

Professors compensate by giving homework and tests that are much easier than the material “covered” in the course, and then grading the homework and tests on a scale that requires little understanding. We assume that the problem is with the students rather than with communication: that the students either just don’t have what it takes, or else just don’t care.

Outsiders are amazed at this phenomenon, but within the mathematical community, we dismiss it with shrugs.

47. **Riemann and robust notation.** Figures 2.8 and 2.9, which reproduce the first two pages of the final manuscript of Riemann’s celebrated 1859 paper [171] illustrate how robust mathematical notation is. If a counterexample or theorem is originally due to Hilbert or Cauchy, there is a good chance we reproduce it with exactly the same syntax.
48. **Hoffman’s art.** Experimental visualization in low-dimensional geometry was one of the primary motivations for Klaus Peters’ decision in the early nineties to establish the journal *Experimental Mathematics*—see <http://www.expmath.org>. This is now a flourishing, high quality journal with several Fields medalists on its board. Some of the better examples of geometric visualization are due to David Hoffman—see the web sites <http://www.gang.umass.edu> and [http://emsh.calarts.edu/~mathart/portfolio/SPD\\_Costa\\_portfolio.html](http://emsh.calarts.edu/~mathart/portfolio/SPD_Costa_portfolio.html).
49. **Euler, Goldbach and the birth of  $\zeta$ .** What follows is a direct transcription of correspondence between Euler and Goldbach [124], leading to the origin of the zeta function and multi-zeta values.
- (a) **59. Goldbach an Euler, Moskau, 24. Dez. 1742** [footnote: AAL: F.136, Op. 2, Nr.8, Blatt 54–55]  
 [...] Als ich neulich die vermeinten summas der beiden letzteren serierum in meinem vorigen Schreiben wieder betrachtet, habe ich alsofort wahrgenommen, daß selbige aus einem bloßem Schreibfehler entstanden, von welchem es aber in der Tat heißet: Si non errasset, fecerat ille minus<sup>1</sup>.

---

<sup>1</sup>Frei zitiert nach Marcus Valerius Martialis, I, 21,9



denn ich bin durch diese Gelegenheit auf summationes aliarum serierum geraten, die ich sonst kaum gesucht, viel weniger gefunden haben würde.

Ich halte dafür, daß es ein problema problematum ist, die summam huius:

$$/54r/1 + \frac{1}{2^n} \left(1 + \frac{1}{2^m}\right) + \frac{1}{3^n} \left(1 + \frac{1}{2^m} + \frac{1}{3^m}\right) + \frac{1}{4^n} \left(1 + \frac{1}{2^m} + \frac{1}{3^m} + \frac{1}{4^m}\right) + etc.$$

in den casibus zu finden, wo  $m$  et  $n$  nicht numeri integri pares et sibi aequales sind, doch gibt es casus, da die summa angegeben werden kann, exempli gr[atia], si  $m = 1$ ,  $n = 3$ , denn es ist

$$1 + \frac{1}{2^3} \left(1 + \frac{1}{2}\right) + \frac{1}{3^3} \left(1 + \frac{1}{2} + \frac{1}{3}\right) + \frac{1}{4^3} \left(1 + \frac{1}{2} + \frac{1}{3} + \frac{1}{4}\right) + etc. = \frac{\pi^4}{72}.$$

(wenn  $\pi$  gewöhnlichermaßen für die peripheriam circuli, cuius diameter = 1, genommen wird), hingegen weiß ich die summas serierum

$$A \cdots 1 + \frac{1}{2^5} \left(1 + \frac{1}{2}\right) + \frac{1}{3^5} \left(1 + \frac{1}{2} + \frac{1}{3}\right) + \frac{1}{4^5} \left(1 + \frac{1}{2} + \frac{1}{3} + \frac{1}{4}\right) + etc.$$

$$B \cdots 1 + \frac{1}{2^4} \left(1 + \frac{1}{2}\right) + \frac{1}{3^4} \left(1 + \frac{1}{2} + \frac{1}{3}\right) + \frac{1}{4^4} \left(1 + \frac{1}{2} + \frac{1}{3} + \frac{1}{4}\right) + etc.$$

noch nicht, ob ich gleich weiß, daß  $2A + B = \frac{19\pi^6}{2 \cdot 5 \cdot 7 \cdot 3^4}$ , wie ich denn auch die summam der folgenden beiden /55/ serierum  $C + D$  allezeit finden kann, si  $m$  et  $n$  sint numeri pares quicunque

$$C \cdots 1 + \frac{1}{2^n} \left(1 + \frac{1}{2^m}\right) + \frac{1}{3^n} \left(1 + \frac{1}{2^m} + \frac{1}{3^m}\right) + etc.$$

$$D \cdots 1 + \frac{1}{2^m} \left(1 + \frac{1}{2^n}\right) + \frac{1}{3^m} \left(1 + \frac{1}{2^n} + \frac{1}{3^n}\right) + etc.$$

Übrigens beziehe ich mich auf mein voriges Schreiben und verbleibe nächst herzlicher Anwünschung eines glücklichen neuen Jahres[...]

(b) “*Lisez Euler, lisez Euler, c’est notre maitre a tous.*”

This letter is the first in which Goldbach precisely formulates the series which sparked Euler’s further investigations into what would

become known as the Zeta-function. These investigations were apparently due to a serendipitous mistake. The first sentence above translates as follows:

When I recently considered further the indicated sums of the last two series in my previous letter, I realized immediately that the same series arose due to a mere writing error, from which indeed the saying goes, “Had one not erred, one would have achieved less.” (*Si non errasset, fecerat ille minus*). [Translation thanks to Martin Matmüller, scientific collaborator of Euler’s *Opera Omnia*, vol. IVA4, Birkhuser Verlag.]

(c) Confirm Goldbach’s evaluation ( $m = 3, n = 1$ ):

$$\zeta(3, 1) + \zeta(4) = \frac{\pi^4}{72}$$

(in the language of Section 3.4 of the second volume of this work).

#### 50. Gauss: the next generation.

Gauss’s second son, Eugene, emigrated to the United States in 1830, enlisted in the army, and later went into business in Missouri. Eugene is said to have had some of his father’s gift for languages and the ability to perform prodigious arithmetic calculations, which he did for recreation after his sight failed him in old age. (David T. Lykken [145])

#### 51. Flexibility.

When the facts change, I change my mind. What do you do, sir? (John Maynard Keynes, quoted in *The Economist*, Dec. 18, 1999, pg. 47)

#### 52. Mathematics and the aesthetic impulse.

If my teachers had begun by telling me that mathematics was pure play with presuppositions, and wholly in the air, I might

have become a good mathematician, because I am happy enough in the realm of essence. But they were overworked drudges, and I was largely inattentive, and inclined lazily to attribute to incapacity in myself or to a literary temperament that dullness which perhaps was due simply to lack of initiation. (George Santayana, *Persons and Places* [177, pg. 238–9])

# Chapter 3

## Pi and Its Friends

I am ashamed to tell you to how many figures I carried these computations, having no other business at the time.

Issac Newton, 1666

The desire, as well as the need, to calculate ever more accurate values of  $\pi$ , the ratio of the circumference of a circle to its diameter, has challenged mathematicians for many centuries. In recent years,  $\pi$  computations have provided some fascinating examples of computational mathematics. But first, a little historical background, which we have condensed from a 1997 article [17]. Additional historical details are available from the fascinating new book *Pi Unleashed* [7], and from [50, 34]. We should also mention the MacTutor web site, maintained by the Department of Mathematics at the University of St. Andrews in Scotland: <http://www-gap.dcs.st-and.ac.uk/history>. This is particularly good for biographical information about historical mathematicians, including those involved with  $\pi$ .

### 3.1 A Short History of Pi

About 2000 BCE, the Babylonians used the approximation  $3\frac{1}{8} = 3.125$ . At this same time or earlier, according to an ancient Egyptian document, Egyptians assumed that a circle with diameter nine has the same area as a square of side eight, which implies  $\pi = \frac{256}{81} = 3.1604\dots$  Some have argued from the following biblical passage that the ancient Hebrews used  $\pi = 3$ :

Also, he made a molten sea of ten cubits from brim to brim, round in compass, and five cubits the height thereof; and a line of thirty cubits did compass it round about (1 Kings 7:23; see also 2 Chron. 4:2).

The first rigorous mathematical calculation of the value of  $\pi$  was due to Archimedes (ca. 250 BCE), who used a scheme based on inscribed and circumscribed polygons to obtain the bounds  $3\frac{10}{71} < \pi < 3\frac{1}{7}$ . Indeed, Archimedes' scheme constitutes the first true algorithm for  $\pi$ , in the sense that it is a well-defined scheme that is capable of producing an arbitrarily accurate value for the constant. This scheme can be stated as a simple recursion, as follows. Set  $a_0 = 2\sqrt{3}$  and  $b_0 = 3$ . Then define

$$\begin{aligned} a_{n+1} &= \frac{2a_n b_n}{a_n + b_n} \\ b_{n+1} &= \sqrt{a_{n+1} b_n}. \end{aligned} \tag{3.1.1}$$

This recursion converges to  $\pi$ , with the error decreasing by a factor of approximately four with each iteration. Variations of Archimedes' geometrical scheme were the basis for all high-accuracy calculations of  $\pi$  for the next 1,800 years. For example, the Chinese mathematician Tsu Chung-Chih used a variation of this method to compute  $\pi$  correct to seven digits in the fifth century CE.

Little progress was made in Europe during the dark ages, but one very significant advance in arithmetic was made in India. This was the invention, in about 450 CE, of our modern positional, zero-based decimal arithmetic system—the *Indo-Arabic* system. This discovery greatly accelerated arithmetic in general, and computing  $\pi$  in particular. The Indo-Arabic system was introduced into Europe in 1000 CE, but due to resistance from several quarters, centuries elapsed before it finally enjoyed widespread use. The resistance ranged from accountants who didn't want their livelihood upset to clerics who regarded the system as "diabolical," on account of what they incorrectly assumed was its origin in the Arabic world. European commerce resisted the system until the 18th century, and even in scientific circles, its usage was limited until the 17th century [119]. The difficulty of doing arithmetic prior to the Indo-Arabic system is indicated by the words of counsel given to a wealthy German merchant in the 16th century, who was considering where to send his son for college:

If you only want him to be able to cope with addition and subtraction, then any French or German university will do. But if you are intent

on your son going on to multiplication and division—assuming that he has sufficient gifts—then you will have to send him to Italy. [119, pg. 577]

In the 17th century, Newton and Leibniz discovered calculus, and this powerful new tool was quickly exploited to find new formulas for  $\pi$ . One early calculus-based formula comes from the integral

$$\begin{aligned}\arctan x &= \int_0^x \frac{dt}{1+t^2} = \int_0^x (1 - t^2 + t^4 - t^6 + \dots) dt \\ &= x - \frac{x^3}{3} + \frac{x^5}{5} - \frac{x^7}{7} + \frac{x^9}{9} - \dots.\end{aligned}\tag{3.1.2}$$

Substituting  $x = 1$  gives the well-known Gregory–Leibniz formula

$$\pi/4 = 1 - 1/3 + 1/5 - 1/7 + 1/9 - 1/11 + \dots\tag{3.1.3}$$

(recall our discussion of Gregory’s series in Section 2.2). This particular series is practically useless for computing  $\pi$ —it converges so slowly that hundreds of terms would be required to compute even two correct digits. However, by employing a trigonometric identity

$$\pi/4 = \arctan(1/2) + \arctan(1/3),\tag{3.1.4}$$

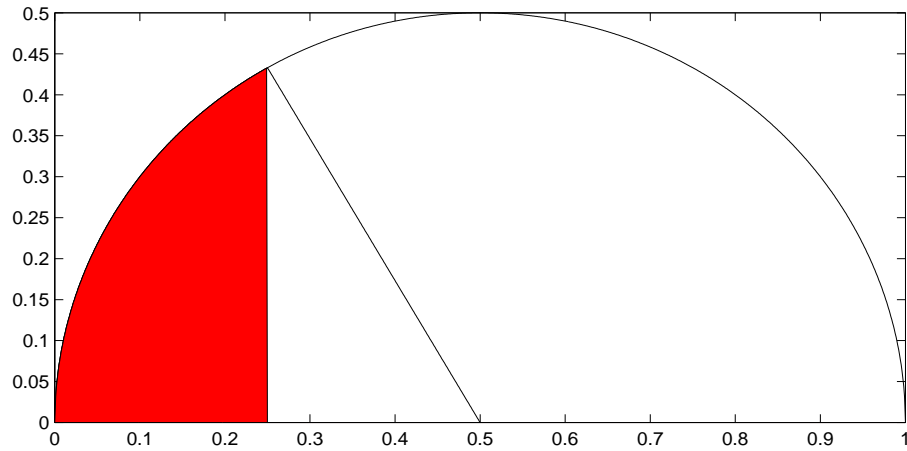
attributed to Euler in 1738, you can obtain

$$\begin{aligned}\pi/4 &= \frac{1}{2} - \frac{1}{3 \cdot 2^3} + \frac{1}{5 \cdot 2^5} - \frac{1}{7 \cdot 2^7} + \dots \\ &\quad + \frac{1}{3} - \frac{1}{3 \cdot 3^3} + \frac{1}{5 \cdot 3^5} - \frac{1}{7 \cdot 3^7} + \dots,\end{aligned}\tag{3.1.5}$$

which converges much more rapidly. An even faster formula, discovered a generation earlier by Machin, can be obtained using the identity

$$\pi/4 = 4 \arctan(1/5) - \arctan(1/239)\tag{3.1.6}$$

in a similar way. This formula was used in numerous computations of  $\pi$ , culminating with Shanks’ computation of  $\pi$  to 707 decimal digits accuracy in 1874 (although it was later found that this result was in error after the 527th decimal place).

Figure 3.1: Newton's scheme to compute  $\pi$ .

Newton himself discovered a different formula. He began by considering the area  $A$  of the left-most region under the circle as shown in Figure 3.1. On one hand, this is merely the difference between the area of the circular sector, which is  $\pi/24$ , and the area of the triangle, which is  $\sqrt{3}/32$ . On the other hand,  $A$  can be written as the integral

$$A = \int_0^{1/4} \sqrt{x-x^2} dx. \quad (3.1.7)$$

By employing his own binomial theorem, Newton wrote (3.1.7) as:

$$\begin{aligned} A &= \int_0^{1/4} x^{1/2}(1-x)^{1/2} dx \\ &= \int_0^{1/4} x^{1/2} \left(1 - x/2 - x^2/8 - x^3/16 - 5x^4/128 - \dots\right) dx \\ &= \int_0^{1/4} \left(x^{1/2} - \frac{x^{3/2}}{2} - \frac{x^{5/2}}{8} - \frac{x^{7/2}}{16} - \frac{5x^{9/2}}{128} - \dots\right) dx. \end{aligned} \quad (3.1.8)$$

Integrating this series term-by-term and combining the above results yielded the formula

$$\pi = \frac{3\sqrt{3}}{4} + 24 \left( \frac{1}{3 \cdot 8} - \frac{1}{5 \cdot 32} - \frac{1}{7 \cdot 128} - \frac{1}{9 \cdot 512} - \dots \right). \quad (3.1.9)$$

Newton personally used this formula to compute  $\pi$ . He recorded 15 digits in his diary, but as he later confessed, “I am ashamed to tell you to how many figures I carried these computations, having no other business at the time.”

One of the more interesting figures in the history of  $\pi$  calculations is Johan Zacharias Dase, who lived in Vienna in the mid-1800s. According to the MacTutor web site (see above), Dase once demonstrated his computational skill by multiplying  $79532853 \times 93758479 = 7456879327810587$  in just 54 seconds; two 20-digit numbers required 6 minutes; two 40-digit numbers required 40 minutes; two 100-digit numbers required 8 hours 45 minutes. After being shown how to use the formula

$$\pi/4 = \arctan(1/2) + \arctan(1/5) + \arctan(1/8), \quad (3.1.10)$$

he calculated  $\pi$  to 200 places in his head. He later calculated a seven-digit logarithm table, and extended a table of integer factorizations to 10,000,000. Gauss requested that Dase be permitted to assist him, but Dase died shortly afterward.

Along this line, recent historical research has uncovered several non-European mathematicians who did notable early work on  $\pi$  that until recently was not well known. For example, the Chinese mathematician Tsu Chung-Chih used a variation of Archimedes’ method to compute  $\pi$  correct to seven digits in the fifth century. A millennium later Al-Kāshi in Samarkand, “who could calculate as eagles can fly,” computed  $2\pi$  in sexagesimal as

$$6 + \frac{16}{60} + \frac{59}{60^2} + \frac{28}{60^3} + \frac{01}{60^4} + \frac{34}{60^5} + \frac{51}{60^6} + \frac{46}{60^7} + \frac{14}{60^8} + \frac{50}{60^9}$$

(accurate to seventeen decimal places) using  $3 \cdot 2^{28}$ -gons. Another remarkable but little-known mathematician was Madhava of Sangamagramma, who lived in India in the late 1300s and early 1400s. Madhava derived the Gregory-Leibniz series for  $\pi$  (3.1.3) at least 200 years before Gregory or Leibniz found it. He also derived the variant series

$$\pi = \sqrt{12} \left( 1 - \frac{1}{3 \cdot 3} + \frac{1}{5 \cdot 3^2} - \frac{1}{7 \cdot 3^3} + \cdots \right), \quad (3.1.11)$$

and used it to compute  $\pi$  to 11 decimal places. For additional information, see [34] and the MacTutor web site.



One motivation for computations of  $\pi$  was very much in the spirit of modern experimental mathematics: to see if the decimal expansion of  $\pi$  repeats, which would mean that  $\pi$  is the ratio of two integers (i.e., rational), or to recognize  $\pi$  as some algebraic constant. The question of the rationality of  $\pi$  was settled in the late 1700s, when Lambert and Legendre proved that the constant is irrational. The question of whether  $\pi$  is algebraic was settled in 1882, when Lindemann proved that it is transcendental. Lindemann's proof also settled, once and for all, the ancient Greek question of whether the circle could be squared with ruler and compass: It cannot, because numbers that are the lengths of lines that can be constructed using ruler and compasses (often called constructible numbers) are necessarily algebraic, and squaring the circle is equivalent to constructing the value  $\pi$ .

In 1945, Ferguson utilized some mechanical calculating machines to compute  $\pi$  to 530 decimal digits, using the formula

$$\frac{\pi}{4} = 3 \arctan\left(\frac{1}{4}\right) + \arctan\left(\frac{1}{20}\right) + \arctan\left(\frac{1}{1985}\right), \quad (3.1.12)$$

and over the next two years increased this to 808 digits. With this result, they discovered that Shanks' 1874 computation was wrong after the first 527 places.

While such devices saved considerable labor compared with pencil-and-paper arithmetic, they were still a far cry from modern high-speed computers—each operand had to be manually typed, for example. The first truly electronic computation of  $\pi$  was performed in 1949 on the original ENIAC. This calculation was suggested by John von Neumann, who wished to study the digits of  $\pi$  and  $e$ . Computing 2037 decimal places of  $\pi$  on the ENIAC required 70 hours. A similar calculation today could be performed in a fraction of second on a personal computer.

Later computer calculations were further accelerated by the discovery of advanced algorithms for performing the required high-precision arithmetic operations. For example, in 1965 it was found that the newly-discovered fast Fourier transform (FFT) could be used to perform high-precision multiplications much more rapidly than conventional schemes. These methods dramatically lowered the computer time required for computing  $\pi$  and other mathematical constants to high precision. These techniques will be discussed more in Section 6.2. See also [10] and [58].

In spite of these advances, until the 1970s all computer evaluations of  $\pi$  still employed classical formulas, usually one of the Machin-type formulas. Some

Name	Year	Digits	Value
Babylonians	2000? BCE	1	$3.125 (3\frac{1}{8})$
Egyptians	2000? BCE	1	$3.16045 (4(\frac{8}{9})^2)$
China	1200? BCE	1	3
Bible (1 Kings 7:23)	550? BCE	1	3
Archimedes	250? BCE	3	3.1418 (ave.)
Chang Hong	130 CE	1	$3.1622 (\sqrt{10})$
Ptolemy	150 CE	3	3.14166
Wang Fau	250 CE	1	$3.15555 (\frac{142}{45})$
Liu Hui	263 CE	5	3.14159
Siddhanta	380 CE	3	3.1416
Tsu Ch'ung Chi	480 CE	7	$3.1415926 (355/113)$
Aryabhata	499 CE	4	$3.14156 (62832/2000)$
Brahmagupta	640 CE	1	$3.162277 (= \sqrt{10})$
Al-Khowarizmi	800 CE	4	3.1416
Fibonacci	1220 CE	3	3.141818
Madhava	1400 CE	11	3.14159265359
Al-Kāshi	1429 CE	14	3.14159265358979
Otho	1573 CE	6	3.1415929
Viéte	1593 CE	9	$3.1415926536 (ave.)$
Romanus	1593 CE	15	3.141592653589793
Van Ceulen	1596 CE	20	3.14159265358979323846
Van Ceulen	1615 CE	35	

Table 3.1: Pre-calculus era  $\pi$  calculations.

Name	Year	Digits	
Newton	1665	16	
Sharp	1699	71	
Seki Kowa	1700	10	
Kamata	1730	25	
Machin	1706	100	
De Lagny	1719	127	(112 correct)
Takebe	1723	41	
Matsunaga	1739	50	
Von Vega	1794	140	(136 correct)
Rutherford	1824	208	(152 correct)
Strassnitzky and Dase	1844	200	
Clausen	1847	248	
Lehmann	1853	261	
Rutherford	1853	440	
Shanks	1874	707	(527 correct)

Table 3.2: Calculus era  $\pi$  calculations.

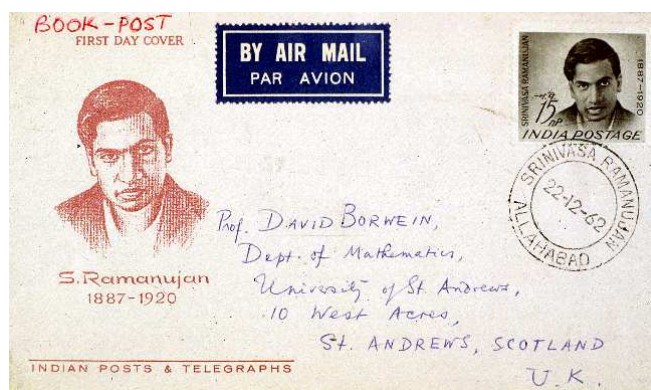


Figure 3.2: Ramanujan's 75th birthday stamp.

new infinite series formulas were discovered by Ramanujan around 1910, but these were not well known until quite recently when his writings were widely published. Ramanujan's related mathematics may be followed in [109, 51, 37]. One of these series is the remarkable formula

$$\frac{1}{\pi} = \frac{2\sqrt{2}}{9801} \sum_{k=0}^{\infty} \frac{(4k)!(1103 + 26390k)}{(k!)^4 396^{4k}}. \quad (3.1.13)$$

Each term of this series produces an additional *eight* correct digits in the result. Gosper used this formula to compute 17 million digits of  $\pi$  in 1985. Gosper also computed the first 17 million terms of the continued fraction expansion of  $\pi$ . At about the same time, David and Gregory Chudnovsky found the following variation of Ramanujan's formula:

$$\frac{1}{\pi} = 12 \sum_{k=0}^{\infty} \frac{(-1)^k (6k)!(13591409 + 545140134k)}{(3k)!(k!)^3 640320^{3k+3/2}}. \quad (3.1.14)$$

Each term of this series produces an additional 14 correct digits. The Chudnovskys implemented this formula using a clever scheme that enabled them to utilize the results of an initial level of precision to extend the calculation to even higher precision. They used this method in several large calculations of  $\pi$ , culminating with a computation to over four billion decimal digits in 1994.

Along this line, it is interesting to note that the Ramanujan-type series (see [50, pg. 188])

$$\frac{1}{\pi} = \sum_{n=0}^{\infty} \left( \frac{\binom{2n}{n}}{16^n} \right)^3 \frac{42n + 5}{16} \quad (3.1.15)$$

permits one to compute the billionth binary digit of  $1/\pi$  without computing the first half of the series.

While the Ramanujan and Chudnovsky series are considerably more efficient than the classical formulas, they share with them the property that the number of terms one must compute increases linearly with the number of digits desired in the result. In other words, if you want to compute  $\pi$  to twice as many digits, you have to evaluate twice as many terms of the series.

In 1976, Eugene Salamin and Richard Brent independently discovered an algorithm for  $\pi$  based on the arithmetic-geometric mean (AGM) and some ideas originally due to Gauss in the 1800s (although for some reason Gauss never saw the connection to computing  $\pi$ ). The Salamin–Brent algorithm may be stated as follows. Set  $a_0 = 1$ ,  $b_0 = 1/\sqrt{2}$  and  $s_0 = 1/2$ . Calculate

$$\begin{aligned} a_k &= \frac{a_{k-1} + b_{k-1}}{2} \\ b_k &= \sqrt{a_{k-1}b_{k-1}} \\ c_k &= a_k^2 - b_k^2 \\ s_k &= s_{k-1} - 2^k c_k \\ p_k &= \frac{2a_k^2}{s_k}. \end{aligned} \quad (3.1.16)$$

Then  $p_k$  converges *quadratically* to  $\pi$ : Each iteration of this algorithm approximately *doubles* the number of correct digits—successive iterations produce 1, 4,

9, 20, 42, 85, 173, 347, and 697 correct decimal digits of  $\pi$ . Twenty-five iterations are sufficient to compute  $\pi$  to over 45 million decimal digit accuracy. However, each of these iterations must be performed using a level of numeric precision that is at least as high as that desired for the final result.

Beginning in 1985, one of the present authors (Jonathan Borwein) and his brother Peter Borwein discovered some additional algorithms of this type [50]. One is as follows. Set  $a_0 = 1/3$  and  $s_0 = (\sqrt{3} - 1)/2$ . Iterate

$$\begin{aligned} r_{k+1} &= \frac{3}{1 + 2(1 - s_k^3)^{1/3}} \\ s_{k+1} &= \frac{r_{k+1} - 1}{2} \\ a_{k+1} &= r_{k+1}^2 a_k - 3^k (r_{k+1}^2 - 1). \end{aligned} \quad (3.1.17)$$

Then  $1/a_k$  converges *cubically* to  $\pi$ —each iteration approximately triples the number of correct digits. Another algorithm is as follows: Set  $a_0 = 6 - 4\sqrt{2}$  and  $y_0 = \sqrt{2} - 1$ . Iterate

$$\begin{aligned} y_{k+1} &= \frac{1 - (1 - y_k^4)^{1/4}}{1 + (1 - y_k^4)^{1/4}} \\ a_{k+1} &= a_k (1 + y_{k+1})^4 - 2^{2k+3} y_{k+1} (1 + y_{k+1} + y_{k+1}^2). \end{aligned} \quad (3.1.18)$$

Then  $a_k$  converges *quartically* to  $1/\pi$ . This particular algorithm, together with the Salamin–Brent scheme, has been employed by Yasumasa Kanada of the University of Tokyo in several computations of  $\pi$  over the past 15 years or so, including his 1999 computation of  $\pi$  to more than 206 billion decimal digits.

Daniel Shanks, who in 1961 computed  $\pi$  to over 100,000 digits, once declared that a billion digit computation would be “forever impossible.” But both Kanada and the Chudnovskys computed more than one billion digits in 1989. Similarly, the intuitionist mathematicians Brouwer and Heyting once asserted the “impossibility” of ever knowing whether the sequence “0123456789” appears in the decimal expansion of  $\pi$  [116]. This sequence was found in 1997 by Kanada, beginning at position 17,387,594,880. Even as late as 1989, British mathematical physicist Roger Penrose ventured, in the first edition of his book *The Emperor’s New Mind*, that we are not likely to know whether a string of ten consecutive sevens occurs in the decimal expansion of  $\pi$  [162, pg. 115]. By the time his book was published, Kanada had already found a string of ten consecutive sixes

in his 480-million-digit computation of  $\pi$ . When one of the present authors mentioned this to Penrose in 1990, he replied that he was “startled to learn how far the combination of human mathematical ingenuity with computer technology has enabled the calculation of the decimal expansion of  $\pi$  to be carried out.” Accordingly, he changed his text to “twenty consecutive sevens,” which appeared in subsequent printings of the book. This was just in time, as a string of ten consecutive sevens was found by Kanada in 1997, beginning at position 22,869,046,249.

As this book was nearing completion (2003), the authors learned that Kanada, with a team consisting of Y. Ushiro of Hitachi, H. Kuroda, and M. Kudoh of the University of Tokyo, and the assistance of nine others from Hitachi, have now computed  $\pi$  to over 1.24 *trillion* decimal digits. Kanada and his team first computed  $\pi$  in hexadecimal (base 16) to 1,030,700,000,000 places, using the following two arctangent relations for  $\pi$ :

$$\begin{aligned}\pi &= 48 \arctan \frac{1}{49} + 128 \arctan \frac{1}{57} - 20 \arctan \frac{1}{239} + 48 \arctan \frac{1}{110443} \\ \pi &= 176 \arctan \frac{1}{57} + 28 \arctan \frac{1}{239} - 48 \arctan \frac{1}{682} + 96 \arctan \frac{1}{12943}.\end{aligned}\tag{3.1.19}$$

The first formula was found in 1982 by K. Takano, a high school teacher and song writer. The second formula was found by F. C. W. Störmer in 1896.

Kanada and his team evaluated these formulas using a scheme analogous to that employed by Gosper and the Chudnovskys, in that they were able to avoid explicitly storing the multiprecision numbers involved. This resulted in a scheme that is roughly competitive in efficiency compared to the Salamin-Brent and Borwein quartic algorithms they had previously used, yet with a significantly lower total memory requirement. In particular, they were able to perform their latest computation on a system with 1 Tbyte ( $10^{12}$  bytes) main memory, the same as with their previous computation, yet obtain six times as many digits.

After Kanada and his team verified that the hexadecimal digit strings produced by these two computations were in agreement, they performed an additional check by directly computing 20 hexadecimal digits beginning at position 1,000,000,000,001. This calculation, which employed an algorithm that we shall describe in Section 3.4, required 21 hours, much less than the time required for the first step. The result of this calculation, B4466E8D21 5388C4E014, perfectly

agreed with the corresponding digits produced by the two arctan formulas. At this point they converted their hexadecimal value of  $\pi$  to decimal, and converted back to hexadecimal as a check. These conversions employed a numerical approach similar to that used in the main and verification calculations. The entire computation, including hexadecimal and decimal evaluations and checks, required roughly 600 hours on their 64-node Hitachi parallel supercomputer. The main segment of the computation ran at nearly 1 Tflop/s (i.e., one trillion floating-point operations per second), although this performance rate was slightly lower than the rate of their previous calculation of 206 billion digits. Full details will appear in an upcoming paper [126].

According to Kanada, the ten decimal digits ending in position one trillion are 6680122702, while the ten hexadecimal digits ending in position one trillion are 3F89341CD5. Some data on the frequencies of digits in  $\pi$ , based on Kanada's computations, are given in Section 4.1. Additional information of this sort is available from Kanada's web site: <http://www.super-computing.org>. Additional historical background on record-breaking computations of  $\pi$  is available at [http://www.cecm.sfu.ca/personal/jborwein/pi\\_cover.html](http://www.cecm.sfu.ca/personal/jborwein/pi_cover.html). A listing of some milestones in the history of the computation of  $\pi$  is given in Tables 3.1, 3.2, and 3.3.

In retrospect, one might wonder why in antiquity  $\pi$  was not measured to an accuracy in excess of  $22/7$ . One conjecture is that it reflects not an inability to do so, but instead a very different mind set to a modern (Baconian) experimental one. Along this line, Marty Gerriets, an acquaintance of one of the present authors, may provide an explanation:

I work at a youth emergency shelter and my students (right now) are 11 to 17. My youngest were as accurate as my oldest. They measured circles that I had drawn on paper . . . and used rulers to measure the string. It was probably coincidental, but they were consistently coming up with 3.1, 3.15, 3.11. That's pretty good for string and rulers. You have to understand that they didn't have a clue that we were finding pi. These students have very poor educational backgrounds. Those that had heard of it thought it was some magic number that only geometry teachers know about. The concept that is just a ratio between diameter and circumference had never been presented to any of them. For my convenience, I had made the circles with diameters of 2, 3, 4, 5 and 7. I had them measure and record the diameter,



Name	Year	Digits
Ferguson	1946	620
Ferguson	1947	710
Ferguson and Wrench	1947	808
Smith and Wrench	1949	1,120
Reitwiesner et al. (ENIAC)	1949	2,037
Nicholson and Jeanel	1954	3,092
Felton	1957	7,480
Genuys	1958	10,000
Felton	1958	10,021
Guilloud	1959	16,167
Shanks and Wrench	1961	100,265
Guilloud and Filliatre	1966	250,000
Guilloud and Dichampt	1967	500,000
Guilloud and Bouyer	1973	1,001,250
Miyoshi and Kanada	1981	2,000,036
Guilloud	1982	2,000,050
Tamura	1982	2,097,144
Tamura and Kanada	1982	4,194,288
Tamura and Kanada	1982	8,388,576
Kanada, Yoshino and Tamura	1982	16,777,206
Ushiro and Kanada	Oct. 1983	10,013,395
Gosper	Oct. 1985	17,526,200
Bailey	Jan. 1986	29,360,111
Kanada and Tamura	Sep. 1986	33,554,414
Kanada and Tamura	Oct. 1986	67,108,839
Kanada, Tamura, Kubo, et. al	Jan. 1987	134,217,700
Kanada and Tamura	Jan. 1988	201,326,551
Chudnovskys	May 1989	480,000,000
Chudnovskys	Jun. 1989	525,229,270
Kanada and Tamura	Jul. 1989	536,870,898
Kanada and Tamura	Nov. 1989	1,073,741,799
Chudnovskys	Aug. 1989	1,011,196,691
Chudnovskys	Aug. 1991	2,260,000,000
Chudnovskys	May 1994	4,044,000,000
Takahashi and Kanada	Jun. 1995	3,221,225,466
Kanada	Aug. 1995	4,294,967,286
Kanada	Oct. 1995	6,442,450,938
Kanada and Takahashi	Jun. 1997	51,539,600,000
Kanada and Takahashi	Sep. 1999	206,158,430,000
Kanada, Ushiro, Kuroda, Kudoh	Dec. 2002	1,241,100,000,000

Table 3.3: Digital era  $\pi$  calculations.

then loop the string around the circle and record the circumference. Then I had them go back and find the ratio in each case. Of course, it came to about 3 in each case and they were astounded! From there we went to a historical perspective of pi and they created their own formula for finding circumference.

In the same vein, one likely reason that Gauss and Ramanujan did not further develop the ideas in their identities for  $\pi$  is that an iterative algorithm, as opposed to explicit results, was not satisfactory for them (especially Ramanujan). Ramanujan would much have preferred formulae such as  $\pi \approx 3 \log 5280/\sqrt{67}$  or  $\pi \approx 3 \log 640320/\sqrt{163}$ .

## 3.2 Fascination with Pi

One might wonder what has been the motivation behind these modern computations of  $\pi$ , given that questions such as the irrationality and transcendence of  $\pi$  were settled more than 100 years ago. Doubtless one motivation is simply the raw challenge of harnessing the stupendous power now available in modern computer systems for this classical computational problem. We should add that programming such calculations is definitely not trivial, especially on large, distributed memory computer systems. There have been several practical spin-off benefits from these efforts. For example, some new techniques for performing the fast Fourier transform (FFT), which is heavily used in modern science and engineering computing, had their roots in attempts to accelerate computations of  $\pi$ .

What's more, it has long been recognized that the computation of  $\pi$  is a very strong test of computer system integrity. As we pointed out both in Section 1.3 and in our discussion above on Kanada's calculation, such computations are typically checked by repeating the computation with a different formula, or even with a completely different computation approach. When two independent computations produce the same result, this is rather strong evidence that the computer has performed trillions of arithmetic operations, and transferred trillions of bytes of data between different components of the computer system, without error. In 1985, one of the present authors used two computations of  $\pi$  to test a new Cray-2 supercomputer at NASA's Ames Research Center, and these computations in fact disclosed at least one significant hardware defect. Testing

computer integrity and performance is Yasumasa Kanada's primary motivation in the computations he has done on Hitachi supercomputers.

Beyond purely practical considerations is the continuing interest in the fundamental question of the normality (digit randomness) of  $\pi$ . Kanada, for example, has performed detailed statistical analysis of his computed results of  $\pi$  to see if there are any statistical abnormalities that suggest  $\pi$  is not normal. We shall discuss this question more fully in Section 4.1.

But there is no doubt that one motivation for computations of  $\pi$  is the eternal fascination with this most prominent and mysterious of mathematical constants, a fascination shared both by professional scientists and mathematicians as well as the lay public. At the present time there are hundreds of web sites devoted to  $\pi$  (see, for example, <http://www.piworld.de>), and thousands of online research papers and other documents dealing with  $\pi$ . There are even several Internet-based clubs devoted to  $\pi$ , one of which requires that initiates recite by heart the first 100 digits of  $\pi$  with "respect, fluency, and smoothness" (see <http://pi314.at>). The very popular web-site <http://www.cecm.sfu.ca/pi/yapPing.html> recites  $\pi$  in many different languages.

Fascination with  $\pi$  is also evident in the many recent popular books, television shows, and movies that have mentioned  $\pi$ . In the 1967 *Star Trek* episode "Wolf in the Fold," Spock foiled the evil computer by telling it to compute the last digit of  $\pi$ . A May 1993 segment of *The Simpsons* featured the character Apu boasting that he can recite  $\pi$  to 40,000 digits, and that the last digit is a "1" (one of the present authors was asked by the show's writers for this digit—see Figure 3.3). In November 1996, the MSNBC network aired a news segment about  $\pi$ , including interviews with one of the present authors and several other mathematicians at Simon Fraser University. The 1997 movie *Contact*, starring Jodie Foster, was based on the 1986 book by noted astronomer Carl Sagan. In the book version, the lead female character searched for patterns in the digits of  $\pi$ , and after her mysterious experience found confirmation in the base-11 expansion of  $\pi$  [176]. A 1997 book entitled *The Joy of Pi* sold thousands of copies [39]. A 1998 movie entitled *Pi* began with decimal digits of  $\pi$  displayed on the screen. The 2001 Academy Award-winning movie *A Beautiful Mind*, which portrayed the life of mathematician John Nash, shows  $\pi$  in equations scribbled by the lead actor Russell Crowe (although the  $\pi$  characters here appear to denote the prime-counting function and other nongeometric entities). The 2002 book *Life of Pi* by Canadian author Yann Martel won the Booker Prize, a prestigious British

literary award [152]. Finally, in the 2003 movie *Matrix Reloaded*, the Key Maker warns that a door will be accessible for exactly 314 seconds, a figure that *Time* speculated was a reference to  $\pi$  [77].

Along this line, National Public Radio reported on April 12, 2003 that novelty automatic teller machine withdrawal slips, showing a balance of \$314,159.26, are popular in New York City. One can scribble a note on the back and, apparently innocently, let the intended target be suitably impressed by one's amazingly healthy saving account. Scott Simon, the host of this NPR show, noted the close resemblance to  $\pi$ .

For those who know *The Hitchhiker's Guide to the Galaxy*, it is amusing that 042 occurs at the digits ending at the fifty-billionth decimal place in each of  $\pi$  and  $1/\pi$ —thereby providing an excellent answer to the ultimate question, “What is forty two?”

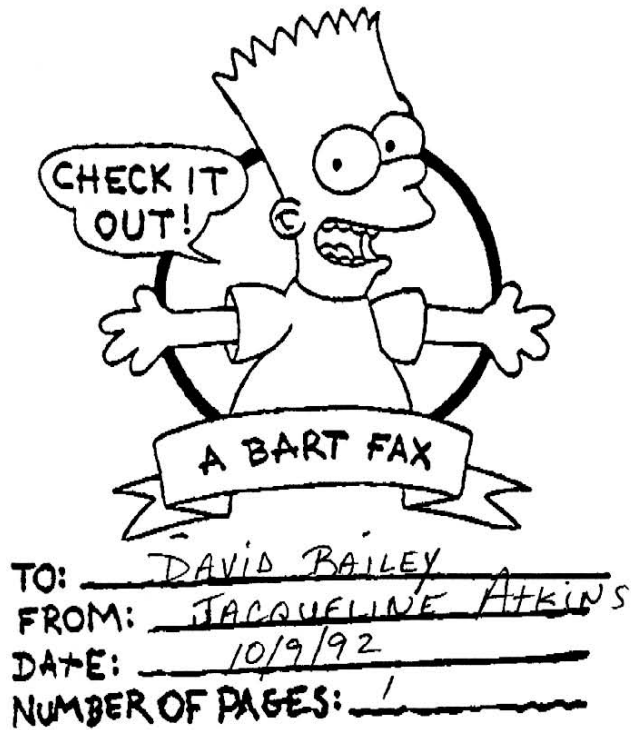
It is hard to imagine  $e$  or  $\log 2$  playing the same role. That said, the publicity surrounding the Millennium problems has led to a spate of recent popular books on the Riemann Hypothesis. The million dollar prize for the Riemann Hypothesis even formed the basis of a recent episode of the television program *Law and Order*, in which student notebooks contained quite accurate formulas.

### 3.3 Behind the Cubic and Quartic Iterations

The genesis of the  $\pi$  algorithms (3.1.13), (3.1.17), (3.1.18) and related material is itself an illustrative example of experimental mathematics. For nonnegative integer  $N$ , the function

$$\alpha(N) = \frac{E'(k_N)}{K(k_N)} - \frac{\pi}{4K^2(k_N)} \quad (3.3.20)$$

had arisen, where  $k_N$  is the  $N$ -th singular value, and  $K$  and  $E'$  are complete elliptic integrals as detailed in the discussion of Legendre's relation in Section 5.6.1. For present purposes, it suffices that (3.3.20) is very easy to compute. For



FAX (310) 203-3852  
 PHONE (310) 203-3959

A Professor at UCLA told me that you might be able to give me the answer to: What is the 40,000<sup>th</sup> digit of  $\pi$ ?

We would like to use the answer in our show. Can you help?

Figure 3.3: Fax from *The Simpsons* show.

example, the first few non-composite values are (to 20 digit accuracy):

$$\begin{aligned}\alpha(1) &\approx 0.49999999999999999999 \\ \alpha(2) &\approx 0.41421356237309504880 \\ \alpha(3) &\approx 0.36602540378443864678 \\ \alpha(5) &\approx 0.33188261099247156221 \\ \alpha(7) &\approx 0.32287565553229529536.\end{aligned}$$

It is obvious that  $\alpha(1) = 1/2$  and easy to spot that  $\alpha(2) = \sqrt{2}-1$ , from which it was quickly observed that  $\alpha(3) = (\sqrt{3}-1)/2$  and that  $\alpha(7) = (\sqrt{7}-2)/2$ , but  $\alpha(5)$  did not appear to be a quadratic. Twenty years ago, such identification was not as easy, and it was only when it occurred to the Borweins that quadratic fields congruent to  $\pm 1 \pmod{4}$  behave differently that they stumbled upon (experimentally) the identity

$$\alpha(5) = \frac{\sqrt{5} - \sqrt{2\sqrt{5}-2}}{2}. \quad (3.3.21)$$

Nowadays, this is almost trivial: A “Minpoly” calculation immediately returns  $29 - 80x - 24x^2 + 16x^4 = 0$  and this has the surd above as its smallest positive root. At this point, the authors could use known results only to prove the value of  $\alpha(1)$ ,  $\alpha(2)$  and  $\alpha(3)$ . Those for  $\alpha(5)$  and  $\alpha(7)$  remained conjectural. There was, however, an empirical family of algorithms for  $\pi$ : let  $\alpha_0 = \alpha(N)$  and  $k_0 = k'_N$  (where  $k' = \sqrt{1-k^2}$ ) and iterate

$$k_{n+1} = \frac{1 - k'_n}{1 + k'_n} \quad (3.3.22)$$

$$\alpha_{n+1} = (1 + k_{n+1})^2 \alpha_n - \sqrt{N} 2^{n+1} k_{n+1}. \quad (3.3.23)$$

Then

$$\lim_{n \rightarrow \infty} \alpha_n^{-1} = \pi. \quad (3.3.24)$$

Again, (3.3.24) was provable for  $N = 1, 2, 3$  and only conjectured for  $N = 5, 7$ . In each case the algorithm *appeared* to converge quadratically to  $\pi$ . On closer inspection while the provable cases were correct to 5,000 digits, the empirical ones agreed with  $\pi$  to roughly 100 places only. Now, in many ways to have

discovered a “natural” number that agreed with  $\pi$  to that level—and no more—would have been more interesting than the alternative. That seemed unlikely, but recoding and rerunning the iterations kept producing identical results.

Twenty years ago, very high-precision calculation was less accessible, and the code was being run in a Berkeley Unix integer package. After about six weeks of effort, it was found that the square root algorithm in the package was badly flawed, but only if run with an odd precision of more than 60 digits! And for idiosyncratic reasons that had only been the case in the two unproven cases. Needless to say, tracing the bug was a salutary and somewhat chastening experience.

### 3.4 Computing Individual Digits of Pi

An outsider might be forgiven for thinking that essentially everything of interest with regards to  $\pi$  has been discovered. For example, this sentiment is suggested in the closing chapters of Beckmann’s 1971 book on the history of  $\pi$  [33, pg. 172]. Ironically, the Salamin–Brent quadratically convergent iteration was discovered only five years later, and the higher-order convergent algorithms followed in the 1980s. In 1990, Rabinowitz and Wagon discovered a “spigot” algorithm for  $\pi$ , which permits successive digits of  $\pi$  (in any desired base) to be computed with a relatively simple recursive algorithm based on the previously generated digits (see [168] and Item 15 at the end of this chapter).

But even insiders are sometimes surprised by a new discovery. Prior to 1996, almost all mathematicians believed that if you want to determine the  $d$ -th digit of  $\pi$ , you have to generate the entire sequence of the first  $d$  digits. (For all of their sophistication and efficiency, the schemes described above all have this property.) But it turns out that this is not true, at least for hexadecimal (base 16) or binary (base 2) digits of  $\pi$ . In 1996, Peter Borwein, Simon Plouffe, and one of the present authors (Bailey) found an algorithm for computing individual hexadecimal or binary digits of  $\pi$  [21]. To be precise, this algorithm:

- (1) directly produces a modest-length string of digits in the hexadecimal or binary expansion of  $\pi$ , beginning at an arbitrary position, without needing to compute any of the previous digits;
- (2) can be implemented easily on any modern computer;

- (3) does not require multiple precision arithmetic software;
- (4) requires very little memory; and
- (5) has a computational cost that grows only slightly faster than the digit position.

Using this algorithm, for example, the one millionth hexadecimal digit (or the four millionth binary digit) of  $\pi$  can be computed in less than a minute on a 2001-era computer. The new algorithm is not fundamentally faster than best-known schemes for computing all digits of  $\pi$  up to some position, but its elegance and simplicity are nonetheless of considerable interest. This scheme is based on the following remarkable new formula for  $\pi$ :

**Theorem 3.4.1**

$$\pi = \sum_{i=0}^{\infty} \frac{1}{16^i} \left( \frac{4}{8i+1} - \frac{2}{8i+4} - \frac{1}{8i+5} - \frac{1}{8i+6} \right). \quad (3.4.25)$$

**Proof.** First note that for any  $k < 8$ ,

$$\begin{aligned} \int_0^{1/\sqrt{2}} \frac{x^{k-1}}{1-x^8} dx &= \int_0^{1/\sqrt{2}} \sum_{i=0}^{\infty} x^{k-1+8i} dx \\ &= \frac{1}{2^{k/2}} \sum_{i=0}^{\infty} \frac{1}{16^i(8i+k)}. \end{aligned} \quad (3.4.26)$$

Thus one can write

$$\begin{aligned} \sum_{i=0}^{\infty} \frac{1}{16^i} \left( \frac{4}{8i+1} - \frac{2}{8i+4} - \frac{1}{8i+5} - \frac{1}{8i+6} \right) \\ = \int_0^{1/\sqrt{2}} \frac{4\sqrt{2} - 8x^3 - 4\sqrt{2}x^4 - 8x^5}{1-x^8} dx, \end{aligned} \quad (3.4.27)$$

which on substituting  $y = \sqrt{2}x$  becomes

$$\begin{aligned} \int_0^1 \frac{16y-16}{y^4-2y^3+4y-4} dy &= \int_0^1 \frac{4y}{y^2-2} dy - \int_0^1 \frac{4y-8}{y^2-2y+2} dy \\ &= \pi. \end{aligned} \quad (3.4.28)$$



□

However, in presenting this formal derivation, we are disguising the actual route taken to the discovery of this formula. This route is a superb example of experimental mathematics in action.

It all began in 1995, when Peter Borwein and Simon Plouffe of Simon Fraser University observed that the following well-known formula for  $\log 2$  permits one to calculate isolated digits in the binary expansion of  $\log 2$ :

$$\log 2 = \sum_{k=0}^{\infty} \frac{1}{k2^k}. \quad (3.4.29)$$

This scheme is as follows. Suppose we wish to compute a few binary digits beginning at position  $d+1$  for some integer  $d > 0$ . This is equivalent to calculating  $\{2^d \log 2\}$ , where  $\{\cdot\}$  denotes fractional part. Thus we can write

$$\begin{aligned} \{2^d \log 2\} &= \left\{ \left\{ \sum_{k=0}^d \frac{2^{d-k}}{k} \right\} + \sum_{k=d+1}^{\infty} \frac{2^{d-k}}{k} \right\} \\ &= \left\{ \left\{ \sum_{k=0}^d \frac{2^{d-k} \bmod k}{k} \right\} + \sum_{k=d+1}^{\infty} \frac{2^{d-k}}{k} \right\}. \end{aligned} \quad (3.4.30)$$

We are justified in inserting “ $\bmod k$ ” in the numerator of the first summation, because we are only interested in the fractional part of the quotient when divided by  $k$ .

Now the key observation is this: The numerator of the first sum in Equation (3.4.30), namely  $2^{d-k} \bmod k$ , can be calculated very rapidly by means of the binary algorithm for exponentiation, performed modulo  $k$ . The binary algorithm for exponentiation is merely the formal name for the observation that exponentiation can be economically performed by means of a factorization based on the binary expansion of the exponent. For example, we can write  $3^{17} = (((((3^2)^2)^2)^2) \cdot 3$ , thus producing the result in only 5 multiplications, instead of the usual 16. According to Knuth, this technique dates back at least to 200 BCE [131, pg. 461]. In our application, we need to obtain the exponentiation result modulo a positive integer  $k$ . This can be done very efficiently by reducing modulo  $k$  the intermediate multiplication result at each step of the binary algorithm for exponentiation. A formal statement of this scheme is as follows:

**Algorithm 1** *Binary algorithm for exponentiation modulo  $k$ .*

To compute  $r = b^n \bmod k$ , where  $r, b, n$  and  $k$  are positive integers: First set  $t$  to be the largest power of two such that  $t \leq n$ , and set  $r = 1$ . Then

```

A: if  $n \geq t$  then  $r \leftarrow br \bmod k$ ;    $n \leftarrow n - t$ ;   endif
 $t \leftarrow t/2$ 
if  $t \geq 1$  then  $r \leftarrow r^2 \bmod k$ ;   go to A;   endif

```

Note that the above algorithm is performed entirely with positive integers that do not exceed  $k^2$  in size. Thus ordinary 64-bit floating-point or integer arithmetic, available on almost all modern computers, suffices for even rather large calculations. 128-bit floating-point arithmetic (double-double or quad precision), available at least in software on many systems (see Section 6.2.1), suffices for the largest computations currently feasible.

We can now present the algorithm for computing individual binary digits of  $\log 2$ .

**Algorithm 2** *Individual digit algorithm for  $\log 2$ .*

To compute the  $(d + 1)$ -th binary digit of  $\log 2$ : Given an integer  $d > 0$ , (1) calculate each numerator of the first sum in Equation (3.4.30), using Algorithm 1, implemented using ordinary 64-bit integer or floating-point arithmetic; (2) divide each numerator by the respective value of  $k$ , again using ordinary floating-point arithmetic; (3) sum the terms of the first summation, while discarding any integer parts; (4) evaluate the second summation as written using floating-point arithmetic—only a few terms are necessary since it rapidly converges; and (5) add the result of the first and second summations, discarding any integer part. The resulting fraction, when expressed in binary, gives the first few digits of the binary expansion of  $\log 2$  beginning at position  $d + 1$ .  $\square$

As soon as Borwein and Plouffe found this algorithm, they began seeking other mathematical constants that shared this property. It was clear that any constant  $\alpha$  of the form

$$\alpha = \sum_{k=0}^{\infty} \frac{p(k)}{q(k)2^k}, \quad (3.4.31)$$

where  $p(k)$  and  $q(k)$  are integer polynomials, with  $\deg p < \deg q$  and  $q$  having no zeroes at nonnegative integer arguments, is in this class. Further, any rational linear combination of such constants also shares this property. Checks of various mathematical references eventually uncovered about 25 constants that possessed series expansions of the form given by equation (3.4.31).

As you might suppose, the question of whether  $\pi$  also shares this property did not escape these researchers. Unfortunately, exhaustive searches of the mathematical literature did not uncover any formula for  $\pi$  of the requisite form. But given the fact that any rational linear combination of constants with this property also shares this property, Borwein and Plouffe performed integer relation searches to see if a formula of this type existed for  $\pi$ . This was done, using computer programs written by one of the present authors (Bailey), which implement the “PSLQ” integer relation algorithm in high-precision, floating-point arithmetic [92, 13]. We will discuss the PSLQ algorithm and related techniques more in Section 6.3.

In particular, these three researchers sought an integer relation for the real vector  $(\alpha_1, \alpha_2, \dots, \alpha_n)$ , where  $\alpha_1 = \pi$  and  $(\alpha_i, 2 \leq i \leq n)$  is the collection of constants of the requisite form gleaned from the literature, each computed to several hundred decimal digit precision. To be precise, they sought an  $n$ -long vector of integers  $(a_i)$  such that  $\sum_i a_i \alpha_i = 0$ , to within a very small “epsilon.” After a month or two of computation, with numerous restarts using new  $\alpha$  vectors (when additional formulas were found in the literature) the identity (3.4.25) was finally uncovered. The actual formula found by the computation was:

$$\pi = 4F(1/4, 5/4; 1; -1/4) + 2 \arctan(1/2) - \log 5, \quad (3.4.32)$$

where  $F(1/4, 5/4; 1; -1/4) = 0.955933837\dots$  is a hypergeometric function evaluation. Reducing this expression to summation form yields the new  $\pi$  formula:

$$\pi = \sum_{i=0}^{\infty} \frac{1}{16^i} \left( \frac{4}{8i+1} - \frac{2}{8i+4} - \frac{1}{8i+5} - \frac{1}{8i+6} \right). \quad (3.4.33)$$

To return briefly to the derivation of Formula (3.4.33), let us point out that it was discovered not by formal reasoning, or even by computer-based symbolic processing, but instead by numerical computations using a high-precision implementation of the PSLQ integer relation algorithm. It is most likely the first instance in history of the discovery of a new formula for  $\pi$  by a computer. We

might mention that, in retrospect, Formula (3.4.33) could be found much more quickly, by seeking integer relations in the vector  $(\pi, S_1, S_2, \dots, S_8)$ , where

$$S_j = \sum_{k=0}^{\infty} \frac{1}{16^k(8k+j)}. \quad (3.4.34)$$

Such a calculation could be done in a few seconds on a computer, even if one did not know in advance to use 16 in the denominator and 9 terms in the search, but instead had to stumble on these parameters by trial and error. But this observation is, as they say, 20-20 hindsight. The process of real mathematical discovery is often far more tortuous and less elegant than the polished version typically presented in textbooks and research journals.

It should be clear at this point that the scheme for computing individual hexadecimal digits of  $\pi$  is very similar to Algorithm 2. For completeness, we state it as follows:

**Algorithm 3** *Individual digit algorithm for  $\pi$ .*

To compute the  $(d+1)$ -th hexadecimal digit of  $\pi$ : Given an integer  $d > 0$ , we can write

$$\{16^d \pi\} = \{4\{16^d S_1\} - 2\{16^d S_4\} - \{16^d S_5\} - \{16^d S_6\}\}, \quad (3.4.35)$$

using the  $S_j$  notation of Equation (3.4.34). Now apply Algorithm 2, with

$$\begin{aligned} \{16^d S_j\} &= \left\{ \left\{ \sum_{k=0}^d \frac{16^{d-k}}{8k+j} \right\} + \sum_{k=d+1}^{\infty} \frac{16^{d-k}}{8k+j} \right\} \\ &= \left\{ \left\{ \sum_{k=0}^d \frac{16^{d-k} \bmod 8k+j}{8k+j} \right\} + \sum_{k=d+1}^{\infty} \frac{16^{d-k}}{8k+j} \right\} \end{aligned} \quad (3.4.36)$$

instead of equation (3.4.30), to compute  $\{16^d S_j\}$  for  $j = 1, 4, 5, 6$ . Combine these four results, discarding integer parts, as shown in (3.4.35). The resulting fraction, when expressed in hexadecimal notation, gives the hex digit of  $\pi$  in position  $d+1$ , plus a few more correct digits.  $\square$

As with Algorithm 2, multiple-precision arithmetic software is not required—ordinary 64-bit or 128-bit floating-point arithmetic suffices even for some rather

large computations. We have omitted here some numerical details for large computations—see [21]. Sample implementations in both C and Fortran-90 are available from the web site <http://www.expmath.info>.

One mystery that remains unanswered is why Formula (3.4.33) was not discovered long ago. As you can see from the above proof, there is nothing very sophisticated about its derivation. There is no fundamental reason why Euler, for example, or Gauss or Ramanujan, could not have discovered it. Perhaps the answer is that its discovery was a case of “reverse mathematical engineering.” Lacking a motivation to find such a formula, mathematicians of previous eras had no reason to derive one. But this still doesn’t answer the question of why the algorithm for computing individual digits of  $\log 2$  had not been discovered before—it is based on a formula, namely Equation (3.4.29), that has been known for centuries.

Needless to say, Algorithm 3 has been implemented by numerous researchers. In 1997, Fabrice Bellard of INRIA computed 152 binary digits of  $\pi$  starting at the trillionth binary digit position. The computation took 12 days on 20 workstations working in parallel over the Internet. His scheme is actually based on the following variant of 3.4.33:

$$\pi = 4 \sum_{k=0}^{\infty} \frac{(-1)^k}{4^k(2k+1)} - \frac{1}{64} \sum_{k=0}^{\infty} \frac{(-1)^k}{1024^k} \left( \frac{32}{4k+1} + \frac{8}{4k+2} + \frac{1}{4k+3} \right). \quad (3.4.37)$$

This formula permits individual hex or binary digits of  $\pi$  to be calculated roughly 43% faster than (3.4.25).

A year later, Colin Percival, then a 17-year-old student at Simon Fraser University, utilized a network of 25 machines to calculate binary digits in the neighborhood of position 5 trillion, and then in the neighborhood of 40 trillion. In September 2000, he found that the quadrillionth binary digit is “0,” based on a computation that required 250 CPU-years of run time, carried out using 1,734 machines in 56 countries. Table 3.4 gives some results known as of this writing.

One question that immediately arises in the wake of this discovery is whether or not there is a formula of this type and an associated computational scheme to compute individual *decimal* digits of  $\pi$ . Searches conducted by numerous researchers have been unfruitful. Now it appears that there is no nonbinary

Position	Hex Digits Beginning at This Position
$10^6$	26C65E52CB4593
$10^7$	17AF5863EFED8D
$10^8$	ECB840E21926EC
$10^9$	85895585A0428B
$10^{10}$	921C73C6838FB2
$10^{11}$	9C381872D27596
$1.25 \times 10^{12}$	07E45733CC790B
$2.5 \times 10^{14}$	E6216B069CB6C1

Table 3.4: Computed hexadecimal digits of  $\pi$ .

formula of this type—this is ruled out by a new result co-authored by one of the present authors (see Section 3.7) [49]. However, none of this removes the possibility that there exists some completely different approach that permits rapid computation of individual decimal digits of  $\pi$ . Also, as we will see in the next section, there do exist formulas for certain other constants that admit individual digit calculation schemes in various nonbinary bases (including base ten).

### 3.5 Unpacking the BBP Formula for Pi

It is worth asking “why” the formula

$$\pi = \sum_{i=0}^{\infty} \frac{1}{16^i} \left( \frac{4}{8i+1} - \frac{2}{8i+4} - \frac{1}{8i+5} - \frac{1}{8i+6} \right) \quad (3.5.38)$$

exists. As observed above, this identity is equivalent to, and can be proved by establishing:

$$\pi = \int_0^{1/\sqrt{2}} \frac{4\sqrt{2} - 8x^3 - 4\sqrt{2}x^4 - 8x^5}{1-x^8} dx.$$

The present version of *Maple* evaluates this integral to

$$-2 \log 2 + 2 \log(2 - \sqrt{2}) + \pi + 2 \log(2 + \sqrt{2}), \quad (3.5.39)$$

which simplifies to  $\pi$ . In any event, one can ask what the individual series in (3.5.38) comprise. So consider

$$S_b = \sum_{k=0}^{\infty} \frac{1}{16^k(8k+b)}$$

for  $1 \leq b \leq 8$  and the corresponding normalized integrals

$$I(b) = 2^{b/2} \int_0^{1/\sqrt{2}} \frac{x^{b-1}}{1-x^8} dx. \quad (3.5.40)$$

Again, *Maple* provides closed forms for  $I(b)$  in which the basic quantities seem to be the following:  $\arctan(2)$ ,  $\arctan(1/2)$ ,  $\sqrt{2} \arctan(1/\sqrt{2})$ ,  $\log(2)$ ,  $\log(3)$ ,  $\log(5)$ , and  $\log(\sqrt{2} \pm 1)$ . At this point one may use integer relation methods and obtain:

$$\begin{aligned} S_1 &= \frac{\pi}{8} + \frac{\log 5}{8} - \frac{\sqrt{2} \log(\sqrt{2}-1)}{4} - \frac{\arctan(1/2)}{4} + \frac{\sqrt{2} \arctan(\sqrt{2}/2)}{4} \\ S_2 &= \frac{\log(3)}{4} + \frac{\arctan(1/2)}{2} \\ S_3 &= \frac{\pi}{4} - \frac{\sqrt{2} \log(\sqrt{2}-1)}{2} - \frac{\arctan(1/2)}{2} - \frac{\sqrt{2} \arctan(\sqrt{2}/2)}{2} - \frac{\log 5}{4} \\ S_4 &= \frac{\log 5}{2} - \frac{\log 3}{2} \\ S_5 &= -\frac{\pi}{2} - \sqrt{2} \log(\sqrt{2}-1) + \arctan(1/2) + \sqrt{2} \arctan(\sqrt{2}/2) - \frac{\log 5}{2} \\ S_6 &= \log 3 - 2 \arctan(1/2) \\ S_7 &= -\pi + \log 5 - 2\sqrt{2} \log(\sqrt{2}-1) + 2 \arctan(1/2) - 2\sqrt{2} \arctan(\sqrt{2}/2) \\ S_8 &= 8 \log 2 - 2 \log 5 - 2 \log 3. \end{aligned} \quad (3.5.41)$$

Thus the “simple” hexadecimal formula (3.5.38) is actually a molecule made up of more subtle hexadecimal atoms: with the final bond coming from the simple identity  $\arctan 2 + \arctan(1/2) = \pi/2$ . As an immediate consequence, one obtains the formula  $\arctan(1/2) = S_2 - S_6/4$ .

Furthermore, the facts that

$$\begin{aligned} \operatorname{Im} \left( \log \left( 1 - \frac{1-i}{x} \right) \right) &= \arctan \left( \frac{1}{1-x} \right) \\ 2 \arctan(1/3) + \arctan(1/7) &= \arctan(1/2) + \arctan(1/3) \\ &= \arctan 1 = \pi/4 \end{aligned} \quad (3.5.42)$$

allow one to write directly a base-64 series for  $\arctan(1/3)$  (using  $x = 4$ ) and a base-1024 series for  $\arctan(1/7)$  (using  $x = 8$ ). This yields the identity

$$\begin{aligned} \frac{\pi}{4} &= \frac{1}{16} \sum_{n=0}^{\infty} \frac{(-1)^n}{64^n} \left( \frac{8}{4n+1} + \frac{4}{4n+2} + \frac{1}{4n+3} \right) \\ &+ \frac{1}{256} \sum_{n=0}^{\infty} \frac{(-1)^n}{1024^n} \left( \frac{32}{4n+1} + \frac{8}{4n+2} + \frac{1}{4n+3} \right), \end{aligned} \quad (3.5.43)$$

which is similar to, although distinct from, the identity used by Bellard and Percival in their computations.

### 3.6 Other BBP-Type Formulas

A formula of the type mentioned in the previous sections, namely

$$\alpha = \sum_{k=0}^{\infty} \frac{p(k)}{b^k q(k)}, \quad (3.6.44)$$

is now referred to as a BBP-type formula, named after the initials of the authors of the 1997 paper where the  $\pi$  hex digit algorithm appeared [21]. For a constant  $\alpha$  given by a formula of this type, it is clear that individual base- $b$  digits can be calculated, using the scheme similar to the ones outlined in the previous section. The paper [21] includes formulas of this type for several other constants. Since then, a large number of other BBP-type formulas have been discovered.

Most of these identities were discovered using an experimental approach, using PSLQ searches. Others were found as the result of educated guesses based on experimentally obtained results. In each case, these formulas have been formally established, although the proofs are not always as simple as the proof of Theorem 3.4.1. We present these results, in part, to underscore the fact that the approach used to find the new formula for  $\pi$  has very broad applicability.

A sampling of the known binary BBP-type formulas (i.e., formulas with a base  $b = 2^p$  for some integer  $p$ ) is shown in Table 3.5. Some nonbinary BBP-type formulas are shown in Table 3.6. These formulas are derived from several sources: [21, 61, 62]. An updated collection is available at [16]. The constant  $G$  that appears in Table 3.5 is Catalan's constant, namely  $G = 1 - 1/3^2 + 1/5^2 - 1/7^2 + \dots = 0.9159655941 \dots$



$$\begin{aligned}
\pi\sqrt{3} &= \frac{9}{32} \sum_{k=0}^{\infty} \frac{1}{64^k} \left( \frac{16}{6k+1} + \frac{8}{6k+2} - \frac{2}{6k+4} - \frac{1}{6k+5} \right) \\
\pi^2 &= \frac{9}{8} \sum_{k=0}^{\infty} \frac{1}{64^k} \left( \frac{16}{(6k+1)^2} - \frac{24}{(6k+2)^2} - \frac{8}{(6k+3)^2} - \frac{6}{(6k+4)^2} \right. \\
&\quad \left. + \frac{1}{(6k+5)^2} \right) \\
\log^2 2 &= \frac{1}{32} \sum_{k=0}^{\infty} \frac{1}{64^k} \left( \frac{64}{(6k+1)^2} - \frac{160}{(6k+2)^2} - \frac{56}{(6k+3)^2} - \frac{40}{(6k+4)^2} \right. \\
&\quad \left. + \frac{4}{(6k+5)^2} - \frac{1}{(6k+6)^2} \right) \\
\pi \log 2 &= \frac{1}{256} \sum_{k=0}^{\infty} \frac{1}{4096^k} \left( \frac{4096}{(24k+1)^2} - \frac{8192}{(24k+2)^2} - \frac{26112}{(24k+3)^2} + \frac{15360}{(24k+4)^2} \right. \\
&\quad - \frac{1024}{(24k+5)^2} + \frac{9984}{(24k+6)^2} + \frac{11520}{(24k+8)^2} + \frac{2368}{(24k+9)^2} - \frac{512}{(24k+10)^2} \\
&\quad + \frac{768}{(24k+12)^2} - \frac{64}{(24k+13)^2} + \frac{408}{(24k+15)^2} + \frac{720}{(24k+16)^2} \\
&\quad \left. + \frac{16}{(24k+17)^2} + \frac{196}{(24k+18)^2} + \frac{60}{(24k+20)^2} - \frac{37}{(24k+21)^2} \right) \\
G &= \frac{1}{1024} \sum_{k=0}^{\infty} \frac{1}{4096^k} \left( \frac{3072}{(24k+1)^2} - \frac{3072}{(24k+2)^2} - \frac{23040}{(24k+3)^2} + \frac{12288}{(24k+4)^2} \right. \\
&\quad - \frac{768}{(24k+5)^2} + \frac{9216}{(24k+6)^2} + \frac{10368}{(24k+8)^2} + \frac{2496}{(24k+9)^2} - \frac{192}{(24k+10)^2} \\
&\quad + \frac{768}{(24k+12)^2} - \frac{48}{(24k+13)^2} + \frac{360}{(24k+15)^2} + \frac{648}{(24k+16)^2} \\
&\quad \left. + \frac{12}{(24k+17)^2} + \frac{168}{(24k+18)^2} + \frac{48}{(24k+20)^2} - \frac{39}{(24k+21)^2} \right)
\end{aligned}$$

Table 3.5: Binary BBP-type formulas.

$$\begin{aligned}
\log 2 &= \frac{2}{3} \sum_{k=0}^{\infty} \frac{1}{9^k(2k+1)} \\
\pi\sqrt{3} &= \frac{1}{9} \sum_{k=0}^{\infty} \frac{1}{729^k} \left( \frac{81}{12k+1} - \frac{54}{12k+2} - \frac{9}{12k+4} - \frac{12}{12k+6} \right. \\
&\quad \left. - \frac{3}{12k+7} - \frac{2}{12k+8} - \frac{1}{12k+10} \right) \\
\log 3 &= \frac{1}{729} \sum_{k=0}^{\infty} \frac{1}{729^k} \left( \frac{729}{6k+1} + \frac{81}{6k+2} + \frac{81}{6k+3} + \frac{9}{6k+4} \right. \\
&\quad \left. + \frac{9}{6k+5} + \frac{1}{6k+6} \right) \\
\pi^2 &= \frac{2}{27} \sum_{k=0}^{\infty} \frac{1}{729^k} \left( \frac{243}{(12k+1)^2} - \frac{405}{(12k+2)^2} - \frac{81}{(12k+4)^2} \right. \\
&\quad - \frac{27}{(12k+5)^2} - \frac{72}{(12k+6)^2} - \frac{9}{(12k+7)^2} - \frac{9}{(12k+8)^2} \\
&\quad \left. - \frac{5}{(12k+10)^2} + \frac{1}{(12k+11)^2} \right) \\
\log\left(\frac{9}{10}\right) &= \frac{-1}{10} \sum_{k=0}^{\infty} \frac{1}{k10^k} \\
\log\left(\frac{1111111111}{387420489}\right) &= \frac{1}{10^8} \sum_{k=0}^{\infty} \frac{1}{10^{10k}} \left( \frac{10^8}{10k+1} + \frac{10^7}{10k+2} + \frac{10^6}{10k+3} \right. \\
&\quad + \frac{10^5}{10k+4} + \frac{10^4}{10k+5} + \frac{10^3}{10k+6} + \frac{10^2}{10k+7} \\
&\quad \left. + \frac{10}{10k+8} + \frac{1}{10k+9} \right) \\
\frac{25}{2} \log\left(\frac{781}{256} \left(\frac{57-5\sqrt{5}}{57+5\sqrt{5}}\right)^{\sqrt{5}}\right) &= \sum_{k=0}^{\infty} \frac{1}{5^{5k}} \left( \frac{5}{5k+2} + \frac{1}{5k+3} \right) \tag{3.6.45}
\end{aligned}$$

Table 3.6: Nonbinary BBP-type formulas.

In addition to the formulas in Tables 3.5 and 3.6, there are two other classes of constants known to possess binary BBP-type formulas. The first is logarithms of certain integers. Clearly,  $\log n$  can be written with a binary BBP formula (i.e. a formula with  $b = 2^m$  for some integer  $m$ ) provided  $n$  factors completely using primes whose logarithms have binary BBP formulas—one merely combines the individual series for the different primes into a single binary BBP formula. We have seen that the logarithm of the prime 2 possesses a binary BBP formula, and so does  $\log 3$ , by the following reasoning:

$$\begin{aligned} \log 3 &= 2 \log 2 + \log \left(1 - \frac{1}{4}\right) = 2 \sum_{k=1}^{\infty} \frac{1}{k2^k} - \sum_{k=1}^{\infty} \frac{1}{k4^k} \\ &= \frac{1}{2} \sum_{k=0}^{\infty} \frac{1}{4^k} \left(\frac{2}{2k+1} + \frac{1}{2k+2}\right) - \frac{1}{4} \sum_{k=0}^{\infty} \frac{1}{4^k} \left(\frac{2}{2k+2}\right) \\ &= \sum_{k=0}^{\infty} \frac{1}{4^k} \left(\frac{1}{2k+1}\right). \end{aligned} \tag{3.6.46}$$

In a similar manner, it can be shown, by examining the factorization of  $2^n + 1$  and  $2^n - 1$ , where  $n$  is an integer, that numerous other primes have this property. Some additional primes can be obtained by noting that the real part of the Taylor series expansion of

$$\alpha = \log \left(1 \pm \frac{(1+i)^k}{2^n}\right) \tag{3.6.47}$$

yields a BBP-type formula. See [16] for details.

The logarithms of the following primes are now known to possess binary BBP formulas [69]:

$$\begin{aligned} &2, 3, 5, 7, 11, 13, 17, 19, 29, 31, 37, 41, 43, 61, 73, 109, 113, 127, 151, \\ &241, 257, 331, 337, 397, 683, 1321, 1429, 1613, 2113, 2731, 5419, 8191, \\ &14449, 26317, 38737, 43691, 61681, 65537, 87211, 131071, 174763, \\ &246241, 262657, 268501, 279073, 312709, 524287, 525313, 599479, \\ &2796203, 4327489, 7416361, 15790321, 18837001, 22366891 \end{aligned} \tag{3.6.48}$$

This list is certainly not complete, and it is unknown whether or not all primes have this property, or even whether the list of such primes is finite or infinite.

One can also obtain BBP-type formulas in nonbinary bases for the logarithms of certain integers and rational numbers. One example is given by the base ten formula for  $\log(9/10)$  in Table 3.6. This has been used to compute the ten billionth decimal digit of  $\log(9/10)$  [21].

One additional class of binary BBP-type formulas that we will mention here is arctangents of certain rational numbers. We present here the results of experimental searches, using the PSLQ integer relation algorithm, which we have subsequently established formally. The formal derivation of these results proceeds as follows. Consider the set of rationals given by  $q = |\operatorname{Im}(T)/\operatorname{Re}(T)|$  or  $|\operatorname{Re}(T)/\operatorname{Im}(T)|$ , where

$$T = \prod_{k=1}^m \left(1 \pm \frac{i}{2^{t_k}}\right)^{u_k} \left(1 \pm \frac{1+i}{2^{v_k}}\right)^{w_k} \quad (3.6.49)$$

for various  $m$ -long nonnegative integer vectors  $t$ ,  $u$ ,  $v$ ,  $w$  and choices of signs as shown [50, pg. 344]. For example, setting  $t = (1, 1)$ ,  $u = (1, 1)$ ,  $v = (1, 3)$ ,  $w = (1, 1)$ , with signs  $(1, -1, -1, 1)$ , gives the result  $T = 25/32 - 5i/8$ , which yields  $q = 4/5$ . Indeed, one can obtain the formula

$$\begin{aligned} \arctan\left(\frac{4}{5}\right) &= \frac{1}{2^{17}} \sum_{k=0}^{\infty} \frac{1}{2^{20k}} \left( \frac{524288}{40k+2} - \frac{393216}{40k+4} - \frac{491520}{40k+5} + \frac{163840}{40k+8} \right. \\ &\quad + \frac{32768}{40k+10} - \frac{24576}{40k+12} + \frac{5120}{40k+15} + \frac{10240}{40k+16} + \frac{2048}{40k+18} \\ &\quad + \frac{1024}{40k+20} + \frac{640}{40k+24} + \frac{480}{40k+25} + \frac{128}{40k+26} - \frac{96}{40k+28} \\ &\quad \left. + \frac{40}{40k+32} + \frac{8}{40k+34} - \frac{5}{40k+35} - \frac{6}{40k+36} \right). \quad (3.6.50) \end{aligned}$$

The set of rationals for which BBP formulas can be obtained in this way can be further expanded by applying the formula

$$\tan(r+s) = \frac{\tan r + \tan s}{1 - \tan(rs)}, \quad (3.6.51)$$

for rationals  $r$  and  $s$  for which binary BBP-type formulas are found. By applying these methods, it can be shown that binary BBP formulas exist for the arctangents of the following rational numbers. Only those rationals with numerators

< denominators  $\leq 25$  are listed here.

$$\begin{aligned}
 & 1/2, 1/3, 2/3, 1/4, 3/4, 1/5, 2/5, 3/5, 4/5, 1/6, 5/6, 1/7, 3/7, \\
 & 4/7, 5/7, 6/7, 1/8, 7/8, 1/9, 2/9, 7/9, 8/9, 3/10, 1/11, 2/11, \\
 & 3/11, 7/11, 8/11, 10/11, 1/12, 5/12, 1/13, 4/13, 6/13, 7/13, 9/13, \\
 & 11/13, 12/13, 3/14, 5/14, 1/15, 4/15, 8/15, 1/16, 7/16, 11/16, 13/16, \\
 & 15/16, 1/17, 4/17, 6/17, 7/17, 9/17, 11/17, 15/17, 16/17, 1/18, 13/18, \\
 & 3/19, 4/19, 6/19, 7/19, 8/19, 9/19, 11/19, 17/19, 9/20, 1/21, 13/21, \\
 & 16/21, 20/21, 3/22, 7/22, 9/22, 19/22, 21/22, 2/23, 4/23, 6/23, 7/23, \\
 & 9/23, 10/23, 11/23, 14/23, 15/23, 7/24, 11/24, 23/24, 1/25, 2/25, \\
 & 13/25, 19/25, 21/25 \tag{3.6.52}
 \end{aligned}$$

Note that not all “small” rationals appear in this list. As it turns out, by applying the methods given in the paper [49] (see the next section), one can rule out the possibility of Machin-type BBP formulas (as described in Section 3.6) for the arctangents of  $2/7$ ,  $3/8$ ,  $5/8$ ,  $4/9$ , and  $5/9$ . Thus we believe the above list to be complete for rationals with numerators and denominators up to ten. Beyond this level, we do not know for sure whether this list is complete, or whether applying formula (3.6.49), together with addition and subtraction formulas, generates all possible rationals possessing binary BBP-type formulas.

One can obtain BBP formulas in nonbinary bases for the arctangents of certain rational numbers by employing an appropriate variant of formula (3.6.49).

### 3.7 Does Pi Have a Nonbinary BBP Formula?

As we mentioned above, from the day that the BBP-formula for  $\pi$  was discovered, many researchers have wondered whether there exist BBP-type formulas that would permit computation of individual digits in bases other than powers of two (such as base ten). This is not such a far-fetched possibility, because both base-2 and base-3 formulas are known for  $\pi^2$ , as well as for  $\log 2$  (see Tables 3.5 and 3.6). But extensive computations failed to find any nonbinary formulas for  $\pi$ .

Recently one of the present authors, together with David Borwein (Jon’s father) and William Galway, established that there are no nonbinary Machin-type arctangent formulas for  $\pi$ . We believe that if there is no nonbinary Machin-

type arctangent formula for  $\pi$ , then there is no nonbinary BBP-type formula of any form for  $\pi$ . We will summarize this result here. Full details and other related results can be found in [49].

We say that the integer  $b > 1$  is not a proper power if it cannot be written as  $c^m$  for any integers  $c$  and  $m > 1$ . We will use the notation  $\text{ord}_p(z)$  to denote the  $p$ -adic order of the rational  $z \in \mathbb{Q}$ . In particular,  $\text{ord}_p(p) = 1$  for prime  $p$ , while  $\text{ord}_p(q) = 0$  for primes  $q \neq p$ , and  $\text{ord}_p(wz) = \text{ord}_p(w) + \text{ord}_p(z)$ . The notation  $\nu_b(p)$  will mean the order of the integer  $b$  in the multiplicative group of the integers modulo  $p$ . We will say that  $p$  is a primitive prime factor of  $b^m - 1$  if  $m$  is the least integer such that  $p | (b^m - 1)$ . Thus  $p$  is a primitive prime factor of  $b^m - 1$  provided  $\nu_b(p) = m$ . Given the Gaussian integer  $z \in \mathbb{Q}[i]$  and the rational prime  $p \equiv 1 \pmod{4}$ , let  $\theta_p(z)$  denote  $\text{ord}_{\mathfrak{p}}(z) - \text{ord}_{\bar{\mathfrak{p}}}(z)$ , where  $\mathfrak{p}$  and  $\bar{\mathfrak{p}}$  are the two conjugate Gaussian primes dividing  $p$ , and where we require  $0 < \Im(\mathfrak{p}) < \mathcal{R}(\mathfrak{p})$  to make the definition of  $\theta_p$  unambiguous. Note that

$$\theta_p(wz) = \theta_p(w) + \theta_p(z). \quad (3.7.53)$$

Given  $\kappa \in \mathbb{R}$ , with  $2 \leq b \in \mathbb{Z}$  and  $b$  not a proper power, we say that  $\kappa$  has a  $\mathbb{Z}$ -linear or  $\mathbb{Q}$ -linear *Machin-type BBP arctangent formula* to the base  $b$  if and only if  $\kappa$  can be written as a  $\mathbb{Z}$ -linear or  $\mathbb{Q}$ -linear combination (respectively) of generators of the form

$$\arctan\left(\frac{1}{b^m}\right) = \Im \log\left(1 + \frac{i}{b^m}\right) = b^m \sum_{k=0}^{\infty} \frac{(-1)^k}{b^{2mk}(2k+1)}. \quad (3.7.54)$$

We will also use the following theorem, first proved by Bang in 1886:

**Theorem 3.7.1** *The only cases where  $b^m - 1$  has no primitive prime factor(s) are when  $b = 2$ ,  $m = 6$ ,  $b^m - 1 = 3^2 \cdot 7$ ; and when  $b = 2^N - 1$ ,  $N \in \mathbb{Z}$ ,  $m = 2$ ,  $b^m - 1 = 2^{N+1}(2^{N-1} - 1)$ .*

We can now state the main result of this section:

**Theorem 3.7.2** *Given  $b > 2$  and not a proper power, then there is no  $\mathbb{Q}$ -linear Machin-type BBP arctangent formula for  $\pi$ .*

**Proof.** It follows immediately from the definition of a  $Q$ -linear Machin-type BBP arctangent formula that any such formula has the form

$$\pi = \frac{1}{n} \sum_{m=1}^M n_m \Im \log(b^m - i) \quad (3.7.55)$$

where  $n > 0 \in Z$ ,  $n_m \in Z$ , and  $M \geq 1$ ,  $n_M \neq 0$ . This implies that

$$\prod_{m=1}^M (b^m - i)^{n_m} \in e^{ni\pi} Q^\times = Q^\times \quad (3.7.56)$$

For any  $b > 2$  and not a proper power we have  $M_b \leq 2$ , so it follows from Bang's Theorem that  $b^{4M} - 1$  has a primitive prime factor, say  $p$ . Furthermore,  $p$  must be odd, since  $p = 2$  can only be a *primitive* prime factor of  $b^m - 1$  when  $b$  is odd and  $m = 1$ . Since  $p$  is a primitive prime factor, it does not divide  $b^{2M} - 1$ , and so  $p$  must divide  $b^{2M} + 1 = (b^M + i)(b^M - i)$ . We cannot have both  $p|b^M + i$  and  $p|b^M - i$ , since this would give the contradiction that  $p|(b^M + i) - (b^M - i) = 2i$ . It follows that  $p \equiv 1 \pmod{4}$ , and that  $p$  factors as  $p = \mathfrak{p}\bar{\mathfrak{p}}$  over  $Z[i]$ , with exactly one of  $\mathfrak{p}$ ,  $\bar{\mathfrak{p}}$  dividing  $b^M - i$ . Referring to the definition of  $\theta$ , we see that we must have  $\theta_p(b^M - i) \neq 0$ . Furthermore, for any  $m < M$ , neither  $\mathfrak{p}$  nor  $\bar{\mathfrak{p}}$  can divide  $b^m - i$  since this would imply  $p | b^{4m} - 1$ ,  $4m < 4M$ , contradicting the fact that  $p$  is a primitive prime factor of  $b^{4M} - 1$ . So for  $m < M$ , we have  $\theta_p(b^m - i) = 0$ . Referring to equation (3.7.55), using Equation (3.7.53) and the fact that  $n_M \neq 0$ , we get the contradiction

$$0 \neq n_M \theta_p(b^M - i) = \sum_{m=1}^M n_m \theta_p(b^m - i) = \theta_p(Q^\times) = 0. \quad (3.7.57)$$

Thus our assumption that there was a  $b$ -ary Machin-type BBP arctangent formula for  $\pi$  must be false.  $\square$

### 3.8 Commentary and Additional Examples

1. **The ENIAC Integrator and Calculator.** ENIAC, built in 1946 at the University of Pennsylvania, had 18,000 vacuum tubes, 6,000 switches,



Figure 3.4: The ENIAC computer.

10,000 capacitors, 70,000 resistors, 1,500 relays, was 10 feet tall, occupied 1,800 square feet, and weighed 30 tons. ENIAC could perform 5,000 arithmetic operations per second—1,000 times faster than any earlier machine, but a far cry from today's leading-edge microprocessors, which can perform more than four billion operations per second. The first stored-memory computer, ENIAC could store 200 digits, which again is a far cry from the hundreds of megabytes in a modern personal computer system. Data flowed from one accumulator to the next, and after each accumulator finished a calculation, it communicated its results to the next in line. The accumulators were connected to each other manually. A photo is shown in Figure 3.4.

2. **Four approximations to pi.** Here are two well known, but fascinating, approximations to  $\pi$ :

$$\pi \approx \frac{3}{\sqrt{163}} \log(640320),$$



correct to 15 decimal places, and

$$\pi \approx \frac{3}{\sqrt{67}} \log(5280),$$

correct to 9 decimal places. Both rely on somewhat deeper number theory (see Section 1.4 in the second volume). Here are two nice algebraic  $\pi$  approximations:

$$\pi \approx 66 \frac{\sqrt{2}}{33\sqrt{29} - 148}$$

and

$$\pi \approx \frac{63}{25} \frac{17 + 15\sqrt{5}}{7 + 15\sqrt{5}}.$$

3. **An arctan series for pi.** Find rational coefficients  $a_i$  such that the identity

$$\begin{aligned} \pi = & a_1 \arctan \frac{1}{390112} + a_2 \arctan \frac{1}{485298} \\ & + a_3 \arctan \frac{1}{683982} + a_4 \arctan \frac{1}{1984933} \\ & + a_5 \arctan \frac{1}{2478328} + a_6 \arctan \frac{1}{3449051} \\ & + a_7 \arctan \frac{1}{18975991} + a_8 \arctan \frac{1}{22709274} \\ & + a_9 \arctan \frac{1}{24208144} + a_{10} \arctan \frac{1}{201229582} \\ & + a_{11} \arctan \frac{1}{2189376182} \end{aligned}$$

holds [7, pg. 75]. Also show that an identity with even simpler coefficients exists if  $\arctan 1/239$  is included as one of the terms on the RHS. Hint: Use an integer relation program (see Section 6.3), or try the tools at one of these sites: <http://www.cecm.sfu.ca/projects/IntegerRelations> or <http://www.expmath.info>.

4. **Ballantine's series for pi.** A formula of Euler for  $\operatorname{arccot}$  is

$$x \sum_{n=0}^{\infty} \frac{(n!)^2 4^n}{(2n+1)! (x^2+1)^{n+1}} = \operatorname{arctan} \left( \frac{1}{x} \right) \quad (3.8.58)$$

As observed by Ballantine in 1939, ([34]) this allows one to rewrite the variant of Machin's formula, used by Guilloud and Bouyer in 1973 to compute a million digits of  $\pi$ ,

$$\frac{\pi}{4} = 12 \arctan \frac{1}{18} + 8 \arctan \frac{1}{57} - 5 \arctan \frac{1}{239} \quad (3.8.59)$$

in the neat form

$$\begin{aligned} \pi &= 864 \sum_{n=0}^{\infty} \frac{(n!)^2 4^n}{(2n+1)! 325^{n+1}} + 1824 \sum_{n=0}^{\infty} \frac{(n!)^2 4^n}{(2n+1)! 3250^{n+1}} \\ &\quad - 20 \arctan \frac{1}{239}, \end{aligned} \quad (3.8.60)$$

where the terms of the second series are just decimal shifts of the terms of the first.

5. **Convergence rates for pi formulas.** Analyze the rates of convergence of Archimedes iteration (3.1.1), the Salamin-Brent iteration (3.1.16), the Borwein cubic iteration (3.1.17) and the Borwein quartic iteration (3.1.18), by means of explicit computations. Use the high-precision arithmetic facility built into *Maple* or *Mathematica*, or write your own C++ or Fortran-90 code using the ARPREC arbitrary precision software available at <http://www.expmath.info>, or the GNU multiprecision software available at <http://www.gnu.org/software/gmp/gmp.html>. Such iterations are discussed more in Sections 5.6.2 and 5.6.3.
6. **Biblical pi.** As noted in Section 3.1, the Biblical passages 1 Kings 7:23 and 2 Chronicles 4:2 indicate that  $\pi = 3$ . In spite of the fact the context of these verses clearly suggests an informal approximation, not a precise statement of mathematical fact, this discrepancy has been a source of consternation among Biblical literalists for centuries. For example, an 18th-century German Bible commentary attempted to explain away this discrepancy using the imaginative (if pathetic) suggestion that the circular pool in Solomon's temple (clearly described in 2 Chron. 4:2 as "round in compass") was instead hexagonal in shape [33, pg. 75–76]. Even today, some are still unwilling to accept that the Bible could simply be mistaken here. One evangelical scholar, for example, writes:

However, the recorded dimensions are still no problem if we consider the shape of the vessel. In 1 Kings 7:26, we read that its “brim was made like the brim of a cup, as a lily blossom.” Hence, the sea was not a regular cylinder, but had an outward curving rim. Although we do not know the exact points on the vessel where the measurements were taken, the main part of the sea always will be somewhat smaller than the 10 cubits measured “from brim to brim.” [150]

Another geometric difficulty in these Biblical passages is that 1 Kings 7:26 (three verses after it gives its dimensions) gives a volume of 2,000 “baths” for the basin, while 2 Chron. 4:5 gives the figure 3,000 baths. 2 Chron. 4:2 gives the height of the basin as five cubits. Using the accepted conversions that one “cubit” is roughly 46 cm, and that one “bath” is roughly 23 liters, then assuming Solomon’s pool was cylindrical in shape, we obtain an actual volume of roughly 1660 baths. If the basin was rounded on the bottom, then its volume was even lower than this.

7. **Exponentiation of pi.** Arguably the most accessible transcendental number to compute is  $e^\pi$ . The following iteration computes  $D$  digits in  $\log(D)$  steps.

**Algorithm 4** *Computation of  $\exp(\pi)$ .*

Set  $k_0 = 1/\sqrt{2}$  and for  $n < N$  iterate

$$k'_n = \sqrt{1 - k_n^2}, \quad k_{n+1} = \frac{1 - k'_n}{1 + k'_n}.$$

Then return

$$\left(\frac{k_N}{4}\right)^{-1/2^N}.$$

Some care needs to be taken with guard digits. □

8. **Algorithms for Gamma values.** An algorithm for  $\pi$  may be viewed as an algorithm for  $\Gamma\left(\frac{1}{2}\right)$ , and there is a quite analogous iteration for  $\Gamma$  at the values  $1/3$ ,  $2/3$ ,  $1/4$ ,  $3/4$ ,  $1/6$ , and  $5/6$ . This, in turn, allows rapid computation of  $\Gamma\left(\frac{k}{24}\right)$ , for all integer  $k$ . We illustrate with:

**Algorithm 5** *Computation of  $\Gamma\left(\frac{1}{4}\right)$ .*

Let  $x_0 = 2^{1/2}, y_1 = 2^{1/4}$ . Let

$$x_{n+1} = \frac{\sqrt{x_n} + 1/\sqrt{x_n}}{2} \quad (3.8.61)$$

$$y_{n+1} = \frac{y_n \sqrt{x_n} + 1/\sqrt{x_n}}{y_n + 1}. \quad (3.8.62)$$

Then

$$\Gamma^4\left(\frac{1}{4}\right) = 16(1 + \sqrt{2})^3 \prod_{n=1}^{\infty} x_n^{-1} \left(\frac{1 + x_n}{1 + y_n}\right)^3.$$

This yields a quadratically convergent iteration for  $\Gamma\left(\frac{1}{4}\right)$ .  $\square$

No such iteration is known for  $\Gamma\left(\frac{1}{5}\right)$ ; see [50, 46].

9. **An integral representation of Euler's constant.** While it is known that  $\Gamma\left(\frac{1}{3}\right)$  and  $\Gamma\left(\frac{1}{4}\right)$  are transcendental, the status of *Euler's constant*

$$\gamma = \lim_{n \rightarrow \infty} \sum_{k=1}^n \frac{1}{k} - \log(n) \quad (3.8.63)$$

is unsettled.

Problem: Show that

$$\gamma = \int_0^{\infty} \left( \frac{1}{e^t - 1} - \frac{1}{t e^t} \right) dt.$$

10. **Computation of Euler's constant.** Perhaps the most efficient method of computation, due to Brent and MacMillan ([50, pg. 336]), is based on Bessel function identities. It allows one to show that if  $\gamma$  is rational it must have a denominator with millions of digits. The underlying identity, known to Euler, is

$$\gamma + \log(z/2) = \frac{S_0(z) - K_0(z)}{I_0(z)} \quad (3.8.64)$$

where  $I_\nu(z) = \sum_{k=0}^{\infty} (z/2)^{2k+\nu} / (k! \Gamma(k + \nu + 1))$ , while  $K_0(z) = \partial I_\nu(0) / \partial \nu$  and  $S_0(z) = \sum_{k=0}^{\infty} (\sum_{j=1}^k 1/j) (z/2)^{2k} / (k!)^2$ .

An algorithm follows from knowing the first terms of the asymptotic expansion for  $K_0$  and  $I_0$ . It is

**Algorithm 6** *Computation of  $K_0$  and  $I_0$ :*

$A_0 = -\log(n)$ ,  $B_0 = 1$ ,  $U_0 = A_0$ ,  $V_0 = 1$ , and for  $k = 1, 2, \dots$ ,

$$\begin{aligned} B_k &= B_{k-1} n^2 / k^2, & A_k &= (A_{k-1} n^2 / k + B_k) / k, \\ U_k &= U_{k-1} + A_k, & V_k &= V_{k-1} + B_k. \end{aligned} \quad (3.8.65)$$

Terminate when  $U_k$  and  $V_k$  no longer change, and return  $\gamma \approx U_k / V_k$ . With  $\log(n)$  computed efficiently, this scheme takes  $O(D)$  storage and approximately  $2.07 D$  steps to compute  $\gamma$  to  $D$  decimal places.  $\square$

11. **Buffon's needle.** Suppose we have a lined sheet of paper, and a needle that is precisely as long as the distance between the lines. Compute the probability that the needle "thrown at random" on the sheet of paper will lie on a line. Answer:  $2/\pi$ . Although this is certainly not a good way to calculate  $\pi$  (millions of trials would be required to obtain just a few digits), it is an instructive example of how  $\pi$  arises in unlikely settings (see also the next two exercises). Some additional discussion of this problem, plus a computer-based tool that allows one to perform these trials, is available at the URL <http://www.mste.uiuc.edu/reese/buffon/buffon.html>.
12. **Putnam problem 1993-B3.** If two real numbers  $x$  and  $y$  are generated uniformly at random in  $(0, 1)$ , what is the probability that the nearest integer to  $x/y$  is even? Hint: Ignoring negligible events, for this to occur either  $0 < x/y < 1/2$  or  $(4n - 1)/2 < x/y < (4n + 1)/2$ . The first occurs in a triangle of area  $1/4$  and the subsequent in triangles of area  $1/(4n - 1) - 1/(4n + 1)$ . Now apply the Gregory-Leibniz formula. Answer: The probability is  $(5 - \pi)/4 \approx 0.4646018$ .
13. **Number-theory probabilities.** Prove (a) The probability that an integer is square-free is  $6/\pi^2$ . (b) The probability that two integers are

relatively prime is also  $6/\pi^2$ . This is a good example of  $\pi$  appearing in a number-theory setting. See [110].

14. **The irrationality of  $\pi$ .** We reproduce in extenso Ivan Niven's 1947 very concise proof that  $\pi$  is irrational [157].

Let  $\pi = a/b$ , the quotient of positive integers. We define the polynomials

$$f(x) = \frac{x^n(a - bx)^n}{n!}$$

$$F(x) = f(x) - f^{(2)}(x) + f^{(4)}(x) - \dots + (-1)^n f^{(2n)}(x),$$

the positive integer  $n$  being specified later. Since  $n!f(x)$  has integral coefficients and terms in  $x$  of degree not less than  $n$ ,  $f(x)$  and its derivatives  $f^{(j)}(x)$  have integral values for  $x = 0$ ; also for  $x = \pi = a/b$ , since  $f(x) = f(a/b - x)$ . By elementary calculus we have

$$\frac{d}{dx}\{F'(x) \sin x - F(x) \cos x\} = F''(x) \sin x + F'(x) \cos x = f(x) \sin x$$

and

$$\int_0^\pi f(x) \sin x dx = [F'(x) \sin x - F(x) \cos x]_0^\pi = F(\pi) + F(0). \quad (3.8.66)$$

Now  $F(\pi) + F(0)$  is an *integer*, since  $f^{(j)}(0)$  and  $f^{(j)}(\pi)$  are integers. But for  $0 < x < \pi$ ,

$$0 < f(x) \sin x < \frac{\pi^n a^n}{n!},$$

so that the integral in (3.8.66) is *positive but arbitrarily small* for  $n$  sufficiently large. Thus (3.8.66) is false, and so is our assumption that  $\pi$  is rational.  $\square$

This proof gives a good taste of the ingredients of more subtle irrationality and transcendence proofs.

15. **A spigot algorithm for  $e$  and  $\pi$ .** A spigot method for a numerical constant is one that can produce digits one by one ("drop by drop") [168]. This is especially easy for  $e$  as carries are not a big issue.

- (a) The following algorithm, due to Rabinowitz and Wagon, generates successive digits of  $e$ . Initialize an array  $A$  of length  $n + 1$  to 1. Then repeat the following  $n - 1$  times: (a) multiply each entry in  $A$  by ten; (b) Starting from the right, reduce the  $i$ -th entry of  $A$  modulo  $i + 1$ , carrying the quotient of the division one place left. The final quotient produced is the next digit of  $e$ . This algorithm is based on the following formula, which is simply a restatement of  $e = \sum 1/i!$ .

$$e = 1 + \frac{1}{1} \left( 1 + \frac{1}{2} \left( 1 + \frac{1}{3} \left( 1 + \frac{1}{4} \left( 1 + \frac{1}{5} (1 + \dots) \right) \right) \right) \right).$$

- (b) Implement a parallel spigot algorithm for  $\pi$ , based on showing that:

$$\pi = 2 + \frac{1}{3} \left( 2 + \frac{2}{5} \left( 2 + \frac{3}{7} \left( 2 + \dots \left( 2 + \frac{k}{2k+1} \dots \right) \right) \right) \right).$$

The last term can be approximated by  $2 + 4k/(2k + 1)$  where  $k = \log_2(10)n$  to produce  $n$  digits of  $\pi$  “drop by drop.” If one wishes to run the algorithm without a specified end, one must take more care.

One may view the iteration (4.2.7), which we will study in Chapter 4, as a spectacular (albeit unproven) spigot algorithm for  $\pi$  base 16.

16. **Wagon’s BBP identity.** Determine the range of validity of the following identity, which is due to Stan Wagon:

$$\pi + 4 \arctan z + 2 \log \left( \frac{1 - 2z - z^2}{z^2 + 1} \right) = \sum_{k=0}^{\infty} \frac{1}{16^k} \left( \frac{4(z+1)^{1+8k}}{1+8k} - \frac{2(z+1)^{4+8k}}{4+8k} - \frac{(z+1)^{5+8k}}{5+8k} - \frac{(z+1)^{6+8k}}{6+8k} \right).$$

17. **Monte Carlo calculation of pi.** Monte Carlo simulation was pioneered during the Manhattan project by Stanislaw Ulam and others, who recognized that this scheme permitted simulations beyond the reach of conventional methods on the systems then available. We illustrate here a Monte Carlo calculation of  $\pi$ , which is a poor method to compute  $\pi$ , but illustrative of this general class of computation. Nowadays, Monte Carlo methods are quite popular because they are well suited to parallel computation on systems such as “Beowulf” clusters.

- (a) Design and implement a Monte Carlo simulation for  $\pi$ , based on generating pairs of uniformly distributed numbers in the unit square and testing whether they lie inside the unit circle. Use the pseudorandom number generator  $x_0 = 314159$  and  $x_n = cx_{n-1} \bmod 2^{32}$ , where  $c = 5^9 = 1953125$ . This generator is of the well known class of linear congruential generators and has period  $2^{30}$  [131, pg. 21]. It can be easily implemented on a computer using IEEE 64-bit “double” datatype, since the largest integer that can arise here is less than  $2^{53}$ . Variations with longer periods can easily be designed, although the implementation is not as convenient. The results of this generator are normalized, by  $2^{32}$  in this case, to produce results in the unit interval.
- (b) Extend your program to run on a parallel computer system, with the property that your parallel program generates the same overall scheme of pseudorandom numbers, and thus gets the same result for  $\pi$ , as a serial implementation (you may for convenience assume that  $n$ , the total number of pseudorandom numbers generated, is evenly divisible by  $p$ , the number of processors). This is a very desirable feature of a parallel program, because it allows you to certify your parallel results by comparing them with a conventional single-processor run, and it permits you to take advantage of a range of system sizes. Hint: Design the program so that processor  $k$  (where processors are numbered from 0 to  $p - 1$ ) generates the  $m = n/p$  members of the sequence  $(x_{km}, x_{km+1}, x_{km+2}, \dots, x_{km+m-1})$ . Note that the starting value  $x_{km}$  for processor  $k$  can be directly computed as  $x_{km} = 5^{9km}x_0 \bmod 2^{32}$ . This exponentiation modulo  $2^{32}$  may be performed by using Algorithm 1, implemented with 128-bit floating-point or “double-double” arithmetic (see Section 6.2.1).
- (c) Generate the following sequence of pseudorandom numbers and experimentally determine what distribution they satisfy (i.e., by computing means, standard deviations, graphs, etc.): Let  $x_1$  and  $x_2$  be a pair of uniform  $(0, 1)$  pseudorandom numbers generated as described above. Set  $v = \sqrt{x_1^2 + x_2^2}$  and  $w = \sqrt{-2 \log v/v}$ . Then produce the results  $y_1 = wx_1$  and  $y_2 = wx_2$ .

18. **Life of Pi.** At the end of his story, Piscine (Pi) Molitor [152, pp.316–7] writes



I am a person who believes in form, in harmony of order. Where we can, we must give things a meaningful shape. For example—I wonder—could you tell my jumbled story in exactly one hundred chapters, not one more, not one less? I'll tell you, that's one thing I hate about my nickname, the way that number runs on forever. It's important in life to conclude things properly. Only then can you let go.

We may not share the sentiment, but we should celebrate that Pi knows  $\pi$  to be irrational.

# Chapter 4

## Normality of Numbers

Anyone who wants to make a name for himself can examine the major issue of whether  $\pi$  is *normal*, or perhaps more accurately, whether  $\pi$  is *not normal*.

Jörg Arndt and Christoph Haenel, *Pi Unleashed*, 2001

In this chapter, we address a fundamental problem of mathematics, a paradox of sorts: Whereas on one hand it can be proven that “almost all” real numbers are normal, and whereas it appears from experimental analysis that many of the fundamental constants of mathematics are normal to commonly used number bases, as yet there are no proofs, nor even any solid reason why we should observe this behavior. What we shall show here is that the theory of BBP constants, which as we have seen is a classic case study of experimental mathematics in action, opens a pathway into the investigation of normality, and in fact has already yielded some intriguing results.

### 4.1 Normality: A Stubborn Question

Given a real number  $\alpha$  and an integer  $b > 2$ , we say that  $\alpha$  is *b-normal* or *normal base b* if *every* sequence of  $k$  consecutive digits in the base- $b$  expansion of  $\alpha$  appears with limiting frequency  $b^{-k}$ . In other words, if a constant is 10-normal, then the limiting frequency of “3” (or any other single digit) in its decimal expansion is 1/10, the limiting frequency of “58” (or any other two-digit

pair) is  $1/100$ , and so forth. We say that a real number  $\alpha$  is *absolutely normal* if it is  $b$ -normal for all integers  $b > 1$  simultaneously.

In spite of these strong conditions, it is well known from measure theory that the set of absolutely normal real numbers in the unit interval has measure one, or in other words, that almost all real numbers are absolutely normal (see Exercise 1 at the end of this chapter). Further, from numerous analyses of computed digits, it appears that many of the fundamental constants of mathematics are normal to commonly used number bases. By “fundamental constants,” we include  $\pi$ ,  $e$ ,  $\sqrt{2}$ , the golden mean  $\tau = (1 + \sqrt{5})/2$ , as well as  $\log n$  and the Riemann zeta function  $\zeta(n)$  for positive integers  $n > 1$ , and many others. For example, it is a reasonable conjecture that *every* irrational algebraic number is absolutely normal, since there is no known example of an irrational algebraic number whose decimal expansion (or expansion in any other base) appears to have skewed digit-string frequencies.

Decimal values are given for a variety of well known mathematical constants in Table 4.1 [66, 94]. In addition to the widely recognized constants such as  $\pi$  and  $e$ , we have listed Catalan’s constant ( $G$ ), Euler’s constant ( $\gamma$ ), an evaluation of the elliptic integral of the first kind  $K(1/\sqrt{2})$ , an evaluation of an elliptic integral of the second kind  $E(1/\sqrt{2})$ , Feigenbaum’s  $\alpha$  and  $\delta$  constants, Khintchine’s constant  $\mathcal{K}$ , and Madelung’s constant  $\mathcal{M}_3$ . Binary values for some of these constants, as well as Chaitin’s  $\Omega$  constant (from the field of computational complexity) [66], are given in Table 4.2. As you can see, none of the expansions in either table exhibits any evident “pattern.”

The digits of  $\pi$  have been studied more than any other single constant, in part because of the widespread fascination with  $\pi$ . Along this line, Yasumasa Kanada of the University of Tokyo has tabulated the number of occurrences of the ten decimal digits “0” through “9” in the first one trillion decimal digits of  $\pi$ . These counts are shown in Table 4.3. For reasons given in Section 3.4, binary (or hexadecimal) digits of  $\pi$  are also of considerable interest. To that end, Kanada has also tabulated the number of occurrences of the 16 hexadecimal digits “0” through “F,” as they appear in the first one trillion hexadecimal digits. These counts are shown in Table 4.4. As you can see, both the decimal and hexadecimal single-digit counts are entirely reasonable.

Some readers may be amused by the PiSearch utility, which is available at <http://pi.nersc.gov>. This online tool permits one to enter one’s name (or any other modest-length alphabetic string, or any modest-length hexadecimal string)

Constant	Value
$\sqrt{2}$	1.4142135623730950488 ...
$\sqrt{3}$	1.7320508075688772935 ...
$\sqrt{5}$	2.2360679774997896964 ...
$\phi = \frac{\sqrt{5}-1}{2}$	0.61803398874989484820 ...
$\pi$	3.1415926535897932385 ...
$1/\pi$	0.31830988618379067153 ...
$e$	2.7182818284590452354 ...
$1/e$	0.36787944117144232160 ...
$e^\pi$	23.140692632779269007 ...
$\log 2$	0.69314718055994530942 ...
$\log 10$	2.3025850929940456840 ...
$\log_2 10$	3.3219280948873623478 ...
$\log_{10} 2$	0.30102999566398119522 ...
$\log_2 3$	1.5849625007211561815 ...
$\zeta(2)$	1.6449340668482264365 ...
$\zeta(3)$	1.2020569031595942854 ...
$\zeta(5)$	1.0369277551433699263 ...
$G$	0.91596559417721901505 ...
$\gamma$	0.57721566490153286061 ...
$\Gamma(1/2) = \sqrt{\pi}$	1.7724538509055160273 ...
$\Gamma(1/3)$	2.6789385347077476337 ...
$\Gamma(1/4)$	3.6256099082219083121 ...
$K(1/\sqrt{2})$	1.8540746773013719184 ...
$E(1/\sqrt{2})$	1.3506438810476755025 ...
$\alpha_f$	4.669201609102990 ...
$\delta_f$	2.502907875095892 ...
$\mathcal{K}$	2.6854520010653064453 ...
$\mathcal{M}_3$	1.7475645946331821903 ...

Table 4.1: Decimal values of various mathematical constants.

Constant	Value
$\pi$	11.001001000011111101101010100010001000010110100011000010001...
$e$	10.101101111110000101010001011000101000101011101101001010100...
$\sqrt{2}$	1.0110101000001001111001100110011111110011101111001100100100...
$\sqrt{3}$	1.1011101101100111101011101000010110000100110010101010011100...
$\log 2$	0.101100010111001000010111111011111010001110011110111100110...
$\log 3$	1.0001100100111110101001111010101011010000001100001010100101...
$\Omega$	0.0000001000000100001000001000011101110011001001111000100100...

Table 4.2: Binary values of various mathematical constants.

Digit	Occurrences
0	99999485134
1	99999945664
2	100000480057
3	99999787805
4	100000357857
5	99999671008
6	99999807503
7	99999818723
8	100000791469
9	99999854780
Total	100000000000

Table 4.3: Statistics for the first trillion decimal digits of  $\pi$ .

Digit	Occurrences
0	62499881108
1	62500212206
2	62499924780
3	62500188844
4	62499807368
5	62500007205
6	62499925426
7	62499878794
8	62500216752
9	62500120671
A	62500266095
B	62499955595
C	62500188610
D	62499613666
E	62499875079
F	62499937801
Total	1000000000000

Table 4.4: Statistics for the first trillion hexadecimal digits of  $\pi$ .

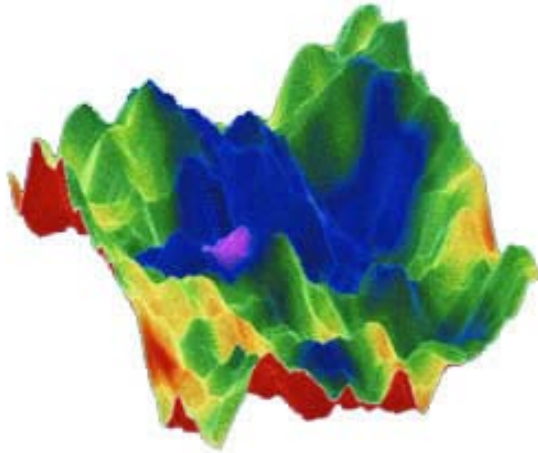


Figure 4.1: A random walk based on one million digits of  $\pi$ .

and see if it appears encoded in the first several billion binary digits of  $\pi$ . Along this line, a graphic based on a random walk of the first million decimal digits of  $\pi$ , courtesy of David and Gregory Chudnovsky, is shown in Figure 4.1. It maps the digit stream to a surface in ways similar to those used by Mandelbrot and others.

As we mentioned in Section 3.2, the question of whether  $\pi$ , in particular, or, say,  $\sqrt{2}$ , is normal or not has intrigued mathematicians for centuries. But in spite of centuries of effort, not a single one of the fundamental constants of mathematics has ever been proven to be  $b$ -normal for any integer  $b$ , much less for all integer bases simultaneously. And this is not for lack of trying—some very good mathematicians have seriously investigated this problem, but to no avail. Even much weaker results, such as the digit “1” appears with nonzero limiting frequency in the binary expansion of  $\pi$ , and the digit “5” appears infinitely often in the decimal expansion of  $\sqrt{2}$ , have heretofore remained beyond the reach of modern mathematics.

One result in this area is the following. Let  $f(n) = \sum_{1 \leq j \leq n} \lfloor \log_{10} j \rfloor$ . Then the Champernowne number,

$$\sum_{n=1}^{\infty} \frac{n}{10^{n+f(n)}} = 0.12345678910111213141516171819202122232425 \dots,$$

where the positive integers are concatenated in a decimal value, is known to be 10-normal. There are similar constants and normality results for other number bases. However, no one, to the authors' knowledge, has ever argued that this constant and its relatives are "natural" or "fundamental" constants.

Consequences of a proof in this area would definitely be interesting. For starters, such a proof would immediately provide an inexhaustible source of provably reliable pseudorandom numbers for numerical or scientific experimentation. We also would obtain the mind-boggling but uncontestable consequence that if  $\pi$ , for example, is shown to be 2-normal, then the entire text of the Bible, the Koran and the works of William Shakespeare, as well as the full L<sup>A</sup>T<sub>E</sub>X source text for this book, must all be contained somewhere in the binary expansion of  $\pi$ , where consecutive blocks of eight bits (two hexadecimal digits) each represent one ASCII character. Unfortunately, this would not be much help to librarians or archivists, since every conceivable misprint of each of these books would also be contained in the binary digits of  $\pi$ .

Before continuing, we should mention the "first digit" principle, also known as Benford's principle. In the 1880s, Simon Newcomb observed a pattern in the first digits of logarithm tables: A "1" is significantly more likely to occur than "2," a "2" more than a "3," and so on. In other words, the collection of first digits of data in logarithm tables certainly does not reflect the statistics expected of 10-normal numbers. In the 20th century, Frank Benford rediscovered this phenomenon, noting that it applies to many types of numerical data, ranging from values of physical constants to census data to the stock market. One can deduce this principle by observing that natural laws surely cannot be dependent on our choice of units, and thus must be scale-independent. This suggests that we view numerical data on a logarithmic scale. In the logarithmic sense, a leading "1" appears roughly 30% of the time (since  $\log_{10} 2 - \log_{10} 1 = 0.30102999\dots$ ), a "2" appears roughly 17.6% of the time (since  $\log_{10} 3 - \log_{10} 2 = 0.1760912\dots$ ), and so on. More recently scientists have applied Benford's principle in diverse ways, including fraud detection in business accounting [118].

## 4.2 BBP Constants and Normality

Until recently, the BBP formulas mentioned in Sections 3.4 and 3.6 were assigned by some to the realm of "recreational" mathematics—interesting but of no serious consequence. But the history of mathematics has seen many instances where



results once thought to be idle curiosities were later found to have significant consequences. This now appears to be the case with the theory of BBP-type constants.

What we shall establish below, in a nutshell, is that the 16-normality of  $\pi$  (which, of course, is equivalent to the 2-normality of  $\pi$ ), as well as the normality of numerous other irrational constants that possess BBP-type formulas, can be reduced to a certain plausible conjecture in the theory of chaotic sequences. At this time we do not know the full implications of this result. It may be the first salvo in the resolution of this age-old mathematical question, or it may be merely a case of reducing one very difficult mathematical problem to another. But at the least, this result appears to lay out a structure—a “roadmap” of sorts—for the analysis of this question. Thus it seems worthy of investigation.

We shall also establish that a certain well-defined class of real numbers, uncountably infinite in number, *is* indeed  $b$ -normal for certain bases  $b$ . This result is not dependent on any unproven conjecture. We will also present some results on the digit densities of algebraic irrationals. All of these recent results are direct descendants of the theory of BBP-type constants that we have presented in Sections 3.4 and 3.6.

The results for BBP-type constants derive from the following observation, which was given in a recent paper by one of the present authors and Richard Crandall [24]. Here we define the norm  $\|\alpha\|$  for  $\alpha \in [0, 1)$  as  $\|\alpha\| = \min(\alpha, 1 - \alpha)$ . With this definition,  $\|\alpha - \beta\|$  measures the shortest distance between  $\alpha$  and  $\beta$  on the unit circumference circle in the natural way. Suppose  $\alpha$  is given by a BBP-type formula, namely

$$\alpha = \sum_{k=0}^{\infty} \frac{p(k)}{b^k q(k)}, \quad (4.2.1)$$

where  $p$  and  $q$  are polynomials with integer coefficients, with  $0 \leq \deg p < \deg q$ , and with  $q$  having no zeroes at nonnegative integer arguments. Now define the recursive sequence  $(x_n)$  as  $x_0 = 0$ , and

$$x_n = \left\{ bx_{n-1} + \frac{p(n)}{q(n)} \right\}, \quad (4.2.2)$$

where the notation  $\{\cdot\}$  denotes the fractional part as before. Recall from Section 3.4 that we can write the base- $b$  expansion of  $\alpha$  beginning at position  $n+1$ , which

we denote  $\alpha_n$ , as

$$\begin{aligned}\alpha_n &= \{b^n \alpha\} = \left\{ \sum_{k=0}^{\infty} \frac{b^{n-k} p(k)}{q(k)} \right\} \\ &= \left\{ \left\{ \sum_{k=0}^n \frac{b^{n-k} p(k)}{q(k)} \right\} + \sum_{k=n+1}^{\infty} \frac{b^{n-k} p(k)}{q(k)} \right\}.\end{aligned}\quad (4.2.3)$$

Now observe that the sequence  $(x_n)$  generates the first part of this expression. In particular, given  $\epsilon > 0$ , assume that  $n$  is sufficiently large such that  $p(k)/q(k) < \epsilon$  for all  $k \geq n$ . Then we can write, for all sufficiently large  $n$ ,

$$\begin{aligned}\|x_n - \alpha_n\| &\leq \left| \sum_{k=n+1}^{\infty} \frac{b^{n-k} p(k)}{q(k)} \right| \\ &\leq \epsilon \sum_{k=n+1}^{\infty} b^{n-k} = \frac{\epsilon}{b-1} \leq \epsilon.\end{aligned}\quad (4.2.4)$$

With this argument, we have established the following, which we observe is also true if the expression  $p(k)/q(k)$  is replaced by any more general sequence  $r(k)$  that tends to zero for large  $k$ :

**Theorem 4.2.1** *Let  $\alpha$  be a BBP-type constant as defined above, with  $\alpha_n$  the base- $b$  expansion of  $\alpha$  beginning at position  $n+1$ , and  $(x_n)$  the BBP sequence associated with  $\alpha$ , as given in (4.2.2) above. Then  $\|x_n - \alpha_n\| \rightarrow 0$  as  $n \rightarrow \infty$ .*

In other words, the *BBP sequence* associated with  $\alpha$ , as given in Formula (4.2.2), is a close approximation to the sequence  $(\alpha_n)$  of shifted digit expansions, so much so that if one has a property such as equidistribution in the unit interval, then the other does also (this will be made precise in Section 4.5). We now state a hypothesis, which is believed to be true, based on experimental evidence, but which is not yet proven:

**Hypothesis 4.2.2 (Bailey-Crandall).** *Let  $p(x)$  and  $q(x)$  be polynomials with integer coefficients, with  $0 \leq \deg p < \deg q$ , and with  $q$  having no zeroes for nonnegative integer arguments. Let  $b \geq 2$  be an integer, and let  $r_n = p(n)/q(n)$ . Then the sequence  $x = (x_0, x_1, x_2, \dots)$  determined by the iteration  $x_0 = 0$ , and*

$$x_n = \{bx_{n-1} + r_n\} \quad (4.2.5)$$

*either has a finite attractor or is equidistributed in  $[0, 1)$ .*

We shall define “equidistributed” and “finite attractor” in Section 4.5. In the meantime we rely on intuition.

**Theorem 4.2.3** *Assuming Hypothesis 4.2.2, any constant  $\alpha$  given by a formula of the type  $\alpha = \sum_{k \geq 0} p(k)/(b^k q(k))$ , with  $p(k)$  and  $q(k)$  polynomials as given in Hypothesis 4.2.2, is either rational or  $b$ -normal.*

The proof of this theorem is given in Section 4.5. We should note here that even if a particular instance of Hypothesis 4.2.2 could be established, it would have significant consequences. For example, if it could be established that the simple iteration given by  $x_0 = 0$  and

$$x_n = \left\{ 2x_{n-1} + \frac{1}{n} \right\} \quad (4.2.6)$$

is equidistributed in  $[0, 1)$ , then it would follow from Theorem 4.2.3 that  $\log 2$  is 2-normal. Observe that this sequence is simply the BBP sequence for  $\log 2$ . In a similar vein, if it could be established that the iteration given by  $x_0 = 0$  and

$$x_n = \left\{ 16x_{n-1} + \frac{120n^2 - 89n + 16}{512n^4 - 1024n^3 + 712n^2 - 206n + 21} \right\} \quad (4.2.7)$$

is equidistributed in  $[0, 1)$ , then it would follow that  $\pi$  is 16-normal (and so is 2-normal also). This is the BBP sequence for  $\pi$ . The fractional term here is obtained by combining the four fractions in the BBP formula for  $\pi$ , namely Equation (3.4.25), into one fraction, and then shifting the index by one.

Before continuing, we wish to mention a curious phenomenon. Suppose we compute the binary sequence  $y_n = \lfloor 2x_n \rfloor$ , where  $(x_n)$  is the sequence associated with  $\log 2$  as given in Equation 4.2.6. In other words,  $(y_n)$  is the binary sequence defined as  $y_n = 0$  if  $x_n < 1/2$  and  $y_n = 1$  if  $x_n \geq 1/2$ . Theorem 4.2.1 tells us, in effect, that  $(y_n)$  eventually should agree very well with the true sequence of binary digits of  $\log 2$ . In explicit computations, we have found that the sequence  $(y_n)$  disagrees with 15 of the first 200 binary digits of  $\log 2$ , but in only one position over the range 5000 to 8000.

As noted above, the BBP sequence for  $\pi$  is  $x_0 = 0$ , and  $x_n$  as given in Equation (4.2.7). In a similar manner as with  $\log 2$ , we can compute the hexadecimal digit sequence  $y_n = \lfloor 16x_n \rfloor$ . In other words, we can divide the unit

interval into 16 equal subintervals, labeled  $(0, 1, 2, 3, \dots, 15)$ , and set  $y_n$  to be the label of the subinterval in which  $x_n$  lies. When this is done, a remarkable phenomenon occurs: The sequence  $(y_n)$  appears to perfectly (not just approximately) produce the hexadecimal expansion of  $\pi$ . In explicit computations, the first 1,000,000 hexadecimal digits generated by this sequence are identical with the first 1,000,000 hexadecimal digits of  $\pi - 3$ . (This is a fairly difficult computation, as it requires roughly  $n^2$  bit-operations, and is not easily performed on a parallel computer system.)

**Conjecture 4.2.4** *The sequence  $(\lfloor 16x_n \rfloor)$ , where  $(x_n)$  is the sequence of iterates defined in equation (4.2.7), precisely generates the hexadecimal expansion of  $\pi - 3$ .*

Evidently, this phenomenon arises from the fact that in the sequence associated with  $\pi$ , the perturbation term  $r_n = p(n)/q(n)$  is summable, whereas the corresponding expression for  $\log 2$ , namely  $1/n$ , is not summable. In particular, note that expression (4.2.4) for  $\alpha = \pi$  gives

$$\begin{aligned} \|\alpha_n - x_n\| &\leq \sum_{k=n+1}^{\infty} \frac{120k^2 - 89k + 16}{16^{j-n}(512k^4 - 1024k^3 + 712k^2 - 206k + 21)} \\ &\approx \frac{120(n+1)^2 - 89(n+1) + 16}{16(512(n+1)^4 - 1024(n+1)^3 + 712(n+1)^2 - 206(n+1) + 21)} \end{aligned} \quad (4.2.8)$$

so that

$$\sum_{n=1}^{\infty} \|\alpha_n - x_n\| \leq 0.01579\dots \quad (4.2.9)$$

For the sake of heuristic argument, let us assume for the moment that the  $\alpha_n$  are independent, uniformly distributed random variables in  $(0, 1)$ , and let  $\delta_n = \|\alpha_n - x_n\|$ . Note that an error (i.e. an instance where  $x_n$  lies in a different subinterval of the unit interval than  $\alpha_n$ ) can only occur when  $\alpha_n$  is within  $\delta_n$  of one of the points  $(0, 1/16, 2/16, \dots, 15/16)$ . Since  $x_n < \alpha_n$  for all  $n$  (where  $<$  is interpreted in the wrapped sense when  $x_n$  is slightly less than one), this event has probability  $16\delta_n$ . Then the fact that the sum (4.2.9) has a finite

value implies, by the first Borel-Cantelli lemma, that there can only be finitely many errors [38, pg. 153]. The comparable figure for  $\log 2$  is infinite, which implies by the second Borel-Cantelli lemma that discrepancies can be expected to appear indefinitely, but with decreasing frequency. Further, the small value of the sum (4.2.9) suggests that it is unlikely that any errors will be observed. If instead of summing (4.2.9) from one to infinity, we instead sum from 1,000,001 to infinity (since we have computationally verified that there are no errors in the first 1,000,000 elements), then we obtain  $1.465 \times 10^{-8}$ , which suggests that it is very unlikely that any errors will ever occur.

### 4.3 A Class of Provably Normal Constants

We now summarize an intriguing recent development in this arena, due to one of the present authors and Richard Crandall, which offers additional hope that the BBP approach may eventually yield the long-sought proof of normality for  $\pi$ ,  $\log 2$ , and other BBP-type constants [25]. In the previous section, we noted that the 2-normality of

$$\log 2 = \sum_{n=1}^{\infty} \frac{1}{n2^n} \quad (4.3.10)$$

rests on the (unproven) conjecture that the iteration given by  $x_0 = 0$  and  $x_n = \{2x_{n-1} + 1/n\}$  is equidistributed in the unit interval. We now consider the class of constants where the summation defining  $\log 2$ , namely (4.3.10), is taken over a certain subset of the positive integers:

$$\alpha_{b,c} = \sum_{n=c^k > 1} \frac{1}{nb^n} = \sum_{k=1}^{\infty} \frac{1}{c^k b^{c^k}}, \quad (4.3.11)$$

where  $b > 1$  and  $c > 2$  are integers. The simplest instance of this class is

$$\begin{aligned} \alpha_{2,3} &= \sum_{n=3^k > 1} \frac{1}{n2^n} = \sum_{k=1}^{\infty} \frac{1}{3^k 2^{3^k}} & (4.3.12) \\ &= 0.0418836808315029850712528986245716824260967584654857 \dots_{10} \\ &= 0.0AB8E38F684BDA12F684BF35BA781948B0FCD6E9E06522C3F35B \dots_{16}. \end{aligned}$$

We first prove the following interesting fact:

**Theorem 4.3.1** *Each of the constants  $\alpha_{b,c}$ , where  $b > 1$  and  $c > 2$  are integers, is transcendental.*

**Proof.** A famous theorem due to Roth states [173] that if  $|P/Q - \alpha| < 1/Q^{2+\epsilon}$  admits infinitely many rational solutions  $P/Q$  (i.e., if  $\alpha$  is approximable to degree  $2 + \epsilon$  for some  $\epsilon > 0$ ), then  $\alpha$  is transcendental. We show here that  $\alpha_{b,c}$  is approximable to degree  $c - \delta$ . Fix a  $k$  and write

$$\alpha_{b,c} = P/Q + \sum_{n>k} \frac{1}{c^n b^{c^n}}, \tag{4.3.13}$$

where  $\gcd(P, Q) = 1$  and  $Q = c^k b^{c^k}$ . The sum over  $n$  gives

$$|\alpha_{b,c} - P/Q| < \frac{2}{c^{k+1}(Q/c^k)^c} < \frac{c^{kc}}{Q^c}. \tag{4.3.14}$$

Now  $c^k \log b + k \log c = \log Q$ , so that  $c^k < \log Q / \log b$ , and we can write

$$c^{kc} < (\log Q / \log b)^c = Q^{c(\log \log Q - \log \log b) / \log Q}. \tag{4.3.15}$$

Thus for any fixed  $\delta > 0$ ,

$$|\alpha_{b,c} - P/Q| < \frac{1}{Q^{c(1+\log \log b / \log Q - \log \log Q / \log Q)}} < \frac{1}{Q^{c-\delta}}, \tag{4.3.16}$$

for all sufficiently large  $k$ . □

Consider now the BBP sequence associated with  $\alpha_{2,3}$ , namely the sequence defined by  $x_0 = 0$ , and

$$x_n = \{2x_{n-1} + r_n\}, \tag{4.3.17}$$

where  $r_n = 1/n$  if  $n = 3^k$ , and  $r_n = 0$  otherwise. Successive iterates of this sequence are:

$$\begin{aligned} &0, 0, 0, \frac{1}{3}, \frac{2}{3}, \frac{1}{3}, \frac{2}{3}, \frac{1}{3}, \frac{2}{3}, \\ &\frac{4}{9}, \frac{8}{9}, \frac{7}{9}, \frac{5}{9}, \frac{1}{9}, \frac{2}{9}, \frac{4}{9}, \frac{8}{9}, \frac{7}{9}, \frac{5}{9}, \frac{1}{9}, \frac{2}{9}, \frac{4}{9}, \frac{8}{9}, \frac{7}{9}, \frac{5}{9}, \frac{1}{9}, \frac{2}{9}, \\ &\frac{13}{27}, \frac{26}{27}, \frac{25}{27}, \frac{23}{27}, \frac{19}{27}, \frac{11}{27}, \frac{22}{27}, \frac{17}{27}, \frac{7}{27}, \frac{14}{27}, \frac{1}{27}, \frac{2}{27}, \frac{4}{27}, \frac{8}{27}, \frac{16}{27}, \frac{5}{27}, \frac{10}{27}, \frac{20}{27}, \\ &\frac{13}{27}, \frac{26}{27}, \frac{25}{27}, \frac{23}{27}, \frac{19}{27}, \frac{11}{27}, \frac{22}{27}, \frac{17}{27}, \frac{7}{27}, \frac{14}{27}, \frac{1}{27}, \frac{2}{27}, \frac{4}{27}, \frac{8}{27}, \frac{16}{27}, \frac{5}{27}, \frac{10}{27}, \frac{20}{27}, \\ &\frac{13}{27}, \frac{26}{27}, \frac{25}{27}, \frac{23}{27}, \frac{19}{27}, \frac{11}{27}, \frac{22}{27}, \frac{17}{27}, \frac{7}{27}, \frac{14}{27}, \frac{1}{27}, \frac{2}{27}, \frac{4}{27}, \frac{8}{27}, \frac{16}{27}, \frac{5}{27}, \frac{10}{27}, \frac{20}{27} \end{aligned}$$

A pattern is clear: The sequence consists of a concatenation of triply-repeated segments, each consisting of fractions whose denominators are successively higher powers of 3, and whose numerators range over all integers less than the denominator that are coprime to the denominator. Indeed, the successive numerators in each subsequence are given by the simple linear congruential pseudorandom number generator  $z_n = 2z_{n-1} \pmod{3^j}$  for a fixed  $j$ .

What we have observed is that the question of the equidistribution of the sequence  $(x_n)$  (and, hence, the question of the normality of  $\alpha_{2,3}$ ) reduces to the behavior of a concatenation of normalized pseudorandom sequences of a type (namely linear congruential) that have been studied in mathematical literature, and which in fact are widely implemented for use by scientists and engineers. These observations lead to a rigorous proof of normality for many of these constants. In particular, we obtain the result that each of the constants

$$\alpha_{b,c} = \sum_{n=c^k > 1} \frac{1}{nb^n} = \sum_{k=1}^{\infty} \frac{1}{c^k b c^k}, \quad (4.3.18)$$

where  $b > 1$ , and  $c$  is co-prime to  $b$ , is  $b$ -normal. This result was first given in [25]. We present here a significantly simpler proof, based on the following lemma:

**Lemma 4.3.2 (Hot spot lemma.)** *The constant  $\alpha$  is  $b$ -normal if and only if there exists a constant  $M$  such that for every interval  $(c, d)$ ,*

$$\limsup_{n \rightarrow \infty} \frac{\#\{0 \leq j < n \mid \{b^j \alpha\} \in (c, d)\}}{n(d-c)} \leq M. \quad (4.3.19)$$

This result is proved in [135, pg. 77]. A stronger result is given in [29].

**Theorem 4.3.3** *Each of the constants*

$$\alpha_{b,c} = \sum_{k=1}^{\infty} \frac{1}{c^k b c^k}, \quad (4.3.20)$$

*where  $b > 1$ , and  $c$  is co-prime to  $b$ , is  $b$ -normal.*

**Proof.** For convenience, we will establish the result for  $\alpha = \alpha_{2,3}$ . Let  $(x_n)$  be the BBP sequence associated with  $\alpha$ , as given in (4.3.18), and let  $\alpha_n = \{2^n\alpha\}$  be the tail of the binary expansion of  $\alpha$  after the first  $n$  bits. First note, as in the proof of Theorem 4.2.1, that

$$\|x_n - \alpha_n\| = \left| \sum_{k=n+1}^{\infty} 2^{n-k} r_k \right| < \frac{1}{n}. \quad (4.3.21)$$

Given  $n$ , let  $m$  be the largest power of 3 less than  $n$ , and assume that  $n$  is large enough so that  $n > m > 2/(d-c)$ . The interval  $(c - 1/n, d + 1/n)$  contains no more than  $\lfloor m(d-c) \rfloor + 2$  multiples of  $1/m$ , and thus can contain at most three times this many occurrences of  $x_j$  in the first  $n$  elements. Thus we can write

$$\frac{\#\{0 \leq j < n \mid \alpha_j \in (c, d)\}}{n(d-c)} \leq \frac{\#\{0 \leq j < n \mid x_j \in (c - 1/n, d + 1/n)\}}{n(d-c)} \quad (4.3.22)$$

$$\leq \frac{3\lfloor m(d-c) \rfloor + 2}{n(d-c)} < \frac{3\lfloor m(d-c) \rfloor + 2}{m(d-c)} \quad (4.3.23)$$

$$= 3 + \frac{6}{m(d-c)} < 6. \quad (4.3.24)$$

In other words, for all  $n > 6(d-c)$ , no subinterval of  $[0, 1)$  contains more than six times as many elements of the sequence  $(\{2^j\alpha\}, 0 \leq j < n)$  as it “should.” By Lemma 4.3.2, this establishes that  $\alpha$  is 2-normal.  $\square$

We should add that  $\alpha_{2,3}$  was actually proven 2-normal by Stoneham in 1970 [183]. The new BBP-based approach described above now makes it possible to extend these and other previous results to a much larger class. For example, it is shown in [25] that all constants of the form

$$\alpha_{2,3}(r) = \sum_{k=1}^{\infty} \frac{1}{3^k 2^{3^k + r_k}}, \quad (4.3.25)$$

where  $r_k$  is the  $k$ -th bit in the binary expansion of  $r \in (0, 1)$ , are 2-normal and transcendental. It is fairly easy to show that if  $r \neq s$ , then  $\alpha_{2,3}(r) \neq \alpha_{2,3}(s)$ . In other words, for every real number  $r \in (0, 1)$ , there corresponds a distinct real constant  $\alpha_{2,3}(r)$ . Since there are uncountably many reals in  $(0, 1)$ , it follows



that there are uncountably many 2-normal, transcendental numbers of the form given by (4.3.25).

As an amusing sidelight, these  $\alpha$  constants also possess the rapid individual digit computation property possessed by conventional BBP-type constants. Furthermore, in this case the computation is extraordinarily rapid. For example, the googol-th (i.e., the  $10^{100}$ -th) binary digit of  $\alpha_{2,3}$  can be calculated in less than one second on a 2001-era computer. This digit is 0. The first ten hexadecimal digits starting at this position are 2205896E7B.

## 4.4 Algebraic Irrationals

Although there is no proof of this assertion, as far as we can tell, there are no BBP-type formulas for algebraic irrationals such as  $\sqrt{2}$ ,  $\sqrt[3]{10}$  and  $\phi = (1+\sqrt{5})/2$ . Thus it appears that the theory we have developed above to address the question of normality for BBP-type constants is inapplicable to the case of algebraic irrationals. Yet from all available experimental evidence, these constants are also  $b$ -normal for all commonly used number bases. In any event, even basic questions such as whether the decimal expansion of  $\sqrt{2}$  has infinitely many fives remain unanswered.

Recently, some results on the density of ones in binary expansions of certain algebraic irrationals were obtained by the present authors, Richard Crandall, and Carl Pomerance [19]. These results are certainly much weaker than full normality, but are worth mentioning given the paucity of results in this arena.

If  $a$  is an integer, we shall define  $B(a)$  as the number of ones in the binary representation of  $a$ . If  $x$  is real,  $0 < x < 1$ , then by  $B_n(x)$ , we shall mean the number of ones in the first  $n$  binary digits of  $x$  after the “decimal” point. Note that if  $x$  is a binary fraction (i.e.,  $x = c/2^d$  for integers  $c$  and  $d$ ), then  $x$  has two valid binary expansions—one terminating in zeroes and the other terminating in ones. Thus in such instances below, we shall clarify which expansion we mean. For a rational number  $r$  with  $0 < r < 1$ , we define the function  $C(r) = 1$  in the case that  $r$  is a binary fraction; otherwise,  $C(r) = c/d$ , where  $c$  is the number of ones in the terminally repeating block of digits in the binary expansion of  $r$ , and where  $d$  is the length of the block of repeating digits.

For real numbers  $x$  and  $y$ , with  $0 < x < 1$  and  $0 < y < 1$ , we use the notation  $(A_k(x, y, n), 0 \leq k < 2n)$  to denote the acyclic convolution of the first  $n$  binary digits of  $x$  and the first  $n$  binary digits of  $y$ . To be specific, let the binary digits

of  $x$  and  $y$  be indexed starting with 0. Then

$$A_k(x, y, n) = \sum_{j=0}^{n-1} x_j y_{k-j}, \quad 0 \leq k < 2n,$$

where by the subscript  $k - j$  we mean  $k - j + n$  if  $k < j$ .

**Theorem 4.4.1** *Let  $r$  be a rational number,  $0 < r < 1$ , and suppose that  $s = \sqrt{r}$  is irrational. Then*

$$\liminf_{n \rightarrow \infty} \frac{B_n(s)}{\sqrt{C(r)n}} \geq 1,$$

We first prove two lemmas.

**Lemma 4.4.2** *For any two positive integers  $a$  and  $b$ ,  $B(a + b) \leq B(a) + B(b)$ , and  $B(ab) \leq B(a)B(b)$ . The same is true for reals, namely  $B_n(x + y) \leq B_n(x) + B_n(y)$  and  $B_n(xy) \leq B_n(x)B_n(y)$ , provided we first trim  $x$  and  $y$  to  $n$  bits.*

**Proof.** The assertion on sums is easily shown by verifying it for one- and two-bit values of  $a$  and  $b$ , and then arguing by induction on the number of bits in  $a$  and  $b$ . The assertion on products follows from the first by observing that the process of binary multiplication is merely a sequence of binary additions with shifts by appending zeroes.  $\square$

**Lemma 4.4.3** *For  $x$  and  $y$  with  $0 < x, y < 1$ , define*

$$z_n(x, y) = \sum_{k=0}^{2n-1} A_k(x, y, n) 2^{-k-1}. \quad (4.4.26)$$

*Then  $|z_n(x, y) - xy| < 2^{1-n}$ . In other words, the binary expansion of  $z_n(x, y)$  agrees with  $x \cdot y$  up to approximately  $n$  digits (it might be slightly fewer because of trailing ones, but the bound still holds).*

**Proof.** This follows by the simple observation that the acyclic convolution reproduces the process of multiplication of two binary fractions. The error in using only the first  $n$  bits of the two operands can be found as follows. Write  $x = \alpha_n + \beta_n$ , where  $\alpha_n$  is the value obtained by using only the first  $n$  bits of the binary expansion of  $x$ , with  $\beta < 2^{-n}$ , and similarly write  $y = \gamma_n + \delta_n$ . Since  $z_n(x, y) = \alpha_n \gamma_n$  precisely, we have

$$|z_n(x, y) - xy| < \alpha_n \delta_n + \gamma_n \beta_n + \beta_n \delta_n < 2 \cdot 2^{-n} = 2^{1-n}. \quad (4.4.27)$$

Note that  $z_n(x, y)$  is always an underestimate of  $xy$ , except in the case where  $x$  and  $y$  are binary fractions represented using the expansion terminating in zeroes, in which case  $z_n(x, y) = xy$  exactly for all sufficiently large  $n$ .  $\square$

**Proof of Theorem 4.4.1.** In the hypothesis of Theorem 4.4.1,  $r$  is rational, but  $s = \sqrt{r}$  is irrational. Since  $s$  is irrational,  $|z_n(s, s) - r| < 2^{1-n}$ , and further since the binary expansion of  $s$  has infinitely many ones (and it cannot terminate in zeroes or ones),  $z_n(s, s)$  will always underestimate  $s^2 = r$ . In the case that  $r$  is a binary fraction, this underestimation means that the approximations  $z_n(s, s)$  cannot possibly replicate the binary expansion that terminates in zeroes, but they can and therefore must replicate with increasing fidelity the expansion that terminates in ones. In the case that  $r$  is not a binary fraction, the approximations  $z_n(s, s)$  must replicate with increasing fidelity the unique binary expansion of  $r$ . Here by “replicate with increasing fidelity,” we mean that the relevant binary expansions of  $z_n(s, s)$  and  $r$  agree for all digits up to some position  $n - d$  for some fixed integer  $d$ .

Here we observe that Lemma 4.4.2 also applies to the real fraction multiplication (i.e., convolution) process described in Lemma 4.4.3. Thus  $B_n(z_n(s, s)) \leq B_n^2(s)$ . Now since  $z_n(s, s)$  replicates with increasing fidelity the expansion of  $r$ , we must have, at the least,

$$\liminf_{n \rightarrow \infty} \frac{B_n(z_n(s, s))}{n} = \liminf_{n \rightarrow \infty} \frac{B_n(r)}{n}. \quad (4.4.28)$$

But since  $B_n(r)/n$  converges to  $C(r)$ , this means that given  $\epsilon > 0$ ,

$$n(B_n(s)/n)^2 \geq B_n(z_n(s, s))/n \geq B_n(r)/n - \epsilon \geq C(r) - 2\epsilon \quad (4.4.29)$$

for all sufficiently large  $n$ , which establishes our result.  $\square$

Theorem 4.4.1 only applies to irrational square roots of rational numbers. In a new paper (2003) by the present authors, Richard Crandall, and Carl Pomerance, this result has been significantly strengthened to the following:

**Theorem 4.4.4** *For positive real algebraic  $x$  of degree  $d > 1$  and for any  $\epsilon > 0$  there exists some  $N$  (depending on  $x$  and  $\epsilon$ ) such that for all  $n > N$ ,*

$$B_n(x) > (1 - \epsilon)(2a_d)^{-1/d}n^{1/d},$$

where  $a_d > 0$  is the leading coefficient of the minimum integer polynomial for  $y = x/2^{\lfloor \log_2 x \rfloor}$ .

The proof of this result is given in [19]. Interestingly, this proof involves “BBP tails,” or, in other words, BBP sums beyond certain digit positions, driving home once more the usefulness of the BBP idea.

Whereas with Theorem 4.4.1, we had that the number of ones in the first  $n$  bits of an irrational square root is asymptotically at least a constant times  $\sqrt{n}$ , with Theorem 4.4.4 we have a constant times the  $d$ -th root of  $n$ , where  $d$  is the algebraic degree. So, for example, the number of ones in the first  $n$  bits of  $\sqrt[3]{2}$  must be at least some constant times  $\sqrt[3]{n}$ . Obviously we do not believe that this is the strongest result possible. All experimental evidence suggests that the limiting density of ones in the binary expansion of any algebraic irrational is not only greater than zero, but in fact is precisely  $1/2$ , and further that the density of each of the four two-bit combinations is  $1/4$ , etc. But lacking formal proof, such a supposition could conceivably be incorrect.

## 4.5 Periodic Attractors and Normality

In this section, we shall prove the result mentioned above, namely Theorem 4.2.3, which implies normality for irrational BBP-type constants, conditional on Hypothesis 4.2.2. For some additional results and full details, consult [24], from which we have adapted this material. First, some preliminaries:

As before, we define the norm  $\|\alpha\|$  for  $\alpha \in [0, 1)$  as  $\|\alpha\| = \min(\alpha, 1 - \alpha)$ . With this definition,  $\|\alpha - \beta\|$  measures the shortest distance between  $\alpha$  and  $\beta$  on the unit circumference circle in the natural way. A simple result, which we will term the “dilated-norm rule,” is the following: If  $0 \leq \delta \leq 1/(2\|z\|)$ , then because  $\|\delta z\|$  is now bounded above by  $1/2$ , we have  $\|\delta z\| = \delta\|z\|$ .

A base- $b$  expansion  $\alpha = 0.\alpha_1\alpha_2\alpha_3\cdots$ , where each  $\alpha_j$  is an integer in  $[0, b-1]$ , is taken to be unique for  $\alpha$ . When competing expansions exist, as in decimal  $0.1000\dots = 0.0999\dots$ , we select the variant with trailing zeros. We define the frequency (when it exists) with which a given finite digit string  $(d_1d_2\cdots d_k)$  appears in  $\alpha$  to be the limit as  $N \rightarrow \infty$  of the number of instances where  $\alpha_j = d_1, \alpha_{j+1} = d_2, \dots, \alpha_{j+k-1} = d_k, j \leq N+1-k$ , divided by  $N$ .

For a sequence  $x = (x_0, x_1, \dots)$  of real numbers in  $[0, 1)$ , we define the counting function

$$C(x, c, d, N) = \#(x_j \in [c, d) : j < N).$$

In other words, the  $C$  function gives the count of the first  $N$  elements of the sequence  $x$  that lie in the interval  $[c, d)$ . We now introduce a standard definition from the literature:

**Definition 4.5.1** *A sequence  $x$  in  $[0, 1)$  is said to be equidistributed if for any  $0 \leq c < d < 1$  we have*

$$\lim_{N \rightarrow \infty} \frac{C(x, c, d, N)}{N} = d - c.$$

This definition is identical to that of “uniform distribution modulo 1,” as given in [135, p.1]. The following two basic results are taken from that source:

**Theorem 4.5.2** *Let  $(x_n)$  be equidistributed. If a sequence  $(y_n)$  has the property that  $\{y_n\} \rightarrow \beta$  (constant  $\beta$ ) as  $n \rightarrow \infty$ , then the sequence  $(\{x_n + y_n\})$  is likewise equidistributed. In particular, if  $y_n \rightarrow 0$ , then  $(\{x_n + y_n\})$  is equidistributed.*

**Theorem 4.5.3** *A number  $\alpha$  is  $b$ -normal if and only if the sequence  $(\{b^n\alpha\} : n = 1, 2, 3, \dots)$  is equidistributed.*

We now define the notions of *finite attractor* and *periodic attractor*:

**Definition 4.5.4** *A sequence  $x = (x_n)$  in  $[0, 1)$  is said to have a finite attractor  $W = (w_0, w_1, \dots, w_{P-1})$  if for any  $\epsilon > 0$ , there is some  $K = K(\epsilon)$  such that for all  $k \geq 0$ , we have  $\|x_{K+k} - w_{t(k)}\| < \epsilon$ , for some function  $t(k)$ , with  $0 \leq t(k) < P$ .*

**Definition 4.5.5** *A sequence  $x = (x_n)$  in  $[0, 1)$  is said to have a periodic attractor  $W = (w_0, w_1, \dots, w_{P-1})$ , if for any  $\epsilon > 0$ , there is some  $K = K(\epsilon)$  such that for any  $k \geq 0$ , we have  $\|x_{K+k} - w_{k \bmod P}\| < \epsilon$ .*

**Theorem 4.5.6** *Assume a sequence  $(y_n)$  has the property that  $y_n \rightarrow \beta$  (with  $\beta$  constant) as  $n \rightarrow \infty$ . Then a sequence  $(x_n)$  in  $[0, 1)$  has a finite attractor (alternatively, a periodic attractor) if and only if  $(\{x_n + y_n\})$  does.*

**Proof.** This follows immediately from the  $\epsilon$ -restriction in the definitions of finite attractor and periodic attractor.  $\square$

We will now show that in certain cases of interest here, these two notions of attractor coincide:

**Theorem 4.5.7** *Let  $\alpha$  be real and assume an integer base  $b \geq 2$ . If the sequence  $x = (\{b^n \alpha\})$  has a finite attractor  $W$ , then  $W$  is a periodic attractor, and the structure of the attractor  $W$  is necessarily*

$$W = (w_0, \{bw_0\}, \{b^2w_0\}, \dots, \{b^{P-1}w_0\}).$$

Moreover, each element in  $W$  is rational.

**Proof.** Let  $W = \{w_0, w_1, \dots, w_{P-1}\}$  be the finite attractor for  $x$ . Let  $d = \min_{0 \leq i, j < P} (||w_i - w_j||)$ , and choose  $\epsilon < d/(4b)$ . Let  $W_\epsilon$  be the set of all  $z$  in  $[0, 1)$  such that  $||z - w_i|| < \epsilon$  for some  $0 \leq i < P$ . Then we know that there is some  $K'(\epsilon)$  such that for all  $k > K'$  we have  $x_k \in W_\epsilon$ . Let  $K$  be the first  $k > K'$ , such that  $||x_k - w_0|| < \epsilon$ . We then have:

$$||x_{K+1} - bw_0|| = ||bx_K - bw_0|| = b||x_K - w_0|| < b\epsilon < d/4, (4.5.30)$$

where the second equality follows from the bounded dilation rule mentioned at the start of the present section.

It follows that  $x_{K+1}$  is within  $b\epsilon$  of  $\{bw_0\}$ , and similarly  $x_{K+k+1}$  is within  $b\epsilon$  of  $\{bw_0\}$  whenever  $x_{K+k}$  is within  $\epsilon$  of  $w_0$ , which must occur infinitely often. Since there can be at most one element of the attractor set  $W$  in the region of size  $d/4$  about  $\{bw_0\}$ , and since the choice of  $\epsilon$  above was arbitrary, we conclude that  $bw_0$  must be the element of  $W$  in that region. We can for notational convenience assume that  $w_1 = \{bw_0\}$ . Then  $||x_{K+1} - w_1|| < \epsilon$ , and the argument can be repeated to show that  $x_{K+2}$  is close to  $w_2 = \{b^2w_0\}$ , etc., and finally that  $x_{K+P-1}$  is close to  $w_{P-1} = \{b^{P-1}w_0\}$ . It then follows that the member of  $W$  which  $x_{K+P}$  is close to must be  $w_0$ , since otherwise the  $\epsilon$  region around  $w_0$  would never be visited again by the  $x$  sequence and thus  $w_0$  could not be a

member of the attractor set. Therefore,  $W = (w_0, \{bw_0\}, \dots, \{b^{P-1}w_0\})$ , and  $W$  is a periodic attractor for the  $x$  sequence. Rationality of the attractor points is demonstrated by noting the periodicity condition  $w_0 = \{b^P w_0\}$ , which implies that for some integer  $m$ , we have  $w_0 = m/(b^P - 1)$ , and similarly for the other  $w_i \in W$ .  $\square$

**Theorem 4.5.8** *If the sequence  $(x_n)$  as defined by  $x_0 = 0$ ,  $x_n = \{bx_{n-1} + r_n\}$ , where  $b$  is a positive integer,  $r_n \rightarrow 0$  and  $r_n \neq 0$ , has a finite attractor  $W$ , then  $W$  is a periodic attractor, and each element of  $W$  is rational.*

**Proof.** Let  $W = \{w_0, w_1, \dots, w_{P-1}\}$  be the finite attractor for  $x$ . Let  $d = \min_{0 \leq i, j < P} (||w_i - w_j||)$ , and choose  $\epsilon < d/(4b + 4)$ . Let  $W_\epsilon$  be the set of all  $z$  in  $[0, 1)$  such that  $||z - w_i|| < \epsilon$  for some  $0 \leq i < P$ . Then we know that there is some  $K'(\epsilon)$  such that for all  $k > K'$ , we have  $x_k \in W_\epsilon$  and  $|r_k| < \epsilon$ . Let  $K$  be the first  $k > K'$ , such that  $||x_k - w_0|| < \epsilon$ . We then have (again we use the dilated-norm rule from the start of the present section):

$$\begin{aligned} ||x_{K+1} - bw_0|| &= ||bx_K + r_{K+1} - bw_0|| \leq b||x_K - w_0|| + \epsilon \\ &< (b + 1)\epsilon < d/4. \end{aligned} \tag{4.5.31}$$

The remainder of the proof of this result follows the second paragraph of the proof of Theorem 4.5.7.  $\square$

Now we are in a position to establish the following result:

**Theorem 4.5.9** *The sequence  $(\{b^n \alpha\})$  has a finite attractor if and only if  $\alpha$  is rational.*

**Proof.** Assume that the sequence  $(\{b^n \alpha\})$  has a finite attractor. By Theorem 4.5.7 it then has a periodic attractor. In the definition of a periodic attractor, let  $K$  be the index corresponding to  $\epsilon = 1/(4b)$ , and set  $h = ||x_K - w_0||$ . Suppose  $h > 0$ . Then let  $m = \lfloor \log_b(\epsilon/h) \rfloor$ , and note that  $b^m h \leq \epsilon < b^{m+1} h < b\epsilon < 1/4$ . Thus we can write (once again using the dilated-norm rule):

$$\begin{aligned} ||x_{K+m+1} - w_{m+1 \pmod P}|| &= ||b^{m+1}x_K - b^{m+1}w_0|| = b^{m+1}||x_K - w_0|| \\ &= b^{m+1}h > \epsilon. \end{aligned} \tag{4.5.32}$$

But this contradicts the definition of the periodic attractor. Thus we conclude that  $h = 0$ , so that  $x_{K+k} = w_{k \bmod P}$  for all  $k \geq 0$ . In other words, after at most  $K$  initial digits, the base- $b$  expansion of  $\alpha$  repeats with period  $P$ , so that  $\alpha$  is rational. As for the converse,  $\alpha = p/q$  rational implies the sequence  $(b^n p/q) = ((pb^n) \bmod q)/q$  is periodic, having in fact the period 1 for  $\alpha = 0$  and, for  $p/q$  in lowest terms, the period of the powers of  $b$  modulo  $q$ .  $\square$

We may now prove Theorem 4.2.3.

**Theorem 4.2.3.** *Assuming Hypothesis 4.2.2, every irrational constant with a BBP-type formula for a given base  $b$  is  $b$ -normal.*

**Proof.** Assume Hypothesis 4.2.2. Then suppose a real number  $\alpha$  possesses a formula of the form

$$\alpha = \sum_{k=1}^{\infty} \frac{1}{b^k} \frac{p(k)}{q(k)}, \quad (4.5.33)$$

where  $p(k)$  and  $q(k)$  are integer polynomials, with  $\deg p < \deg q$  and with  $q(k)$  not having any zeroes for positive integers  $k$ . We first observe that  $\alpha$  is rational if and only if the sequence  $(x_n)$ , where

$$x_n = \left\{ bx_{n-1} + \frac{p(n)}{q(n)} \right\}$$

has a finite attractor. To see this, we know from Theorem 4.5.9 that the sequence  $(\{b^n \alpha\})$  has a periodic attractor if and only if  $\alpha$  is rational. Following the BBP strategy, we can write

$$\{b^n \alpha\} = \left\{ \sum_{k=1}^n \frac{b^{n-k} p(k)}{q(k)} + \sum_{k=n+1}^{\infty} \frac{b^{n-k} p(k)}{q(k)} \right\} \quad (4.5.34)$$

$$= \{x_n + t_n\}, \quad (4.5.35)$$

where  $x$  is defined by  $x_0 = 0$  and the recursion

$$x_n = bx_{n-1} + \frac{p(n)}{q(n)},$$

with the sequence  $t$  given by

$$t_n = \sum_{k=1}^{\infty} \frac{1}{b^k} \frac{p(k+n)}{q(k+n)}.$$



Provided that  $\deg p < \deg q$  as in Hypothesis 4.2.2, given any  $\epsilon$  there is some  $n$  such that  $|p(k+n)/q(k+n)| < \epsilon$  for all  $k \geq 1$ . For such  $n$ , we have  $|t_n| < \epsilon \sum_{k \geq 1} b^{-k} = \epsilon/(b-1) \leq \epsilon$ . Thus  $t_n$  converges to zero as  $n \rightarrow \infty$ . Hence, it follows from Theorem 4.5.9 that  $(x_n)$  has a periodic attractor if and only if  $\alpha$  is rational.  $\square$

## 4.6 Commentary and Additional Examples

1. **Borel's normal number theorem.** This theorem asserts that almost every real number has normally distributed digits in any or all bases. It is a special case of the *strong law of large numbers* applied to Bernoulli trials, or of the first *Borel-Cantelli lemma* that if a sequence of events  $A_n$  is such that  $\sum_n P(A_n)$  converges, then  $\limsup_n A_n (= \bigcup_n \bigcap_{m > n} A_m) = \emptyset$ .
  - (a) Show that the real numbers in  $I = [0, 1]$  with no fives in prime positions base ten form a negligible set.
  - (b) Prove the Borel normal law for binary numbers. Hint: Let  $s_n(\omega) = \sum_{k=1}^n r_k(\omega)$  where

$$r_k(\omega) = \begin{cases} +1, & d_k = 1 \\ -1, & d_k = 0, \end{cases}$$

and  $d_k = d_k(\omega)$  is the  $k$ -th digit of  $\omega$ . Note that  $s_n(\omega)/n \rightarrow 0$  if and only if  $\sum_{k \leq n} d_k(\omega)/n \rightarrow 1/2$ . Show for  $\epsilon > 0$  that

$$P[\omega : |s_n(\omega)| \geq n\epsilon] \leq \frac{1}{n^4 \epsilon^4} \int_0^1 s_n^4 d\omega.$$

Then show that

$$\int_0^1 s_n^4 d\omega = n + 3n(n-1)$$

and deduce the  $P[\omega : |s_n(\omega)| \geq n\epsilon] \leq 3/(n^2 \epsilon^4)$ . If we choose  $\epsilon_n$  appropriately  $\sum_n P[\omega : |s_n(\omega)| \geq n\epsilon_n]$  is finite. Conclude that  $s_n(\omega)/n \rightarrow 0$  for almost all  $\omega$  and observe that this is the asserted conclusion.

- (c) Generalize this to an arbitrary base. Conclude that almost every number is absolutely normal. In 1916, Sierpinski gave an “effective” construction of an absolutely normal number. This has recently been expressed recursively in [32]. There are still no “natural” examples of absolutely normal numbers, unless one considers Chaitin’s constant to be one.

The second *Borel-Cantelli lemma* shows that for a sequence of *independent* events when  $\sum_n P(A_n)$  diverges, then  $\limsup_n A_n = 1$ . Together they provide a “zero-one” law. Full details are in [38].

2. **The normal numbers are small topologically.** We note that while the absolutely normal numbers comprise a set of full measure, Salt proved in 1966 that it is of the *first Baire category*. Indeed, let  $b, r$ , and  $n$  be integers  $0 \leq r < b$ ,  $b > 1$ , and for  $0 \leq x < 1$ , let  $N_n(r, x)$  denote the number of occurrences of  $r$  in the first  $n$  terms of the base- $b$  expansion of  $x$ , and let  $L(r, x)$  denote the set of limit points of the sequence whose  $n$ th term is  $N(r, x)/n$ . Then  $L(r, x) = [0, 1]$  for all  $x \in [0, 1)$  with the exception of a set of the first category (i.e., a countable union of nowhere dense sets). By contrast, the *Liouville numbers* form a set of transcendental numbers of measure zero whose complement is first category [50, pg. 352].
3. **Rational times normal is normal.**

**Theorem 4.6.1** *Whenever  $\alpha$  is normal to base  $b$ , then so is  $r\alpha$  for every nonzero positive rational  $r$ .*

**Proof.** This result is an example application of the hot spot lemma, namely Lemma 4.3.2. First, suppose that  $\alpha$  is normal, and consider  $p\alpha$  for a positive integer  $p$ . Let  $\{\cdot\}$  denote fractional part. Then  $\{b^j p\alpha\} \in [x, y]$  implies that one of the following  $p$  mutually exclusive conditions must hold:

$$\begin{aligned} \{b^j \alpha\} &\in [x/p, y/p) \\ \{b^j \alpha\} &\in [x/p + 1/p, y/p + 1/p) \\ \{b^j \alpha\} &\in [x/p + 2/p, y/p + 2/p) \\ &\dots \quad \dots \\ \{b^j \alpha\} &\in [x/p + (p-1)/p, y/p + (p-1)/p). \end{aligned}$$

Since  $\alpha$  is normal, the limiting frequency of each of the above is  $(y - x)/p$ . Thus the limiting frequency of  $\{b^j p \alpha\} \in [x, y)$  is  $p$  times this value, or  $y - x$ . This establishes that  $p\alpha$  is normal.

Now suppose that  $\alpha$  is normal, and consider  $\alpha/p$  for a positive integer  $p$ . We can assume that  $y - x < 1/p$ , because otherwise we can take  $C = 2p$  in the condition of Theorem 4.6.1. Then  $\{b^j \alpha/p\} \in [x, y)$  implies  $\{b^j \alpha\} \in [\{px\}, \{py\})$ , where we understand that in some cases  $\{px\} > \{py\}$ , due to “wrapping” around the unit interval, in which case we take this to mean the union of the two intervals  $[0, \{py\})$  and  $[\{px\}, 1)$ . However, in either case, the total length is  $p(y - x)$ , so that the limiting frequency of  $\{b^j \alpha\}$  in this set is  $p(y - x)$ . Thus we can write

$$\limsup_{n \rightarrow \infty} \frac{\#\{0 \leq j \leq n-1 : \{b^j \alpha/p\} \in [x, \gamma)\}}{n} \leq p(\gamma - x),$$

where we must use  $\leq$  since whereas  $\{b^j \alpha/p\} \in [x, \gamma)$  implies  $\{b^j \alpha\} \in [\{px\}, \{py\})$ , the converse is not true. But this is good enough for Theorem 4.3.2, which then implies that  $\alpha/p$  is normal. This proof is taken from [28]. See also exercise 8.9 in [135, pg. 77].  $\square$

#### 4. Rational plus normal is normal.

**Theorem 4.6.2** *Whenever  $\alpha$  is normal to base  $b$ , then so is  $r + \alpha$  for every nonzero positive rational  $r$ .*

**Proof.** This proof is left as an exercise (see proof of Theorem 4.6.1 above).

#### 5. Proof that Champernowne’s constant is 10-normal. Using the hot spot lemma, one can fairly easily prove that Champernowne’s number $C = 0.12345678910111213141516\dots$ is 10-normal:

- (a) Observe that the decimal expansion of  $C$  consists of a concatenation of sections of  $k$ -long integers, with section  $k$  having length  $L_k = 9k10^{k-1}$ .
- (b) Within section  $k$ , observe, by experimental calculations if desired, that among all  $j$ -long strings of decimal digits with  $j \leq k$ , the all-ones string appears most frequently, and this string appears  $A_{j,k} =$

$(j + 9k)10^{k-j-1}$  times. Prove that this expression (or a somewhat larger upper bound) holds. Hint: Enumerate the number of ways a  $j$ -long string of ones is entirely contained within a  $k$ -digit integer, and also the number of instances where a  $j$ -long string of ones spills from one  $k$ -digit integer to the next, in the Champernowne expansion.

- (c) Let  $n_k$  denote the positions of the successive ends of sections in the Champernowne expansion, i.e.,  $n_k = L_1 + L_2 + \cdots + L_k$ . Calculate an upper bound for  $c_{j,k} = A_{j,1} + A_{j,2} + \cdots + A_{j,k}$ , the total count of appearances of the  $j$ -long all-one string through the end of section  $k$ .
- (d) Apply the hot spot lemma, namely Lemma 4.3.2, to prove that  $C$  is 10-normal. Hint: Note that for  $n_{k-1} < n < n_k$ , the count of appearances of any  $j$ -long string in the first  $n$  digits is no greater than  $c_{j,k}$ , and that for  $n$  in this range,  $n > L_{k-1}$ . Also observe that any given interval contains a subinterval with decimal fraction endpoints, with length at least  $1/10$  of the length of the interval.

6. **Density of binary digits in algebraic irrationals.** In some analysis of the density of binary digits in the expansions of certain algebraic irrationals, the present authors, together with Richard Crandall and Carl Pomerance, attempted to estimate the following function, as  $R$  goes to infinity:

$$U_d(R) = \sum_{m=1}^{\infty} \frac{\binom{R+m+d-1}{d-1}}{2^m} \quad (r, D \in \mathbb{N}).$$

- (a) It is easy to discover that  $U_d(R)$  satisfies the recursion

$$f(R, d) - 2f(R, d-1) = \binom{R+d-1}{d-1}. \quad (4.6.36)$$

Now *Maple* can “tell” one that

$$U_d(R) = \frac{1}{2} \binom{R+d}{d-1} F\left(1, d+R+1, R+2; \frac{1}{2}\right), \quad (4.6.37)$$

and a computer-algebra-assisted use of Euler’s integral for the hypergeometric function simplifies to

$$\sum_{m=1}^{\infty} \frac{\binom{R+m+d-1}{d-1}}{2^m} = \frac{\int_0^1 (2-x)^{d-1} x^R dx}{\beta(R+1, d)} \quad (4.6.38)$$

(the  $\beta$  function is discussed in Section 5.4). From (4.6.38), it is fairly easy to obtain precise asymptotics and estimates as needed. However, it is more relevant that it is now easy by computer to show that the right side of (4.6.38), call this  $V_d(R)$ , satisfies (4.6.36). As

$$V_1(R) = (R+1) \int_0^1 x^R dx = \sum_{m=1}^{\infty} \frac{1}{2^m} = U_1(R) = 1,$$

we are done, in what is an exemplary use of experimental math.

- (b) Problem. Compute a good upper bound for  $U_d(R)$  for large  $R$  and determine its asymptotic behaviour in  $R$ .

Answer. We assume  $R > d$  and begin by writing

$$\mathcal{K}(R, d) = \int_0^1 (2-x)^{d-1} x^R dx, \quad (4.6.39)$$

so that

$$V_d(R) = \frac{\mathcal{K}(R, d)}{\beta(R+1, d)}. \quad (4.6.40)$$

Observe with a change of variables

$$\begin{aligned} \frac{1}{R+1} &< \mathcal{K}(R, d) = \int_0^1 (1+t)^{d-1} (1-t)^R dt \\ &= \int_0^1 (1-t^2)^{d-1} (1-t)^{R+1-d} dt \\ &< \int_0^1 (1-t)^{R+1-d} dt = \frac{1}{R+2-d}, \end{aligned}$$

where the lower bound comes from estimating the original integral. In a similar manner,

$$\begin{aligned} \frac{1}{\beta(R+1, d)} &= \frac{\Gamma(R+1+d)}{\Gamma(R+1)\Gamma(d)} = \frac{1}{(d-1)!} \prod_{k=1}^d \left(1 + \frac{k}{R}\right) \\ &< \frac{R^d}{(d-1)!} \left(1 + \frac{d}{R}\right)^d < \frac{R^d}{(d-1)!} e^{d/R}. \end{aligned}$$

Combining these last two inequalities yields

$$U_d(R) < \left[ \frac{R}{R+2-d} e^{d/R} \right] \frac{R^{(d-1)}}{(d-1)!},$$

for  $R > d$ . If we call this last bracketed expression  $\gamma_d(R)$ , we see that

$$\frac{R}{R+2} \leq \gamma_d(R) < \left( 1 + \frac{1}{m} \right) e^{\frac{1}{m+1}}$$

for  $R > md, m \geq 1$ . Thus, globally the constant is bounded above by  $2\sqrt{e}$  and asymptotically in  $R$  the upper and lower bounds tend to 1. This also shows

$$\limsup_{R \rightarrow \infty} \frac{U_d(R)}{(R^{(d-1)})/(d-1)!} \leq 1.$$

We leave it to the reader to show, using Stirling's formula, that indeed

$$\lim_{R \rightarrow \infty} \frac{U_d(R)}{(R^{(d-1)})/(d-1)!} = 1.$$

(c) Show that

$$U_d(R) = \sum_{j=0}^{d-1} \binom{R+d}{j}.$$

See [19] for further details.

7. **Sierpinski's number.** As noted in this chapter, in 1916 Sierpinski gave an "effective" construction of an absolutely normal number,  $\sigma$ , recently expressed recursively in [32]. That said, the present algorithm does not allow computation of any digits of  $\sigma$ . It would be very interesting to rectify this state of affairs.
8. **A normal number as a pseudorandom number generator.** Recall the constant  $\alpha_{2,3}$  defined in Section 4.3, which is known to be 2-normal. The binary expansion of  $\alpha_{2,3}$  is a good pseudorandom bit generator, and the sequence  $y_n = \{2^n \alpha_{2,3}\}$  of shifted fractional parts is a good  $(0, 1)$  uniform pseudorandom number generator. In particular, the sequence  $y_{53n}$  can be used to generate independent 64-bit IEEE floating-point iterates, each with 53 bits of  $\alpha_{2,3}$ , with a scheme that can be easily implemented in any programming language and is well suited for parallel computation:

- (a) Observe that in the sequence (4.3.18), each section of length  $2 \cdot 3^k$  starts with  $\lfloor 3^k/2 \rfloor / 3^k$ .
- (b) Note that if  $m$  is the largest power of three less than  $n$ , then

$$x_n = \frac{2^{n-m} \lfloor m/2 \rfloor \bmod m}{m}.$$

The exponentiation here can be done using Algorithm 1 (Section 3.4), implemented using 128-bit integer or double-double floating-point arithmetic (see Section 6.2.1).

- (c) Observe that if  $y_n$  is the sequence of shifted fractional parts as defined above, then  $x_n$  is identical to  $y_n$ , to within the 53-bit accuracy of IEEE double floating-point arithmetic, provided that  $n$  is not within 100 of a power of three. Hint: Recall the proof of Theorem 4.2.1.
- (d) Observe that the numerators of successive elements of the sequence  $x_{53n}$  (within a  $2m/3$ -long segment) can be generated using the iteration  $x_{53(n+1)} = qx_{53n} \bmod m$ , where  $q = 2^{53}$ .
- (e) Implement this scheme on a computer, generating  $n$  independent IEEE 64-bit floating-point values in  $(0, 1)$ , each with successive 53-bit sections of  $\alpha_{2,3}$ , beginning at a given starting index  $a$  in the binary expansion of  $\alpha_{2,3}$ , subject to the assumed restriction that the range  $a \leq j < a + n$  does not include any value within 100 of a power of 3.
- (f) Show that the period of the resulting pseudorandom number generator is  $2m/3$ , where  $m$  is the largest power of three less than or equal to the starting index  $a$ . Hint: See [131, pg. 21]. What range should  $a$  be chosen in, so as to insure that the period is as large as possible, yet not so large as to result in numeric overflows, given the computer arithmetic being used (say 128-bit integer or double-double floating-point—see Section 6.2.1)?
- (g) Implement this scheme on a parallel computer, preserving the property that the overall sequence of pseudo-random numbers generated by all processors is identical to that generated by a single-processor program. This property is important in parallel computing not only for debugging purposes, but also to permit flexibility in the number of processors used. Hint: Note that each processor can independently

calculate the starting value of its assigned segment of the sequence. See also Exercise 17 of Chapter 3.

9. **Chaitin's (Omega) constant.** Chaitin's constant is perhaps the most compact information theoretic way to encode undecidability via the *halting problem for Turing machines*. One common accounting of this is as follows. Fix a prefix-free universal Turing machine  $U$ : (i.e., if instances  $U(p)$  and  $U(q)$  each halt, then neither  $p$  nor  $q$  is a prefix of the other.) Then Chaitin's Omega is defined by

$$\Omega = \sum_{\{U(p) \text{ halts}\}} 2^{-|p|}.$$

Despite being intrinsically incomputable, the first 64 binary digits of this version of Chaitin's constant are

0000001000000100000110001000011010001111110010111011101101000010000,

as has been recently established in [66].

10. **Transcendentality and irrationality.**

- (a) Prove that  $\log_a b$  and  $a^{1/b}$  are irrational except in the obvious cases.  
 (b) Prove that  $e$  is irrational by considering

$$\left| \frac{p}{q} - \sum_{n=0}^q \frac{1}{n!} \right| < \frac{1}{q \cdot q!}.$$

- (c) More generally, prove that  $\sum_{n \geq 0} (\pm 1)_n / n!$  is irrational for all choices of  $(\pm 1)_n$ , thus explicitly constructing continuum-many irrational numbers.  
 (d) Establish *Liouville's theorem* that if  $\alpha$  is algebraic of degree  $n$ , then for all  $\varepsilon > 0$  and all  $c > 0$

$$0 < \left| \alpha - \frac{p}{q} \right| < \frac{c}{q^{n+\varepsilon}}$$

has only finitely many solutions with  $p$  and  $q$  integer.



(e) Deduce that  $\sum_{n \geq 0} (\pm 1)_n / 10^{n!}$  is transcendental for all choices of  $(\pm 1)_n$ , thus explicitly constructing an uncountably infinite class of transcendental numbers.

11. **The Copeland-Erdős constant.** This is 0.23571113171923 . . . , obtained by concatenating the primes. It is sequence A033308 in Sloane's online sequence encyclopedia. This and other information may be found by entering the decimal in the Inverse Symbolic Calculator. In 1946, Copeland and Erdős showed that it is 10-normal.
12. **Paul Erdős, a life that added up to something.** In 1996 the mathematical community lost a brilliant and beloved colleague. Here are excerpts from an essay written by Charles Krauthammer of the *Washington Post* [133]:

One of the most extraordinary minds of our time has left. "Left" is the word Paul Erdős, a prodigiously gifted and productive mathematician, used for "died." "Died" is the word he used to signify "stopped doing math." Erdős never died. He continued doing math, notoriously a young person's field, right until the day he died Friday, Sept. 20. He was 83. . . .

He had no home, no family, no possessions, no address. He went from math conference to math conference, from university to university, knocking on the doors of mathematicians throughout the world, declaring "My brain is open" and moving in. His colleagues, grateful for a few days collaboration with Erdős—his mathematical breadth was as impressive as his depth—took him in.

Erdős traveled with two suitcases, each half-full. One had a few clothes; the other, mathematical papers. He owned nothing else. Nothing. His friends took care of the affairs of everyday life for him—checkbook, tax returns, food. He did numbers. . . .

His *Washington Post* obituary ends with this abrupt and rather painful line: "He leaves no immediate survivors."

But in reality he did: hundreds of scientific collaborators and

1,500 mathematical papers produced with them. An astonishing legacy in a field where a lifetime product of 50 papers is considered extraordinary.

Mathematicians tend to bloom early and die early. The great Indian genius, Srinivasa Ramanujan, died at 32. The great French mathematician, Evariste Galois died at 21. (In a duel. The night before, it is said, he stayed up all night writing down everything he knew. Premonition?) And those who don't literally die young, die young in Erdős' sense. By 30, they've lost it.

Erdős didn't. He began his work early. At 20 he discovered a proof for a classic theorem of number theory (that between any number and its double must lie a prime, i.e., indivisible, number). He remained fecund till his death. Indeed, his friend and benefactor, Dr. (of math, of course) Ron Graham, estimates that perhaps 50 new Erdős papers are still to appear.

Erdős was unusual in yet one other respect. The notion of the itinerant, eccentric genius, totally absorbed in his own world of thought, is a cliché that almost always attaches to the adjective "anti-social." From Bobby Fischer to Howard Hughes, obsession and misanthropy seem to go together.

Not so Erdős. He was gentle, open and generous with others. He believed in making mathematics a social activity. Indeed, he was the most prolifically collaborative mathematician in history. Hundreds of colleagues who have published with him or been advised by him can trace some breakthrough or insight to an evening with Erdős, brain open.

The picture of Paul Erdős in Figure 4.2 captures him very well. It is shown at Grossman's *Erdős Number Project* (see <http://www.acs.oakland.edu/~grossman/erdoshp.html>). Using this online tool, one finds, for example, that there are three Borweins with "Erdős number two," which connotes that they each wrote a paper with some direct co-author of Erdős, but not with Erdős himself. One also finds that Roland Girgensohn has Erdős number two, and that David Bailey has Erdős number three.

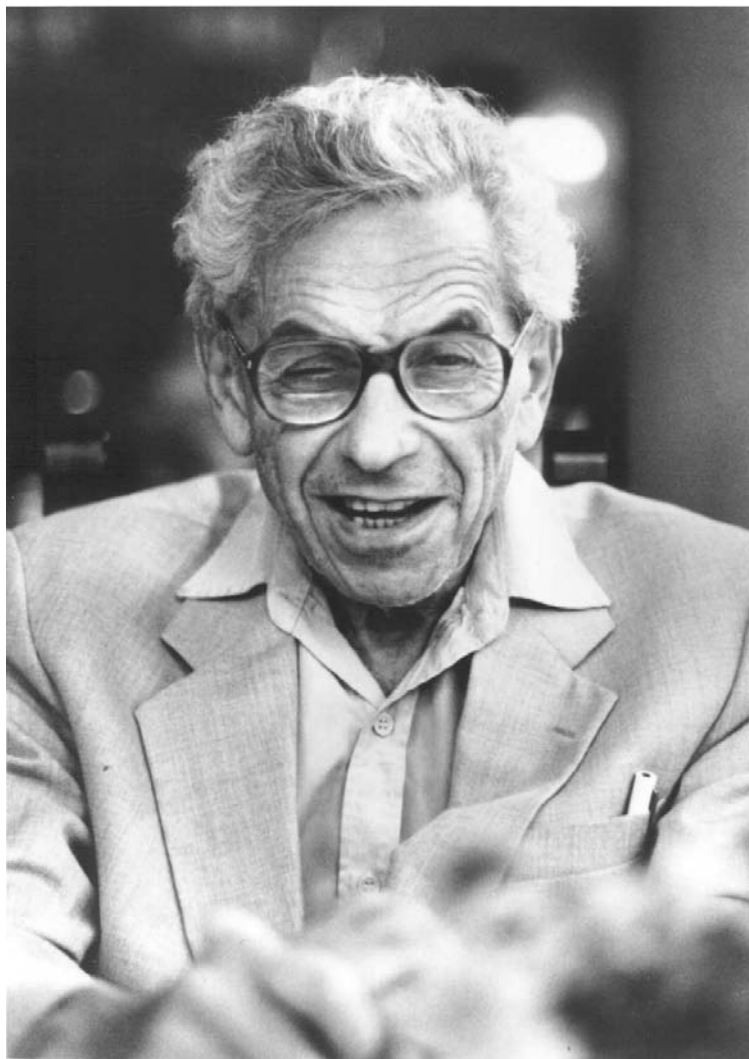


Figure 4.2: A 1993 photo of Paul Erdős.

# Chapter 5

## The Power of Constructive Proofs I

A proof is a proof. What kind of a proof? It's a proof. A proof is a proof. And when you have a good proof, it's because it's proven.

Jean Chretien, Prime Minister of Canada, on the need for evidence of weaponry in Iraq, *CBC News*, Sept. 5, 2002.

The advent of the modern computer mathematics systems makes much previously elusive material tangible and certainly helps drive a taste for the concrete. Even when not explicitly using such software, a constructive computational approach to mathematical proofs and derivations is often significantly more enlightening and intuition-promoting than purely abstract approaches. Illustrating both the power and the pleasure of these constructive approaches is the aim of this chapter.

### 5.1 The Fundamental Theorem of Algebra

We will start this chapter with a proof of the Fundamental Theorem of Algebra, as an experimental mathematician might construct. Let  $p$  be a polynomial defined on the complex plane  $\mathbb{C}$ :

$$p(x) = a_0 + a_1x + a_2x^2 + \cdots + a_nx^n. \quad (5.1.1)$$

Suppose that  $p(x)$  has no roots in  $\mathbb{C}$ . For some sufficiently large  $s$ , we can assume  $|p(x)| > 1$  for all  $x$  with  $|x| \geq s$ , because for sufficiently large  $x$ , the  $x^n$  term dominates. Since  $|p(x)|$  is continuous on the interior and boundary of the circle with radius  $s$ , it follows from fundamental continuity axioms that  $|p(x)|$  achieves a minimum value at a point  $t$  in this (solid) circle. We will see numerous examples in subsequent sections of this “min/max” or “variational” approach in action. Now write

$$q(z) = p(z - t) = q_0 + q_1z + \cdots + q_nz^n. \quad (5.1.2)$$

Presumably the polynomial  $q$  has a global minimum  $M$  at 0, and  $M = |q_0| = |q(0)| > 0$ .

Our proof strategy is to construct some point  $x$ , close to the origin, such that  $|q(x)| < |q(0)|$ , thus contradicting the presumption that  $q$  has a global minimum at zero. One idea that immediately suggests itself to the experimental mathematician is to employ *Newton's method*, for finding roots of a function  $f$ . This is the well-known scheme in which we choose  $x_0$  to be some initial value, and then iterate

$$x_{k+1} = x_k - \frac{f(x_k)}{f'(x_k)} \quad (5.1.3)$$

(see Section 6.2.5). One detail that needs to be observed here is how to handle the case where the derivative in Newton's iteration (5.1.3) is zero. In that case we simply adapt a higher-order approximation from the Taylor's series of the function  $f(x)$ . In particular, if the first  $m$  derivatives of  $f(x)$  at  $x_k$  are zero, but not the  $(m + 1)$ -st derivative, then we can write

$$x_{k+1} = x_k + \left( \frac{-m!f(x_k)}{f^{(m)}(x_k)} \right)^{1/m}, \quad (5.1.4)$$

where the  $m$ -th root of a complex number  $z = ae^{ib}$  is calculated using the well-known formula

$$z^{1/m} = a^{1/m} (\cos(b/m) + i \sin(b/m)). \quad (5.1.5)$$

In general we have no global guarantee that  $x_{k+1}$  is closer to the root than  $x_k$ , but for the purposes of our present proof strategy, we do not require this—we only

require some point  $x$  whose function value is slightly smaller than the function value at zero. If Newton's method gives us merely a direction in the complex plane for which the function value decreases in magnitude (a *descent direction*), then by moving a small distance in that direction, we hope to achieve our goal of exhibiting a complex  $x$  such that  $|q(x)| < |q(0)|$ . This is the strategy we will pursue.

So suppose that the first nonzero coefficient of  $q$ , following  $q_0$ , is  $q_m$  (if there is no nonzero coefficient following  $q_0$ , then the polynomial reduces to a constant):

$$q(x) = q_0 + q_m x^m + q_{m+1} x^{m+1} + \cdots + q_n x^n. \quad (5.1.6)$$

Note that  $q^{(m)}(x) = m!q_m$ . Thus, in accordance with (5.1.4), we will choose  $x$  to be the complex number

$$x = r \left( \frac{-q_0}{q_m} \right)^{1/m} \quad (5.1.7)$$

for some small positive real  $r$  that we will specify below. Now we can write

$$\begin{aligned} q(x) &= q_0 - q_0 r^m + q_{m+1} r^{m+1} \left( \frac{-q_0}{q_m} \right)^{1+1/m} + \cdots + q_n r^n \left( \frac{-q_0}{q_m} \right)^{n/m} \\ &= q_0(1 - r^m) + \mathcal{E}, \end{aligned} \quad (5.1.8)$$

where the error term  $\mathcal{E}$  satisfies

$$|\mathcal{E}| \leq \frac{r^{m+1} \max |q_i|}{1 - s} \left| \frac{q_0}{q_m} \right|^{1+1/m} \quad (5.1.9)$$

for  $s = r|q_0/q_m|$ , which is smaller than one when  $r$  is sufficiently small. In this way,  $\mathcal{E}$  can be made arbitrarily small in ratio to  $|q_0|r^m$  by choosing  $r$  small enough. Thus, for such an  $r$ , we have  $|q(x)| = |q_0(1 - r^m) + \mathcal{E}| < |q_0|$ , which contradicts our original assumption that  $q(x)$  has a global minimum at zero.

This establishes the Fundamental Theorem of Algebra. Continuing by induction, we have constructively proven:

**Theorem 5.1.1** *Every complex polynomial of degree  $n$  has exactly  $n$  complex roots.*

More importantly, we have gained some modest insight into how to practically locate roots. Indeed, a careful scheme along these lines can lead to a fairly robust numerical procedure for finding polynomial roots (this is discussed more in Section 7.3 of the second volume of this work).

### 5.1.1 Liouville's Theorem and the Sine Product

Of course, depending on what tools we consider *elementary*, other proofs may seem equally or more transparent. Suppose we are familiar with the following generalization of the Fundamental Theorem of Algebra, an application of the Cauchy-Goursat theorem:

**Theorem 5.1.2 (Liouville).** *Every entire function in the complex plane with polynomial growth of degree at most  $n$  is a polynomial of degree at most  $n$ .*

Then the assumption that

$$p(x) = a_0 + a_1x + a_2x^2 + \cdots + a_nx^n$$

never vanishes leads almost immediately to the Fundamental Theorem of Algebra.

We illustrate the use of Theorem 5.1.2 by deriving the product formula for the sin function (5.4.41). This will be discussed more completely in Section 5.4. We begin by considering

$$\sigma(z) = \pi z \prod_{n=1}^{\infty} \left(1 - \frac{z^2}{n^2}\right), \quad (5.1.10)$$

for  $z \in \mathbb{C}$ , and analyzing the quotient

$$q(z) = \frac{\sigma(z)}{\sin(\pi z)}. \quad (5.1.11)$$

We rewrite  $\sigma$  as

$$\sigma(z) = \pi z(1-z) \prod_{n=1}^{\infty} \left(1 + \frac{z(1-z)}{n(n+1)}\right),$$

by gathering the terms in  $1 + z/n$  and  $1 - z/(n+1)$  together. This shows that  $\sigma(1-z) = \sigma(z)$  and so that  $\sigma(z+1) = -\sigma(z)$ . Thus,  $q$  is entire as both numerator and denominator have simple zeros at the integers and are pole free. Moreover,

$$q(z+1) = q(z),$$

and

$$q(n) = q(0) = \lim_{z \rightarrow 0} \frac{\sin(z)}{z} = 1, \quad (5.1.12)$$

for each integer  $n$ . It remains to bound  $q$ . First, we write

$$\begin{aligned} \left| \prod_{n=1}^{\infty} \left( 1 - \frac{z^2}{n^2} \right) \right| &\leq \exp \left( \sum_{k=1}^{\infty} \log \left( 1 + \frac{|z|^2}{k^2} \right) \right) \\ &\leq \exp \left( \log(1 + |z|^2) + \int_1^{\infty} \log \left( 1 + \frac{|z|^2}{u^2} \right) du \right) \\ &= \exp(\pi|z| - 2|z| \arctan(1/|z|)) \leq \exp(\pi|z|). \end{aligned} \quad (5.1.13)$$

Then we observe that, as  $q$  has period 1, it suffices to establish that

$$|q(z)| \leq a|z| + b, \text{ when } -1/2 \leq \operatorname{Re}(z) \leq 1/2 \quad (5.1.14)$$

for some  $a, b > 0$ . Once (5.1.14) is established, Theorem 5.1.2 ensures that  $q$  is at most linear, but as  $q$  takes the value 1 infinitely often,  $q(z) \equiv 1$  and we are done. To establish Equation (5.1.14), let us write  $x$  and  $y$  for the real and imaginary parts of  $z$ . We note that

$$\inf_{|x|, |y| \leq 1/2} \left| \frac{\sin z}{z} \right| > 0 \quad (5.1.15)$$

as  $\sin$  has zeroes only at the integers, while for  $|x| \leq 1/2$  and  $|y| \geq 1/2$  we have

$$|\sin(\pi z)| \geq |\sinh(\pi|y|)| \geq c \exp(\pi|z|) - d, \quad (5.1.16)$$

for  $c, d > 0$ . Combining these last two estimates with that of (5.1.10) produces (5.1.14).

Even if we distrust and only partially understand Theorem 5.1.2, and we perform the steps in the previous proof informally, we finish with greater security that the result is true and more reason to search for a compelling and constructive proof.

That said, computer algebra systems are not perfect and this seems an appropriate place to highlight one of the issues most vexing for symbolic computation—selecting the right branch of the inverse of a function.



### 5.1.2 Roots and Branches

We start with two “failures”:

**Example 5.1.3** *Simplifying  $\sqrt{x}$ .*

Evaluating asymptotic expansions for the Airy function (the first solution of  $w'' - zw = 0$ , which one can see in the rainbow), Rob Corless and Dave Jeffrey could not get correct values for negative  $x$ , in *Maple* version 4 (circa 1988).

The problem turned out to be in the calculation of

$$\sqrt{-x}.$$

*Maple* was “simplifying” automatically

$$\sqrt{-x} \rightarrow i\sqrt{x},$$

and therefore when it tried substituting, say,  $x = -4$  it computed as follows.

$$\sqrt{-x} \rightarrow i\sqrt{x}; \quad i\sqrt{x} \rightarrow i\sqrt{-4}; \quad i\sqrt{-4} \rightarrow i^2\sqrt{4} \rightarrow -2.$$

Of course, they wanted  $\sqrt{-(-4)} = 2$ . Corless and Jeffrey’s quick fix was to program

```
temp:= -x; sqrt(temp);
```

The slow fix was to identify a set of problems called generically by *Maple* developers as “The Square Root Bug.” Of course, humans routinely write  $\sqrt{x^2} = x$  and worry about it afterwards.  $\square$

**Example 5.1.4** *Simplifying  $\sin(\arcsin(x))$ .*

One computer algebra system simplified  $\text{SIN}(\text{ASIN}(5/4)) \rightarrow -5/4$  using the “standard” formulae:

$$\arcsin z = \arctan \left( \frac{z}{\sqrt{1-z^2}} \right) \quad (5.1.17)$$

$$\sin(\arctan z) = \frac{z}{\sqrt{z^2+1}}, \quad (5.1.18)$$

and internally used (5.1.17) to convert inverse sine to inverse tangent:

$$\arcsin(5/4) = \arctan(-5i/3).$$

The system then used (5.1.18), with the complex number, and got the wrong sign:

$$\frac{-5i/3}{\sqrt{5^2(-1)/3^2 + 1}} = \frac{-5i/3}{\sqrt{-16/9}} = -5/4.$$

This time the cure was to modify

$$\begin{aligned} \arcsin z &= \arctan\left(\frac{z}{\sqrt{1-z^2}}\right) \\ \sin(\arctan z) &= \frac{z}{\sqrt{z^2+1}} = z \frac{1}{\sqrt{z^2+1}} \end{aligned}$$

to

$$\arcsin z = \arctan\left(\frac{z}{\sqrt{1-z^2}}\right) \quad (5.1.19)$$

$$\sin(\arctan z) = z\sqrt{\frac{1}{z^2+1}}, \quad (5.1.20)$$

and to ensure that

$$\sqrt{\frac{1}{z}} \text{ was not automatically simplified to } \frac{1}{\sqrt{z}}.$$

□

Additionally, the different systems and reference books do not necessarily agree on what the domain of arcsin is (even, for example, in different editions of Abramowitz and Stegun [1]), or where the branch cut for log occurs.

These and other examples show how errors may be caused by any of the following

$$\begin{aligned} \sqrt{xy} &\rightarrow \sqrt{x}\sqrt{y}, & \sqrt{x^2} &\rightarrow x, \\ \sqrt{\frac{1}{z}} &\rightarrow \frac{1}{\sqrt{z}}, & \log(e^z) &\rightarrow z, \\ \log(1/z) &\rightarrow -\log z, & \arctan z &= \arcsin\left(\frac{z}{\sqrt{z^2+1}}\right), \end{aligned}$$

being used in simplifications.

The source of the problems is that in the complex plane, functions such as log, square root, are multi-valued. In each problematic transformation, at least one multi-valued function is being manipulated in the complex plane using a law that applies to the single-valued function defined for real, positive arguments. The issue becomes how to treat multivalued functions (i) in human mathematics, and (ii) in a computer system.

Many solutions have been proposed:

**Example 5.1.5** *Simplifying*  $\log AB$ .

Consider the relation

$$\log(AB) = \log A + \log B .$$

*Maple* proves this is wrong by substituting  $A = B = -1$ .

```
> eq:= log(A*B)=log(A)+log(B);
   eq := ln(A B) = ln(A) + ln(B)
> subs({A=-1,B=-1},eq);
   ln(1) = 2 ln(-1)
> simplify(%);
   0 = 2 I Pi
```

Many people are quite unhappy with this. They reply either (i) Abramowitz and Stegun define the multivalued function  $\text{Log}$  (a Principal Value) and this is the function to use: the correct equation should be

$$\text{Log}(AB) = \text{Log}A + \text{Log}B;$$

or (ii) they observe that *the trouble is that Maple is not working on the Riemann surface*, and if it did, all its problems would be solved.

To explain these responses, several terms need to be specified. A *multivalued function* is a function (e.g., log, inverse sine, square root) that can, in principle, have multiple values associated with it. A function returning *multiple values* is a multivalued function that is defined to return more than one value. The values might be returned as a set or in some other data structure. A *principal-valued function* is a selection of multivalued function that specifies a single preferred value. Corless and Jeffrey then asked whether  $\log(-1)$  represents the single

number  $i\pi$ ? All the solutions  $z$  of the equation  $\exp(z) = -1$ , as the set  $\{i\pi + 2i\pi k \mid k \in \mathbb{Z}\}$ ? Some object in between these two, perhaps a single number  $i\pi + 2i\pi k$  with the value of  $k$  being fixed later?

A traditional response was provided by Carathéodory writing on the equation

$$\log z_1 z_2 = \log z_1 + \log z_2, \quad (5.1.21)$$

for complex  $z_1, z_2$ .

The equation merely states that the sum of one of the (infinitely many) logarithms of  $z_1$  and one of the (infinitely many) logarithms of  $z_2$  can be found among the (infinitely many) logarithms of  $z_1 z_2$ , and conversely every logarithm of  $z_1 z_2$  can be represented as a sum of this kind (with a suitable choice of  $\log z_1$  and  $\log z_2$ ).

The above definition is not an equality between sets. Consider a special case of  $3 \log z = \log z^3$ .

$$3 \log(-1) = \log[(-1)^3] = \log(-1).$$

The left-hand side is the set  $\{3i\pi + 6i\pi k\}$ , but the right-hand is the set  $\{i\pi + 2i\pi k\}$ . Also, mathematically,  $\log z$  does not have a unique value, while computer algebra systems accept a specific value for  $z$  and return a unique  $\log z$ . From another point of view, the value of  $\log z$  is no longer determined solely by the value of  $z$ : The value to be given to  $\log z$  is also determined by the context. Thus, if the left-hand side  $\log(-1)$  equals  $i\pi$ , then the right-hand  $\log(-1)$  equals  $3i\pi$ .

Multiple values are not computational as working with the solution of the cubic equation  $z^3 + 3qz - 2r = 0$  shows. In keeping with Carathéodory, the solution by radicals, ascribed to Cardano (1501–1576) and Tartaglia (1499–1557), is written as

$$z = \left(r + (r^2 + q^3)^{1/2}\right)^{1/3} + \left(r - (r^2 + q^3)^{1/2}\right)^{1/3}.$$

Each square root has two values and each cube root has three, and each term in the addition takes an independent value. Even allowing for symmetry,  $z$  is elected from a set of *more than three* possible  $z$ .

Jeffrey writes:

Traditionally, this inefficiency has not bothered human mathematicians, who patiently try all the solutions until the three correct ones are verified, but it is not desirable in a computer system.

He likewise argues against using Riemann surfaces (others disagree) and ends up recommending the use of principal values in computer systems, with labeled branches to express multivaluedness better, and the use of the so-called *unwinding number*, [122, 76], to keep transformations correct. In conclusion he appropriately notes that there is reluctance to give up the old formulae

$$\left( \left( \left( \left( (z^2)^{1/2} \right)^3 \right)^{1/3} \right)^4 \right)^{1/4} = z ??$$

and that there are limits to what computer algebra systems can introduce, unilaterally, in the way of new mathematical ideas and notation. We refer the reader to [75] for a fuller discussion of the issues in “reasoning about the elementary functions of complex analysis.”

When a program returns a patently negative value for an area or a complex number in a real integral, any alert user will be flagged. While numerical checking will help, it is much more invidious when simplification of an expression takes a wrong branch internally, but returns, say,  $\log(7/9)$  when the right answer is  $\log(4/5)$ .

Suffice it to say that the user of any system should be alert to such issues and on the inherent difficulties in addressing the underlying problems.

## 5.2 The Uncertainty Principle

In this section, we illustrate in detail how one may be led to a larger discovery by directed computation and visualization.

Some readers may be familiar with the uncertainty principle from quantum mechanics, which is often expressed as the fact that the position and momentum of a subatomic-scale particle cannot simultaneously be prescribed or measured to arbitrary accuracy. Other readers may be familiar with the uncertainty principle from signal processing theory, which is often expressed as the fact that a signal cannot simultaneously be “time-limited” and “frequency-limited.” Remarkably, the precise mathematical formulations of these two principles are identical.

Richard Crandall illustrates the uncertainty principle with a “thought experiment.” Imagine that a tuning fork is struck, then very quickly silenced—just a brief burst of sound. If the duration  $\Delta t$  is short enough, the perceived pitch would be ambiguous, so that the “uncertainty” in the frequency  $\Delta f$  would be large. More precisely, the spectrum of such a sound, if viewed on a spectrum analyzer, would be rather broad. If the duration is long, then the listener (and the spectrum analyzer) can hear enough cycles of the signal to gauge the pitch quite accurately, in which case  $\Delta f$  is small—the spectrum would be sharply peaked at the tuning fork’s natural frequency. Thus we can write  $\Delta t \Delta f > C$  for some constant  $C$ . This is the basic idea of the uncertainty principle.

The most common proof of the uncertainty principle is not terribly difficult, although it is hardly intuitive (and certainly not very enlightening) at first reading. We will sketch it here, assuming some elementary facts of analysis. Let us consider a real, continuously differentiable,  $L^2$  function  $f(t)$ , which further satisfies  $f(t)\sqrt{t} \rightarrow 0$  as  $|t| \rightarrow \infty$ . For convenience, we will also assume here that  $f(-t) = f(t)$ , as this insures that the Fourier transform  $\hat{f}(x)$  of  $f(t)$  is purely real, although in general this condition is not necessary. Define (see also Section 2.2 of the second volume):

$$\begin{aligned} E(f) &= \int_{-\infty}^{\infty} f^2(t) dt \\ V(f) &= \int_{-\infty}^{\infty} t^2 f^2(t) dt \\ \hat{f}(x) &= \int_{-\infty}^{\infty} f(t) e^{-itx} dt \\ Q(f) &= \frac{V(f)}{E(f)} \cdot \frac{V(\hat{f})}{E(\hat{f})}. \end{aligned} \tag{5.2.22}$$

**Theorem 5.2.1** *With the above assumptions and definitions,  $Q(f) \geq 1/4$ , with equality if and only if  $f(t) = ae^{-(bt)^2/2}$  for real constants  $a$  and  $b$ .*

**Proof.** By applying the integral form of the Schwarz inequality [4, pg. 256] (see also formula (5.3.32)), to the functions  $tf(t)$  and  $f'(t)$ , we can write

$$\left| \int_{-\infty}^{\infty} tf(t)f'(t) dt \right|^2 \leq \left[ \int_{-\infty}^{\infty} t^2 f^2(t) dt \right] \left[ \int_{-\infty}^{\infty} (f'(t))^2 dt \right]. \tag{5.2.23}$$

Furthermore,

$$\int_{-\infty}^{\infty} tf(t)f'(t) dt = \frac{1}{2} \int_{-\infty}^{\infty} t \frac{df^2(t)}{dt} dt = -\frac{1}{2}E(f) \quad (5.2.24)$$

by applying integration by parts. Let  $g(t) = f'(t)$ . By noting that  $\hat{g}(x) = -ix\hat{f}(x)$ , and by applying Parseval's identity [161, pg. 65] to  $f(t)$  and  $g(t)$ , we obtain, respectively,

$$\begin{aligned} \int_{-\infty}^{\infty} |f(t)|^2 dt &= \frac{1}{2\pi} \int_{-\infty}^{\infty} |\hat{f}(x)|^2 dx \\ \int_{-\infty}^{\infty} |f'(t)|^2 dt &= \frac{1}{2\pi} \int_{-\infty}^{\infty} x^2 |\hat{f}(x)|^2 dx. \end{aligned} \quad (5.2.25)$$

Combining these results, we obtain our desired inequality. Recall that equality in the Schwarz inequality occurs only when the two functions ( $tf(t)$  and  $f'(t)$  in this case) are linear scalings of each other, i.e.,  $f'(t) = ct f(t)$  for some  $c$ . By solving this elementary differential equation, we conclude that the minimum value  $1/4$  is achieved if and only if  $f(t) = ae^{-(bt)^2/2}$  for constants  $a$  and  $b$ . By the way, it is worth noting that  $Q(f)$  is unaffected by a linear scaling of either the function argument  $t$  or the function value  $f(t)$ .  $\square$

Now let us approach this problem as an experimental mathematician might. As was mentioned above, it is natural when studying Fourier transforms (particularly in the context of signal processing) to consider the “dispersion” of a function and to compare this with the dispersion of its Fourier transform. Noting what appears to be an inverse relationship between these two quantities, we are naturally led to consider the expression

$$Q(f) = \frac{V(f)}{E(f)} \cdot \frac{V(\hat{f})}{E(\hat{f})}. \quad (5.2.26)$$

With the assistance of *Maple* or *Mathematica*, we can readily work out some examples, as shown in Table 5.1. We note that each of the entries in the last column of the table is in the range  $(1/4, 1/2)$ . For example, the expression in the bottom right corner of the table is approximately 0.26329. Can we get any lower?

$f(t)$	Interval	$\hat{f}(x)$	$E(f)$	$V(f)$	$V(\hat{f})$	$Q(f)$
$1 - t \operatorname{sgn} t$	$[-1, 1]$	$2(1 - \cos x)/x^2$	$2/3$	$1/15$	$4\pi$	$3/10$
$1 - t^2$	$[-1, 1]$	$4(\sin x - x \cos x)/x^3$	$16/15$	$16/105$	$16\pi/3$	$5/14$
$1/(1 + t^2)$	$[-\infty, \infty]$	$\pi \exp(-x \operatorname{sgn} x)$	$\pi/2$	$\pi/2$	$\pi^2/2$	$1/2$
$e^{- t }$	$[-\infty, \infty]$	$2/(1 + x^2)$	$1$	$1/2$	$2\pi$	$1/2$
$1 + \cos t$	$[-\pi, \pi]$	$2 \sin(\pi x)/(x - x^3)$	$3\pi$	$\pi^3 - 15\pi/2$	$2\pi^2$	$(\pi^2 - 15/2)/9$

Table 5.1:  $Q$  values for various functions.

To further study this problem with an experimental approach, we employ numerical methods, since these permit the above calculations to be performed for a wide variety of functions, analytic or tabular. We can do this by approximating the Fourier transform  $\hat{f}(x)$  by means of a step-function approximation to the integral, as follows (see also Section 6.1). Assume that  $f(t)$  is zero (or sufficiently small) outside  $[-a/2, a/2]$ , and its Fourier transform  $\hat{f}(x)$  is zero (or sufficiently small) outside  $[-b/2, b/2]$ . Select an even integer  $n$  so that  $ab = 2\pi n$ . This usually requires that we adjust  $a$  and  $b$  to be larger than the minimum values needed, particularly if a large value of  $n$  is needed for high accuracy in the approximation.  $n$  is typically chosen to be a power of two. Then we can write

$$\begin{aligned} \hat{f}\left(\frac{bk}{n}\right) &= \int_{-\infty}^{\infty} f(t)e^{-itbk/n} dt \\ &\approx \frac{a}{n} \sum_{j=-\frac{n}{2}+1}^{\frac{n}{2}+1} f\left(\frac{aj}{n}\right) e^{-i(aj/n)(bk/n)} \end{aligned} \tag{5.2.27}$$

$$= \frac{a}{n} \sum_{j=-\frac{n}{2}+1}^{\frac{n}{2}+1} f\left(\frac{aj}{n}\right) e^{-2\pi ijk/n} \tag{5.2.28}$$

for  $-n/2 + 1 \leq k \leq n/2$ . If we adopt the convention that  $f(t) = f(t - a)$  for  $t > a/2$ , and  $\hat{f}(x) = \hat{f}(x - b)$  for  $x > b/2$ , then we can write

$$\hat{f}\left(\frac{bk}{n}\right) = \frac{a}{n} \sum_{j=0}^{n-1} f\left(\frac{aj}{n}\right) e^{-2\pi ijk/n}, \tag{5.2.29}$$

for  $0 \leq k \leq n - 1$ , which is now in a form suitable for the application of widely available fast Fourier transform (FFT) routines.



The integrals in the definition of  $Q$  can now be computed using similar step-function approximations, as follows:

$$\begin{aligned}
E(f) &= \int_{-\infty}^{\infty} f^2(t) dt \approx \frac{a}{n} \sum_{j=-\frac{n}{2}+1}^{\frac{n}{2}+1} f^2\left(\frac{aj}{n}\right) \\
V(f) &= \int_{-\infty}^{\infty} t^2 f^2(t) dt \approx \frac{a}{n} \sum_{j=-\frac{n}{2}+1}^{\frac{n}{2}+1} \left(\frac{aj}{n}\right)^2 f^2\left(\frac{aj}{n}\right) \\
E(\hat{f}) &= \int_{-\infty}^{\infty} \hat{f}^2(x) dx \approx \frac{b}{n} \sum_{j=-\frac{n}{2}+1}^{\frac{n}{2}+1} \hat{f}^2\left(\frac{bj}{n}\right) \\
V(\hat{f}) &= \int_{-\infty}^{\infty} x^2 \hat{f}^2(x) dx \approx \frac{b}{n} \sum_{j=-\frac{n}{2}+1}^{\frac{n}{2}+1} \left(\frac{bj}{n}\right)^2 \hat{f}^2\left(\frac{bj}{n}\right), \quad (5.2.30)
\end{aligned}$$

where we use the results of the FFT scheme for the indicated array of  $\hat{f}$  values.

Given an equispaced array  $(t_k)$  with spacing  $d$ , and a corresponding array  $(y_k)$ , define  $f(t)$  to be the function given by parabolic interpolation in the  $(y_k)$  array. To be precise, this means that for  $t$  in the interval  $(t_k, t_{k+1})$ , define  $f(t) = y_k - h(3y_k - 4y_{k+1} + y_{k+2})/2 + h^2(y_k - 2y_{k+1} + y_{k+2})/2$ , where  $h = (t - t_k)/d$ . We can now employ a simple search strategy to explore the space of continuous functions to find one that minimizes  $Q$ :

**Algorithm 7** *To find a  $Q$ -minimizing function  $f(t)$ :*

Given  $c$  and  $m$ , first construct the  $(m + 3)$ -long abscissa array  $(t_k = ck/m)$ , for  $0 \leq k \leq m + 2$ , which has spacing  $d = c/m$ . Define the tent-function array  $(y_k = 1 - k/m)$  for  $0 \leq k \leq m$ , with  $y_{m+1} = y_{m+2} = 0$ . Define  $f(t)$  as above based on the  $(y_k)$  array. For negative  $t$ , define  $f(t) = f(-t)$ , and define  $f(t) = 0$  for  $|t| > c$ . Initially set  $\delta$  to be some small value, and calculate  $Q_{min} = Q(f)$ , utilizing the scheme described above, for  $f(t)$  as just defined. Now iterate as follows: Starting with  $k = 1$  (not  $k = 0$ ), increase  $y_k$  by  $\delta$  (leaving other  $y_j$  unchanged) and calculate  $Q(f)$  for  $f(t)$  associated with this altered  $y$  array. If  $Q(f) < Q_{min}$ , then replace  $Q_{min}$  by  $Q(f)$ , and replace the  $y$  array with the altered  $y$  array; if not, then decrease  $y_k$  by  $\delta$ , and replace as before if

$Q(f) < Q_{min}$ . Either way, increment  $k$  and perform similar tests until the case  $k = m$  is complete. If for some  $k$  the test  $Q(f) < Q_{min}$  was true, then repeat beginning with  $k = 1$ . If not, then reduce  $\delta$  by a factor of two and start again. Terminate when  $\delta$  is sufficiently small.  $\square$

In our implementation of this procedure, we use the parameters  $a = 128$ ,  $b = 64\pi$  and  $n = 4096$  for the FFT and integration procedures (so that  $ab = 2\pi n$  as required in the FFT). In the search procedure, we set  $c = 8$  and  $m = 32$ . We set the initial  $\delta = 1/16$ , and we terminate when  $\delta < 10^{-6}$ . This requires about 30 minutes on a 2003-era personal computer or workstation. An acceptable, but less accurate, result is obtained in just a minute or so by using the termination condition  $\delta < 10^{-4}$ .

The final experimentally determined  $Q$ -minimizing function  $f(t)$  that we obtain is plotted in Figure 5.1. Needless to say, its shape strongly suggests a Gaussian probability curve. Actually, we have graphed both  $f(t)$  and the function  $e^{-(bt)^2/2}$ , where  $b = 0.45446177$ , on the same plot—they are identical to the precision of the plot! We graphed the difference between these two functions in Figure 5.2. The value of  $Q$  corresponding to this  $f(t)$  is 0.25001158, which indeed is very close to  $1/4$ .

If you do calculations of this sort, you may discover a curious fact. When the numeric computational scheme described above (which is based on discrete, step-function approximations for the various integrals involved) is applied to compute  $Q$  for a scaled Gaussian function (namely a function of the form  $ae^{-(bt)^2/2}$ ), the result of this calculation is very accurate. The value of  $Q$  produced by our program is  $1/4$ , correct to 15 decimal places (using ordinary IEEE double floating-point arithmetic), even with a modest value of  $n$ . For other types of functions, with a step-function interval of  $10^{-4}$  or so, typically only six or seven digits are good.

What is happening here is that we have experimentally uncovered an important and nontrivial fact of numerical analysis: If a function  $f(t)$  and all its derivatives approach zero as  $t$  approaches the endpoints of the interval  $[a, b]$  (which condition is certainly satisfied by the Gaussian probability density function on  $(-\infty, \infty)$ ), then calculating the definite integral of  $f(t)$  on  $[a, b]$  by means of a simple step-function or trapezoidal approximation scheme is extraordinarily accurate [9, pg. 280]. This fact, which is a consequence of the Euler-Maclaurin summation formula, is the basis of some new robust and high-precision numerical

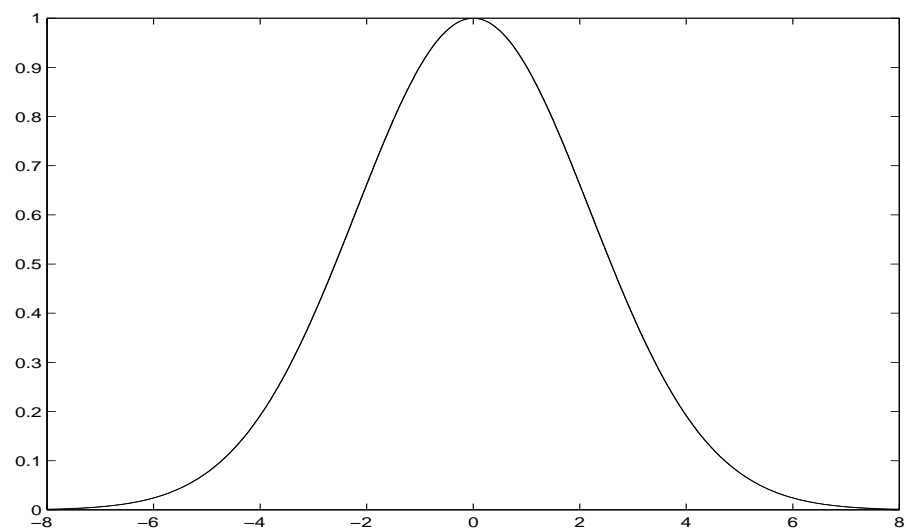


Figure 5.1: Experimental  $Q$ -minimizing function and matching Gaussian.

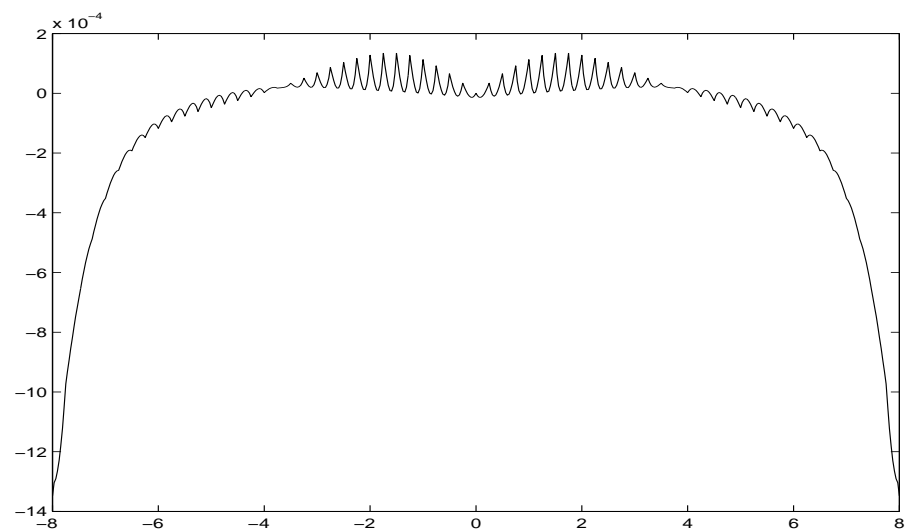


Figure 5.2: Difference between  $Q$ -minimizing function and matching Gaussian.

quadrature algorithms. We shall discuss these schemes, plus the Euler-Maclaurin summation formula, in Section 7.4 of the second volume.

### 5.3 A Concrete Approach to Inequalities

Discovering and establishing inequalities underpins much of mathematics. It has been claimed that most of analytic number theory is the Cauchy-Schwarz inequality in various disguises. Its ubiquitous extension is Hölder's inequality. For  $1 \leq p, q \leq \infty$ , we define the  $p$ -norm of a sequence in  $\mathbb{R}^n$  by

$$\|x\|_p = \sqrt[p]{\sum_{k=1}^n |x_k|^p}.$$

For  $p = \infty$ , the norm is  $\|x\|_\infty = \max_{k=1}^n |x_k|$ , which coincides with  $\lim_{p \rightarrow \infty} \|x\|_p$ . We use  $\langle x, y \rangle$  for the dot-product.

**Theorem 5.3.1** *If  $1/p + 1/q = 1$  then*

$$\langle x, y \rangle \leq \|x\|_p \|y\|_q, \quad (5.3.31)$$

*with equality only if  $x_k$  and  $y_k$  are "aligned" (i.e.,  $x = cy$  for some scalar  $c$ ).*

**Proof.** We may assume  $x$  and  $y$  are nonzero. We proceed in several steps. First, we establish the *Fenchel-Young inequality*

$$\frac{1}{p} |x|^p + \frac{1}{q} |y|^q \geq xy$$

for real  $x$  and  $y$ . This can be done by observing that it suffices to assume  $x, y$  are positive and to use calculus on the function  $x \mapsto (x^p)/p + (y^q)/q - xy$ , which has a critical point at 0 and is convex. Then we have

$$\sum_k \left( \frac{1}{p} x_k^p + \frac{1}{q} y_k^q \right) \geq \sum_k x_k y_k,$$

which is

$$\frac{1}{p} \|x\|_p^p + \frac{1}{q} \|y\|_q^q \geq \langle x, y \rangle.$$

In particular if  $z = x/\|x\|_p^p$ , and  $w = y/\|y\|_q^q$ , we deduce that  $\langle z, w \rangle \leq 1$  as each of  $w, z$  are unit length and  $1/p + 1/q = 1$ . This establishes (5.3.31).

We leave the proof of the equality statement to the reader. They emerge on analyzing when the Fenchel-Young inequality is an equality.  $\square$

The corresponding integral inequality

$$\int_R x(t)y(t) dt \leq \left[ \int_R |x(t)|^p dt \right]^{1/p} \left[ \int_R |y(t)|^q dt \right]^{1/q} \quad (5.3.32)$$

is proven in the same way.

When  $p = q = 2$ , we recover the Cauchy-Schwarz inequality. In general for a function  $f : \mathbb{R}^n \mapsto \mathbb{R} \cup \infty$ , we may define the (*Fenchel-Legendre*) *conjugate* by

$$f^*(y) = \sup_x \langle x, y \rangle - f(x).$$

The conjugate is always convex and lower semicontinuous. First used, less explicitly, by Legendre in the study of adjoint systems of ODE's,  $f^*$  plays a role like the Fourier conjugate in that for convex (lower-semi) continuous functions  $f = f^{**}$ , [54]. For example,  $f = x \rightarrow x \log(x) - x$  has  $f^* = \exp$  and  $f^{**} = f$ . Hence, symbolically performing the code below actually proves that  $f$  is convex. Moreover, the definitional inequality  $f(x) + f^*(y) \geq \langle x, y \rangle$  is what we used in the proof. As we observed in Section 5.1.2, computer algebra systems struggle with branches (i.e., multiple inverses).

Thus, the following *Maple* code

```
conj:=proc(f,y) local x,g:g:=y*x-f(x);
simplify(subs(x=solve(diff(g,x),x)[1],g),symbolic);end;
```

has to be massaged to deal with  $x^p/p$ , but is fine with  $x^4/4$ . We have illustrated that convexity is in some sense easy to handle computationally. A much more sophisticated version of the above is detailed next.

**Example 5.3.2** *A symmetric pair of functions [54].*

(a) Given real  $\gamma_1, \gamma_2, \dots, \gamma_n > 0$ , define  $h : \mathbb{R}^n \rightarrow (-\infty, +\infty]$  by

$$h(x) = \begin{cases} \prod_{i=1}^n x_i^{-\gamma_i} & \text{if } x \in \mathbb{R}_{++}^n \\ +\infty & \text{otherwise.} \end{cases}$$

By writing  $g(x) = \exp(\log g(x))$  and using a composition formula, one can prove that

$$h^*(y) = \begin{cases} -(\gamma + 1) \prod_{i=1}^n \left( \frac{-y_i}{\gamma_i} \right)^{\gamma_i/(\gamma+1)} & \text{if } -y \in \mathbb{R}_+^n \\ +\infty & \text{otherwise,} \end{cases}$$

where  $\gamma = \sum_i \gamma_i$ .

(b) Given real  $\alpha_1, \alpha_2, \dots, \alpha_n > 0$ , define  $\alpha = \sum_i \alpha_i$  and suppose a real  $\mu$  satisfies  $\mu > \alpha + 1$ . Now define a function  $f : \mathbb{R}^n \times \mathbb{R} \rightarrow (-\infty, +\infty]$  by

$$f(x, s) = \begin{cases} \mu^{-1} s^\mu \prod_i x_i^{-\alpha_i} & \text{if } x \in \mathbb{R}_{++}^n, s \in \mathbb{R}_+ \\ +\infty & \text{otherwise.} \end{cases}$$

Using part (a) one may then prove that

$$f^*(y, t) = \begin{cases} \rho \nu^{-1} t^\nu \prod_i (-y_i)^{-\beta_i} & \text{if } -y \in \mathbb{R}_{++}^n, t \in \mathbb{R}_+ \\ +\infty & \text{otherwise} \end{cases}$$

for constants

$$\nu = \frac{\mu}{\mu - (\alpha + 1)}, \quad \beta_i = \frac{\alpha_i}{\mu - (\alpha + 1)}, \quad \rho = \prod_i \left( \frac{\alpha_i}{\mu} \right)^{\beta_i}.$$

The punch line is that on deducing  $f = f^{**}$ , we explicitly represent  $f$  as a convex function. Indeed, one can now show that  $f$  is strictly convex, since  $f^*$  is differentiable.

□

## 5.4 The Gamma Function

The Gamma function naturally extends the factorial and will appear throughout this chapter. We begin with a characterization that again exploits convexity. The  $\Gamma$ -function is usually defined as an integral

$$\Gamma(x) = \int_0^{\infty} \exp(-t)t^{x-1} dt \quad (5.4.33)$$

for  $\operatorname{Re}(x) > 0$ .

**Theorem 5.4.1 (Bohr-Mollerup).** *The  $\Gamma$ -function is the unique function  $f : (0, \infty) \rightarrow (0, \infty)$  with the following three properties:*

1.  $f(1) = 1$ ;
2.  $f(x+1) = xf(x)$ ;
3.  $f$  is log-convex (i.e.,  $x \rightarrow \log(f(x))$  is convex).

**Proof.** We sketch the proof. First we show the integral in (5.4.33) does indeed have the requisite three properties. The first two are easy integration exercises (which can be automated in *Maple* or *Mathematica*). For the key property 3, we use the integral version of Hölder's inequality (5.3.32) to show that

$$\sqrt{\Gamma(x)\Gamma(y)} \leq \Gamma\left(\frac{x+y}{2}\right).$$

Conversely, we set  $g(x) = \log(f(x))$ , we observe that  $g(n+1) = \log(n!)$ , and use the convexity estimate

$$x \log(n) \leq g(n+1+x) - g(n+1) \leq x \log(n+1) \quad (5.4.34)$$

for  $0 < x \leq 1$ .

Hence, for  $0 < x \leq 1$ ,

$$0 \leq g(x) - \log \frac{n! n^x}{x(x+1) \cdots (x+n)} \leq x \log \left(x + \frac{1}{n}\right). \quad (5.4.35)$$

It follows using  $f(x+1) = xf(x)$ , that for all  $x > 0$

$$f(x) = \lim_{n \rightarrow \infty} \frac{n! n^x}{x(x+1) \cdots (x+n)} = \Gamma(x). \quad (5.4.36)$$

□

Equation (5.4.36) often proves useful in its own right, as we shall see in (5.4.37) and in (5.4.44).

Another nice characterization of the Gamma function, which in practice often seems more difficult to apply, is:

**Theorem 5.4.2** *If  $f$  is a complex valued function on  $(a, \infty]$  such that  $f(x+1) = xf(x)$  for all  $x > a$  and*

$$\lim_{n \rightarrow \infty} \frac{f(t+n+1)}{n! n^t} = 1 \quad (5.4.37)$$

for  $0 \leq t < 1$ , then  $f = \Gamma$  on  $(a, \infty]$ .

**Proof.** The fact that  $\Gamma$  satisfies (5.4.37) is (5.4.36). Conversely, if  $f$  and  $g$  satisfy (5.4.37), then consider  $h = f/g$  which satisfies  $\lim_{n \rightarrow \infty} h(t+n+1) = 1$ . Since  $h(x) = h(t+n+1)$ , the two solutions must both coincide with  $\Gamma$ , on  $(a, \infty]$ .

□

We illustrate the use of the Bohr-Mollerup Theorem as follows:

**Example 5.4.3** *The  $\beta$ -function.*

The  $\beta$ -function is defined by the integral

$$\beta(x, y) = \int_0^1 t^{x-1} (1-t)^{y-1} dt \quad (5.4.38)$$

for  $\operatorname{Re}(x), \operatorname{Re}(y) > 0$ . It has the following striking evaluation, often established by using polar coordinates and double integrals (we shall see a third approach in the next section):

$$\beta(x, y) = \frac{\Gamma(x) \Gamma(y)}{\Gamma(x+y)}. \quad (5.4.39)$$



This we shall now deduce from the Bohr-Mollerup result. We define  $f = x \rightarrow \beta(x, y) \Gamma(x + y) / \Gamma(y)$  and proceed to verify the three properties. Point one is an easy integration exercise. For point two, we need to show  $(1 + t/s)\beta(s + 1, t) = \beta(s, t)$ . We observe that  $\beta(s, t + 1) = \beta(s, t) - \beta(s + 1, t)$  which we combine with  $\beta(s + 1, t) = (s/t)\beta(s, t + 1)$ . It remains to show  $f$  is log-convex as again follows by an application of Hölder's inequality. Thus  $f = \Gamma$  as required. In particular, we now obtain the famous evaluation

$$\Gamma\left(\frac{1}{2}\right) = \sqrt{\beta\left(\frac{1}{2}, \frac{1}{2}\right)} = \sqrt{\pi}, \quad (5.4.40)$$

as follows on making the change of variables  $t = \cos(s)^2$  in the definition of the  $\beta$ -function. In the next subsection we shall link this formula to the volumes of  $p$ -balls in  $\mathbb{R}^n$ .  $\square$

The power of this approach is that it provides a generally applicable paradigm and that all or most can be automated or checked computationally—even when the general case may prove too hard to program, special cases can be checked symbolically or numerically, and add security! A second illustration is:

**Example 5.4.4** *The  $\Gamma$ -duplication formula.*

$$\Gamma(2s) = 2^{2s-1} \Gamma(s) \Gamma(s + 1/2) / \sqrt{\pi}.$$

This is established in the same way by considering

$$f(s) = \sqrt{\pi} \Gamma(2s) / (2^{2s-1} \Gamma(s + 1/2)).$$

In this case log-convexity has already been established and the other steps are easy in light of (5.4.40).

To complete this section, we derive the classical product for the sine function and relate it to the  $\Gamma$ -function.  $\square$

**Example 5.4.5** *The sin-product formula.*

$$\sin(\pi x) = \pi x \prod_{n=1}^{\infty} \left(1 - \frac{x^2}{n^2}\right) \quad (5.4.41)$$

and

$$\frac{\pi}{\sin(\pi x)} = \Gamma(x)\Gamma(1-x). \quad (5.4.42)$$

Now (5.4.42) follows directly from (5.4.41) and (5.4.36). We outline a proof of (5.4.41) whose full details are in [184]. We examine the polynomial

$$P_m(x) = \frac{1}{2i} \left[ \left(1 + \pi i \frac{x}{m}\right)^m - \left(1 - \pi i \frac{x}{m}\right)^m \right] \rightarrow_m \sin(\pi x),$$

and check that the roots of  $P_{2m}$  are  $0, \pm(2k/\pi) \tan(j\pi/(2k))$  for  $0 < j < k$ . As the coefficient of  $x$  in  $P_{2m}$  is  $\pi$ , we may write

$$P_{2m}(x) = x \prod_{j=1}^{k-1} \left(1 - \frac{\pi^2 x^2}{4k^2 \tan^2(j\pi/(2k))}\right). \quad (5.4.43)$$

Finally, one must take care to legitimate the passage to the limit in (5.4.43), using the fact that  $\tan(x)/x \rightarrow 1$  as  $x \rightarrow 0$ .  $\square$

One may heuristically derive (5.4.41) as Euler did by considering  $\sin(\pi x)$  as an “infinite” polynomial and rewriting it as a product in terms of the roots  $0, \{1/n^2\}$ . It is thus plausible that

$$\frac{\sin(\pi x)}{x} = c \prod_{n=1}^{\infty} \left(1 - \frac{x^2}{n^2}\right), \quad (5.4.44)$$

and Euler argues that as with a polynomial,  $c$  should be the value at zero and the coefficient of the linear term (in  $x^2$ ) should be the sum of the roots:  $\sum_n n^{-2} = \pi^2/6$ . The argument above made this both less obvious and more rigorous. Also, we emphasize that each step may be checked or discovered formally in many computer algebra packages.

### 5.4.1 Volumes of Balls

The volume of the unit ball in the  $\|\cdot\|_p$ -norm,  $V_n(p)$ , was first determined by Dirichlet by explicitly evaluating the iterated integrals. He obtained

$$V_n(p) = 2^n \frac{\Gamma(1 + \frac{1}{p})^n}{\Gamma(1 + \frac{n}{p})}, \quad (5.4.45)$$

([5], Section 1.8). We note that the  $\Gamma$ -function again proves invaluable, and that even the special case when  $p = 2$ ,

$$V_n = 2^n \frac{\Gamma(\frac{3}{2})^n}{\Gamma(1 + \frac{n}{2})} = \frac{\Gamma(\frac{1}{2})^n}{\Gamma(1 + \frac{n}{2})},$$

gives a formula more concise than that recorded in most analysis books. One can try graphing this as a function of  $p$  for  $n = 2, 3, \dots$ , and observing the limiting cases  $p = 1, \infty$ .

The following *Maple* code derives this formula as an iterated integral for arbitrary  $p$  and fixed  $n$ . The intermediate steps again give beta function values. The code can easily be converted into a human proof valid for arbitrary  $n$ .

```

vol := proc(n)
  local f,x,i,ul,u,j,t; global p;
  p := evaln(p);
  if n=1 then 2 else
    f := (1-add(x[i]^p,i=1..n-1))^(1/p);
    for i from n-1 by -1 to 1 do
      f := subs(x[i]=t,f); f := int(f,t);
      ul := 1-add(x[j]^p,j=1..i-1); u := ul^(1/p);
      f := subs(t^p=ul,f); f := subs(t=u,f);
      f := map(normal,f); f := simplify(f);
    od;
    2^n*f;
  fi; end:

```

It is also fun to determine in which dimension the volume of the Euclidean ball is maximized.

Let us recall the *Laplace transform* of a function  $f$  defined by

$$L(z) = L(f)(z) = \int_0^\infty e^{-zt} f(t) dt$$

for  $\operatorname{Re}(z) > 0$ . We finish the section with another, induction-free, Laplace transform-based, proof of (5.4.45). We shall evaluate

$$v_p(y) = \operatorname{Vol}_n(\{x : \|x\|_p^p \leq y\}),$$

for  $y > 0$  as follows. First, observe that

$$v_p(y) = v_p(1) y^{n/p}.$$

We now compute

$$L(v_p)(1) = v_p(1) L(y \mapsto y^{n/p})(1) = v_p(1) \Gamma\left(1 + \frac{n}{p}\right). \quad (5.4.46)$$

To evaluate the left-hand side we write

$$\begin{aligned} L(v_p)(1) &= \int_0^\infty e^{-t} \left( \int_{\|x\|_p^p \leq t} dx \right) dt \\ &= \int_{\mathbb{R}^n} \left( \int_{\|x\|_p^p}^\infty e^{-t} dt \right) dx = \int_{\mathbb{R}^n} e^{-\|x\|_p^p} dx \\ &= \prod_{k=1}^n \int_{\mathbb{R}} e^{-|x_k|^p} dx_k = 2^n \Gamma\left(1 + \frac{1}{p}\right)^n. \end{aligned} \quad (5.4.47)$$

Now comparing (5.4.46) and (5.4.47) recovers

$$v_p(1) = 2^n \frac{\Gamma(1 + \frac{1}{p})^n}{\Gamma(1 + \frac{n}{p})}$$

as claimed. Moreover, most if not all of this can be performed by a good computer algebra integral transform package.

## 5.5 Stirling's Formula

One of the most fundamental asymptotic formula in analysis is the so-called *Stirling's formula* (actually due to De Moivre) for the factorial, and more generally, for the  $\Gamma$  function.

### Theorem 5.5.1

$$\lim_{n \rightarrow \infty} \frac{n!}{\sqrt{2\pi n} (n/e)^n} = 1 \quad (5.5.48)$$

and moreover, for  $n > 1$

$$1 + \frac{1}{12n + 12} < \frac{n!}{\sqrt{2\pi n} (n/e)^n} < 1 + \frac{1}{12n - 24}. \quad (5.5.49)$$

While (5.5.48) yields an excellent relative error formula, (5.5.49) shows that the absolute error will be large. For example, with  $n = 100$ , the ratio is approximately 0.9997, but the absolute error of the order of  $10^{155}$ . With more work, (5.5.49) can be made significantly more precise, but the version given is often more than sufficient.

**Proof.** To deduce the estimate we consider  $\log n!$  and argue as follows. By *the trapezoidal formula* (see Section 7.4 of the second volume), we have

$$\begin{aligned} n \log(n) - n + 1 &= \sum_{k=1}^{n-1} \int_k^{k+1} \log(x) dx \\ &= \frac{1}{2} \sum_{k=1}^{n-1} [\log(k) + \log(k+1)] + \frac{1}{12} \sum_{k=1}^n e_k^{-2}, \end{aligned} \quad (5.5.50)$$

where  $e_k$  lies in  $[k, k+1]$ . If we denote this last series by  $R_n$ , it is easy to see by the comparison test that it converges to a finite number  $R$ . We may rewrite (5.5.50) as  $(n+1/2) \log(n) - n - \log(n!) \rightarrow R-1$ , or equivalently that  $r_n = n^{-n-1/2} \exp(n) n!$  converges to a limit  $r$ , and so also

$$r = \lim_{n \rightarrow \infty} \frac{r_n^2}{r_{2n}} = \lim_{n \rightarrow \infty} \frac{\Gamma(n+1)^2 4^n \sqrt{2}}{\Gamma(2n+1) \sqrt{n}}. \quad (5.5.51)$$

Squaring, and taking the limit, we obtain

$$r^2 = 4 \prod_{k=1}^{\infty} \frac{4k^2}{4k^2 - 1} = 2\pi, \quad (5.5.52)$$

where this last identity is one of Wallis' famous formulae. It may be derived from Euler's product formula for  $\pi$  (5.4.41) with  $x = 1/2$ , or by repeatedly integrating  $\int_0^{\pi/2} \sin^{2n}(t) dt$  by parts. This establishes (5.5.48).

Now with a little more care, one can obtain (5.5.49) by estimating that  $1/12(n+1) < R - R_n \leq 1/12(n-1)$ , and considering the first few terms of the Taylor series for  $\exp$ .  $\square$

A very important consequence, derived along the way, is that the central binomial coefficient  $\binom{2n}{n} \sim 4^n / \sqrt{\pi n}$ , and you may equally easily estimate other binomial coefficients such as  $\binom{3n}{n}$ , of the form examined elsewhere in this book.

It is interesting to consider how to be led to the correct result. First, to see that  $(n/e)^n \sqrt{n}$  is the right estimate and then to discover that  $r = \sqrt{2\pi}$ , especially given that the convergence is only  $O(1/n)$ .

## 5.6 Derivative Methods of Evaluation

Another potent and concrete way of establishing an identity is by obtaining an appropriate differential equation. For example consider

$$f(x) = \left( \int_0^x e^{-s^2} ds \right)^2$$

$$g(x) = \int_0^1 \frac{\exp(-x^2(1+t^2))}{1+t^2} dt.$$

The derivative of  $f + g$  is zero: In *Maple*,

```
f:=x->Int(exp(-s^2),s=0..x)^2;
g:=x->Int(exp(-x^2*(1+t^2))/(1+t^2),t=0..1);
with(student):d:=changevar(s=x*t,diff(f(x),x),t)+diff(g(x),x);
d:=expand(d);
```

shows this. Hence,  $f(s) + g(s)$  is constant for  $0 \leq s \leq \infty$  and so, on justifying the limit at  $\infty$ ,

$$\left( \int_0^\infty \exp(-t^2) dt \right)^2 = f(\infty) = g(0) = \arctan(1) = \frac{\pi}{4}.$$

The change of variables  $t^2 = x$  shows that this evaluation of the normal distribution agrees with  $\Gamma(1/2) = \sqrt{\pi}$ .

In similar fashion, we may evaluate

$$F(y) = \int_0^\infty \exp(-x^2) \cos(2xy) dx$$

by checking that it satisfies the differential equation  $F'(y) + 2yF(y) = 0$ . We obtain

$$F(y) = \frac{\sqrt{\pi}}{2} \exp(-y^2),$$

since we have just evaluated  $F(0) = \sqrt{\pi}/2$ , and again this is all easily implementable.

### 5.6.1 Legendre's Relation and Applications

We now turn to a study of the *complete elliptic integral of the first and second kinds*

$$\begin{aligned} K(k) &= \int_0^{\pi/2} \frac{1}{\sqrt{1-k^2 \sin^2(s)}} ds \\ E(k) &= \int_0^{\pi/2} \sqrt{1-k^2 \sin^2(s)} ds \end{aligned} \quad (5.6.53)$$

for  $k$  in  $[0, 1]$ .

We denote the *complementary integrals* by  $K'(k) = K(k')$ ,  $E'(k) = E(k')$  with  $k' = \sqrt{1-k^2}$ . A beautiful and important relationship follows:

**Theorem 5.6.1 (Legendre's Relation).** *For all  $0 < k < 1$ ,*

$$E(k) K(k') + K(k) E'(k) - K(k) K'(k) = \frac{\pi}{2}.$$

**Proof.** It is easy to check that  $K$  and  $E$  satisfy the coupled second order differential equation:

$$\frac{dK(k)}{dk} = \frac{E(k) - (1-k^2)K(k)}{k(1-k^2)} \quad \frac{dE(k)}{dk} = \frac{E(k) - K(k)}{k}.$$

These allow us to check, by hand or computer, that

$$f(k) = E(k) K'(k) + K(k) E'(k) - K(k) K'(k)$$

has derivative zero. It remains to determine the constant. The easiest way is to evaluate  $\lim_{t \rightarrow 0} f(t)$ . We observe that  $K(0) = E(0) = \pi/2$  with error of order  $k$ , and that  $E'(0) = 1$ ; since  $K'(k) \sim \log(4/k)$ , we determine that

$$\lim_{t \rightarrow 0} f(t) = \frac{\pi}{2} + \lim_{t \rightarrow 0} (K(t) - E(t))K'(t) = \frac{\pi}{2}.$$

□

Legendre, in a tour de force, actually determined the constant by evaluating the four quantities at the *third singular value*  $k_3 = (\sqrt{3}-1)/\sqrt{8}$ . This is discussed in detail in [50].

In a similar manner, we next evaluate  $K(1/\sqrt{2})$  and  $E(1/\sqrt{2})$ .

```

alias(K=EllipticK,E=EllipticE):k0:=1/sqrt(2);
K_0:=Int(1/sqrt(1-k0^2*sin(s)^2),s=0..Pi/2);
K1:=changevar(t=sin(s),K_0,t);assume(x>0,x<1):
K1:=simplify(changevar(x^2=t^2/(2-t^2),K1,x),symbolic);
K1:=simplify(changevar(u=x^4,K1,u));

```

The changes of variables ( $s \rightarrow t \rightarrow x \rightarrow u$ ) given in this *Maple* code reduces to

$$K\left(\frac{1}{\sqrt{2}}\right) = \frac{1}{2\sqrt{2}}\beta\left(\frac{1}{2}, \frac{1}{4}\right) = \frac{\sqrt{\pi}\Gamma(\frac{1}{4})}{\sqrt{8}\Gamma(\frac{3}{4})} = \frac{\Gamma^2(\frac{1}{4})}{4\sqrt{\pi}},$$

on using (5.4.39), (5.4.40) and (5.4.42). By combining similar computations with (5.6.53), we may deduce that

$$E\left(\frac{1}{\sqrt{2}}\right) = \frac{4\Gamma^2(\frac{3}{4}) + \Gamma^2(\frac{1}{4})}{8\sqrt{\pi}}.$$

### 5.6.2 The AGM Iteration

The arithmetic-mean of Gauss and Legendre has already been introduced in previous chapters. Our goal in this section is to establish the basic properties of the iteration. As we noted, Gauss's specific discovery was that the reciprocal of the integral

$$\frac{2}{\pi} \int_0^1 \frac{dt}{\sqrt{1-t^4}} \tag{5.6.54}$$

agreed numerically with the limit of the rapidly convergent *arithmetic-mean iteration* given by  $a_0 = 1$ ,  $b_0 = \sqrt{2}$  and computing

$$a_{n+1} = \frac{a_n + b_n}{2}, \quad b_{n+1} = \sqrt{a_n b_n}. \tag{5.6.55}$$

Indeed, the sequences  $a_n, b_n$  have a common limit 1.1981402347355922074...

It is a nice exercise to show that for all positive initial values, the iteration converges quadratically to a shared limit which we denote  $M(a, b)$ , see [50]. It was the genius of Gauss that saw the limit. The next result gives a proof especially suited to computer verification or discovery.



**Theorem 5.6.2**

$$\frac{\pi/2}{M(a,b)} = \int_0^{\pi/2} \frac{dt}{\sqrt{a^2 \cos^2(t) + b^2 \sin^2(t)}}. \quad (5.6.56)$$

**Proof.** We first show that

$$I(a,b) = \frac{\pi/2}{M(a,b)}$$

satisfies

$$I(a,b) = I(A(a,b), G(a,b)), \quad (5.6.57)$$

where  $A(a,b) = (a+b)/2$  and  $G(a,b) = \sqrt{ab}$ . Once this is shown, we have

$$I(a,b) = I(a_n, b_n) = \cdots = I(M(a,b), M(a,b)) = \frac{\pi/2}{M(a,b)},$$

as required since  $I(c,c) = \pi/(2c)$ , as it is easy to justify exchanging the limit and the integral.

To establish (5.6.57), consider the following *Maple* code.

```
assume(a>0,b>0):
J:=simplify(changevar(s=b*tan(t),I(a,b),s),symbolic);
J1:=changevar(u=(s-a*b/s)/2,J,u); lprint(%);
j:=4/(a^2+(u-(u^2+a*b)^(1/2))^2)^(1/2)/
((u-(u^2+a*b)^(1/2))^2+b^2)^(1/2)
*(u-(u^2+a*b)^(1/2))^2/((u-(u^2+a*b)^(1/2))^2+a*b);
j0:=1/sqrt((A(a,b)^2+u^2)*(G(a,b)^2+u^2));
```

The code implements the change of variables  $s = b \tan(t)$  and  $u = (s - ab/s)/2$ . The most problematic step is the verification that the consequent integrand,  $j0/j$ , simplifies as required. This is achieved by a *resultant* computation:

```
S:=simplify((j0/j)^2,radical);
p1 := R^2 - (u^2+a*b);
p2 := subs((u^2+a*b)^(1/2) = R, Zero - (numer(S) - denom(S)));
resultant(p1,p2,R); simplify(series(j/j0,u,8));
```

and for peace of mind, the series of  $j/j_0$  can be computed!  $\square$

Placing together (5.6.54) and (5.6.56), we have succeeded in establishing the discovery of Gauss in the form

$$M(\sqrt{2}, 1) = \sqrt{\frac{2}{\pi}} \Gamma^2\left(\frac{3}{4}\right) = 1.1981402347355922074399224922803238 \dots,$$

using five steps of the iteration.

It is also possible to discover the form of the substitution  $u = (s - ab/s)/2$ , by taking a general quadratic form and solving for the parameters.

### 5.6.3 Other Mean Iterations

Finally, the more general study of means [50] also provides many opportunities for experimentation and computation. Consider a *mean iteration*, in which we start with  $a_0 > 0$  and  $b_0 > 0$  and iterate the discrete dynamical system

$$a_{n+1} = M(a_n, b_n), \quad b_{n+1} = N(a_n, b_n)$$

for two strict *means*, that is  $a > b$  implies  $a > M(a, b) > b$ . This is a *Gaussian iteration*. The limit exists and is denoted  $(M \otimes N)(a, b)$ , and exhibits quadratic convergence when the means are symmetric. The triumph of the AGM is that we identify quadratic convergence to an important nonelementary function.

As a second example, consider

$$a_{n+1} = \frac{a_n + \sqrt{a_n b_n}}{2}, \quad b_{n+1} = \frac{b_n + \sqrt{a_n b_n}}{2}.$$

While it is hard to guess the (linearly convergent) limit, once told that it is the *logarithmic mean*

$$\mathcal{L}(a, b) = \frac{a - b}{\log(a) - \log(b)},$$

proof of this fact reduces to showing the identity

$$\mathcal{L}(a, b) = \mathcal{L}\left(\frac{a + \sqrt{ab}}{2}, \frac{b + \sqrt{ab}}{2}\right). \quad (5.6.58)$$

The following *Maple* code implements the iteration:

```
L:=proc(a,b,n) local k,c,d; c:=a;d:=b; for k to n
do(c,d):=(M(c,d),M(d,c)) od;c;end;
```

Thus, a hundred iterates yields

$$L(2., 1., 100.) = 1.4426950408889634073599246810014976 \dots,$$

while

$$\mathcal{L}(2, 1) = 1.4426950408889634073599246810018921 \dots,$$

which agree to 30 places.

This illustrates the *invariance principle* asserting that limit is the unique mean  $f$  with  $f(a_n, b_n) = f(a_{n+1}, b_{n+1})$ . In this case the means are not symmetric and convergence is only linear.

Similarly, we met the *Archimedian iteration* in Chapter 3. In this language, it is due to Gauss, Pfaff, Schwab, and Borchardt in the 19th century and establishes that

$$a_{n+1} = \frac{2a_n b_n}{a_n + b_n}, \quad b_{n+1} = \sqrt{a_{n+1} b_n}$$

converges to

$$\mathcal{B}(a, b) = \frac{\sqrt{b^2 - a^2}}{\arccos(a/b)}$$

for  $0 < a < b$ , and to

$$\mathcal{B}(a, b) = \frac{\sqrt{a^2 - b^2}}{\operatorname{arccosh}(a/b)}$$

for  $0 < b < a$ . Armed with the invariance principle, the reader or her computer should be able to prove this assertion, and to determine the (linear) rate of convergence.

#### 5.6.4 Gregory and Euler

We conclude this section by providing computer-verifiable proofs of

$$\frac{\pi}{4} = \sum_{k=1}^{\infty} \frac{(-1)^{k-1}}{2k-1} \tag{5.6.59}$$

and

$$\frac{\pi^2}{6} = \sum_{k=1}^{\infty} \frac{1}{k^2}. \quad (5.6.60)$$

For the former, consider

$$\cot(x) = \frac{\cot(x/2) - \cot(\pi/2 - x/2)}{2} \quad (5.6.61)$$

and then observe that a telescoping identity yields

$$2^{-n} \sum_{k=1}^{2^n-1} (-1)^k \cot\left(\frac{(2k+1)\pi}{4}\right) = 1.$$

Recall that  $2^n \tan(a/2^n) \rightarrow a$ , as  $n \rightarrow \infty$ .

As always, we need to justify taking limits. The justification follows from Tannery's theorem (a form of dominated convergence) and we have proven Gregory's formula (5.6.59).

Similarly, check that

$$\csc^2(x) = \frac{\csc^2(x/2) + \csc^2(\pi/2 + x/2)}{4}, \quad (5.6.62)$$

and that a telescoping sum yields

$$2 \cdot 4^{-n} \sum_{k=1}^{2^{n-1}-1} \csc^2\left(\frac{(2k+1)\pi}{2^{n+1}}\right) = 1.$$

Again, taking limits carefully establishes

$$\frac{\pi^2}{8} = \sum_{k=1}^{\infty} \frac{1}{(2k-1)^2},$$

which is equivalent to (5.6.60), on separating odd and even terms.

While the technical details are important, from our perspective it is more interesting that we can perform all the steps experimentally, and worry about their legitimacy at the end, once we "know the truth!"

## 5.7 Commentary and Additional Examples

1. **A 3-D maximum problem.** Determine the maximum of

$$f : (x, y, z) \mapsto \frac{\sin(x)^2}{x} + \frac{\sin(y)^2}{y} + \frac{\sin(z)^2}{z}$$

for nonnegative  $x + y + z = \pi$ . [Taken from [106].]

Solution: The unique maximum can be seen numerically to be  $x = y = z = 1.047197551196597\dots$ , which is clearly  $\pi/3$ , and by symmetry one does expect that  $x = y = z$ . Then observe that for all variables in  $[0, \pi/2]$ ,  $f$  is strictly concave and so the critical point is a strict maximum value of  $27/(4/\pi)$ . Otherwise we may assume  $x > \pi/2$  and  $0 < y, z \leq \pi/2$ . Let  $g$  denote  $x \mapsto \sin(x)^2/x$ . Hence, by monotonicity and concavity  $f(x, y, z) \leq g(\pi/2) + f(0, y, z) < g(\pi/2) + 2g(\pi/4) = 6/\pi < 27/(4/\pi)$ .

2. **An infinite product formula for the Gamma function.** Show that

$$\Gamma(x) = \frac{e^{-\gamma x}}{x} \prod_{n=1}^{\infty} \left[ \frac{e^{x/n}}{1 + \frac{x}{n}} \right],$$

for  $x > 0$ . Hint: Use the definition of  $\gamma$  and the product obtained in the conclusion of the Bohr-Mollerup Theorem.

3. **A radially invariant generalization of arctan(1).** For each positive integer  $n$ ,

$$\int_{[0,1]^n} \frac{dx_1 \cdots dx_n}{(1 + x_1^2 + \cdots + x_n^2)^{(n+1)/2}} = \frac{\pi^{(n+1)/2}}{(n+1)2^n \Gamma(\frac{n+1}{2})}.$$

Reason: Let

$$A = \int_{[0,1]^n} \frac{dx_1 \cdots dx_n}{(1 + x_1^2 + \cdots + x_n^2)^{(n+1)/2}}$$

$$J = \int_{[0,\infty)^n} \frac{dx_1 \cdots dx_n}{(1 + x_1^2 + \cdots + x_n^2)^{(n+1)/2}}.$$

Each variable is maximal in one  $n$ th of the  $n$ -dimensional hypercube. So symmetry and setting  $y_n = 1/x_n$ ,  $y_j = x_j y_n$  for  $1 \leq j \leq n-1$  gives

$$\begin{aligned}
 J &= n \int_0^\infty \int_{[0, y_n]^{n-1}} \frac{dy_1 \cdots dy_n}{(1 + y_1^2 + \cdots + y_n^2)^{(n+1)/2}} \\
 &= n \int_0^\infty \int_{[0, 1]^{n-1}} \frac{dx_1 \cdots dx_n}{(1 + x_1^2 + \cdots + x_n^2)^{(n+1)/2}} \\
 &= n \int_0^1 \int_{[0, 1]^{n-1}} \frac{dx_1 \cdots dx_n}{(1 + x_1^2 + \cdots + x_n^2)^{(n+1)/2}} \\
 &+ n \int_1^\infty \int_{[0, 1]^{n-1}} \frac{dx_1 \cdots dx_n}{(1 + x_1^2 + \cdots + x_n^2)^{(n+1)/2}} \\
 &= nA + n \int_0^1 \int_{[0, y_n]^{n-1}} \frac{dy_1 \cdots dy_n}{(1 + y_1^2 + \cdots + y_n^2)^{(n+1)/2}} \\
 &= (n+1)A.
 \end{aligned}$$

Thus

$$A = \frac{J}{n+1} = \frac{1}{(n+1)2^n} \int_{-\infty}^\infty \cdots \int_{-\infty}^\infty \frac{dy_1 \cdots dy_n}{(1 + y_1^2 + \cdots + y_n^2)^{(n+1)/2}}.$$

We may now apply a standard result for radially invariant integrals over  $\mathbb{R}^n$ , in which we denote the surface measure of the unit  $n$ -sphere by  $\sigma(S^{n-1}) = 2\pi^{n/2}/\Gamma(n/2)$ . We have

$$\begin{aligned}
 \int_{-\infty}^\infty \cdots \int_{-\infty}^\infty \frac{dy_1 \cdots dy_n}{(1 + y_1^2 + \cdots + y_n^2)^{(n+1)/2}} &= \sigma(S^{n-1}) \int_0^\infty \frac{r^{n-1}}{(1 + r^2)^{(n+1)/2}} dr, \\
 &= \frac{\pi^{n/2}}{\Gamma(n/2)} \int_0^\infty \frac{u^{n/2-1}}{(1 + u)^{(n+1)/2}} du \\
 &= \frac{\pi^{(n+1)/2}}{\Gamma((n+1)/2)},
 \end{aligned}$$

via the beta integral, and the result follows.

4. **The Psi function.** Many sums can be expressed in terms of the  $\Psi$  function and its derivatives, also known as *polygamma functions*. This is because

$$\Psi(z) = \frac{\Gamma'(z)}{\Gamma(z)} = \sum_{n=1}^{\infty} \frac{z}{n(z+n)} - \frac{1}{z} - \gamma,$$

where the first equality is definitional.

(a) Show that  $\Psi(1+z) = \Psi(z) + 1/z$ , that  $\Psi(1) = -\gamma$ ,  $\Psi(1/2) = -\gamma - 2 \log 2$ , and  $\Psi(n) = \sum_{k=1}^n 1/k - \gamma$ . Obtain a similar formula for  $\Psi(n + 1/2)$ , for  $n \in \mathbb{N}$ .

(b) Prove that

$$\Psi(z) - \Psi(1-z) = \pi \cot(\pi z).$$

(c) Obtain a duplication formula for  $\Psi$  from that of  $\Gamma$ .

(d) Show that

$$\Psi(z) = \int_0^\infty \frac{e^{-t} - e^{-zt}}{1 - e^{-t}} dt - \gamma$$

and that for  $n = 1, 2, 3, \dots$

$$\Psi^{(n)}(z) = (-1)^{n+1} \int_0^\infty \frac{t^n e^{-zt}}{1 - e^{-t}} dt$$

for  $\operatorname{Re} z > 0$ .

5. **Derangements.** A derangement is a permutation of  $n$  symbols that leaves no symbol in the same place. (a) Show that the number of derangements of  $n$  symbols,  $d_n$ , satisfies the recursion

$$d_n = (n-1)d_{n-1} + (n-1)d_{n-2}$$

and has  $d_1 = 0, d_2 = 1$ . Hence, find (b) an exponential generating function  $\mathcal{E}$ , and (c) a closed form for  $d_n$ .

Answer: (b)

$$\frac{\exp(-x)}{1-x} = \sum_{k=1}^{\infty} \frac{d_k}{k!}.$$

(c) Hence,

$$d_n = n! \sum_{k=0}^n \frac{(-1)^k}{k!}.$$

Hint: (a) Consider whether two objects are interchanged or not. (b) The differential equation satisfied by  $\mathcal{E}$  is  $(1-x)\mathcal{E}' = x\mathcal{E}$ ; now deduce (c). Alternatively, write  $d_n = n!b_n$  so that  $b_n - b_{n-1} = -(b_{n-1} - b_{n-2})/n$  and solve directly.

6. **An ODE pair.** Find a first order system of differential equations satisfied by

$$a(x) = \int_0^\infty \frac{e^{-t} \sin(xt)}{\sqrt{t}} dt \quad (5.7.63)$$

$$b(x) = \int_0^\infty \frac{e^{-t} \cos(xt)}{\sqrt{t}} dt \quad (5.7.64)$$

for  $x > 0$  and so, or otherwise, solve for  $a$  and  $b$  explicitly. Answer:

$$\sqrt{\frac{\pi (\sqrt{1+x^2} \pm 1)}{2(1+x^2)}},$$

are the two functions.

7. **Maximum volume of unit ball.** Find the integer dimension  $N$  where the volume of the unit ball is maximized. For what *real* dimension is the volume maximized?
8. **A simple recursion.** Consider the recursion with  $a_0 = e, a_1 = 2e$  and

$$na_n = 2(a_{n-1} + a_{n-2}).$$

Show that the ordinary generating function satisfies  $\mathcal{O}' = 2(1+x)\mathcal{O}$ , and thus  $\mathcal{O}(x) = \exp((1+x)^2)$ . Hence, show

$$a_{2n} = \sum_{m \geq 0} \frac{1}{(n+m)!} \binom{2n+2m}{2n},$$

and find a similar formula for the odd terms.

9. **Base-four sequences.** There are  $4^n$  base-four sequence of length  $n$ ? How many are there with an odd number of zeros? Answer: There are  $2^{n-1}(2^n - 1)$  such sequences. Hint: The number is

$$\sum_{\pi} \frac{1}{m_0! m_1! m_2! m_3!}$$

summed over all partitions of  $n = m_0 + m_1 + m_2 + m_3$  with  $m_0$  odd. The exponential generating function is  $\mathcal{E}(x) = \sinh(x)(\exp(x))^3$ .



10. **Putnam problem 1997–B5.** Let  $\tau_0 = 1$  and  $\tau_n = 2^{\tau_{n-1}}$ , for  $n \geq 1$ . Show that  $\tau_n \equiv \tau_{n-1} \pmod{n}$ . Hint: First, try some examples using *Maple* or *Mathematica*. Then use Fermat's little theorem to show inductively the stronger claim that  $\tau_m \equiv \tau_{n-1} \pmod{n}$  for  $m \geq n - 1$ .
11. Show that under very general conditions on  $f$ , the identity

$$\int_{-\infty}^{\infty} f(x) dx = \int_{-\infty}^{\infty} f(x - 1/x) dx$$

holds. Use this to prove

$$\int_{-\infty}^{\infty} e^{-x^2 - 1/x^2} dx = \frac{\sqrt{\pi}}{e^2}.$$

12. **Berkeley problem 3.2.3.** Solve the differential equation

$$\frac{d^2}{dx^2} y(x) - 2 \frac{d}{dx} y(x) + y(x) = \sin(x)$$

with initial conditions  $y(0) = 1$  and  $y'(0) = 0$ . Hint: Both *Maple* and *Mathematica* can solve this directly. Answer:  $y(x) = (1/2)e^x - (1/2)x e^x + (1/2)\cos(x)$ .

13. **A dilogarithm identity.** The *dilogarithm* is a natural extension of  $-\log(1-x)$  defined by

$$\text{Li}_2(x) = \sum_{n=1}^{\infty} \frac{x^n}{n^2}.$$

Prove, by differentiation, the functional equation

$$\text{Li}_2(x) + \text{Li}_2(1-x) - \log(x)\log(1-x) = \frac{\pi^2}{6}.$$

Hence, show

$$\text{Li}_2\left(\frac{1}{2}\right) = \sum_{n=1}^{\infty} \frac{1}{2^n n^2} = \frac{\pi^2}{12} - \frac{1}{2} \log^2(2).$$

There is a corresponding functional equation for the *trilogarithm*  $\text{Li}_3(x) = \sum_{n=1}^{\infty} x^n/n^3$ . The functional equation developed by Landen around 1760 is

$$\text{Li}_3(x) + \text{Li}_3(1-x) + \text{Li}_3\left(\frac{x}{-1+x}\right)$$

$$= \zeta(3) + \zeta(2) \log(1-x) - \frac{1}{2} \log^2(1-x) \log(x) + \frac{1}{6} \log^3(1-x),$$

from which one can obtain closed forms at  $x = 1/2$  and also at  $x = (\sqrt{3} - 5)/2$  (see (5.7.65)).

14. **Forensic mathematics.** Lewin [141, 6.3.3] gives as (6.13) the identity

$$\operatorname{Li}_3\left(\frac{3-\sqrt{5}}{2}\right) = \frac{4}{5}\operatorname{Li}_3(1) + \frac{\pi^2}{15} \log\left(\frac{3-\sqrt{5}}{2}\right) - \frac{1}{12} \log^3\left(\frac{3-\sqrt{5}}{2}\right) \quad (5.7.65)$$

due to Landen (1780), and then writes

These calculations are given in Edward's *Treatise on the Integral Calculus*, Vol. 2, and there is quoted from Landen's memoirs the equation

$$\frac{1}{1^3} + \frac{1}{2^3} + \frac{1}{3^3} + \cdots = \frac{\theta^2}{1^3} + \frac{\theta^4}{2^3} + \frac{\theta^6}{3^3} + \cdots, \quad \text{where } \theta = 2 \sin(\pi/10).$$

This result is obviously incorrect, since the terms on the right are all less than the corresponding terms on the left, but a search through the memoirs has failed to reveal any relation of the above type apart from the trivial one in which  $\theta = 2 \sin(\pi/6)$ , though it is difficult to believe this is the result which was intended. It might be pointed out that since  $\theta^2 = (3 - \sqrt{5})/2$ , such a relation between  $\operatorname{Li}_3(1)$  and  $\operatorname{Li}_3[(3 - \sqrt{5})/2]$ , if one should exist, would enable each to be isolated with the aid of (6.13), and there is no reason to expect that  $\operatorname{Li}_3(1)$  can in fact be calculated in this way. Edward's publishers have not succeeded in throwing any further light on the subject, and it looks as if this little mystery may have to remain unsolved.

The mystery remains. Can integer relation methods shed any new light?

15. **Berkeley problem 6.10.7.** Let  $R$  be a commutative ring with a unit. Show that for all positive integers and all ring elements, the ideal generated by  $a^n - 1$  and  $a^m - 1$  is the same as that generated by  $a^{[n,m]} - 1$ . (Here

$[n, m]$  denotes the gcd of  $n$  and  $m$ ). Hint: Let  $n = sd$  and consider the factorization of  $x^{sd} - 1$ . Conversely, write  $d = xn - ym$  and consider  $a^d - 1 = (a^{xn} - 1) - a^d(a^{ym} - 1)$ .

16. **Berkeley problem 7.1.28.** Let  $\mathcal{P}$  denote all polynomials over  $\mathbb{R}$ . Show that  $p \mapsto p+p'$  is an invertible map on  $\mathcal{P}$ . Hint: Try low-degree polynomial examples.
17. **Berkeley problem 7.5.23.** Let  $M_n^1$  be the  $n \times n$  tridiagonal matrix with  $a_{jj} = 2, a_{ij} = -1$  if  $|i - j| = 1$  and all other entries zero. (a) Find the determinant of  $M_n^1$  and (b) show that all eigenvalues are positive. Hint: Use *Maple* or *Mathematica* to find this determinant. Answer: (a)  $\det(M_n^1) = n + 1$ . (b)  $M_n^1$  is positive definite.
18. **Fibonacci and Lucas numbers in terms of hyperbolic functions.** Show that  $F_n$  is a constant times  $i^{-n} \sinh(ns)$  and  $L_n = 2i^{-n} \cosh(ns)$ . Several formulas are then easy to obtain from the addition formulas for  $\sinh$  and  $\cosh$ .
19. **Berkeley problem 7.6.13.** Is there a real square matrix with  $A^2 + 2A + 5I = 0$ ? Hint: Try some  $2 \times 2$  examples. Answer: Yes.
20. **Generalized arithmetic-geometric mean inequality.** Show that the complex extension of the arithmetic-geometric mean inequality,

$$\left| \frac{a+b}{2} \right| \geq \sqrt{|ab|},$$

holds for complex  $a, b$  whose ratio  $a/b$  lies “outside” a certain cardioid in the complex plane. What can be found as a complex extension of  $(a_1 + a_2 + \cdots + a_n)/n \geq (a_1 a_2 \cdots a_n)^{1/n}$  for positive  $a_1, a_2, \cdots, a_n$ ?

21. **Viéte’s product.**

$$\frac{\sqrt{2}}{2} \frac{\sqrt{2 + \sqrt{2}}}{2} \frac{\sqrt{2 + \sqrt{2 + \sqrt{2}}}}{2} \cdots = \frac{2}{\pi}$$

is considered to be the first truly infinite formula [34].

(a) Consider the following analogue of Viète's formula:

$$\mathcal{P} = \frac{\sqrt{2}}{\sqrt{1 + \sqrt{2}} \sqrt{1 + \frac{\sqrt{2}}{\sqrt{1 + \sqrt{2}}}} \sqrt{1 + \frac{\sqrt{2}}{\sqrt{1 + \frac{\sqrt{2}}{\sqrt{1 + \sqrt{2}}}}}} \cdots .$$

By writing a recursion for  $\mathcal{P}$ , determine the limit to high-precision and attempt to identify it. Then prove the conjectured result.

(b) Apply the same method to Viète's product.

(c) Consider  $t_0 = 0$  and  $t \mapsto (1 + \sqrt{3 - 2t})/2$ .

Hint: (a) Consider the mapping  $t_0 = 1/\sqrt{2}$  and  $t \mapsto \sqrt{2/(1+t)}$ . (c) This should lead to a similar looking formula involving roots of two for  $2\sqrt{3}/\pi$ . Answer: (a)  $K(1/\sqrt{2})/4$ .

22. **Berkeley problem 7.6.14.** Show that a square Hermitian solution to  $A^5 + A^3 + A = 3I$  must be the identity. Hint: The minimal polynomial for  $A$  must have only real roots and divide  $t^5 + t^3 + t - 3$  which has one real root at 1.

23. **Berkeley problem 7.9.16.** Let

$$A = \begin{bmatrix} 2 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 2 \end{bmatrix}.$$

Show that any matrix  $B$  that commutes with  $A$  must be a quadratic polynomial in  $A$ . Hint: Consider the explicit form of a quadratic in  $A$  and the information implicit in  $AB - BA = 0$ . This yields a set of linear equations that can be solved using *Maple* or *Mathematica*.

24. **Two proofs without words.** Show that

$$\sum_{n=1}^{\infty} \frac{2^n [1 - \cos(x/2^n)]^2}{\sin(x/2^{n-1})} = \tan\left(\frac{x}{2}\right) - \frac{x}{2}, \quad |x| < \pi$$

and

$$2 \sum_{n=1}^{\infty} \frac{1 - \cos(x/2^n)}{\sin(x/2^{n-1})} = \tan\left(\frac{x}{2}\right), \quad |x| < \pi.$$

These arose from dissecting a kite (recursively) into a circular sector and triangles [68].

25. **From Ramanujan's lost notebook.** The following examples appeared in Ramanujan's lost notebook [36]:

(a) For  $t \geq 0$  and  $a > 0$ , evaluate

$$\begin{aligned} \mathcal{I}(a, t) &= \int_{-t}^{\infty} \frac{a^x}{\Gamma(x+1)} dx \\ &\quad + \int_0^{\infty} \frac{e^{-ax} x^{t-1}}{\pi^2 + \log^2(x)} \left( \cos(\pi t) - \frac{\sin(\pi t)}{\pi} \log(x) \right) dx. \end{aligned}$$

(b) Hence, for  $s \geq 0$ , evaluate

$$\mathcal{J}(s) = \int_0^{\infty} \frac{s^x}{\Gamma(x+1)} dx + \int_0^{\infty} \frac{e^{-sx}}{x(\pi^2 + \log^2(x))} dx = \mathcal{I}(s, 0).$$

(c) For  $t \geq 0$  and  $a > 0$ , evaluate

$$\begin{aligned} \mathcal{K}(a, t) &= \int_{-t}^{\infty} \frac{a^x}{\Gamma(x+1)} dx \\ &\quad + \frac{\pi}{2} \int_0^{\infty} \left( \frac{e^{i\pi(t+ix)}}{a^{t+ix}} \Gamma(t+ix) + \frac{e^{-i\pi(t-ix)}}{a^{t-ix}} \Gamma(t-ix) \right) dx. \end{aligned}$$

(d) For  $a \geq 0$  and  $0 \leq \lambda < 1$  evaluate

$$\lim_{\epsilon \rightarrow 0} \epsilon \sum_{n=0}^{\infty} \frac{a^{(\lambda+n)\epsilon}}{\Gamma(1 + (\lambda+n)\epsilon)}.$$

Hint: (a) by differentiating  $\mathcal{I}$  with respect to  $t$ , show  $\mathcal{I}(a, t)$  is a function of  $a$  alone. By differentiating  $\mathcal{I}$  with respect to  $a$ , identify the function up to a multiplicative constant. Evaluate  $\mathcal{I}(0, 0)$  to obtain the constant.

(c) More elaborately, show  $\mathcal{K}(t, a)$  satisfies the same differential equation and finally, identify the constant by showing  $\mathcal{K}(1, 0) = \mathcal{J}(1)$ .

For (d), identify the limit as a Riemann integral.

Parts (a), (c) and (d) originate on pages 226 and 227 of the *Lost Notebook*, which was uncovered in Trinity College, Cambridge by George Andrews in 1976. Part (d) is a true part of a false “identity” in the notebook.

In a lecture entitled “25 years with Ramanujan’s Lost Notebook: Some General Observations” at the New York Academy of Sciences, Andrews, an Evan Pugh Professor of Mathematics at Pennsylvania State University, recalled the “heart-pounding excitement” of the moment when he realized he had come across a notebook detailing more than 135 pages of work from the last year in the life of the self-educated mathematical genius Ramanujan, the son of an accountant, who died at age 33 in 1920.

“That’s a moment that’s hard to beat,” Andrews said. “It was one of those great moments in life when for completely undeserved reasons you realize that you have stumbled across something that is extremely important.” In a talk that mixed Ramanujan’s biographical details with his mathematical theories, Andrews stressed the “surprise” that characterized the work of Ramanujan, a national hero in India. “Not many mathematicians are on postage stamps,” [Figure 3.2] he noted.

See [http://members.nyas.org/events/conference/mtg\\_02\\_1205.html](http://members.nyas.org/events/conference/mtg_02_1205.html).



# Chapter 6

## Numerical Techniques I

Far better an approximate answer to the right question, which is often vague, than the exact answer to the wrong question, which can always be made precise.

J. W. Tukey, “The Future of Data Analysis,” *Annals of Mathematical Statistics*, 1962 [187]

So far, we have focused on case studies of experimental mathematics in action. In this chapter, we will examine in more detail some of the underlying computational techniques that are used in this type of research. We will also mention a number of widely available tools for implementing these techniques, either commercially or freely via the Internet.

We shall discuss computational techniques such as high-precision arithmetic and integer relation detection, in part because these techniques are significantly more accessible than other aspects of experimental mathematics, such as symbolic computation. Further, the algorithms we describe here can be implemented in ordinary Java, C, C++, or Fortran-90 code. Such code can be enhanced for particular applications, potentially resulting in computer programs that outperform commercial software for certain tasks. In contrast, symbolic computation requires some rather sophisticated techniques and data structures, and is thus not easily implemented by nonspecialists. What’s more, advanced concepts, such as Groebner bases, are involved, which likely are unfamiliar to the majority of readers. As a result, we will, for the most part, not deal with symbolic computation here. Interested readers are referred to Joel Cohen’s newly published books on the topic [71, 72].



We focus here on *practical* algorithms and techniques. In some cases, there are known techniques that have superior efficiencies or other characteristics, but for various reasons are not considered suitable for practical implementation. We acknowledge the existence of such algorithms but do not, in most cases, devote space to them in these books.

## 6.1 Convolutions and Fourier Transforms

Convolution computations frequently arise in experimental mathematics, in applications ranging from polynomial multiplication, which we will address in Chapter 14 of the second volume, to high-precision multiplication, which we will address in Section 6.2.4, and even the computation of the Riemann zeta function. The “acyclic” or “linear” convolution  $C_k(x, y)$  of two  $n$ -long vectors  $x = (x_n)$  and  $y = (y_n)$  is defined as the  $2n$ -long vector

$$A_k(x, y) = \sum_{i+j=k} x_i y_j \quad 0 \leq k < 2n. \quad (6.1.1)$$

The cyclic convolution of  $x$  and  $y$  is defined as  $n$ -long vector

$$C_k(x, y) = \sum_{i+j=k} x_i y_j \quad 0 \leq k < n \quad (6.1.2)$$

$$= \sum_{j=0}^{n-1} x_j y_{k-j}, \quad (6.1.3)$$

where in this case the indices  $i + j$  and  $k - j$  are interpreted modulo  $n$ —i.e.,  $k - j$  is read as  $k - j + n$  if  $k - j < 0$ .

For small values of  $n$ , these convolutions can be calculated explicitly as shown. For larger  $n$  (typically  $n \geq 64$  or so), significantly faster results can be obtained by employing fast Fourier transforms (FFTs). The FFT is merely an efficient means of computing the discrete Fourier transform (DFT) and the inverse discrete Fourier transform, which are defined for a complex sequence

$z = (z_0, z_1, z_2, \dots, z_{n-1})$  as follows:

$$\begin{aligned} F_k(z) &= \sum_{j=0}^{n-1} z_j e^{-2\pi i j k / n} & 0 \leq k < n \\ F_k^{-1}(z) &= \frac{1}{n} \sum_{j=0}^{n-1} z_j e^{2\pi i j k / n} & 0 \leq k < n. \end{aligned} \quad (6.1.4)$$

An acyclic convolution can be reduced to a cyclic convolution simply by extending the  $n$ -long input vectors  $x$  and  $y$  to length  $2n$  with zeroes. When this is done, we can write

$$A_k(x, y) = \sum_{j=0}^{2n-1} x_j x_{k-j}, \quad 0 \leq k < 2n, \quad (6.1.5)$$

where the subscript  $k - j$  is interpreted as  $k - j + 2n$  when  $k - j < 0$ . Cyclic convolutions of size  $2n$  can in turn be calculated using the DFT as

$$c_k = F_k^{-1}[F_j(a)F_j(b)], \quad 0 \leq k < 2n. \quad (6.1.6)$$

The three DFTs indicated in the above formula, in turn, can be efficiently calculated by employing the FFT. The cost of an  $n$ -point FFT is roughly  $5n \log_2 n$  operations, so the cost of three  $2n$ -point FFTs is roughly  $30n \log_2 n$ . But in this case, where the input data vectors  $x$  and  $y$  are real, special variants of the FFT are widely known that reduce the computational cost to approximately half this figure, or roughly  $15n \log_2 n$ . These operation counts assume that  $n$  is a power of two, since the FFT algorithm is most efficient when  $n$  is a power of two.

We should note here that very efficient FFT library routines (even variants for real data input) are usually supplied by computer vendors as part of their scientific software libraries. There is also some excellent FFT software available from community software libraries, notably the FFTW software at <http://www.fftw.org>. Thus it is usually not necessary for researchers to write their own FFT computation code. For those who do need to investigate further into FFT computation techniques, these are discussed in [11, 189].

The DFT, as well as associated FFT computational techniques, are defined for finite-length sequences. As we have seen in Section 5.2, experimental math-

emicians frequently encounter the continuous Fourier transform, which is defined for a real function  $f(t)$  on  $(-\infty, \infty)$  as

$$\hat{f}(x) = \int_{-\infty}^{\infty} f(t)e^{-itx} dx \quad (6.1.7)$$

(see also Section 2.2 of the second volume). Discrete approximations to  $\hat{f}(x)$  can be found using the DFT and FFTs, as follows. Assume that  $f(t)$  is zero (or sufficiently small) outside the interval  $[-a/2, a/2]$ , and assume its Fourier transform  $\hat{f}(x)$  is zero (or sufficiently small) outside  $[-b/2, b/2]$ . Select an even integer  $n$  so that  $ab = 2\pi n$ . The integer  $n$  is usually chosen to be a power of two, since this choice facilitates the usage of FFTs. A large value of  $n$  is often required to produce sufficient accuracy in the approximations below, so the parameters  $a$  and  $b$  may need to be increased to match the size of  $n$ . Given values of  $a$ ,  $b$ , and  $n$ , we can write

$$\begin{aligned} \hat{f}\left(\frac{bk}{n}\right) &= \int_{-\infty}^{\infty} f(t)e^{-itbk/n} dt \\ &\approx \frac{a}{n} \sum_{j=-\frac{n}{2}+1}^{\frac{n}{2}+1} f\left(\frac{aj}{n}\right) e^{-i(aj/n)(bk/n)} \end{aligned} \quad (6.1.8)$$

$$= \frac{a}{n} \sum_{j=-\frac{n}{2}+1}^{\frac{n}{2}+1} f\left(\frac{aj}{n}\right) e^{-2\pi ijk/n}, \quad (6.1.9)$$

for  $-n/2 + 1 \leq k \leq n/2$ . If we adopt the convention that  $f(t) = f(t - a)$  for  $t > a/2$ , and  $\hat{f}(x) = \hat{f}(x - b)$  for  $x > b/2$ , then we can write

$$\hat{f}\left(\frac{bk}{n}\right) = \frac{a}{n} \sum_{j=0}^{n-1} f\left(\frac{aj}{n}\right) e^{-2\pi ijk/n}, \quad (6.1.10)$$

for  $0 \leq k \leq n - 1$ , which is now in a form suitable for the application of FFTs.

An example of using this method to compute several Fourier transforms is given in Section 5.2.

## 6.2 High-Precision Arithmetic

We have already seen numerous examples of high-precision numerical calculations. Indeed, such computations frequently arise in experimental mathematics. We shall focus here on high-precision floating-point computation. High-precision integer computation is also required in some aspects of mathematical computation, particularly in prime number computations and symbolic manipulations, but as we shall see, many of the algorithms described below are equally applicable to both types of arithmetic. An excellent presentation of high-precision integer arithmetic is given in [78].

At this point in time, almost all computer systems support the IEEE-754 standard for floating-point arithmetic. The IEEE “single” (32-bit) format features roughly seven decimal digit accuracy and a dynamic range of  $10^{\pm 38}$ . The IEEE “double” (64-bit) format features roughly 16 decimal digit accuracy and a dynamic range of  $10^{\pm 308}$ . On many Intel and Intel-compatible processors, an additional “extended” (80-bit) format with roughly 19 digits is available. An IEEE “quad” (128-bit) format with roughly 33 decimal digits has been defined, but it is not yet implemented on any of the widely used microprocessors.

When higher precision arithmetic is required in experimental mathematics, many researchers employ the high-precision arithmetic facilities built into commercial software packages such as *Maple* and *Mathematica*. In both of these products, almost all defined functions can be numerically evaluated to arbitrarily high precision. One weakness of these products is that performance is significantly slower compared to what can be achieved using custom programs and libraries. However, intelligent usage of these custom libraries requires correspondingly greater understanding of the underlying nature of high-precision computation. Either way, some familiarity with this type of computation is essential.

### 6.2.1 Double-Double and Quad-Double Arithmetic

The most frequently needed form of high-precision arithmetic is twice that of the IEEE 64-bit format. As mentioned above, this level of precision is not yet widely available in hardware. However, it can be achieved by utilizing a pair of IEEE 64-bit “double” words, where the first word is the closest 64-bit word to the full value, and the second word is the difference (positive or negative) between

the full value and the first word. Fortunately, this “double-double” arithmetic (with 106 mantissa bits or roughly 32 decimal digits) is both fast and easy to implement. This is due mainly to the thoughtful design of the IEEE floating-point standard, wherein the results of all arithmetic operations are guaranteed to be the closest possible value to the exact result for the given operands.

We briefly state here some key algorithms used in double-double arithmetic. Here  $\oplus$ ,  $\ominus$  and  $\otimes$  denote the result of IEEE 64-bit arithmetic operations.

**Algorithm 8** *Double + double.*

This computes the high- and low-order words of the sum of two IEEE 64-bit values  $a$  and  $b$ .

1.  $s := a \oplus b$ ;
2.  $v := s \ominus a$ ;
3.  $e := (a \ominus (s \ominus v)) \oplus (b \ominus v)$ ;
4. Return  $(s, e)$ . □

**Algorithm 9** *Split.*

This splits an IEEE 64-bit value  $a$  into  $a_{\text{hi}}$  and  $a_{\text{lo}}$ , each with 26 bits of significance and one hidden bit, such that  $a = a_{\text{hi}} + a_{\text{lo}}$ .

1.  $t := (2^{27} + 1) \otimes a$ ;
2.  $a_{\text{hi}} := t \ominus (t \ominus a)$ ;
3.  $a_{\text{lo}} := a \ominus a_{\text{hi}}$ ;
4. Return  $(a_{\text{hi}}, a_{\text{lo}})$ . □

**Algorithm 10** *Double  $\times$  double.*

This computes the high- and low-order words of the product of two IEEE 64-bit values  $a$  and  $b$ .

1.  $p := a \otimes b$ ;
2.  $(a_{\text{hi}}, a_{\text{lo}}) := \text{SPLIT}(a)$ ;
3.  $(b_{\text{hi}}, b_{\text{lo}}) := \text{SPLIT}(b)$ ;
4.  $e := ((a_{\text{hi}} \otimes b_{\text{hi}} \ominus p) \oplus a_{\text{hi}} \otimes b_{\text{lo}} \oplus a_{\text{lo}} \otimes b_{\text{hi}}) \oplus a_{\text{lo}} \otimes b_{\text{lo}}$ ;
5. Return  $(p, e)$ . □

With regards to Algorithm 10, we should note that some processors, notably IBM PowerPC and RS6000 processors and Intel IA-64 processors, have a “fused multiply-add” instruction that greatly simplifies double  $\times$  double operations. In this case, one can simply write  $p := a \otimes b$  and  $e := a \otimes b - p$ . Note, however, that it is often necessary to specify a special compiler option (such as `-qstrict` on IBM systems) to insure that this code is performed as written.

From these basic building blocks, one can construct a complete library of double-double arithmetic operations. For example, to add  $(a_1, a_2)$  to  $(b_1, b_2)$ , it suffices to use Algorithm 8 to add  $a_1$  and  $b_1$ , yielding high- and low-order results, then adding  $a_2 + b_2$  to the low-order result. This scheme gives a double-double sum accurate to within two bits in the low-order word. A fully accurate result can be obtained with somewhat more effort. Double-double multiplication and division can be performed using straightforward adaptations of the classic multiplication and division schemes for decimal arithmetic. Full details are given in [117].

With somewhat more effort, “quad-double” arithmetic can be performed using these same methods. This features 212 mantissa bits or roughly 64 decimal digits. Note however that this arithmetic, like double-double arithmetic, does not extend the dynamic range, which remains roughly  $10^{\pm 308}$ .

A software library implementing the above schemes is available on the Internet. This library, which is written in C, includes not only double-double but also quad-double routines. In addition to the basic arithmetic operations, many transcendental functions are also supported. Included in this package are C++ and Fortran-90 translation modules, which permit the library routines to be called from ordinary C++ and Fortran-90 programs with only minor modifications to the source code, by utilizing operator overloading. Programs that use double-double arithmetic typically run five times slower than the equivalent 64-bit versions, and quad-double versions typically run 20 times slower. This software is available at <http://www.expmath.info>.

### 6.2.2 Arbitrary Precision Arithmetic

By “arbitrary precision” we mean a software facility that permits one to adjust the level of numeric precision over a wide range, typically extending to the equivalent of thousands or possibly even millions of decimal digits. An extended dynamic range is almost always included as well, since such computations often

require a larger range than the  $10^{\pm 308}$  range available with the IEEE double format.

For these levels of precision, the best approach is as follows. Define an arbitrary precision datum to be an  $(n + 4)$ -long string of words. The sign of the first word is the sign  $\pm$  of the datum, and the absolute value of the first word is  $n$ , the number of mantissa words used. The second word contains an exponent  $p$ . Words three through  $n + 2$  are the  $n$  mantissa words  $m_i$ , each of which has an integer value between 0 and  $2^b - 1$ , or in other words  $b$  bits of the mantissa. Finally, words  $n + 3$  and  $n + 4$  are reserved as “scratch” words for various arithmetic routines. One can optionally designate an additional word, placed at the start of the data structure, to specify the amount of memory available for this datum, so as to avoid memory overwrite errors during execution. The value  $A$  represented by this datum is

$$A = \pm(2^{pb}m_1 + 2^{(p-1)b}m_2 + 2^{(p-2)b}m_3 + \dots + 2^{(p-n+1)b}m_n),$$

where it is assumed that  $m_1 \neq 0$  and  $m_n \neq 0$  for nonzero  $A$ . Zero is represented by a string consisting of a sign word and an exponent word, both of which are zero.

There are several variations possible with this general design. One approach is to utilize 64-bit IEEE floating-point words, with  $b = 48$  mantissa bits per word. Addition operations can easily be performed by adding the two vectors of mantissas (suitably shifted to adjust for differences in exponent), and then releasing carries beginning from the last mantissa word back to the first. Multiplication and division can be performed by straightforward adaptations of the long multiplication and long division schemes taught in grammar school, performed modulo  $2^{48}$  instead of modulo 10. The multiplication of two individual 48-bit entities can be performed by 48-bit variations of Algorithms 9 and 10. In fact (and this is a key point for high performance), up to  $2^5 = 32$  such products can be accumulated before needing to release carries, since  $5 + 48 = 53$ , and integers as large as  $2^{53}$  can be accommodated exactly in a 64-bit IEEE word.

This approach was taken in the software package described in [26]. Like the double-double and quad-double libraries mentioned above, this software includes C++ and Fortran-90 translation modules, so that these functions can be invoked from ordinary programs with only minor modifications to the source code. The resulting software is available at the URL <http://www.expmath.info>.

Another approach is to utilize arrays of integer data, with integer arithmetic

operations, since all values in the data structure above are whole numbers. The disadvantage of this approach is that integer arithmetic is not as well standardized in microprocessors as floating-point arithmetic. Also, 64-bit integer multiplication is not yet universally supported in hardware. Thus it is hard to write programs that are both fast and easily portable to different systems. Nonetheless, some integer-based implementations have been successful, notably the GNU package, available at the URL <http://www.gnu.org/software/gmp/gmp.html>.

### 6.2.3 Karatsuba Multiplication

One of the simplest schemes to accelerate multiprecision multiplication, and also one of the most widely used, was originally discovered by the Russian mathematician Karatsuba in 1962 [127]. It is based on the observation that if a high-precision number is split in two, then the full product can be obtained with only three half-sized times half-sized multiplications, instead of four as with conventional multiplication.

In particular, let  $x$  be a number with  $2b$ -bit precision. We can write  $x = x_0 + Bx_1$ , where  $B = 2^{-b}$  and  $x_0$  and  $x_1$  are binary fractions of comparable precision. Let  $y$  similarly be represented as  $y_0 + By_1$ . Then we can compute as follows [78]:

$$t = (x_0 + x_1) \cdot (y_0 + y_1) \quad (6.2.11)$$

$$u = (x_0 - x_1) \cdot (y_0 - y_1) \quad (6.2.12)$$

$$v = x_1 y_1 \quad (6.2.13)$$

$$xy = \frac{1}{2}(t + u) - v + \frac{B}{2}(t - u) + vB^2 \quad (6.2.14)$$

Note that multiplications by  $1/2$  and by  $B$  are merely binary shifts of the “decimal” point. Thus the only expensive computations required here are the three multiplications involved in computing  $t$ ,  $u$ , and  $v$ .

What’s more, this scheme can be applied recursively to perform the three indicated multiplications, until the multiplication problems are reduced to a size that ordinary grammar-school multiplication is preferred. It can be shown that the total number of operations in the recursive Karatsuba scheme is on the order of  $(\log n)^{\log 3 / \log 2}$ , compared with  $\log^2 n$  for classical schemes [78].

In practical implementations, the Karatsuba scheme is typically more economical than efficient conventional schemes beginning at roughly 400 decimal



digit precision. It is not as efficient as FFT-based multiplication, which we will address in the next subsection (above roughly 1,000 decimal digit precision), but it is much easier to implement than a full-fledged FFT-based scheme. What's more, it requires significantly less memory than a FFT-based scheme, for a given precision level. It is interesting to note that in several of Kanadas recent record-breaking computations of  $\pi$ , he used a Karatsuba scheme at the *highest* precision levels, and applied FFT-based schemes only when the Karatsuba recursion reached a certain precision level, where there was sufficient memory to use FFTs. For reasons such as these, the Karatsuba scheme is likely to remain a common fixture in high-precision computation packages.

### 6.2.4 FFT-Based Multiplication

The conventional and Karatsuba multiplication schemes are suitable for up to roughly 1000 decimal digit precision. For even higher levels of precision, significantly faster performance can be achieved by employing a convolution-FFT approach. Suppose we wish to multiply two  $n$ -precision values whose mantissa words (assuming a fixed number of bits per word) are given by  $a = (a_0, a_1, a_2, \dots, a_{n-1})$  and  $b = (b_0, b_1, b_2, \dots, b_{n-1})$ . It is easy to see that the desired result, except for releasing carries, is simply the acyclic convolution of the  $a$  and  $b$  vectors, which can in turn be reduced to cyclic convolutions and then rapidly evaluated using FFTs.

This can be done as follows. First, extend the two vectors to length  $2n$  by appending  $n$  zeroes to each. Then the  $2n$ -long product  $c$  of  $a$  and  $b$  is the cyclic convolution of  $a$  and  $b$ :

$$c_k = \sum_{j=0}^{2n-1} a_j b_{k-j}, \quad 0 \leq k < 2n, \quad (6.2.15)$$

where by  $b_{k-j}$  we mean  $b_{k-j+2n}$  when  $k-j$  is negative. This cyclic convolution is evaluated using FFTs as described earlier in the chapter. The  $c_k$  results from the FFT process are floating-point numbers. Rounding these values to the nearest integer, and then releasing carries beginning at  $c_{2n-1}$  gives the desired multiplication result.

One important issue here is that the FFT-based computation of the  $c_k$  must be accurate enough to reliably recover their integer-valued results. For example, in a one million by one million decimal digit multiplication, the  $c_k$  results are

typically as large as  $2^{113}$ , and these must be reliably rounded to the correct integer values even in the presence of some floating-point roundoff error. A common solution is to split each input word containing 48 mantissa bits into 4 words, each containing 12 mantissa bits, and employ conventional 64-bit IEEE arithmetic in the FFT. The computational cost of this strategy is roughly  $60n \log_2 n$  operations, where  $n$  is the number of input words, but this is still much faster than conventional multiplication (roughly  $30n^2$  operations, assuming Algorithms 8, 9, and 10) for higher levels of precision.

This scheme is suitable for precision levels up to approximately 10,000,000 decimal digits. As a precaution for computations above approximately 1,000,000 digits, it is wise to perform the following validity check: After rounding the results of the final inverse FFT to the nearest integer, also determine the largest difference between these results and the nearest integer. If this maximum difference is larger than, say,  $3/8$ , then the user should be alerted that the results can no longer be certified. To use the FFT-based approach for even higher levels of precision, it is necessary to divide the input data further, say to only 6 mantissa bits per 64-bit word, or to employ double-double arithmetic for the FFT. We might add that the test we describe above has also proven to be an effective computer system integrity test—one of the present authors has uncovered several flaws in computer hardware and software using this test.

Another detail worth mentioning here is that efficient FFTs are most commonly targeted to power-of-two data sizes. If your desired level of precision (measured in 64-bit words) is not a power of two, it is often possible to employ FFTs in the computation, yet avoid having to increase the precision level to the next highest power of two, which would double both the memory requirement and the computational cost. This can best be illustrated by example.

Consider the case  $n = p + 2$ , where  $p = 2^r$  (so that  $x_p, x_{p+1}, y_p$  and  $y_{p+1}$  are potentially nonzero). First, extend the sequences  $x$  and  $y$  with zeroes to length  $2p = 2^{r+1}$  (instead of length  $2^{r+2}$  as in the conventional FFT procedure). Then apply forward and inverse FFTs to the extended sequences  $x$  and  $y$  to produce the circular convolution of these extended sequences, which is:

$$\begin{aligned} z_0 &= x_0y_0 + x_{p-1}y_{p+1} + x_p y_p + x_{p+1}y_{p-1} \\ z_1 &= x_0y_1 + x_1y_0 + x_p y_{p+1} + x_{p+1}y_p \\ z_2 &= x_0y_2 + x_1y_1 + x_2y_0 + x_{p+1}y_{p+1} \\ z_3 &= x_0y_3 + x_1y_2 + x_2y_1 + x_3y_0 \end{aligned}$$

$$\begin{aligned}
\dots &= \dots \\
z_p &= x_0y_p + x_1y_{p-1} + \dots + x_py_0 \\
\dots &= \dots \\
z_{2p-1} &= x_{p-2}y_{p+1} + x_{p-1}y_p + x_py_{p-1} + x_{p+1}y_{p-2}
\end{aligned} \tag{6.2.16}$$

This result differs from the desired  $2n$ -long linear convolution of  $x$  and  $y$  in two respects: (1) the initial three values ( $z_0, z_1$  and  $z_2$ ) are “corrupted” with some additional terms, and (2) the final four members of the sequence are missing. These four missing values are

$$\begin{aligned}
z_{2n-4} &= z_{2p} = x_{p-1}y_{p+1} + x_py_p + x_{p+1}y_{p-1} \\
z_{2n-3} &= z_{2p+1} = x_py_{p+1} + x_{p+1}y_p \\
z_{2n-2} &= z_{2p+2} = x_{p+1}y_{p+1} \\
z_{2n-1} &= z_{2p+3} = 0
\end{aligned} \tag{6.2.17}$$

Ignoring the last zero value, these three expressions are exactly the values that have “corrupted” the first three elements of the desired  $z$  sequence. Thus by separately computing these three expressions, one can correct the  $z$  sequence to the desired  $2n$ -point linear convolution result. Note that these three values can be obtained by computing a linear convolution on the sequences  $\bar{x} = \{x_{p-1}, x_p, x_{p+1}\}$  and  $\bar{y} = \{y_{p-1}, y_p, y_{p+1}\}$  and discarding the first two elements of the six-long result. It is clear from this example that this technique can be extended to evaluate the linear convolution of sequences of size  $n = p + m$  for any  $m < p = 2^r$ , although it is not efficient once  $m$  is larger than about  $2^{r-1}$ . The scheme can be applied recursively.

As it turns out, if one plots the computational cost (the number of arithmetic operations versus problem size) required for this scheme, the result is an intriguing fractal structure that is sometimes called the “devil’s staircase” (see Figure 6.1). For this reason, the above technique is sometimes referred to as the “devil’s convolution.” Additional details of this scheme are given in [14, 79].

Additional results and techniques for using various types of FFTs for high-precision arithmetic can be found in Crandall’s books on mathematical computation [79, 78].

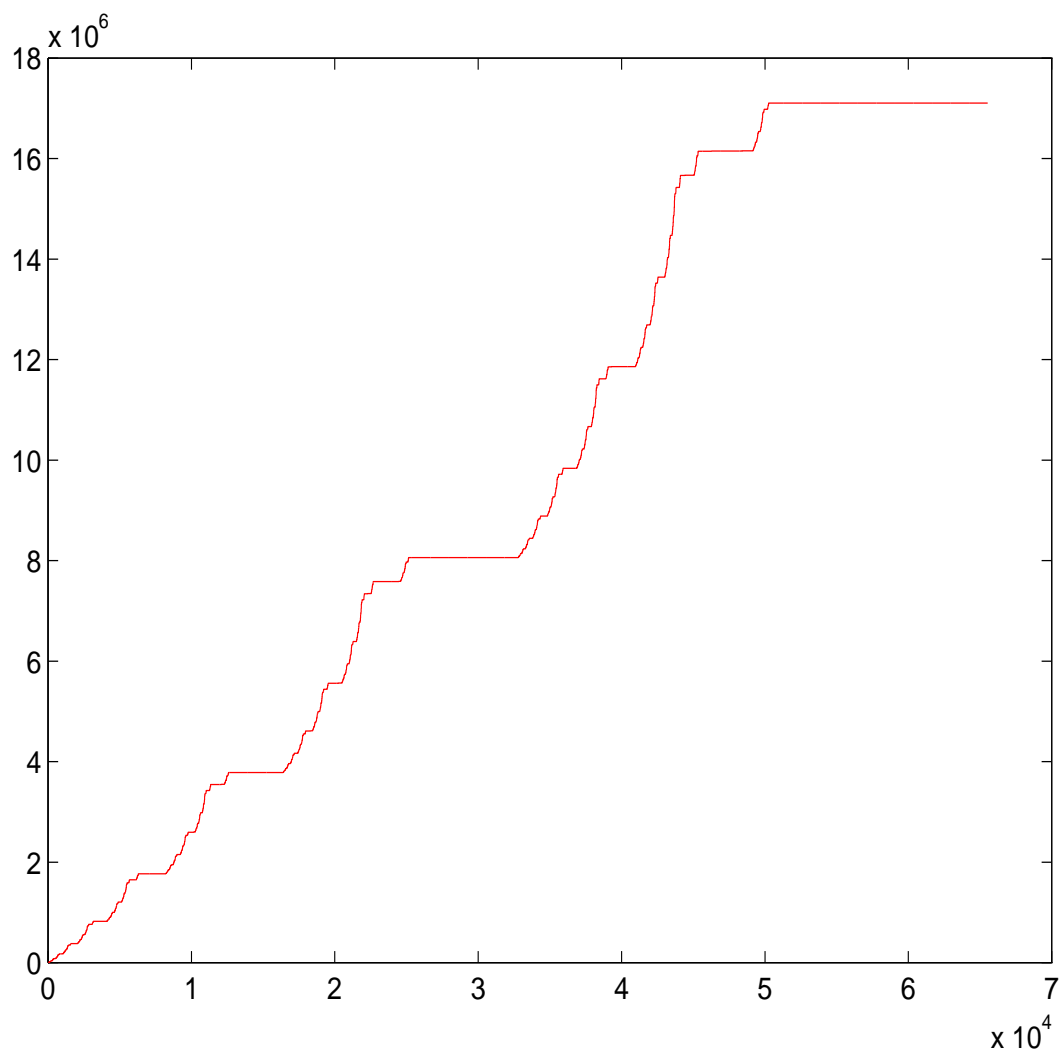


Figure 6.1: Computational cost for the devil's convolution.

### 6.2.5 Algebraic and Transcendental Functions

So far we have focused on multiplication. In this section, we will describe some schemes for high-precision computation of several additional operations, including division and common transcendental functions. We present here only an informal sketch of these methods. See [12, 26] for additional details.

We start with a scheme for extra-high precision division. As we mentioned earlier, for moderate precision levels an adaptation of the conventional grammar school division is adequate. Beyond several hundred digit precision, it is faster to use a Newton iteration scheme, which involves only multiplications, and a few other relatively inexpensive operations. Square roots, cube roots and  $n$ -th roots for integers  $n > 3$  can be found using similar Newton iteration schemes.

**Algorithm 11** *Extra-high precision division.*

To perform  $a/b$ , first set  $x_0$  to the 64-bit approximation of  $1/b$ , and then employ a Newton iteration scheme to obtain  $1/b$  to full precision:

$$x_{k+1} = x_k + x_k(1 - bx_k). \quad (6.2.18)$$

Multiplying the final approximation to  $1/b$  by  $a$  gives the desired quotient. These iterations can be accelerated by utilizing a working precision level that approximately doubles with each iteration. Note that the above scheme involves only addition/subtraction operations and multiplies (which can be done using FFTs). If done carefully, the total computation cost is only slightly more than three times the cost of the final multiplication.

**Algorithm 12** *Square roots.*

Square roots can be computed in a similar manner, using the Newton iteration for  $1/\sqrt{a}$ :

$$x_{k+1} = x_k + \frac{x_k}{2}(1 - ax_k^2). \quad (6.2.19)$$

Multiplying the final approximation to  $1/\sqrt{a}$  by  $a$  gives the square root of  $a$ . Some additional details are given in [26].

**Algorithm 13** *Pi*.

The constant  $\pi$ , which is needed for trigonometric functions, among other things, may be efficiently calculated using the Salamin-Brent algorithm: Set  $a_0 = 1$ ,  $b_0 = 1/\sqrt{2}$ , and  $d_0 = \sqrt{2} - 1/2$ . Then iterate the following arithmetic-geometric mean (AGM) operations beginning with  $k = 1$ :

$$\begin{aligned} a_k &= (a_{k-1} + b_{k-1})/2 \\ b_k &= \sqrt{a_{k-1}b_{k-1}} \\ d_k &= d_{k-1} - 2^k(a_k - b_k)^2. \end{aligned} \quad (6.2.20)$$

Then  $p_k = (a_k + b_k)^2/d_k$  converges quadratically to  $\pi$ —each iteration approximately doubles the number of correct digits. Unfortunately this algorithm is not self-correcting like algorithms based on Newton iterations. Thus all iterations must be done with at least the precision level desired for the final result.

The value of  $\pi$  can also be calculated quite rapidly using some of the infinite series presented in Chapter 3, for example Formulas (3.1.13), (3.1.14), or (3.1.19).

**Algorithm 14** *Exp*.

For moderate levels of precision (up to roughly 1,000 decimal digits),  $e^t$  may be calculated using the following modification of the Taylor's series for  $e^t$ :

$$e^t = \left(1 + r + \frac{r^2}{2!} + \frac{r^3}{3!} + \frac{r^4}{4!} \cdots\right)^{256} 2^n, \quad (6.2.21)$$

where  $r = t'/256$ ,  $t' = t - n \log 2$  and where  $n$  is chosen to minimize the absolute value of  $t'$ . The exponentiation by 256 in this formula is performed by repeated squaring. Reducing  $t$  in this manner significantly accelerates convergence in the above series. On some systems, reducing by 1,024 instead of 256 is even faster.

**Algorithm 15** *Log*.

Given this scheme for  $e^t$ , moderate-precision natural logarithms of an argument  $a$  may then be calculated using the Newton iteration

$$x_{k+1} = x_k + \frac{a - e^{x_k}}{e^{x_k}}, \quad (6.2.22)$$

using a doubling level of precision as before.

**Algorithm 16** *Extra-high precision exp and log.*

For very high precision (beyond several hundred decimal digits), it is better to calculate logarithms by means of a quadratically convergent algorithm originally due to Salamin, as described by Brent in [58]. Inputs  $t$  that are extremely close to 1 are handled using a Taylor series. Otherwise, let  $n$  be the number of bits of precision required in the result. If  $t$  is exactly two, select  $m > n/2$ . Then the following formula gives  $\log 2$  to the required precision:

$$\log 2 = \frac{\pi}{2mA(1, 4/2^m)}. \quad (6.2.23)$$

Here  $A(a, b)$  is the limit of the AGM: Let  $a_0 = a$  and  $b_0 = b$ , and iterate

$$a_{k+1} = (a_k + b_k)/2 \quad b_{k+1} = \sqrt{a_k b_k} \quad (6.2.24)$$

until convergence. For other  $t$ , select  $m$  such that  $s = t2^m > 2^{n/2}$ . Then the following formula gives  $\log t$  to the required precision:

$$\log t = \frac{\pi}{2A(1, 4/s)} - m \log 2. \quad (6.2.25)$$

Given this scheme for logarithms, very high-precision exponentials can be calculated using Newton iterations, using a doubling level of precision as before.

**Algorithm 17** *Sin and cos.*

For modest precision, one can calculate  $\sin t$  via the Taylor's series

$$\sin s = s - \frac{s^3}{3!} + \frac{s^5}{5!} - \frac{s^7}{7!} \cdots, \quad (6.2.26)$$

where  $s = t - a\pi/256$ , and where the integer  $a$  is chosen to minimize the absolute value of  $s$ . Then one calculates

$$\begin{aligned} \sin t &= \sin(s + a\pi/256) \\ \cos t &= \cos(s + a\pi/256), \end{aligned}$$

by applying elementary trigonometric identities for sums. The required values  $\sin(a\pi/256)$  and  $\cos(a\pi/256)$  can be precalculated—they involve compound square roots of 2. Reducing  $t$  in this manner significantly accelerates convergence in the above series for  $\sin s$ . On some systems, using 1,024 subdivisions instead of 256 is even faster. Inverse cosine and sine can then be calculated using Newton iterations.

**Algorithm 18** *Extra-high precision sin and cos.*

For very high precision (above roughly 1,000 decimal digits), sines and cosines, as well as the inverse functions, can be calculated using complex arithmetic versions of the advanced algorithms described above for exp and log (recall that  $e^{ix} = \cos x + i \sin x$ ). See [26] for full details.  $\square$

An arbitrary precision software package implementing the above algorithms, including both basic arithmetic functions (with FFT-based schemes for very high precision) and numerous transcendental functions, is available from the URL <http://www.expmath.info>.

## 6.3 Constant Recognition

Mathematicians have long dreamed of a facility that would recognize a computed numerical constant in an analytic formula in terms of known mathematical functions and constants. With some recent discoveries and algorithmic improvements, that day has arrived. Indeed, we have already seen numerous examples of this methodology in action. We describe here some of the algorithmic and numerical techniques involved.

Let  $x = (x_1, x_2, \dots, x_n)$  be a vector of real numbers.  $x$  is said to possess an integer relation if there exist integers  $a_i$ , not all zero, such that  $a_1x_1 + a_2x_2 + \dots + a_nx_n = 0$ . By an *integer relation algorithm*, we mean a practical computational scheme that can recover the integers  $a_i$ , if they exist, or can produce bounds within which no integer relation exists.

Although the integer relation problem is often regarded to be a relatively “new” problem, it is really a rather old problem. The earliest reference is in the works of Euclid, whose Euclidean algorithm solves this problem in the case  $n = 2$ . A solution for  $n > 2$  was attempted by numerous mathematicians in the 18th, 19th, and 20th centuries. The first general  $n$  integer relation algorithm was discovered in 1977 by Ferguson and Forcade [93].

One approach to integer relation detection is to apply the Lenstra-Lenstra-Lovasz (LLL) lattice reduction algorithm. However, there are difficulties with this approach, notably the selection of a required multiplier—if it is too small, or too large, the LLL solution will not be the desired integer relation. Some of



these difficulties were addressed in the “HJLS” algorithm [111], but this scheme suffers from numerical instability.

### 6.3.1 PSLQ

At the present time, the best integer relation algorithms are variants of the “PSLQ” algorithm [92], which was discovered in 1993 by Helaman Ferguson. PSLQ originally came to prominence with its role in the discovery of the BBP formula for  $\pi$ . In addition to possessing good numerical stability, PSLQ is guaranteed to find a relation in a polynomially bounded number of iterations. The name “PSLQ” derives from its usage of a partial sum of squares vector and a LQ (lower-diagonal-orthogonal) matrix factorization.

We present here a statement of the standard PSLQ algorithm. For additional details, see [92, 15, 23], from which this material has been adapted. Here  $x$  is the  $n$ -long input real vector, and “nint” is the nearest integer function (for half-integer values, select the integer with greater absolute value). One selects the parameter  $\gamma \geq \sqrt{4/3}$  (in our implementations, we select  $\gamma = \sqrt{4/3}$ ).

**Algorithm 19** *Standard PSLQ.*

Initialize:

1. Set the  $n \times n$  matrices  $A$  and  $B$  to the identity.
2. For  $k := 1$  to  $n$  do: set  $s_k := \sqrt{\sum_{j=k}^n x_j^2}$ ; enddo; set  $t = 1/s_1$ ; for  $k := 1$  to  $n$  do: set  $y_k := tx_k$ ;  $s_k := ts_k$ ; enddo.
3. Compute the initial  $n \times (n - 1)$  matrix  $H$  as  $H_{ij} = 0$  if  $i < j$ ,  $H_{jj} := s_{j+1}/s_j$ , and  $H_{ij} := -y_i y_j / (s_j s_{j+1})$  if  $i > j$ .
4. Reduce  $H$ : for  $i := 2$  to  $n$  do: for  $j := i - 1$  to 1 step  $-1$  do: set  $t := \text{nint}(H_{ij}/H_{jj})$ ; and  $y_j := y_j + ty_i$ ; for  $k := 1$  to  $j$  do: set  $H_{ik} := H_{ik} - tH_{jk}$ ; enddo; for  $k := 1$  to  $n$  do: set  $A_{ik} := A_{ik} - tA_{jk}$  and  $B_{kj} := B_{kj} + tB_{ki}$ ; enddo; enddo; enddo.

Iterate the following until an entry of  $y$  is within a reasonable tolerance of zero, or precision has been exhausted:

1. Select  $m$  such that  $\gamma^i |H_{ii}|$  is maximal when  $i = m$ .
2. Exchange the entries of  $y$  indexed  $m$  and  $m + 1$ , the corresponding rows of  $A$  and  $H$ , and the corresponding columns of  $B$ .

3. Remove the corner on  $H$  diagonal: If  $m \leq n-2$  then set  $t_0 := \sqrt{H_{mm}^2 + H_{m,m+1}^2}$ ,  $t_1 := H_{mm}/t_0$  and  $t_2 := H_{m,m+1}/t_0$ ; for  $i := m$  to  $n$  do: set  $t_3 := H_{im}$ ,  $t_4 := H_{i,m+1}$ ,  $H_{im} := t_1 t_3 + t_2 t_4$  and  $H_{i,m+1} := -t_2 t_3 + t_1 t_4$ ; enddo; endif.
4. Reduce  $H$ : for  $i := m+1$  to  $n$  do: for  $j := \min(i-1, m+1)$  to 1 step  $-1$  do: set  $t := \text{nint}(H_{ij}/H_{jj})$  and  $y_j := y_j + t y_i$ ; for  $k := 1$  to  $j$  do: set  $H_{ik} := H_{ik} - t H_{jk}$ ; enddo; for  $k := 1$  to  $n$  do: set  $A_{ik} := A_{ik} - t A_{jk}$  and  $B_{kj} := B_{kj} + t B_{ki}$ ; enddo; enddo; enddo.
5. Norm bound: Compute  $M := 1/\max_j |H_{jj}|$ . Then there can exist no relation vector whose Euclidean norm is less than  $M$ .

Upon completion, the desired relation is found in the column of  $B$  corresponding to the zero entry of  $y$ .  $\square$

High-precision arithmetic must be used for almost all applications of PSLQ. In general, if one wishes to recover a relation of length  $n$ , with coefficients of maximum size  $d$  digits, then the input vector  $x$  must be specified to at least  $nd$  digits, and one must employ floating-point arithmetic accurate to at least  $nd$  digits. PSLQ typically recovers relations when the working precision is only 10 to 15 percent higher than this minimum value.

In the course of the operation of the PSLQ algorithm on a real computer system, the entries of the  $y$  vector gradually decrease in size, with the largest and smallest entries usually differing by no more than two or three orders of magnitude. When a relation is detected by the algorithm, the smallest entry of the  $y$  vector abruptly decreases to roughly the “epsilon” of the working precision (i.e.,  $10^{-p}$ , where  $p$  is the precision level in digits). The detection threshold in the termination test (iteration Step 6) above is typically set to be a few orders of magnitude greater than the epsilon value, in order to allow for reliable relation detection in the presence of some numerical roundoff error. The ratio between the smallest and the largest  $y$  entry when a relation is detected can be taken as a “confidence level” that the relation is a true relation and not an artifact of insufficient numeric precision. Very small ratios at detection, such as  $10^{-100}$ , almost certainly denote a true relation (although, of course, such results are experimental only, and do not constitute rigorous proof).

PSLQ runs reasonably fast for many problems, but in other cases large amounts of computer time are required. For these problems, one can utilize a two-level scheme, wherein most PSLQ iterations are performed using ordinary

64-bit IEEE double arithmetic, with the arbitrary-precision arrays updated as needed. This “two-level” scheme runs as much as 65 times faster on some problems. For very large problems, a three-level scheme yields further acceleration. Details of these methods are given in [23].

### 6.3.2 Multipair PSLQ

Even with the improvements mentioned above, PSLQ runtimes are painfully long for some very large problems of current interest. Thus one is led to consider employing modern parallel computer systems, which have the potential of many times faster performance than single-processor scientific workstations or personal computers.

Unfortunately, PSLQ, unlike most other algorithms used in experimental mathematics, is singularly unsuited for parallel computation. Large PSLQ calculations often require millions of iterations, each of which must be completed before the next begins. Furthermore, within an individual iteration, the key reduction operation (iteration Step 4) has a recursion that inhibits parallel execution, except at the innermost loop level. Thus there is virtually no concurrency to exploit in a parallel implementation.

Recently, a variant of the PSLQ algorithm, known as the “multipair” PSLQ algorithm, was discovered that dramatically reduces the number of sequential iterations, while at the same time exhibiting moderately high concurrency in the major steps of individual iterations. What’s more, it runs somewhat faster than the standard PSLQ even on single-processor systems.

Here  $\gamma \geq \sqrt{4/3}$  and  $\beta < 1/2$  are adjustable parameters (in our implementations, we set  $\gamma = \sqrt{4/3}$  and  $\beta = 2/5$ ).

**Algorithm 20** *Multipair PSLQ.*

Initialize:

1. For  $j := 1$  to  $n$  do: for  $i := 1$  to  $n$  do: if  $i = j$  then set  $A_{ij} := 1$  and  $B_{ij} := 1$  else set  $A_{ij} := 0$  and  $B_{ij} := 0$ ; enddo; enddo.
2. For  $k := 1$  to  $n$  do: set  $s_k := \sqrt{\sum_{j=k}^n x_j^2}$ ; enddo; set  $t = 1/s_1$ ; for  $k := 1$  to  $n$  do: set  $y_k := tx_k$ ;  $s_k := ts_k$ ; enddo.
3. Initial  $H$ : For  $j := 1$  to  $n - 1$  do: for  $i := 1$  to  $j - 1$  do: set  $H_{ij} := 0$ ; enddo; set  $H_{jj} := s_{j+1}/s_j$ ; for  $i := j + 1$  to  $n$  do: set  $H_{ij} := -y_i y_j / (s_j s_{j+1})$ ; enddo; enddo.

Iterate the following until an entry of  $y$  is within a reasonable tolerance of zero, or precision has been exhausted:

1. Sort the entries of the  $(n - 1)$ -long vector  $\{\gamma^i |H_{ii}|\}$  in decreasing order, producing the sort indices.
2. Beginning at the sort index  $m_1$  corresponding to the largest  $\gamma^i |H_{ii}|$ , select pairs of indices  $(m_i, m_i + 1)$ , where  $m_i$  is the sort index. If at any step either  $m_i$  or  $m_i + 1$  has already been selected, pass to the next index in the list. Continue until either  $\beta n$  pairs have been selected, or the list is exhausted. Let  $p$  denote the number of pairs actually selected in this manner.
3. For  $i := 1$  to  $p$  do: exchange the entries of  $y$  indexed  $m_i$  and  $m_i + 1$ , and the corresponding rows of  $A$ ,  $B$  and  $H$ ; enddo.
4. Remove corners on  $H$  diagonal: For  $i := 1$  to  $p$  do: if  $m_i \leq n - 2$  then set  $t_0 := \sqrt{H_{m_i, m_i}^2 + H_{m_i, m_i + 1}^2}$ ,  $t_1 := H_{m_i, m_i} / t_0$  and  $t_2 := H_{m_i, m_i + 1} / t_0$ ; for  $i := m_i$  to  $n$  do: set  $t_3 := H_{i, m_i}$ ;  $t_4 := H_{i, m_i + 1}$ ;  $H_{i, m_i} := t_1 t_3 + t_2 t_4$ ; and  $H_{i, m_i + 1} := -t_2 t_3 + t_1 t_4$ ; enddo; endif; enddo.
5. Reduce  $H$ : For  $i := 2$  to  $n$  do: for  $j := 1$  to  $n - i + 1$  do: set  $l := i + j - 1$ ; for  $k := j + 1$  to  $l - 1$  do: set  $H_{lj} := H_{lj} - T_{lk} H_{kj}$ ; enddo; set  $T_{lj} := \text{nint}(H_{lj} / H_{jj})$  and  $H_{lj} := H_{lj} - T_{lj} H_{jj}$ ; enddo; enddo.
6. Update  $y$ : For  $j := 1$  to  $n - 1$  do: for  $i := j + 1$  to  $n$ : set  $y_j := y_j + T_{ij} y_i$ ; enddo; enddo.
9. Update  $A$  and  $B$ : For  $k := 1$  to  $n$  do: for  $j := 1$  to  $n - 1$ : for  $i := j + 1$  to  $n$  do: set  $A_{ik} := A_{ik} - T_{ij} A_{jk}$  and  $B_{jk} := B_{jk} + T_{ij} B_{ik}$ ; enddo; enddo; enddo.
8. Norm bound: Compute  $M := 1 / \max_j |H_{jj}|$ . Then there can exist no relation vector whose Euclidean norm is less than  $M$ .
9. Termination test: If the largest entry of  $A$  exceeds the level of numeric precision used, then precision is exhausted. If the smallest entry of the  $y$  vector is less than the detection threshold, a relation has been detected and is given in the corresponding row of  $B$ .  $\square$

There are several differences between this algorithm and the standard one-level PSLQ algorithm: (1) there is no reduction step in the initialization; (2) the  $B$  matrix is transposed from the standard PSLQ algorithm; (3) up to  $\beta n$  disjoint pairs (not just a single pair) of adjacent indices are selected in each iteration; (4) the  $H$  reduction loop proceeds along successive lower diagonals of the  $H$  matrix; (5) a  $T$  matrix is employed, which contains the  $t$  multipliers of the standard PSLQ; and (6) the  $y$ ,  $A$ , and  $B$  arrays are not updated with  $H$ , but in separate

loops.

Unlike standard PSLQ, there is no proof that the multipair PSLQ scheme is guaranteed to find a relation within a polynomially bounded number of iterations. On the other hand, we have found, based on experience with a variety of sample problems, that the norm bound increases much more rapidly than in the standard PSLQ. Indeed, it appears that the selection of up to  $\beta n$  disjoint pairs of indices in Step 2 above has the effect of reducing the iteration count by nearly the factor  $\beta n$ . This results in a significant saving in the number of expensive reduction and update steps.

The bottom line is that even on single processor systems, the multipair PSLQ algorithm appears to be significantly faster than the standard PSLQ [23]. Both the standard PSLQ and the multipair PSLQ schemes, including two-level and even three-level variants, have been programmed in C++ and Fortran-90, targeted to the arbitrary precision package mentioned earlier in the chapter. This software is available from the URL <http://www.expmath.info>.

### 6.3.3 Practical Constant Recognition

Practical constant recognition facilities require more than efficient integer relation detection schemes. Usually a sequence of methods are used. An approach that is used in some software is the following (here  $\alpha$  is the input numerical constant) [27].

1. Table look-up. An indexed table of well-known numerical constants is checked to see if it contains  $\alpha$ , to within some tolerance.
2. Algebraic check. The vector  $(1, \alpha, \alpha^2, \dots, \alpha^n)$  is computed and then input to an integer relation finding routine, to see if  $\alpha$  is an algebraic number of degree  $n$  or less. The integer  $n$  is chosen so that the runtime is reasonable for the numeric precision being used.
3. Strided algebraic check. The vector  $(1, \alpha^k, \alpha^{2k}, \dots, \alpha^{nk})$  is computed and then input to an integer relation finding routine, for integers  $k$  up to some maximum level  $m$ . As before, the parameters  $m$  and  $n$  are selected to make runtimes reasonable.
4. Multiplicative check.  $\log |\alpha|$  is computed, and then an integer relation check is done with the natural logarithms of the first few primes, as well as

several other well-known constants, such as  $\pi$ ,  $e$ ,  $\log 2$ ,  $\log 3$ ,  $G$  and  $\gamma$  (here  $G$  is Catalan's constant and  $\gamma$  is Euler's gamma constant.) If a relation is found in this set of numbers, then this means  $\alpha$  is given by a one-term formula of the form

$$\alpha = \pm 2^a 3^b 5^c 7^d 11^e 13^f 17^g 19^h \pi^k e^l (\log 2)^m (\log 3)^n G^o \gamma^p$$

for rationals  $a, b, c, \dots, p$ . The number of primes and the set of other constants used here are quite arbitrary. They are typically chosen for effective constant discovery in a particular environment, under the constraint of reasonable run times.

5. Multiterm constant check. In this test, which is typically done with just double-double or with double precision, numerous integer relation tests are made for sample formulas that involve, for example, up to four terms, each of which has some combination of well-known mathematical constants. Thousands of such searches can be done in a few seconds. Such searches often turn up formulas that are not found using the above searches.

An on-line tool that employs this search strategy is available at the URL <http://www.cecm.sfu.ca/projects/ISC>. A scheme employing a slightly different approach (a Java-based client) is available at <http://psg.cecm.sfu.ca/projects/revenge/client/RevEngClient.html>. Software that performs this type of search strategy is also available as part of the Experimental Mathematician's Toolkit: <http://www.expmath.info>. Another interesting resource in this regard is Finch's recently published *Mathematical Constants* [94].

Maple 9, released in summer 2003, incorporates a quite functional constant recognition tool. Typing `identify(3.146264370)` returns  $3^{1/2} + 2^{1/2}$ . One can expect such features with refinements in future releases of *Maple* and *Mathematica*.

## 6.4 Commentary and Additional Examples

1. **Continued fractions.** We should emphasize how continued fractions expose different structure: The golden mean is the simplest irrational,  $e$  shows great structure,  $[1, 2, 3, 4, \dots]$  is a ratio of Bessel functions, and  $\pi$  remains a complete mystery. Thus, a good rule in constant recognition is

to always convert a decimal expansion to a simple continued fraction or to another base (say, binary), to see if this alternate representation is helpful.

2. **A new formula for Khintchine's constant.** Let  $\{a_i\}$  be the continued fraction expansion of a random real number. The *Gauss-Kuz'min distribution* states that

$$\text{Prob}(a_n = k) = -\log_2 \left[ 1 - \frac{1}{(k+1)^2} \right].$$

Khintchine showed that the limiting geometric mean,  $\lim_k (a_1 a_2 \cdots a_k)^{1/k}$ , exists a.e. and equals

$$\begin{aligned} K_0 &= \prod_{k=1}^{\infty} \left[ 1 + \frac{1}{k(k+2)} \right]^{\log_2 k} = \prod_{k=1}^{\infty} k^{\left[ \log_2 \left( 1 + \frac{1}{k(k+2)} \right) \right]} \\ &= 2.68545200106530644530971483548179569382038229399446 \dots \end{aligned}$$

Little is known about  $K_0$ . Is it rational, algebraic, or related to  $\pi$  or  $e$ ? The formula is very slowly convergent.

For any positive integer  $N$ , we may, however, derive a quite effective identity

$$\log(K_0) \log(2) = \sum_{s=1}^{\infty} \zeta(2s, N) \frac{A_s}{s} - \sum_{k=2}^N \log \left( 1 - \frac{1}{k} \right) \log \left( 1 + \frac{1}{k} \right),$$

where the *Hurwitz zeta function* is defined as

$$\zeta(s, N) = \sum_{n=1}^{\infty} \frac{1}{(n+N)^s} = \zeta(s) - \sum_{n=1}^N n^{-s},$$

while  $A_s = \sum_{k=1}^{2s-1} (-1)^{k-1} / k$  and  $N$  is a free parameter that can be varied for optimization and as an error check. The method has been used to compute  $K_0$  to 7,350 digits—from which it seems to obey the Gauss-Kuz'min distribution [18].

3. **Putnam problem 1998–B5.** Let  $N$  be the “rep-one” integer consisting of 1,998 ones. Find the thousandth digit after the decimal point in  $\sqrt{N}$ .

Answer: The digit is a “1.” Hint: This can easily be verified by employing high-precision arithmetic. Once verified, this can be established by writing

$$10^{1000}\sqrt{N} = \frac{\sqrt{10^{3998} - 10^{2000}}}{3},$$

and so

$$\frac{10^{1999} - 4}{3} < 10^{1000}\sqrt{N} < \frac{10^{1999} + 4}{3}.$$

#### 4. Dynamical systems, numerical analysis, and formal power series.

We present here a brief overview of a surprising connection between numerical analysis of dynamical systems and formal power series [53]. We begin with a simple question: What, exactly, does the fixed time step *forward Euler* numerical method do to the solution of the simple initial value problem

$$\frac{dy}{dt} = y^2, \tag{6.4.27}$$

with  $y(0) = y_0$ ? The numerical procedure is just

$$y_{n+1} = y_n + hy'_n \tag{6.4.28}$$

for integer  $n \geq 0$ , where  $y'_n = y_n^2$  and  $h > 0$  is the chosen time step.

It turns out to be useful to rescale  $y$  and  $t$  so that  $v = hy$  and  $\tau = ht$ , giving

$$\frac{dv}{d\tau} = v^2, \tag{6.4.29}$$

and (6.4.28) becomes

$$v_{n+1} = v_n + v_n^2. \tag{6.4.30}$$

We may then rephrase our question to ask instead what the relationship between  $v_n$  and  $v(\tau)$  is.

The point of view taken in [74] is that of *backward error analysis*. That is, instead of asking for the difference between  $v(n)$  and  $v_n$ , we ask instead if there is another differential equation, say

$$\frac{dw}{d\tau} = B(w)w^2, \tag{6.4.31}$$



whose solution *interpolates*  $v_n$ . That is, we impose the conditions  $w(0) = v_0$  and  $w(\tau + 1) = w(\tau) + w(\tau)^2$  (see Formula (6.4.30), and see if we can find such a function  $B(w)$ ).

It turns out that we can use *the method of modified equations* [105] to find as many terms of the Taylor series for  $B(w)$  as we desire. When we compute the modified equation for (6.4.27) to (say) fifth order, we get

$$\frac{dw}{dt} = \left(1 - w + \frac{3}{2!}w^2 - \frac{16}{3!}w^3 + \frac{124}{4!}w^4 - \frac{1256}{5!}w^5\right)w^2. \quad (6.4.32)$$

Now we see the sequence 1, -1, 3, -16, 124, -1256 appearing. This is sequence M3024 in Sloane's dictionary [181], which points us directly to the very beautiful and useful paper [136]. It tells us that if

$$B(w) = \sum_{n \geq 0} c_n w^n, \quad (6.4.33)$$

then

$$c_n = \frac{1}{n-1} \sum_{i=1}^{n-1} \binom{n-i+1}{i+1} c_{n-i},$$

and this, combined with the functional equation

$$B(w) = \frac{(1+w)^2}{1+2w} B(w+w^2)$$

(which can be iterated to give us two converging infinite products for  $B$ ), allows us to write an efficient program to evaluate  $B(w)$ . We can show that  $B(w)$  has a pole at  $w = -1/2$ ; by mapping backwards, solving  $w + w^2 = -1/2$ , we find two more (complex) poles. Iterating this process finds an infinite number of complex poles, approaching the Julia set for the map  $v \rightarrow v + v^2$  arbitrarily closely; see Figure 6.2.

The Julia set itself approaches the origin arbitrarily closely. That is, there are poles arbitrarily close to the point of expansion of the series given for  $B$ . Thus the series (6.4.33) diverges—but, nonetheless, it can be used to evaluate  $B(w)$  for  $w$  close enough to zero, using the built-in sequence acceleration techniques of *Maple*. This is precisely where the convergent infinite products are slow, and hence, the series is useful. See [74] for details.

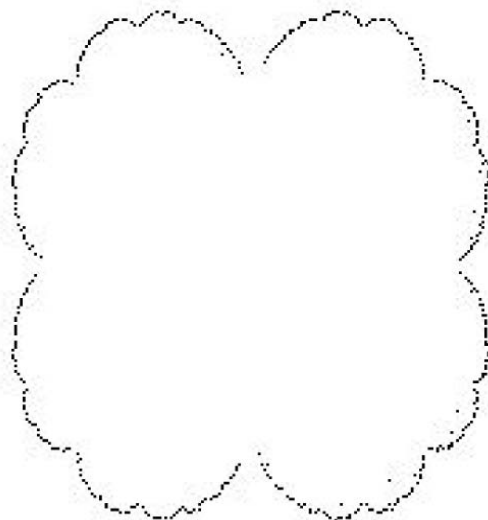


Figure 6.2: The first 16000 poles of  $B(v)$ , approaching the Julia set of  $v \rightarrow v + v^2$ .

5. **Accelerations for Catalan's constant.** (This material is condensed from [57].) Acceleration formulas for slowly convergent series such as Catalan's constant may be based on transformations of the log tangent integral. The simplest acceleration formula of its type is

$$G = \frac{\pi}{8} \log(2 + \sqrt{3}) + \frac{3}{8} \sum_{k=0}^{\infty} \frac{1}{(2k+1)^2 \binom{2k}{k}}, \quad (6.4.34)$$

due to Ramanujan. This is the first of an infinite family of formulae for  $G$ , each of which includes an infinite series whose general term consists of a linear recurrence damped by the summand in (6.4.34). Each series evaluates to a rational linear combination of  $G$  and  $\pi$  times the logarithm of an algebraic unit (i.e., an invertible algebraic integer). The most striking example is

$$G = \frac{\pi}{8} \log \left( \frac{10 + \sqrt{50 - 22\sqrt{5}}}{10 - \sqrt{50 - 22\sqrt{5}}} \right) + \frac{5}{8} \sum_{k=0}^{\infty} \frac{L_{2k+1}}{(2k+1)^2 \binom{2k}{k}}, \quad (6.4.35)$$

where  $L_1 = 1$ ,  $L_2 = 3$ , and  $L_n$  are the Lucas numbers (M0155 in Sloane's Encyclopedia).

The series acceleration results (6.4.34) and (6.4.35) appear as natural consequences of transformation formulae for the log tangent integral, although Ramanujan apparently derived his result (6.4.34) by quite different methods. The connection with log tangent integrals is best explained by the equation

$$G = - \int_0^{\pi/4} \log(\tan \theta) d\theta, \quad (6.4.36)$$

obtained by expanding the integrand into its Fourier cosine series and integrating term by term. Ramanujan's result (6.4.34) arises from the transformation

$$2 \int_0^{\pi/4} \log(\tan \theta) d\theta = 3 \int_0^{\pi/12} \log(\tan \theta) d\theta. \quad (6.4.37)$$

Formula (6.4.35) arises similarly from the transformation

$$\begin{aligned} 2 \int_0^{\pi/4} \log(\tan \theta) d\theta &= 5 \int_0^{3\pi/20} \log(\tan \theta) d\theta \\ &\quad - 5 \int_0^{\pi/20} \log(\tan \theta) d\theta. \end{aligned} \quad (6.4.38)$$

A great virtue of such log tangent identities is that they can be efficiently hunted for by integer relation methods.

## 6. More Catalan formulas. Evaluate

(a)

$$\sum_{n \geq 1} \frac{n}{16^n} \zeta(2n + 1),$$

due to Glaisher (1913), and

(b)

$$\sum_{n \geq 1} (n + 1) \frac{3^n - 1}{4^n} \zeta(n + 2),$$

due to Flajolet and Vardi (1996) in terms of Catalan's constant.

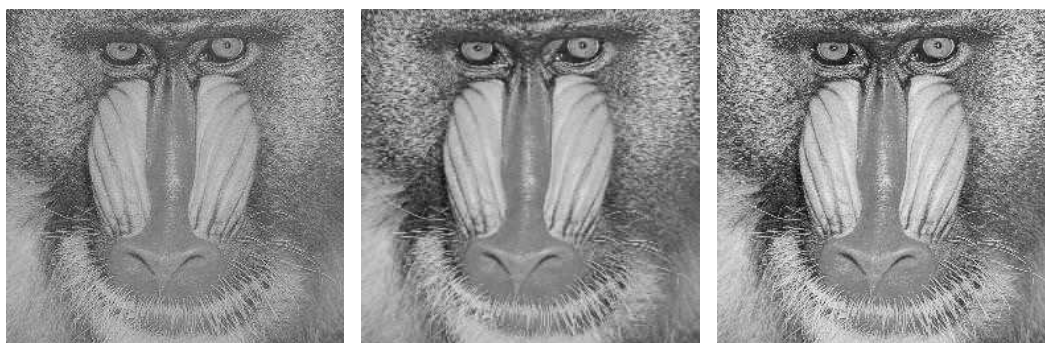


Figure 6.3: Three images quantized at quality 50 (L), 48 (C) and 75 (R).

7. **Continued fraction for Euler's constant.** Compute the simple continued fraction for Euler's constant and see if it appears to satisfy the Gauss-Kuz'min distribution.
8. **The perceptual dipper effect.** Over the past decade, the JPEG image compression standard (see <http://www.jpeg.org>), originally named for the Joint Photographic Experts Group, has become the benchmark for compressing natural still images for use on the internet. This standard has been extremely successful in maintaining a balance between low file size and high subjective image quality.

However, JPEG-encoded images can exhibit unexpected objectionable artifacts. For example, for images already encoded using compression methods based around the *discrete cosine transform* (DCT), further compression by requantization can create objectionable visual artifacts. This phenomenon, called the *JPEG perceptual dipper effect*, is an example of the more general concept of a *visual threshold*, wherein two approximations of the same original image, which differ little from each other in terms of some given numerical measure, differ dramatically in the “quality” of their visual fidelity to the original image.

The dipper effect is illustrated in Figure 6.3. The mandrill image in the center is more compressed than the leftmost, yet appears better than the one on the left, and compares quite favorably to the much less compressed image on the right. The effect can be easily avoided by simple calculations in the process of designing the quantization matrices used to re-quantize

the given image. This is of practical consequence for delivery of images where bandwidth or storage is an issue (such as in delivery of data to cell phones, PDAs and other wireless devices), since it is possible to assess and exploit the dipper effect in software ([31] and [148]). Needless to say, such effects must be seen to be discovered.

9. **The Universal Computer: The road from Leibniz to Turing.** This is an engaging account by Martin Davis of the history of logic and the birth of the modern computer [82]. In his story—chock full of colorful detail, such as Frege’s extreme antisemitism and Gödel’s conversion to full-blooded Platonism—one reads
- (a) “For it is unworthy of excellent men to lose hours like slaves in the labor of calculation which could be safely relegated to anyone else if the machine were used.” (Leibniz)
  - (b) That Poincaré is reputed to have said Cantor’s set theory “would be regarded as a disease from which one had recovered.” and that he did write of Russell, “It is difficult to see that the word *if* acquires when written  $\supset$ , a virtue it did not possess when written *if*.”
  - (c) “To *exist* in mathematics means to be constructed by intuition; and the question whether a certain language is consistent, is not only unimportant in itself, it is also not a test for mathematical existence.” (Brouwer)
  - (d) That Hardy, of Hilbert’s *Entscheidungsproblem* (decision problem), said correctly, “There is of course no such Theorem and this is very fortunate, . . .”
  - (e) When Turing, after showing the unsolvability of the halting problem, went to Princeton to work with Church and others in 1936, von Neumann wrote him a letter of reference that made no mention of his work on computability (von Neumann reportedly said that after what Gödel did he would never read another paper in logic). But by the beginning of World War Two, von Neumann was openly using Turing’s ideas and goals of building practical computing machines. By the end of Turing’s wartime stay at Bletchley Park, he had played a key role in breaking the German codes and was in possession of “a working knowledge of vacuum electronics.”

The last chapters of [82] tell of the development of Turing's computer science influenced Automatic Computing Machine (ACE) and von Neumann and others' more engineering influenced ENIAC, about which Turing may have been thinking when he wrote of "the American tradition of solving one's difficulties by much equipment rather than by thought." By the end of the last century, Turing was recognized along with von Neumann as the intellectual progenitors of modern computing. Davis summarized his account with these words:

"This story underscores the power of ideas and the futility of predicting where they will lead." (Martin Davis)



# Chapter 7

## Making Sense of Experimental Mathematics

### 7.1 Introduction

Philosophers have frequently distinguished mathematics from the physical sciences. While the sciences were constrained to fit themselves via experimentation to the “real” world, mathematicians were allowed more or less free reign within the abstract world of the mind. This picture has served mathematicians well for the past few millennia but the computer has begun to change this. The computer has given us the ability to look at new and unimaginably vast worlds. It has created mathematical worlds that would have remained inaccessible to the unaided human mind, but this access has come at a price. Many of these worlds, at present, can only be known experimentally. The computer has allowed us to fly through the rarefied domains of hyperbolic spaces and examine more than a billion digits of  $\pi$  but experiencing a world and understanding it are two very different phenomena. Like it or not, the world of the mathematician is becoming experimentalized.<sup>1</sup>

The computers of tomorrow promise even stranger worlds to explore. Today, however, most of these explorations into the mathematical wilderness remain isolated illustrations. Heuristic conventions, pictures and diagrams developing

---

<sup>1</sup>This entire chapter is a reprint (with permission) of “Making Sense of Experimental Mathematics,” by J. M. Borwein, P. B. Borwein, R. Girgensohn, and S. Parnes, *Mathematical Intelligencer*, vol. 18 (1996), page 12–18.



in one sub-field often have little content for another. In each sub-field unproven results proliferate but remain conjectures, strongly held beliefs or perhaps mere curiosities passed like folk tales across the Internet. The computer has provided extremely powerful computational and conceptual resources but it is only recently that mathematicians have begun to systematically exploit these abilities. It is our hope that by focusing on experimental mathematics today, we can develop a unifying methodology tomorrow.

### 7.1.1 Our Goals

The genesis of this article was a simple question: “How can one use the computer in dealing with computationally approachable but otherwise intractable problems in mathematics?” We began our current exploration of *experimental mathematics* by examining a number of very long-standing conjectures and strongly held beliefs regarding decimal and continued fraction expansions of certain elementary constants. These questions are uniformly considered to be hopelessly intractable given present mathematical technology. Unified field theory or cancer’s “magic bullet” seem accessible by comparison. But like many of the most tantalizing problems in mathematics their statements are beguilingly simple. Since our experimental approach was unlikely to result in any new discoveries<sup>2</sup>, we focused on two aspects of experimentation: systematization and communication.

For our attempted systematization of *experimental mathematics* we were concerned with producing data that were “completely” reliable and insights that could be quantified and effectively communicated. We initially took as our model experimental physics. We were particularly interested in how physicists verified their results and the efforts they took to guarantee the reliability of their data. The question of reliability is undoubtedly central to mathematicians and here we believe we can draw a useful distinction between experimental physics and mathematics. While it is clearly impossible to extract perfect experimental data from nature such is not the case with mathematics. Indeed, reliability of raw mathematical data is far from the most vexing issue.

Let us turn to our second and primary concern: insight. All experimental sciences turn on the intuitions and insights uncovered through modeling and the

---

<sup>2</sup>We will not discuss the computational difficulties here but there are many non-trivial mathematical and computer-related issues involved in this project.

use of probabilistic, statistical and visual analysis. There is really no other way to proceed, but this process even when applied to mathematics inevitably leads to some considerable loss of exactness.

The communication of insight, whether derived from mathematical experiment or not, is a complex issue. Unlike most experimentalized fields, Mathematics does not have a “vocabulary” tailored to the transmission of condensed data and insight. As in most physics experiments the amount of raw data obtained from mathematical experiments is, in general, too large for anyone to grasp. The collected data needs to be compressed and compartmentalized. To make up for this lack of unifying vocabulary we have borrowed heavily from statistics and data analysis to interpret our results. For now we have used restraint in the presentation of our results in what we hope is an intuitive, friendly and convincing manner. Eventually what will probably be required is a multi-leveled hyper-textual presentation of mathematics, allowing mathematicians from diverse fields to quickly examine and interpret the results of others—without demanding the present level of specialist knowledge. [Not only do mathematicians have trouble communicating with lay audiences, but they have significant difficulty talking to each other. There are hundreds of distinct mathematical languages. The myth of a universal language of mathematics is just that. Many subdisciplines simply can not comprehend each other.]

### 7.1.2 Unifying Themes

We feel that many of these problems can be addressed through the development of a rigorous notion of experimental mathematics. In keeping with the positivist tradition, mathematics is viewed as the most exact of sciences and mathematicians have long taken pride in this. But as mathematics has expanded, many mathematicians have begun to feel constrained by the bonds placed upon us by our collective notion of proof. Mathematics has grown explosively during our century with many of the seminal developments in highly abstract seemingly non-computational areas. This was partly from taste and the power of abstraction but, we would argue, equally much from the lack of an alternative. Many intrinsically more concrete areas were, by 1900, explored to the limits of pre-computer mathematics. Highly computational, even “brute-force” methods were of necessity limited but the computer has changed all that. A re-concretization is now underway. The computer-assisted proofs of the four color theorem are a prime

example of computer-dependent methodology and have been highly controversial despite the fact that such proofs are much more likely to be error free than, say, even the revised proof of Fermat's Last Theorem.

Still, these computerized proofs need offer no insight. The Wilf/Zeilberger algorithms for "hypergeometric" summation and integration, if properly implemented, can rigorously prove very large classes of identities. In effect, the algorithms encapsulate parts of mathematics. The question raised is: "How can one make full use of these very powerful ideas?" Doron Zeilberger has expressed his ideas on experimental mathematics in a paper dealing with what he called "semi-rigorous" mathematics. While his ideas as presented are somewhat controversial, many of his ideas have a great deal of merit.

The last problem is perhaps the most surprising. As mathematics has continued to grow there has been a recognition that the age of the mathematical generalist is long over. What has not been so readily acknowledged is just how specialized mathematics has become. As we have already observed, sub-fields of mathematics have become more and more isolated from each other. At some level, this isolation is inherent but it is imperative that communications between fields should be left as wide open as possible.

As fields mature, speciation occurs. The communication of sophisticated proofs will never transcend all boundaries since many boundaries mark true conceptual difficulties. But experimental mathematics, centering on the use of computers in mathematics, would seem to provide a common ground for the transmission of many insights. And this requires a common meta-language<sup>3</sup>. While such a language may develop largely independent of any conscious direction on the part of the mathematical community, some focused effort on the problems of today will result in fewer growing pains tomorrow.

## 7.2 Experimental Mathematics

### 7.2.1 Journal of

A professor of psychology was exploring the creative process and as one of his subjects chose a mathematician who was world famous for his ability to solve

---

<sup>3</sup>This may not be a fanciful dream as the *Computer Algebra Systems* (CAS) of today are beginning to provide just that.

problems. They gave him a problem to work on. He wrote something down and immediately scribbled it out. He wrote something else down and scribbled it out. The professor asked him to leave everything on the page. He explained that he was interested in the process, the wrong answers and the right answers. The mathematician sat down. Wrote something. The psychology professor waited in anticipation but the mathematician announced he could not proceed without erasing his mistakes. While the mathematician in this situation is undoubtedly fairly idiosyncratic in how he attacks problems there is a strongly felt separation between the creative process of mathematics and the published or finished product.

A current focal point for experimental mathematics is the journal called *Experimental Mathematics*. But does it really seek to change the way we do mathematics, or to change the way we write mathematics? We begin by attempting to extract a definition of “experimental” from the Journal’s introductory article ([87]) “About this Journal” by David Epstein, Silvio Levy and Rafael de la Llave.

The word “experimental” is conceived broadly: many mathematical experiments these days are carried out on computers, but others are still the result of pencil-and-paper work, and there are other experimental techniques, like building physical models. ([87] p. 1)

It seems that almost anything can be conceived of as being experimental. Let us try again.

Experiment has always been, and increasingly is, an important method of mathematical discovery. (Gauss declared that his way of arriving at mathematical truths was “through systematic experimentation”.) Yet this tends to be concealed by the tradition of presenting only elegant, well-rounded and rigorous results. ([87] p. 1)

Now we begin to get closer to the truth. Experimentation is still ill defined but is clearly an important part of the mathematical process. It is clearly not new but by implication must be inelegant, lopsided and lax. We, of course, dispute all three of these points and while we do not reply directly to these charges, we hope the reader will be convinced that there need be no compromises made with respect to the quality of the work.

But what is the journal interested in publishing? Their goal seems to be two-fold.

While we value the theorem-proof method of exposition, and while we do not depart from the established view that a result can only become part of mathematical knowledge once it is supported by a logical proof, we consider it anomalous that an important component of the process of mathematical creation is hidden from public discussion. It is to our loss that most of the mathematical community are almost always unaware of how new results have been discovered. ([87] p. 1)

and

The early sharing of insights increases the possibility that they will lead to theorems: an interesting conjecture is often formulated by a researcher who lacks the techniques to formalize a proof, while those who have the techniques at their fingertips have been looking elsewhere.

It appears that through the journal *Experimental Mathematics* the editors advocate a not undramatic change in writing style. So what does a paper published in that journal look like? A recent example is “Experimental Evaluation of Euler sums” by D. H. Bailey, J. Borwein and R. Girgensohn ([20]). The authors describe how their interest in Euler sums was roused by a surprising discovery:

In April 1993, Enrico Au-Yeung, an undergraduate at the University of Waterloo, brought to the attention of one of us the curious fact that

$$\begin{aligned} \sum_{k=1}^{\infty} \left(1 + \frac{1}{2} + \cdots + \frac{1}{k}\right)^2 k^{-2} &= 4.59987\dots \\ &\approx \frac{17}{4}\zeta(4) = \frac{17\pi^4}{360} \end{aligned}$$

based on a computation to 500,000 terms. This author’s reaction was to compute the value of this constant to a higher level of precision in order to dispel this conjecture. Surprisingly, a computation to 30 and later to 100 decimal digits still affirmed it. ([20] p. 17)

### Serendipity and Experimentation

After Enrico Au-Yeung's serendipitous discovery, D. Bailey, J. Borwein and R. Girgensohn launched a full fledged assault on the problem. This is documented in "Experimental Detection of Euler Sums" (the material below was taken from David Bailey's slides).

#### Experimental Approach

1. Employ an advanced scheme to compute high-precision (100+ digit) numerical values for various constants in a class.
2. Conjecture the form of terms involved in possible closed-form evaluations.
3. Employ an integer relation finding algorithm to determine if an Euler sum value is given by a rational linear combination of the conjectured terms.
4. Attempt to find rigorous proofs of experimental results.
5. Attempt to generalize proofs for specific cases to general classes of Euler sums.

Table 7.1: Serendipity and experimentation

This type of serendipitous discovery must go on all the time, but it needs the flash of insight that will place it in a broader context. It is like a gold nugget waiting to be refined—without a context it would remain a curiosity. The authors now proceeded to provide a context by mounting a full-fledged assault on the problem. They systematically applied an integer relation detection algorithm to large classes of sums of the above type, trying to find evaluations of these sums in terms of zeta functions (see Table 7.1 and 7.2 for details). Some of the experimentally discovered evaluations were then proven rigorously, others remain conjectures. While Au-Yeung's insight may fill us with a sense of amazement, the experimenters' approach appears quite natural and systematic.

The editors of *Experimental Mathematics* are advocating a change in the

way mathematics is written, placing more emphasis on the mathematical process. Imre Lakatos in his influential though controversial book *Proofs and Refutations* [138] advocated a similar change from what he called the deductivist style of proof to the heuristic style of proof. In the deductivist style, the definitions are carefully tailored to the proofs. The proofs are frequently elegant and short. But it is difficult to see what process led to the discovery of the theorem and its proof. The heuristic style maintains the mathematical rigor but again the emphasis is more on process. One does not merely give the definition but perhaps includes a comment on why this definition was chosen and not another. This is clearly an important shift if the editors wish to meet their second objective, the sharing of insights.

### 7.2.2 The Deductivist Style

The major focus of this section is Imre Lakatos's description of the deductivist style in *Proofs and Refutations*. An extreme example of this style is given in the form of a computer generated proof of  $(1 + 1)^n = 2^n$  in Table 7.3.

Euclidean Methodology has developed a certain obligatory style of presentation. I shall refer to this as "deductivist style." This style starts with a painstakingly stated list of *axioms*, *lemmas* and/or *definitions*. The axioms and definitions frequently look artificial and mystifyingly complicated. One is never told how these complications arose. The list of axioms and definitions is followed by the carefully worded *theorems*. These are loaded with heavy-going conditions; it seems impossible that anyone should ever have guessed them. The theorem is followed by the *proof*. ([138] p. 142)

This is the essence of what we have called formal understanding. We know that the results are true because we have gone through the crucible of the mathematical process and what remains is the essence of truth. But the insight and thought processes that led to the result are hidden.

In deductivist style, all propositions are true and all inferences valid. Mathematics is presented as an ever-increasing set of eternal, immutable truths. ([138] p. 142)

## Some Experimental Results

Definitions:

$$\zeta(s) = \sum_{k=1}^{\infty} k^{-s}$$

$$s_h(m, n) = \sum_{k=1}^{\infty} \left(1 + \frac{1}{2} + \cdots + \frac{1}{k}\right)^m (k+1)^{-n} \quad m \geq 1, n \geq 2$$

Some experimentally derived conjectures:

$$s_h(3, 2) = \frac{15}{2}\zeta(5) + \zeta(2)\zeta(3)$$

$$s_h(3, 3) = -\frac{33}{16}\zeta(6) + 2\zeta^2(3)$$

$$s_h(3, 4) = \frac{119}{16}\zeta(7) - \frac{33}{4}\zeta(3)\zeta(4) + 2\zeta(2)\zeta(5)$$

$$s_h(3, 6) = \frac{197}{24}\zeta(9) - \frac{33}{4}\zeta(4)\zeta(5) - \frac{37}{8}\zeta(3)\zeta(6) + \zeta^3(3) + 3\zeta(2)\zeta(7)$$

$$s_h(4, 2) = \frac{859}{24}\zeta(6) + 3\zeta^2(3)$$

We are given the raw data with which to work, carefully organized to give us a glimpse into the investigators' insights on the problem. Note in the first formula for  $s_h(3, 2)$ ,  $3 + 2 = 5$ , on the right hand side of the equation we have  $\zeta(5)$  and  $\zeta(3)\zeta(2)$ .

Some proven Euler sums:

$$s_h(2, 2) = \frac{3}{2}\zeta(4) + \frac{1}{2}\zeta^2(2) = \frac{11\pi^4}{360}$$

$$s_h(2, 4) = \frac{2}{3}\zeta(6) - \frac{1}{3}\zeta(2)\zeta(4) + \frac{1}{3}\zeta^3(2) - \zeta^2(3) = \frac{37\pi^6}{22680} - \zeta^2(3)$$

The proven evaluation for  $s_h(2, 2)$  above implies the truth of Au-Yeung's discovery.

Table 7.2: Some experimental results



### Shrinking or encapsulating mathematics

When one first learns to sum infinite series one is taught to sum geometric series

$$1 + \rho + \rho^2 + \cdots + \rho^k + \cdots = \frac{1}{1 - \rho}$$

when  $|\rho| < 1$ . Next one learns to sum telescoping series. For example if  $f(i) = \frac{1}{i+1} - \frac{1}{i+2}$ , it is not too hard to see that

$$\sum_{i=0}^n f(i) = \left(1 - \frac{1}{2}\right) + \left(\frac{1}{2} - \frac{1}{3}\right) \cdots \left(\frac{1}{n+1} - \frac{1}{n+2}\right) = 1 - \frac{1}{n+2}$$

and in particular that

$$\sum_{i=0}^{\infty} f(i) = 1.$$

The *Wilf-Zeilberger* algorithms employ “creative telescoping” to show that a sum or integral is zero. The algorithms really provide a meta-insight into a broad range of problems involving identities. Unfortunately the proofs produced by the computer, while understandable by most mathematicians are at the same time uninteresting. On the other hand, the existence of WZ proofs for large classes of objects gives us a global insight into these areas.

Table 7.3: Shrinking or encapsulating mathematics

Deductivist style hides the struggle, hides the adventure. The whole story vanishes, the successive tentative formulations of the theorem in the course of the proof-procedure are doomed to oblivion while the end result is exalted into sacred infallibility. ([138] p. 142)

Perhaps the most extreme examples of the deductivist style come out of the computer generated proofs guaranteed by Wilf and Zeilberger’s algorithmic proof theory. It is important to note here that Wilf and Zeilberger transform the problem of proving identities to the more computer oriented problem of solving a system of linear equations with symbolic coefficients.

These WZ proofs (see Table 7.4) are perhaps the ultimate in the deductivist tradition. At present, knowing the WZ proof of an identity amounts to little more (We will discuss the importance of certificates later.) than knowing that the identity is true. In fact, Doron Zeilberger in [195] has advocated leaving only a QED at the end of the statement, the author's seal that he has had the computer perform the calculations needed to prove the identity. The advantage of this approach is that the result is completely encapsulated. Just as one would not worry about how the computer multiplied two huge integers together or inverted a matrix, one now has results whose proofs are uninteresting.

## 7.3 Zeilberger and the Encapsulation of Identity

### 7.3.1 Putting a Price on Reliability

In the last two sections we talked about the importance of communicating insights within the mathematical community. There we focused on the process of mathematical thought but now we want to talk about communicating insights that have not been made rigorous.

We have already briefly talked about Wilf and Zeilberger's algorithmic proof theory and its denial of insight. In this section we will discuss the implications of this theory and D. Zeilberger's philosophy of mathematics as contained in *Theorems for a Price: Tomorrow's Semi-Rigorous Mathematical Culture* ([196]).

It is probably unfortunate but perhaps necessary that the two voices most strongly advocating truly experimental math are also at times the most hyperbolic in their language. We will concentrate mostly on the ideas of Doron Zeilberger but G. J. Chaitin should not and will not be ignored.

We will begin with D. Zeilberger's "Abstract of the future"

We show in a certain precise sense that the Goldbach conjecture is true with probability larger than 0.99999 and that its complete truth could be determined with a budget of 10 billion. ([196] p. 980)

Once people get over the shock of seeing probabilities assigned to truth in mathematics the usual complaint is that the 10 billion is ridiculous. Computers have been getting better and cheaper for years. What can it mean that "the complete

### An Uninteresting(?) Proof

Below is a sample WZ proof of  $(1 + 1)^n = 2^n$  (this proof is a modified version of the output of Doron Zeilberger's original *Maple* program, influenced by the proof in [195]).

Let  $F(n, k) = \binom{n}{k} 2^{-n}$ . We have to show that  $l(n) = \sum_k F(n, k) = 1$ . To do this we will show that  $l(n+1) - l(n) = 0$  for every  $n \geq 0$  and that  $l(0) = 1$ . The second half is trivial since for  $n = 0$ ,  $F(0, 0)$  is equal to 1 and 0 otherwise. The first half is proved by the WZ algorithm.

We construct

$$G(n, k) = \frac{-1}{2^{(n+1)}} \binom{n}{k-1} \left( = \frac{-k}{2(n-k+1)} F(n, k) \right),$$

with the motive that

$$WZ = F(n+1, k) - F(n, k) = G(n, k+1) - G(n, k) \quad (\text{check!}).$$

Summing WZ with respect to  $k$  gives

$$\begin{aligned} \sum_k F(n+1, k) - \sum_k F(n, k) &= \\ \sum_k (G(n, k+1) - G(n, k)) &= 0 \end{aligned}$$

(by telescoping). We have now established that  $l(n+1) - l(n) = 0$  and we are done.

The proof gives little insight into this binomial coefficient identity. However, the algorithms give researchers in other fields direct access to the field of special function identities.

Table 7.4: An uninteresting(?) proof

truth could be determined with a budget of 10 billion?” What is clear from the article is that this is an additive measure of the difficulty of completely solving this problem. If we know that the Riemann hypothesis will be proven if we prove lemmas costing 10 billion, 2 billion and 2 trillion dollars respectively, we can tell at a glance not merely what it would “cost” to prove the hypothesis but also where new ideas will be essential in any proof. ( This assumes that 2 trillion is a lot of “money.”)

The introduction of “cost” leads immediately to consideration of a trend that has over taken the business world and is now intruding rapidly on academia: a focus on productivity and efficiency.

It is a waste of money to get absolute certainty, unless the conjectured identity in question is known to imply the Riemann Hypothesis ([196] p. 980)

We have taken this quote out of its context (Wilf and Zeilberger’s algorithmic proof theory of identities) [196] but even so we think it is indicative of a small but growing group of mathematicians who are asking us to look at not just the benefits of reliability in mathematics but also the associated costs. See for example A. Jaffe and F. Quinn in [120, 121] and G. Chaitin in [67]. Still, we have not dealt with the central question. Why does D. Zeilberger need to introduce probabilistic “truths?” and how might we from a “formalist” perspective not feel this to be a great sacrifice?

### 7.3.2 It’s All About Insight

Why is Zeilberger so willing to give up on absolute truths? The most reasonable answer is that he is pursuing deeper truths. In *Identities in Search of Identities*, Zeilberger advocates an examination of identities for the sake of studying identities. Still as Herb Wilf and others have pointed out it is possible to produce an unlimited number of identities. It is the context, the ability to use and manipulate these identities, that make them interesting. Why then might we think that studying identities for their own sake may lead us down the golden path rather than the garden path?

We are now looking for what might be called meta-mathematical structures. We remove the math from its original context and isolate it, trying to detect new structures. When doing this it is impossible to collect only the relevant

information that will lead to the new discovery. One collects objects (theorems, statistics, conjectures, etc.) that have a reasonable degree of similarity and familiarity and then attempts to eliminate the irrelevant or the untrue (counter examples). We are preparing for some form of eliminative induction. There is a built in stage, where objects are subject to censorship. In this context, it is not unreasonable to introduce objects where one is not sure of their truth, since all the objects, whether proved or not, will be subject to the same degree of scrutiny. Moreover, if these probably true objects fall into the class of reliable (i.e., they fit the new conjecture) objects, it may be possible to find a legitimate proof in the new context. Recall that the fast WZ algorithms transform the problem of proving an identity to one of solving a system of linear equations with symbolic coefficients.

It is very time consuming to solve a system of linear equations with symbolic coefficients. By plugging in specific values for  $n$  and other parameters if present, one gets a system with numerical coefficients, which is much faster to handle. Since it is unlikely that a random system of inhomogeneous linear equations with more equations than unknowns can be solved, the solvability of the system for a number of special values of  $n$  and the other parameters is a very good indication that the identity is indeed true. ([196] p. 980)

Suppose we can solve the system above for ten different assignments for  $n$  and the other parameters but cannot solve the general system. What do we do if we really need this identity? We are in a peculiar position. We have reduced the problem of proving identities involving sums and integrals of proper-hypergeometric terms to the problem of solving a possibly gigantic system of inhomogeneous linear equations with more equations than unknowns. We have an appropriately strong belief that this system has a solution but do not have the resources to uncover this solution.

What can we do with our result? If we agree with G. J. Chaitin, we may want to introduce it as an “axiom.”

I believe that elementary number theory and the rest of mathematics should be pursued more in the spirit of experimental science, and that you should be willing to adopt new principles. I believe that Euclid’s

statement that an axiom is a self-evident truth is a big mistake<sup>4</sup>. The Schrödinger equation certainly isn't a self-evident truth! And the Riemann Hypothesis isn't self-evident either, but it's very useful. A physicist would say that there is ample experimental evidence for the Riemann Hypothesis and would go ahead and take it as a working assumption. ([67] p. 24)

In this case, we have ample experimental evidence for the truth of our identity and we may want to take it as something more than just a working assumption. We may want to introduce it formally into our mathematical system.

## 7.4 Experiment and “Theory”

We have now examined two views of experimental mathematics but we appear to be no closer to a definition than when we began. However, we are now ready to begin in full our exploration of experiment. In *Advice to a Young Scientist*, P. B. Medawar defines four different kinds of experiment: the Kantian, Baconian, Aristotelian, and the Galilean. Mathematics has always participated deeply in the first three categories but has somehow managed to avoid employing the Galilean model. In developing our notion of experimental mathematics we will try to adhere to this Galilean mode as much as possible.

We will begin with the Kantian experiments. Medawar gives as his example:

generating “the classical non-Euclidean geometries (hyperbolic, elliptic) by replacing Euclid's axiom of parallels (or something equivalent to it) with alternative forms.” ([153] pp. 73–74)

It seems clear that mathematicians will have difficulty escaping from the Kantian fold. Even a Platonist must concede that mathematics is only accessible through the human mind and thus at a basic level all mathematics might be considered a Kantian experiment. We can debate whether Euclidean geometry is but an idealization of the geometry of nature (where a point has no length or breadth and a line has length but no breadth?) or nature an imperfect reflection of “pure” geometrical objects, but in either case the objects of interest lie within the minds eye.

---

<sup>4</sup>There is no evidence that Euclid ever made such a statement. However, the statement does have an undeniable emotional appeal.

Similarly, we cannot escape the Baconian experiment. In Medawar's words this

is a contrived as opposed to a natural happening, it "is the consequence of 'trying things out' or even of merely messing about." ([153] p. 69)

Most of the research described as experimental is Baconian in nature and in fact one can argue that all of mathematics proceeds out of Baconian experiments. One tries out a transformation here, an identity there, examines what happens when one weakens this condition or strengthens that one. Even the application of probabilistic arguments in number theory can be seen as a Baconian experiment. The experiments may be well thought out and very likely to succeed but the ultimate criteria of inclusion of the result in the literature is success or failure. If the "messing about" works (e.g., the theorem is proved, the counterexample found) the material is kept; otherwise, it is relegated to the scrap heap.

The Aristotelian experiments are described as demonstrations:

apply electrodes to a frog's sciatic nerve, and lo, the leg kicks; always precede the presentation of the dog's dinner with the ringing of a bell, and lo, the bell alone will soon make the dog dribble. ([153] p. 71)

The results are tailored to demonstrate the theorems, as opposed to the experiments being used to devise and revise the theorems. This may seem to have little to do with mathematics but it has everything to do with pedagogy. The Aristotelian experiment is equivalent to the concrete examples we employ to help explain our definitions, theorems, or the problems assigned to students so they can see how their newly learned tools will work.

The last and most important is the Galilean experiment:

(the) Galilean Experiment is a critical experiment—one that discriminates between possibilities and, in doing so, either gives us confidence in the view we are taking or makes us think it in need of correction. ([153])

Ideally one devises an experiment to distinguish between two or more competing hypotheses. In subjects like medicine the questions are in principal more clear

cut (the Will Roger’s phenomenon or Simpson’s paradox complicates matters <sup>5</sup> Does this medicine work (longevity, quality of life, cost effectiveness, etc.)? Is this treatment better than that one? Unfortunately, these questions are extremely difficult to answer and the model Medawar presents here does not correspond with the current view of experimentation. Since the spectacular “failure” (i.e., it worked beautifully but ultimately was supplanted see [139]) of Newtonian physics it has been widely held that no amount of experimental evidence can prove or disprove a theorem about the world around us and it is widely known that in the real world the models one tests are not true. Medawar acknowledges the difficulty of proving a result but has more confidence than modern philosophers in disproving hypotheses. If experiment cannot distinguish between hypotheses or prove theorems, what can it do? What advantages does it have? Is it necessary?

## 7.5 “Theoretical” Experimentation

While there is an ongoing crisis in mathematics, it is not as severe as the crisis in physics. The untestability of parts of theoretical physics (e.g., string theory) has led to a greater reliance on mathematics for “experimental verification.” This may be in part what led Arthur Jaffe and Frank Quinn to advocate what they have named “Theoretical Mathematics” (note that many mathematicians think they have been doing theoretical mathematics for years) but which we like to think of as “theoretical experimentation.” There are certainly some differences between our ideas and theirs but we believe they are more of emphasis than substance.

Unlike our initial experiment where we are working with and manipulating floating point numbers, “theoretical experimentation” would deal directly with theorems, conjectures, the consequences of introducing new axioms. ... Note that by placing it in the realm of experimentation, we shift the focus from the more general realm of mathematics, which concerns itself with the transmission of both truth and insight, to the realm of experimentation, which primarily deals with the establishment of and transmission of insight. Although it was originally conceived outside the experimental framework, the central problems Jaffe and

---

<sup>5</sup>Simpson’s paradox notes that two data sets can separately support one conclusion while the union of the data supports the opposite conclusion. Will Roger’s phenomenon notes that in a medical study it is possible to transfer a patient from one group to another and improve the statistics of both groups.



Quinn need to deal with are the same. They must attempt to preserve the rigorous core of mathematics, while contributing to an increased understanding of mathematics both formally and intuitively.

As described in Arthur Jaffe and Frank Quinn's *"Theoretical Mathematics": Toward a Cultural Synthesis of Mathematics and Theoretical Physics* it appears to be mainly a call for a loosening of the bonds of rigor. They suggest the creation of a branch of theoretical (experimental) mathematics akin to theoretical physics, where one produces speculative and intuitive works that will later be made reliable through proof. They are concerned about the slow pace of mathematical developments when all the work must be rigorously developed prior to publication. They argue convincingly that a haphazard introduction of conjectorial mathematics will almost undoubtedly result in chaos.

Their solution to the problems involved in the creation of theoretical (experimental) mathematics comes in two parts. They suggest that

theoretical work should be explicitly acknowledged as theoretical and incomplete; in particular, a major share of credit for the final result must be reserved for the rigorous work that validates it. ([120] p.10)

This is meant to ensure that there are incentives for following up and proving the conjectured results.

To guarantee that work in this theoretical mode does not affect the reliability of mathematics in general, they propose a linguistic shift.

Within a paper, standard nomenclature should prevail: in theoretical material, a word like "conjecture" should replace "theorem"; a word like "predict" should replace "show" or "construct"; and expressions such as "motivation" or "supporting argument" should replace "proof." Ideally the title and abstract should contain a word like "theoretical", "speculative", or "conjectural". ([120] p.10)

Still, none of the newly suggested nomenclature would be entirely out of place in a current research paper. Speculative comments have always had and will always have a place in mathematics.

This is clearly an exploratory form of mathematics. But is it truly experimental in any but the Baconian sense? The answer will of course lie in its application. If we accept the description at face value, all we have is a lessening

of rigor, covered by the introduction of a new linguistic structure. More “mathematics” will be produced but it is not clear that this math will be worth more, or even as much as, the math that would have been done without it.

It is not enough to say that mathematical rigor is strangling mathematical productivity. One needs to argue that by relaxing the strictures temporarily one can achieve more. If we view theoretical (experimental) mathematics as a form of Galilean experimentation then in its idealized form “theoretical” (experimental) mathematics should choose between directions (hypotheses) in mathematics. Like any experimental result the answers will not be conclusive, but they will need to be strong enough to be worth acting on.

Writing in this mode, a good theoretical paper should do more than just sketch arguments and motivations. Such a paper should be an extension of the survey paper, defining not what has been done in the field but what the author feels can be done, should be done and might be done, as well as documenting what is known, where the bottlenecks are, etc. In general, we sympathize with the desire to create a “theoretical” mathematics but without a formal structure and methodology it seems unlikely to have the focus required to succeed as a separate field.

One final comment seems in order here. “Theoretical” mathematics, as practiced today, seems a vital and growing institution. Mathematicians now routinely include conjectures and insights with their work (a trend that seems to be growing). This has expanded in haphazard fashion to include algorithms, suggested algorithms and even pseudo algorithms. We would distinguish our vision of “experimental” mathematics from “theoretical” mathematics by an emphasis on the constructive/algorithmic side of mathematics. There are well established ways of dealing with conjectures but the rules for algorithms are less well defined. Unlike most conjectures, algorithms if sufficiently efficacious soon find their way into general use.

While there has been much discussion of setting up standardized data bases to run algorithms on, this has proceeded even more haphazardly. Addressing these issues of reliability would be part of the purview of experimental mathematics. Not only would one get a critical evaluation of these algorithms but by reducing the problems to their algorithmic core, one may facilitate the sharing of insights both within and between disciplines. At its most extreme, a researcher from one discipline may not need to understand anything more than the outline of the algorithm to make important connections between fields.

## 7.6 A Mathematical Experiment

### 7.6.1 Experimentation

We now turn to a more concrete example of a mathematical experiment. Our meta-goal in devising this experiment was to investigate the similarities and differences between experiments in mathematics and in the natural sciences, particularly in physics. We therefore resolved to examine a conjecture which could be approached by collecting and investigating a huge amount of data: the conjecture that every non-rational algebraic number is normal in every base (see Table 7.5). It is important to understand that we did not aim to prove or disprove this conjecture; our aim was to find evidence pointing in one or the other direction. We were hoping to gain insight into the nature of the problem from an experimental perspective.

The actual experiment consisted of computing to 10,000 decimal digits the square roots and cube roots of the positive integers smaller than 1000 and then subjecting these data to certain statistical tests (again, see Table 7.5). Under the hypothesis that the digits of these numbers are uniformly distributed (a much weaker hypothesis than normality of these numbers), we expected the probability values of the statistics to be distributed uniformly between 0 and 1. Our first run showed fairly conclusively that the digits were distributed uniformly. In fact, the Anderson-Darling test, which we used to measure how uniformly distributed our probabilities were suggested that the probabilities might have been “too uniform” to be random. We therefore ran the same tests again, only this time for the first 20,000 decimal digits, hoping to detect some non-randomness in the data. The data were not as interesting on the second run.

### 7.6.2 Verification

It is even more important in mathematics than in the physical sciences that the data under investigation are completely reliable. At first glance it may seem that the increasing reliance of mathematicians on programs such as *Maple* and *Mathematica* has decreased the need for verification. Computers very rarely make arbitrary mistakes in arithmetic and algebra. But all the systems have known and unknown bugs in their programming. It is therefore imperative that we check our results. So what efforts did we take to verify our findings?

### Background on Normality

*Definition.* A real number is normal to the base 10 if every block of digits of length  $k$  occurs with frequency  $1/10^k$ .

*Example:* the Champernowne number

$$0.01234567891011121314 \dots 99100101 \dots$$

is known to be normal base 10.

Except for artificially created examples no numbers have been proven normal in any particular base. If we allow artificial numbers there are no explicit numbers known to be normal in every base<sup>a</sup>

#### *Questions*

- Are all non-rational algebraic numbers normal base 10?
- Do all non-rational algebraic numbers have uniformly distributed digits?

---

<sup>a</sup>G. J. Chaitin in *Randomness and Complexity in Pure Mathematics*, has a number he calls  $\Omega = \sum_p \text{halts } 2^{-|p|}$ , the halting probability, which he notes is “sort of a mathematical pun”, but is normal to all bases. He does this by identifying integers with binary strings representing Turing machines and summing over the programs that stopped (non-trivially, see [67] p.12).

Table 7.5: Background on normality

First of all, we had to make sure that the roots we computed were accurate to at least 10,000 (resp. 20,000) digits. We computed these roots using *Maple* as well as *Mathematica*, having them compute the roots to an accuracy of 10,010 digits. We then did two checks on the computed approximation  $s_n$  to  $\sqrt{n}$ . First, we tested that  $\sqrt{n} \in [s_n - 10^{-10005}, s_n + 10^{-10005}]$  by checking that  $(s_n - 10^{-10005})^2 < n < (s_n + 10^{-10005})^2$ . Second, we tested that the 10,000th through 10,005th digits were not all zeros or nines. This ensures that we actually computed the first 10,000 digits of the decimal expansion of  $\sqrt{n}$ . (We note that *Maple* initially did not give us an accuracy of 10,000 digits for all of the cube roots, so that we had to increase the precision.)

We then had to make sure that we computed the statistics and probability values accurately—or at least to a reasonable precision, since we used asymptotic formulas anyway. We did this by implementing them both in *Maple* and in *Mathematica* and comparing the results. We detected no significant discrepancy.

We claim that these measures reasonably ensure the reliability of our experimental results.

### 7.6.3 Interpretation

Our experimental results support the conjecture that every non-rational algebraic number is normal; more precisely, we have found no evidence against this conjecture. In this section we will describe how we looked at and interpreted the experimental data to arrive at this conclusion. We include only a few examples of how we looked at the data here. In fact, we have only looked at certain aspects of normality and randomness in decimal expansions. Thus our results may be interpreted more narrowly to support the hypothesis that algebraic numbers are normal base 10. A full description will be found in [52].

Our main goal here is to give a quick visual summary that is at once convincing and data rich. These employ some of the most basic tools of visual data analysis and should probably form part of the basic vocabulary of an experimental mathematician. Note that traditionally one would run a test such as the Anderson-Darling test (which we have done) for the continuous uniform distribution and associate a particular probability with each of our sets of probability, but unless the probability values are extremely high or low it is difficult to interpret these statistics.

Experimentally, we want to test graphically the hypothesis of normality and

### Data and Statistics

We looked at the first 10,000 digits after the decimal point of the  $\sqrt{n}$  where  $n < 1000$  is not a perfect square and of  $\sqrt[3]{n}$  where  $n < 1000$  is not a perfect cube.

Tests used:

- $\chi^2$ —to check that each digit occurs 1/10 of the time (discrete uniform distribution base 10).
- Discrete Cramér-von Mises—to check that all groups of 4 consecutive digits occurs 1/10,000 of the time (discrete uniform distribution base 10,000).
- Anderson-Stephens —to check that the power spectrum of the sequence matches that of white noise (periodicity).
- Anderson-Darling—continuous uniform distribution.

Important point: In order for us to claim we have generated any evidence at all either for or against we have made two fairly strong assumptions.

- The first 10,000 digits are representative of the remaining digits.
- These digits behave as far as our statistical tests go like independent random variables.

In fact, for the first and second 10,000 digits our final conclusions are identical. The second assumption is problematic. Since we have beautiful algorithms to calculate these numbers, by most reasonable definitions of independent and random, these digits are neither.

Table 7.6: Data and statistics

randomness (or non-periodicity) for our numbers. Because the statistics themselves do not fall into the nicest of distributions, we have chosen to plot only the associated probabilities. We include two different types of graphs here. A quantile-quantile plot is used to examine the distribution of our data and scatter plots are used to check for correlations between statistics.

The first is a quantile-quantile plot of the chi square base 10 probability values versus a discrete uniform distribution. For this graph we have placed the probabilities obtained from our square roots and plotted them against a perfectly uniform distribution. Finding nothing here is equivalent to seeing that the graph is a straight line with slope one. This is a crude but effective way of seeing the data. The disadvantage is that the data are really plotted along a one dimensional curve and as such it may be impossible to see more subtle patterns.

The other graphs are examples of scatter plots. The first scatter plot shows that nothing interesting is occurring. We are again looking at probability values this time derived from the discrete Cramer-von Mises (CVM) test base 10,000. For each cube root we have plotted the point  $(f_i, s_i)$ , where  $f_i$  is the CVM base 10,000 probability associated with the first 2500 digits of the cube root of  $i$  and  $s_i$  is the probability associated with the next 2500 digits. A look at the graph reveals that we have now plotted our data on a two dimensional surface and there is a lot more “structure” to be seen. Still, it is not hard to convince oneself that there is little or no relationship between the probabilities of the first 2500 digits and the second 2500 digits.

The last graph is similar to the second. Here we have plotted the probabilities associated with the Anderson-Stephens statistic of the first 10,000 digits versus the first 20,000 digits. We expect to find a correlation between these tests since there is a 10,000 digit overlap. In fact, although the effect is slight, one can definitely see the thinning out of points from the upper left hand corner and lower right hand corner.

## 7.7 Conclusion

All the versions of experimental mathematics that we have dealt with so far have two characteristics: their main interest is in expanding our mathematical knowledge as rapidly as possible and none of them stray too far from the mainstream. In many cases this urgency leads to a temporary relaxation of rigor, a relaxation that is well documented and hopefully can be cleaned up afterwards. In other

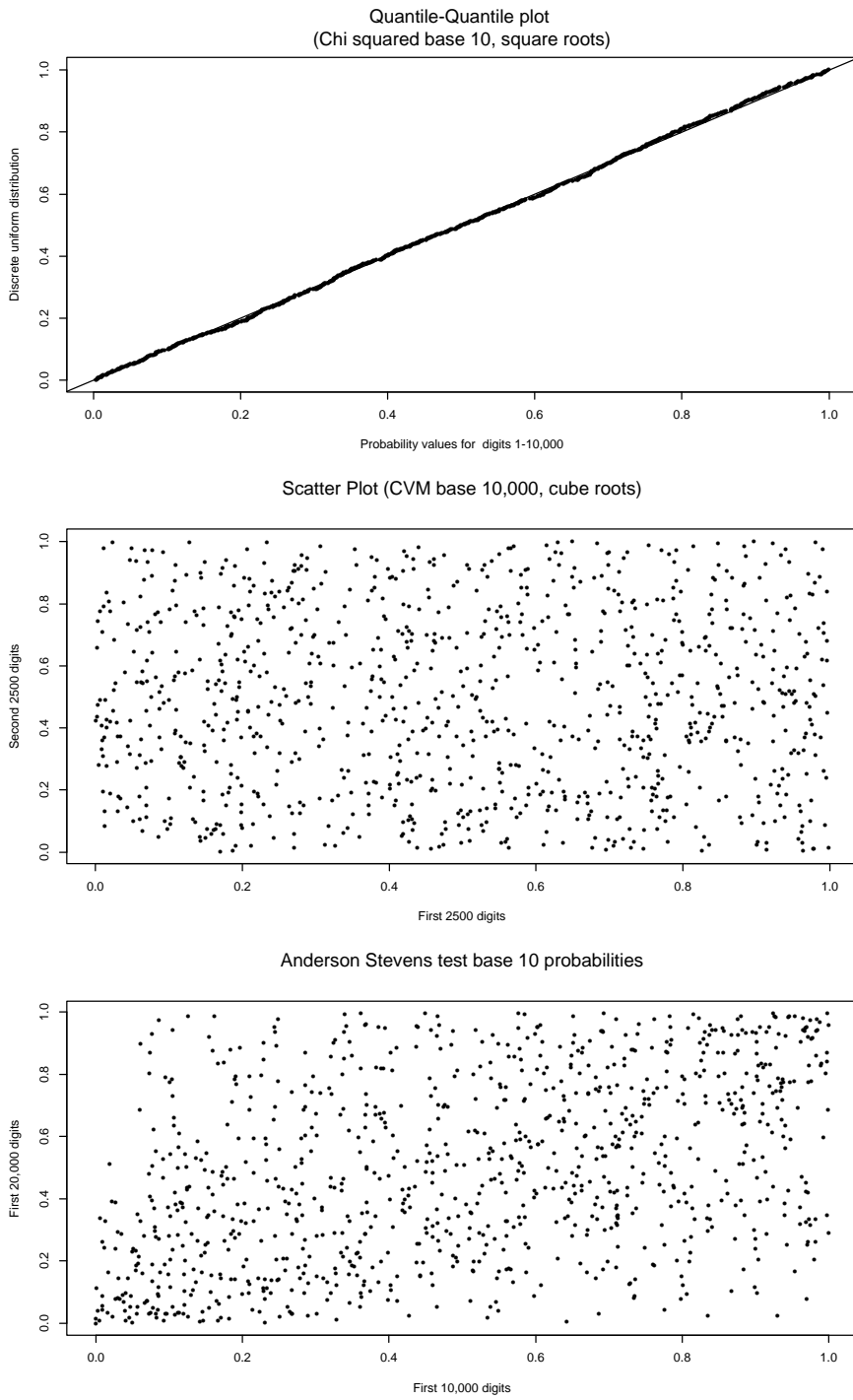


Figure 7.1: Graphical statistics of our experiments



cases it may be intrinsic to the mathematics they wish to explore. When a field has been as wildly successful as mathematics has been in the past few centuries there is a reluctance to change. We have hoped to convince some of the readers that these changes are revolutionary only in the same sense that the earth revolves around the sun.

We conclude with a definition of experimental mathematics.

*Experimental Mathematics* is that branch of mathematics that concerns itself ultimately with the codification and transmission of insights within the mathematical community through the use of experimental (in either the Galilean, Baconian, Aristotelian or Kantian sense) exploration of conjectures and more informal beliefs and a careful analysis of the data acquired in this pursuit.

Results discovered experimentally will, in general, lack some of the rigor associated with mathematics but will provide general insights into mathematical problems to guide further exploration, either experimental or traditional. We have restricted our definition of experimental mathematics to methodological pursuits that in some way mimic Medawar's views of Gallilean experimentation. However, our emphasis on insight also calls for the judicious use of examples (Aristotelian experimentation).

If the mathematical community as a whole was less splintered, we would probably remove the word "codification" from the definition. That is to say that a great deal of time will need to be spent on developing a language for the expression of experimental results. Since there are real communications problems between fields and since the questions to be explored will be difficult, it seems imperative that experimental investigators make every effort to organize their insights and present their data in a manner that will be as widely accessible as possible<sup>6</sup>.

With respect to reliability and rigor, the main tools here are already in place. We need to stress systematization of our exploration. As in our experimental project on normality, it is important to clearly define what has been looked at, how things have been examined, and what confidence the reader should have in the data. Although mathematicians may not like to admit it, ease of use will have to be a primary consideration if experimental results are to be of widescale use. As such, visualization and hypertextual presentations of material

---

<sup>6</sup>It is clear that mechanisms are developing for transmitting insights within fields, even if this is only through personal communications.

will become increasingly important in the future. We began by stealing some of the basic tools of scientific analysis and laying claim to them. As the needs of the community become more apparent one would expect these tools and others to evolve into a form better suited to the particular needs of the mathematical community. Someday, who knows, first year graduate students may be signing up for Experimental Methods in Mathematics I.



# Chapter 8

## A Color Supplement

Here we display various images which could not or should not be seen only in gray scale, as shown elsewhere in these volumes. We also should mention here Tom Banchoff and David Cervone's beautiful interactive collection, *Computer Graphics in Mathematical Research from ICM 1978 to ICMS 2002*. See also <http://www.math.union.edu/local/ICMS-poster>.

1. Figure 8.1 shows a Pascal triangle, where each entry has been reduced modulo two as discussed in Section 2.1, together with a tetrahedral design illustrating the same concept.
2. Figure 8.2 shows three  $10 \times 10$  arrays colored with the digits of the first hundred digits of  $223/71$ ,  $\pi$ , and  $22/7$ . One sees a clear pattern on the right ( $22/7$ ), a more subtle structure on the left ( $223/71$ ), and a "random" coloring in the middle ( $\pi$ ) (see Chapter 1 of the second volume).
3. Figure 8.3 illustrates Apostol's remarkably simple proof that  $\sqrt{2}$  is irrational. We simply continue to draw smaller and smaller integer-sided similar triangles until the area drops below  $1/2$ , when we have a contradiction (see Section 2.9).
4. Figure 8.4 is a photo of Ferguson's "Eight-Fold Way" sculpture, which is at the Mathematical Sciences Research Institute in Berkeley, California, and his "Figure-Eight Knot Complement" sculpture, which is at Mount Holyoke College in South Hadley, Massachusetts. In the "Eight-Fold Way," the ridges and grooves on the white tetrahedral form show how to map the

fundamental domain onto the tetrahedron. If you move your finger along any groove or ridge, alternating left and right turns at each corner, then in eight steps you will return to your starting point. See Section 2.4 and Exercise 33 of Chapter 2 for additional details.

As we mentioned in Section 2.4, the volume of the “Figure-Eight Knot Complement” shape in hyperbolic space is given by a certain infinite series involving binomial coefficients in the denominators. Using Ferguson’s own PSLQ algorithm, Broadhurst later showed that this volume is also given by BBP-type formula, permitting one to rapidly calculate binary digits at arbitrary positions. See Section 2.4 and Exercise 33 of Chapter 2 for additional details.

5. Figure 8.5 shows a snow sculpture rendition of the Costa surface, which was carved in 1999 in an international snow sculpture competition by Helaman Ferguson, with the assistance of Stan Wagon, Dan Schwalbe, Tomas Nemeth, and (posthumously) Alfred Gray. The Costa surface is the simplest example of an infinite family of minimal surfaces. It was discovered visually by David Hoffman, William Meeks III, and James Hoffman, who between 1983 and 1985 studied equations put forth by Celso Costa, a Brazilian graduate student. The unique negative curvature of the Costa surface lends structural stability to the piece. A week after the snow sculpture competition in 1999, all other 14 entries had collapsed, but the “Invisible Handshake” (Ferguson’s name for the sculpture) still stood. See Item 32 of Chapter 2 for additional details.
6. Figure 8.6 is a photo of the “Firmament,” a sculpture of five touching spheres, by John Robinson, presented to Donald Coxeter on his 90th birthday, when the sculpture “Intuition” was unveiled outside the Fields Institute.
7. Figure 8.7 is a reproduction of a first day issue stamp commemorating the 75th birthday of Ramanujan. It was mailed from India with the correct postage added on the back. It arrived safely in Scotland despite the address being David Borwein’s home address.
8. Figure 8.8 is a photo of the octahedral kaleidoscope that Coxeter designed, and which was constructed in a London workshop in the 1920s. Two kalei-

dosscopes were made, one the octahedral model shown here, and another one, an icosahedral model that is not shown here.

9. Figure 8.9 shows the Julia set associated with the modulus of the modular function  $k$ , as described in the exercises to Chapter 4 of the second volume.
10. Figure 8.10 shows a graphic based on a random walk of the first million decimal digits of  $\pi$ , courtesy of David and Gregory Chudnovsky. It maps the digit stream to a surface in ways similar to those used by Mandelbrot and others.
11. Figure 8.11 is a photo of the original ENIAC computer system, from the archives of the Smithsonian Institution.
12. Figure 8.12 illustrates the “experimental topology” of knots, as discussed in Item 37 of Chapter 2 in this book and also in Chapter 3 of the second volume.
13. Figure 8.13 is another example of two knots that were once thought to be distinct, but which are now known to be identical, as mentioned in Chapter 3 of the second volume.
14. Figure 8.14 represents a projection of a regular polyhedron in four dimensions, with 120 faces each of which is a dodecahedron. An artist’s rendering of this polyhedron hangs from the ceiling the Fields Institute in Toronto.
15. Figure 8.15 shows the Riemann surface of the Lambert  $W$  function, respectively, as discussed in Chapter 5 of the second volume.
16. Figure 8.16 shows the Julia set associated with the Newton iteration for the polynomial  $x^3 - 1 = 0$ , as discussed in Chapter 7 of the second volume. What is actually plotted here, encoded in various colors, is the number of iterations required to converge to one of the three roots if you perform complex Newton iterations starting at the given point, using double-double precision arithmetic for extra accuracy. Note that the iterations do not merely converge to the nearest point. Instead, a filamentary structure (a “Julia set”) of measure zero is formed, separating infinitely many disconnected convergence sets.

17. Figure 8.17 is a plot of all roots of polynomials with coefficients in  $(0, 1, -1)$  up to degree  $N = 18$ . These zeros are colored by their local density normalized to the range of densities; from red for low density to yellow for high density. The bands, quite clearly visible in the plot, are unexplained.

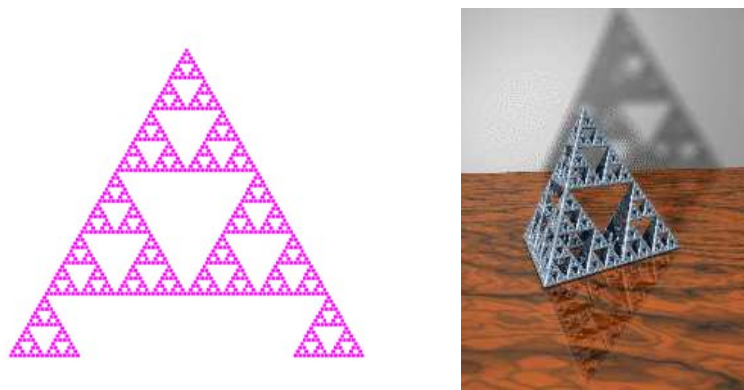
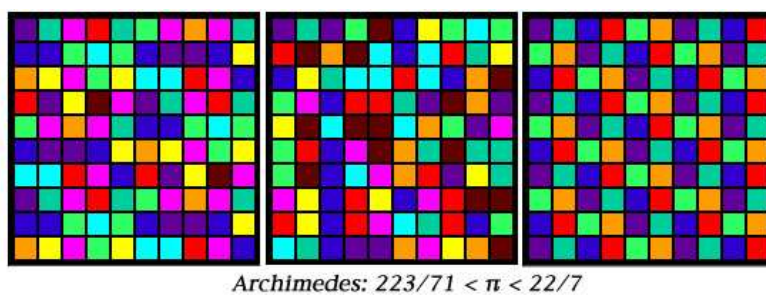


Figure 8.1: Pascal triangles modulo two?



*Archimedes:  $223/71 < \pi < 22/7$*

Figure 8.2: A pictorial proof of Archimedes' inequality

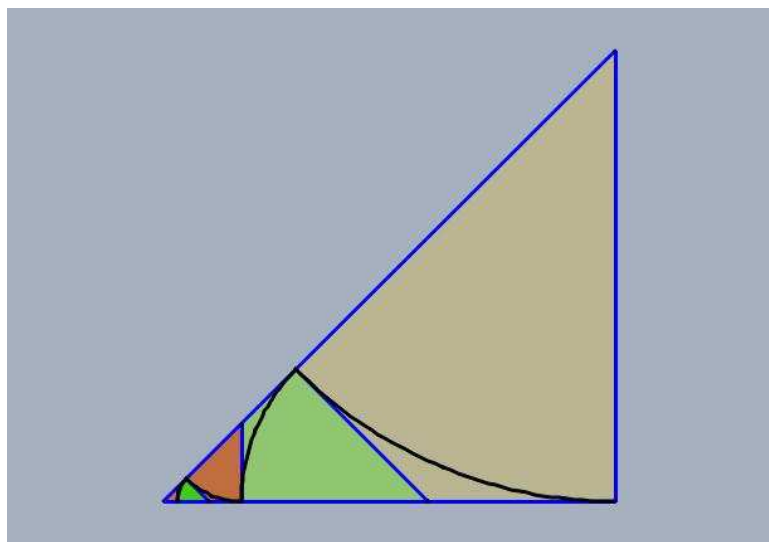


Figure 8.3: Root two is irrational.



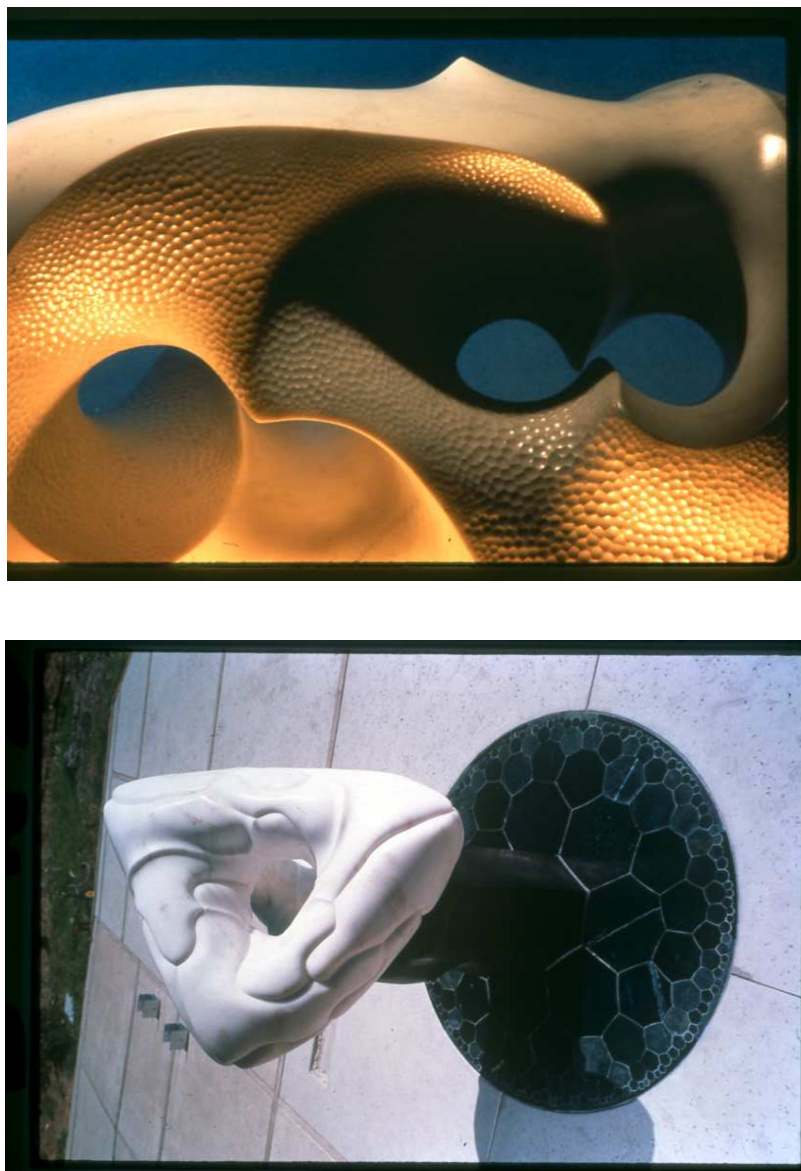


Figure 8.4: Ferguson's "Eight-Fold Way" and "Figure-Eight Knot Complement."

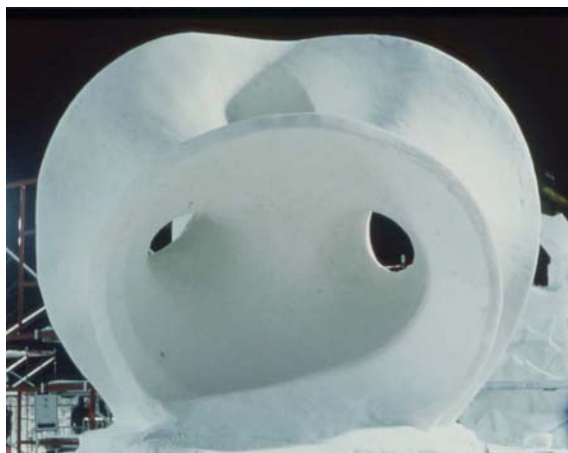


Figure 8.5: Ferguson's snow sculpture of the Costa surface



Figure 8.6: The "Firmament" by John Robinson.

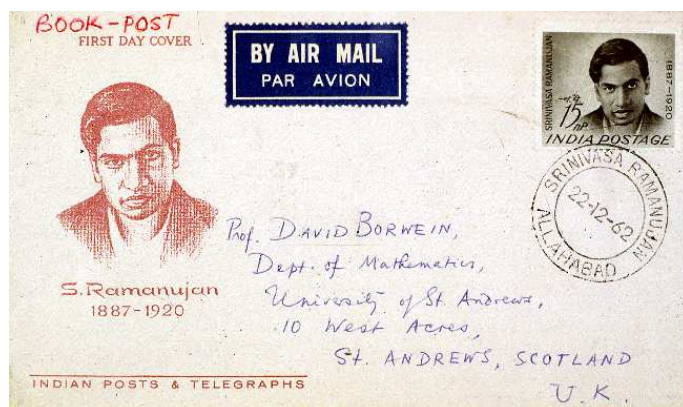


Figure 8.7: Ramanujan's seventy-fifth birthday stamp.



Figure 8.8: Coxeter's own kaleidoscope model.

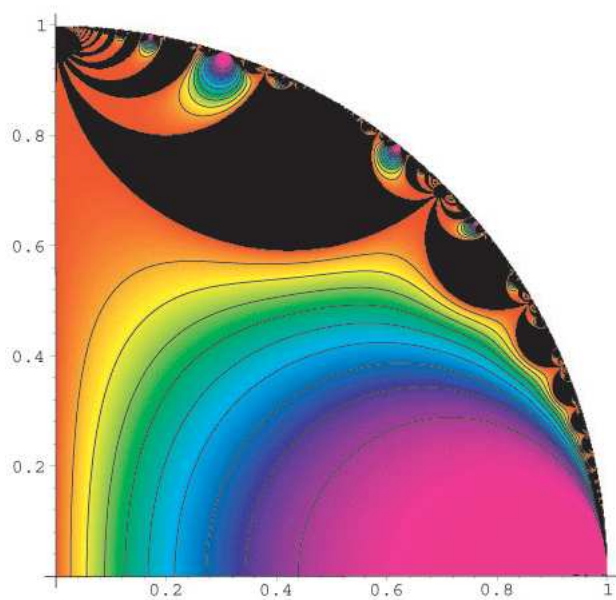


Figure 8.9: The modulus of the modular function  $k$ .

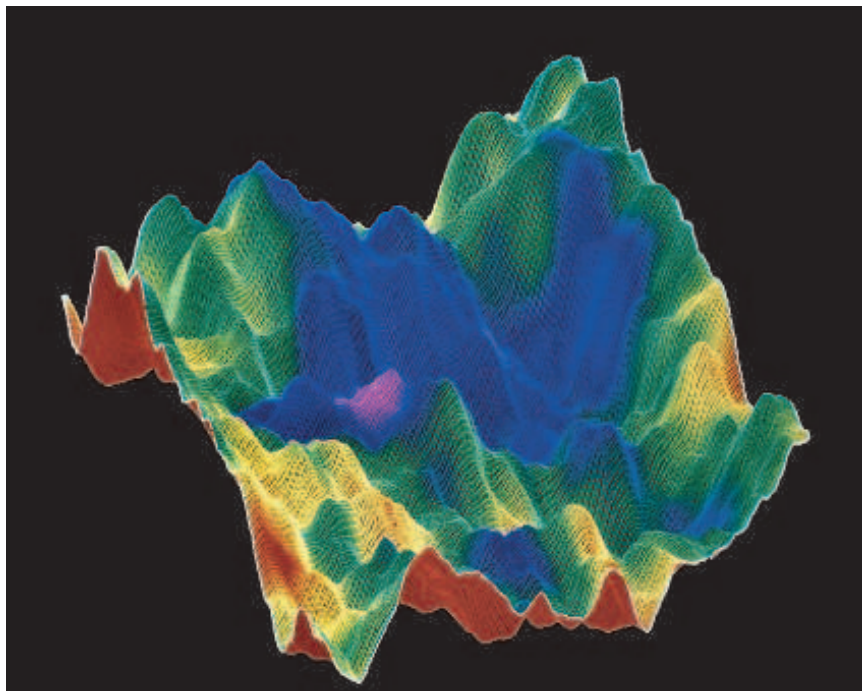


Figure 8.10: A random walk based on one million digits of  $\pi$ .



Figure 8.11: The ENIAC computer.



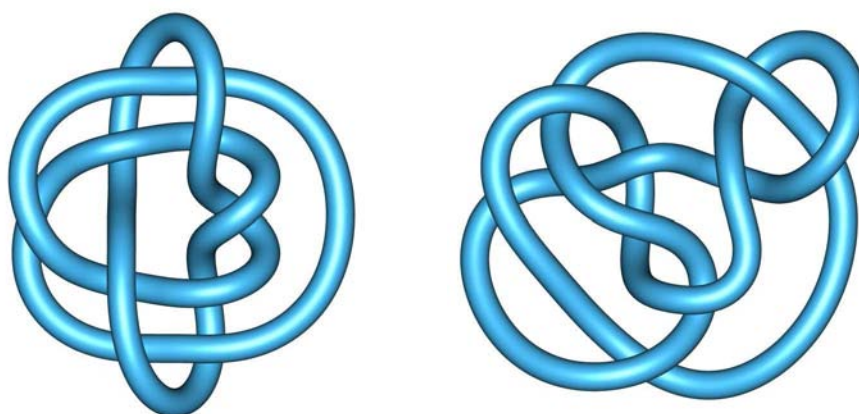


Figure 8.12: The knots  $10_{161}$  and  $10_{162}$ .

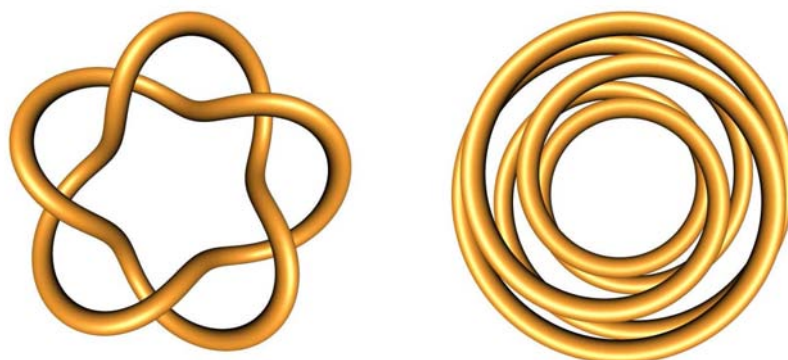


Figure 8.13: How does one identify two knots?

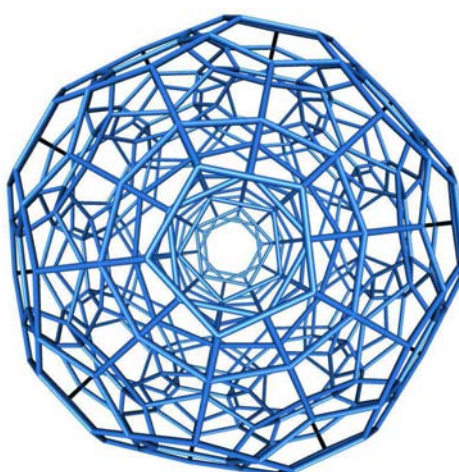


Figure 8.14: Projection of 4-D polyhedron.

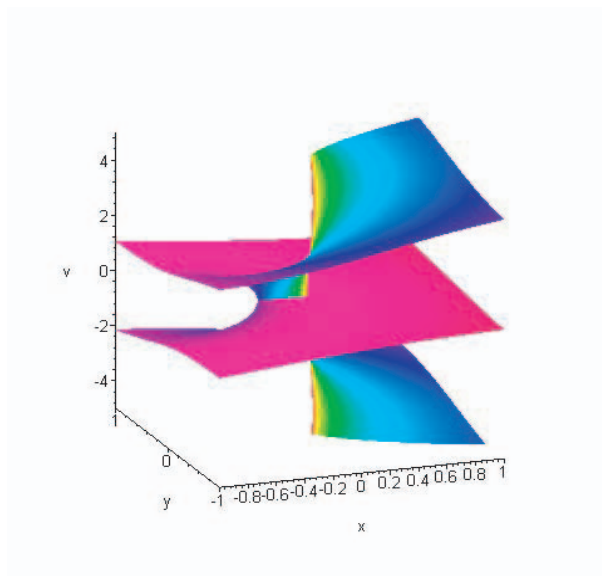


Figure 8.15: The Riemann surface for the Lambert function.

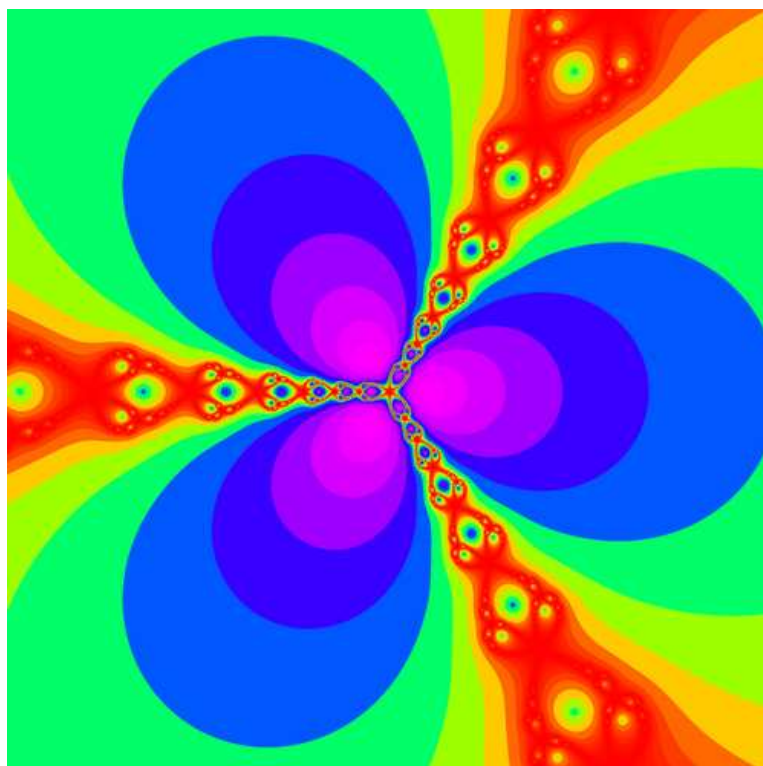


Figure 8.16: Julia set associated with Newton solutions of  $x^3 - 1 = 0$ .

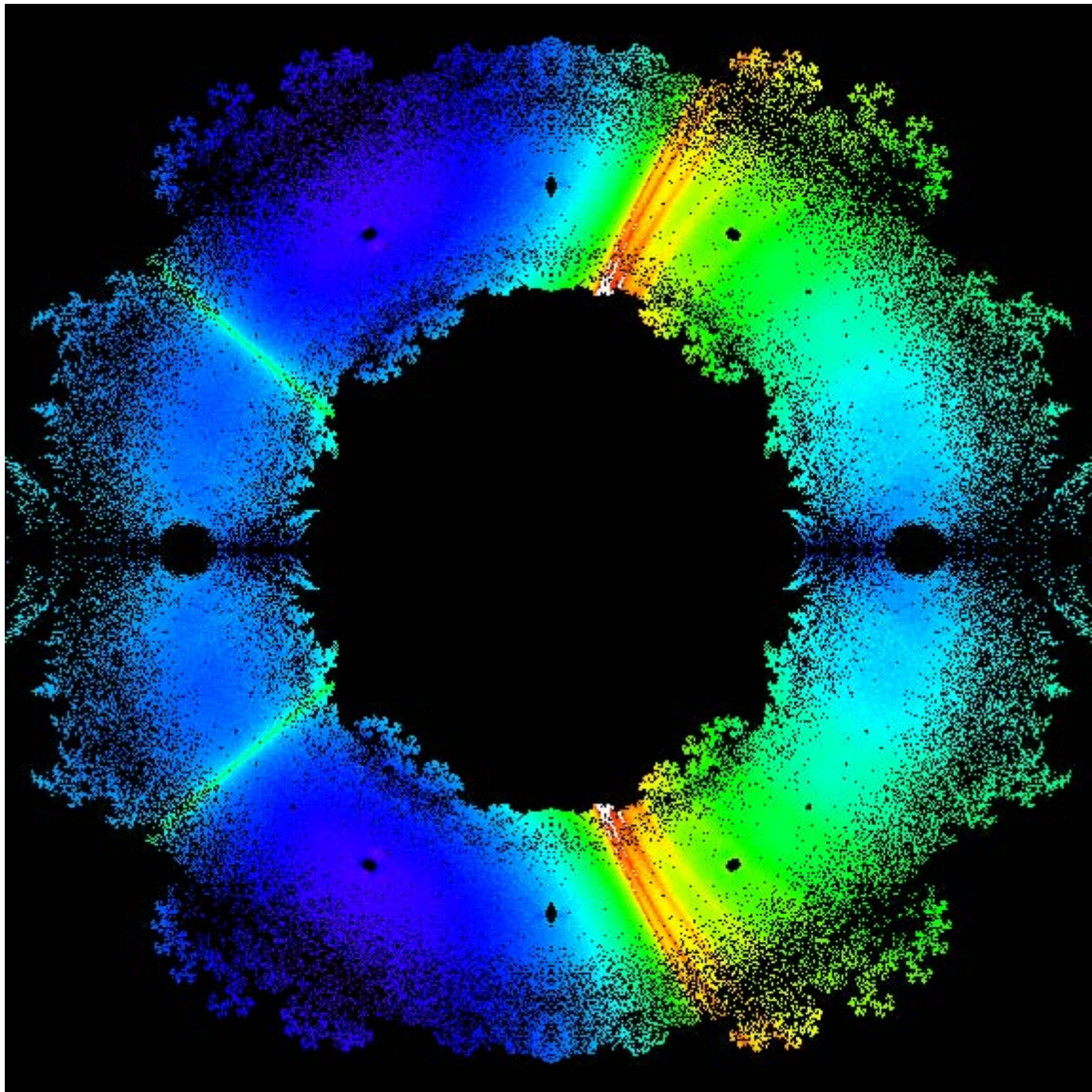


Figure 8.17: Roots of zero-one polynomials.

# Bibliography

- [1] Milton Abramowitz and Irene A. Stegun. *Handbook of Mathematical Functions*. Dover Publications, New York, 1970.
- [2] Elinor Mills Abreu. Gordon Moore Sees Another Decade for Moore's Law. <http://www.cnn.com/2003/TECH/biztech/02/11/moore.law.reut>, Feb. 11, 2003.
- [3] Martin Aigner and Gunter M. Ziegler. *Proofs from THE BOOK* (second edition). Springer-Verlag, Heidelberg, 2002.
- [4] Charalambos D. Aliprantis and Owen Burkinshaw. *Principles of Real Analysis*. Academic Press, New York, 1998.
- [5] George E. Andrews, Richard Askey, and Ranjan Roy. *Special Functions*. Cambridge University Press, Cambridge, 1999.
- [6] Tom Apostol. Irrationality of the Square Root of Two—A Geometric Proof. *American Mathematical Monthly*, 107:841–842, 2000.
- [7] Jorg Arndt and Christoph Haenel. *Pi Unleashed*. Springer-Verlag, Heidelberg, 2001.
- [8] Isaac Asimov and J. A. Shulman, editors. *Isaac Asimov's Book of Science and Nature Quotations*. Weidenfield and Nicolson, London, 1988.
- [9] Kendall E. Atkinson. *An Introduction to Numerical Analysis*. John Wiley and Sons, Hoboken, NJ, 1989.
- [10] David H. Bailey. The Computation of Pi to 29,360,000 Decimal Digits Using Borweins' Quartically Convergent Algorithm. *Mathematics of Computation*, 150:283–296, 1988.
- [11] David H. Bailey. FFTs in External or Hierarchical Memory. *Journal of Supercomputing*, 4:23–35, 1990.



- [12] David H. Bailey. Multiprecision Translation and Execution of Fortran Programs. *ACM Transactions on Mathematical Software*, 19:288–319, 1993.
- [13] David H. Bailey. A Fortran-90 Based Multiprecision System. *ACM Transactions on Mathematical Software*, 21:379–387, 1995.
- [14] David H. Bailey. On the Computational Cost of FFT-Based Linear Convolutions. <http://crd.lbl.gov/~dhbailey/dhbpapers/convops.pdf>, 1996.
- [15] David H. Bailey. Integer Relation Detection. *Computing in Science and Engineering*, 2(1):24–28, 2000.
- [16] David H. Bailey. A Compendium of BBP-Type Formulas for Mathematical Constants. <http://crd.lbl.gov/~dhbailey/dhbpapers/bbp-formulas.pdf>, 2001.
- [17] David H. Bailey, Jonathan M. Borwein, Peter B. Borwein, and Simon Plouffe. The Quest for Pi. *Mathematical Intelligencer*, 19:50–57, 1997.
- [18] David H. Bailey, Jonathan M. Borwein, and Richard E. Crandall. On the Khintchine Constant. *Mathematics of Computation*, 66:417–431, 1997.
- [19] David H. Bailey, Jonathan M. Borwein, Richard E. Crandall, and Carl Pomerance. On the Binary Expansions of Algebraic Numbers. <http://crd.lbl.gov/~dhbailey/dhbpapers/algebraic.pdf>, 2003.
- [20] David H. Bailey, Jonathan M. Borwein, and Roland Girgensohn. Experimental Evaluation of Euler Sums. *Experimental Mathematics*, 3:17–30, 1994.
- [21] David H. Bailey, Peter B. Borwein, and Simon Plouffe. On The Rapid Computation of Various Polylogarithmic Constants. *Mathematics of Computation*, 66:903–913, 1997.
- [22] David H. Bailey and David J. Broadhurst. A Seventeenth-Order Polylogarithm Ladder. <http://crd.lbl.gov/~dhbailey/dhbpapers/ladder.pdf>, 1999.
- [23] David H. Bailey and David J. Broadhurst. Parallel Integer Relation Detection: Techniques and Applications. *Mathematics of Computation*, 70:1719–1736, 2000.
- [24] David H. Bailey and Richard E. Crandall. On the Random Character of Fundamental Constant Expansions. *Experimental Mathematics*, 10:175–190, 2001.

- [25] David H. Bailey and Richard E. Crandall. Random Generators and Normal Numbers. *Experimental Mathematics*, 11:527–546, 2003.
- [26] David H. Bailey, Yozo Hida, Xiaoye S. Li, and Brandon Thompson. ARPREC: An Arbitrary Precision Computation Package. <http://crd.lbl.gov/~dhbailey/dhbpapers/arprec.pdf>, 2002.
- [27] David H. Bailey and Simon Plouffe. Recognizing Numerical Constants. In *The Organic Mathematics Project Proceedings*, volume 20, Ottawa, 1997. Canadian Mathematical Society.
- [28] David H. Bailey and Daniel Rudolph. An Ergodic Proof that Rational Times Normal Is Normal. <http://crd.lbl.gov/~dhbailey/dhbpapers/ratxnormal.pdf>, 2002.
- [29] David H. Bailey and Daniel J. Rudolph. A Strong Hot Spot Theorem. <http://crd.lbl.gov/~dhbailey/dhbpapers/hotspot.pdf>, 2003.
- [30] Michael Barnsley. *Fractals Everywhere*. Morgan Kaufmann, San Francisco, 1993.
- [31] H. Bauschke, C.H. Hamilton, M.S. Macklem, J.S. McMichael, and N.R. Swart. Recombression of JPEG Images by Requantization. *IEEE Transactions on Image Processing*, to appear.
- [32] Veronica Becher and Santiago Figuera. An Example of a Computable Absolutely Normal Number. *Theoretical Computer Science*, 270:947–958, 2002.
- [33] Petr Beckmann. *A History of Pi*. St. Martin's Press, New York, 1971.
- [34] L. Berggren, J. M. Borwein, and P. B. Borwein. *Pi: A Source Book*. Springer-Verlag, Heidelberg, (second edition) 2000.
- [35] David Berlinski. Ground Zero: A Review of The Pleasures of Counting, by T.W. Koerner. *The Sciences*, pages 37–41, Jul./Aug. 1997.
- [36] Bruce C. Berndt. Some Integrals in Ramanujan's Lost Notebook. *Proceedings of the American Mathematical Society*, to appear.
- [37] Bruce C. Berndt and Robert A. Rankin (eds.). *Ramanujan: Essays and Surveys*, volume 22. AMS-LMS History of Mathematics Series, Providence, 2001.

- [38] Patrick Billingsley. *Probability and Measure*. John Wiley, Hoboken, NJ, 1986.
- [39] David Blatner. *The Joy of Pi*. Walker and Company, New York, 1997.
- [40] E. B. Bogomolny and J. P. Keating. Random Matrix Theory and the Riemann Zeros, II: n-Point Correlations. *Nonlinearity*, 9:911–935, 1996.
- [41] Folkmar Bornemann, Dirk Laurie, Jörg Waldvogel, and Stan Wagon. *The SIAM/Oxford 100-digit Challenge*. Society for Industrial and Applied Mathematics, to appear.
- [42] Ephraim J. Borowski and Jonathan M. Borwein. *Dictionary of Mathematics*. Collins, Glasgow, 2002.
- [43] D. Borwein and J. M. Borwein. On Some Intriguing Sums Involving  $\zeta(4)$ . *Proceedings of the American Mathematical Society*, 123:111–118, 1995.
- [44] D. Borwein, J. M. Borwein, and R. Girgensohn. Explicit Evaluation of Euler Sums. *Proceedings of the Edinburgh Mathematical Society*, 38:273–294, 1995.
- [45] J. M. Borwein, D. M. Bradley, D. J. Broadhurst, and P. Lisoněk. Special Values of Multidimensional Polylogarithms. *Transactions of the American Mathematical Society*, 353:907–941, 2001.
- [46] J. M. Borwein and I. J. Zucker. Fast Evaluation of the Gamma Function for Small Rational Fractions using Complete Elliptic Integrals of the First Kind. *IMA Journal on Numerical Analysis*, 12:519–5267, 1992.
- [47] Jonathan Borwein, Peter Borwein, and K. Dilcher. Pi, Euler Numbers and Asymptotic Expansions. *American Mathematical Monthly*, 96:681–687, 1989.
- [48] Jonathan M. Borwein and David H. Bailey. *Experimentation in Mathematics: Computational Paths to Discovery*. A.K. Peters Ltd, Natick, MA, 2003.
- [49] Jonathan M. Borwein, David Borwein, and William F. Galway. Finding and Excluding  $b$ -ary Machin-Type BBP Formulae. <http://www.cecm.sfu.ca/preprints/2002pp.html>, 2002.
- [50] Jonathan M. Borwein and Peter B. Borwein. *Pi and the AGM: A Study in Analytic Number Theory and Computational Complexity*. CMS Series of Monographs and Advanced books in Mathematics, John Wiley, Hoboken, NJ, 1987.

- [51] Jonathan M. Borwein, Peter B. Borwein, and David H. Bailey. Ramanujan, Modular Equations and Pi or How to Compute a Billion Digits of Pi. *American Mathematical Monthly*, 96:201–219, 1989.
- [52] Jonathan M. Borwein, Peter B. Borwein, Roland Girgensohn, and Sheldon Parnes. Experimental Mathematical Investigation of Decimal and Continued Fraction Expansions of Selected Constants. *Manuscript*, 1995.
- [53] Jonathan M. Borwein and Robert Corless. Emerging Tools for Experimental Mathematics. *American Mathematical Monthly*, 106:889–909, 1999.
- [54] Jonathan M. Borwein and Adrian S. Lewis. *Convex Analysis and Nonlinear Optimization. Theory and Examples*. CMS books in Mathematics, Springer-Verlag, vol. 3, Heidelberg, 2000.
- [55] Peter B. Borwein. The Desmic Conjecture. *Journal of Combinatorial Theory, Series A*, 35:1–9, 1983.
- [56] Peter B. Borwein. *Computational Excursions in Analysis and Number Theory*. Springer-Verlag, CMS Books in Mathematics, Heidelberg, 2002.
- [57] David Bradley. A Class of Series Acceleration Formulae for Catalan’s Constant. *The Ramanujan Journal*, 3:159–173, 1999.
- [58] Richard P. Brent. Fast Multiple-Precision Evaluation of Elementary Functions. *Journal of the ACM*, 23:242–251, 1976.
- [59] David Bressoud. *Proofs and Confirmations: The Story of the Alternating Sign Matrix Conjecture*. Mathematical Association of America, Washington, 1999.
- [60] Keith Briggs. A Precise Calculation of the Feigenbaum Constants. *Mathematics of Computation*, 57:435–439, 1991.
- [61] David J. Broadhurst. Massive 3-loop Feynman Diagrams Reducible to SC\* Primitives of Algebras of the Sixth Root of Unity. <http://lanl.arxiv.org/abs/hep-th/9803091>, 1998.
- [62] David J. Broadhurst. Polylogarithmic Ladders, Hypergeometric Series and the Ten Millionth Digits of  $\zeta(3)$  and  $\zeta(5)$ . <http://lanl.arxiv.org/abs/math/9803067>, 1998.
- [63] David J. Broadhurst, John A Gracey, and Dirk Kreimer. Beyond the Triangle and Uniqueness Relations: Non-Zeta Counterterms at Large N from Positive Knots. *Zeitschrift fur Physik*, C75:559–574, 1997.

- [64] David J. Broadhurst and Dirk Kreimer. Association of Multiple Zeta Values with Positive Knots via Feynman Diagrams up to 9 Loops. *Physics Letters*, B383:403–412, 1997.
- [65] Josh Brodie. Perelman Explains Proof to Famous Math Mystery. *The Daily Princetonian*, April 17, 2003.
- [66] C. S. Calude, M. J. Dineen, and C.-K. Shu. Computing a Glimpse of Randomness. *Experimental Mathematics*, 11:361–370, 2002.
- [67] G. J. Chaitin. Randomness and Complexity in Pure Mathematics. *International Journal of Bifurcation and Chaos*, 4:3–15, 1994.
- [68] Marc Chamberland. Proof Without Words—Slicing Kites into Circular Sectors. *Mathematics Magazine*, 73:363, 2000.
- [69] Marc Chamberland. Binary BBP Formulae for Logarithms and Generalized Gaussian-Mersenne Primes. *Manuscript*, 2003.
- [70] Barry Cipra. Minimal Till It Melts. *SIAM News*, page 20, March 1999.
- [71] Joel S. Cohen. *Computer Algebra and Symbolic Computation: Elementary Algorithms*. A. K. Peters, Natick, MA, 2002.
- [72] Joel S. Cohen. *Computer Algebra and Symbolic Computation: Mathematical Methods*. A. K. Peters, Natick, MA, 2003.
- [73] J. Brian Conrey. The Riemann Hypothesis. *Notices of the AMS*, 50:341–353, 2003.
- [74] Robert M. Corless. Error Backward. In *Proceedings of Chaotic Numerics*, volume 172 of *AMS Contemporary Mathematics*, pages 31–62, Providence, 1994.
- [75] Robert M. Corless, James H. Davenport, David J. Jeffrey, Gurjeet Litt, and Stephen M. Watt. *Reasoning About the Elementary Functions of Complex Analysis*, pages 115–126. Springer-Verlag, Heidelberg, 2000.
- [76] Robert M. Corless, Gaston H. Gonnet, David E. G. Hare, David J. Jeffrey, and Donald E. Knuth. On the Lambert W function. *Advances in Computational Mathematics*, 5:329–359, 1996.
- [77] Richard Corliss. Unlocking The Matrix. *Time*, May 12, 2003.

- [78] Richard Crandall and Carl Pomerance. *Prime Numbers: A Computational Perspective*. Springer-Verlag, Heidelberg, 2001.
- [79] Richard E. Crandall. *Topics in Advanced Scientific Computation*. Springer-Verlag, Heidelberg, 1996.
- [80] Richard E. Crandall and Jason S. Papadopoulos. On the Implementation of AKS-Class Primality Tests.  
<http://developer.apple.com/hardware/ve/acgresearch.html>, 2003.
- [81] Charles Darwin. *On the Origin of Species By Means of Natural Selection*. Murray, London, 1859.
- [82] Martin Davis. *The Universal Computer: The Road from Leibniz to Turing*. W. W. Norton and Company, New York, 2000.
- [83] Bart de Smit and Henrik W. Lenstra Jr. The Mathematical Structure of Escher's Print Gallery. *Notices of the American Mathematical Society*, 50:446–451, 2003.
- [84] Paulo Ney de Souza and Jorge-Nuno Silva. *Berkeley Problems in Mathematics*. Springer-Verlag, Heidelberg, 2001.
- [85] Richard Dedekind. *Essays on the Theory of Numbers*. Dover, New York, 1963.
- [86] J. De Doelder. On Some Series Containing  $\psi(x) - \psi(y)$  and  $(\psi(x) - \psi(y))^2$  for Certain Values of  $x$  and  $y$ . *Journal of Computational and Applied Mathematics*, 37:125–141, 1991.
- [87] D. Epstein, S. Levy, and R. de la Llave. About This Journal. *Experimental Mathematics*, 1, 1992.
- [88] David Epstein and Sylvio Levy. Experimentation and Proof in Mathematics. *Notices of the American Mathematical Society*, June, 1995.
- [89] K. Anders Ericsson and Neil Charness. Expert Performance: Its Structure and Aquisition. *American Psychologist*, 49:725–747, 1994.
- [90] J. Fauvel, R. Floyd, and R. Wilson, editors. *Oxford Figures: 800 Years of the Mathematical Sciences*. Oxford University Press, Oxford, 2000.
- [91] Claire Ferguson. *Helaman Ferguson: Mathematics in Stone and Bronze*. Meridian Creative Group, Erie, PA, 1994.

- [92] Helaman R. P. Ferguson, David H. Bailey, and Stephen Arno. Analysis of PSLQ, An Integer Relation Finding Algorithm. *Mathematics of Computation*, 68:351–369, 1999.
- [93] Helaman R. P. Ferguson and Rodney W. Forcade. Generalization of the Euclidean Algorithm for Real Numbers to All Dimensions Higher Than Two. *Bulletin of the American Mathematical Society*, 1:912–914, 1979.
- [94] Stephen Finch. *Mathematical Constants*. Cambridge University Press, Cambridge, 2003.
- [95] Carl F. Gauss. *Disquisitiones Arithmeticae*. Springer-Verlag, Heidelberg, 1986.
- [96] James Gleick. *Chaos: The Making of a New Science*. Viking Penguin, New York, 1987.
- [97] James Gleick. *Genius: The Life and Science of Richard Feynman*. Pantheon Books, New York, 1992.
- [98] Kurt Gödel. *Some Basic Theorems on the Foundations*, volume 3. Oxford University Press, Oxford, 1986–2003.
- [99] L. J. Goldstein. A History of the Prime Number Theorem. *American Mathematical Monthly*, 80:599–615, 1973.
- [100] Judith V. Grabiner. Is Mathematical Truth Time-Dependent? In Thomas Tymoczko, editor, *New Directions in the Philosophy of Mathematics*. Princeton University Press, Princeton, 1998.
- [101] Ronald L. Graham and Henry O. Pollak. Note on a Nonlinear Recurrence Related to  $\sqrt{2}$ . *Mathematics Magazine*, 43:143–145, 1970.
- [102] Andrew Granville. The Arithmetic Properties of Binomial Coefficients. In *The Organic Mathematics Project Proceedings*, volume 20, Ottawa, 1997. Canadian Mathematical Society.
- [103] Andrew Granville. Prime Possibilities and Quantum Chaos. <http://www.msri.org/ext/Emissary/EmissarySpring02.pdf>, 2002.
- [104] Ivor Grattan-Guinness, editor. *From the Calculus to Set Theory, 1630–1910. An Introductory History*. Duckworth, London, 1980.

- [105] D. F. Griffiths and J. M. Sanz-Serna. On the Scope of the Method of Modified Equations. *SIAM Journal of Scientific and Statistical Computing*, 7:994–1008, 1986.
- [106] Francois Guénard and Henri Lemberg. *La Méthode Expérimentale en Mathématiques*. Scopos, Springer-Verlag, Heidelberg, 2001.
- [107] Jacques Hadamard. *The Psychology of Invention in the Mathematical Field*. Dover, New York, 1945.
- [108] Godfrey H. Hardy. *A Mathematician's Apology*. Oxford University Press, Oxford, 1971.
- [109] Godfrey H. Hardy. *Ramanujan*. Chelsea, New York, 1978.
- [110] Godfrey H. Hardy and Edward M. Wright. *An Introduction to the Theory of Numbers*. Oxford University Press, Oxford, 1985.
- [111] J. Hastad, B. Just, J. C. Lagarias, and C. P. Schnorr. Polynomial Time Algorithms for Finding Integer Relations Among Real Numbers. *SIAM Journal of Computing*, 18:859–881, 1989.
- [112] Brian Hayes. The Easiest Hard Problem. *American Scientist*, 90(2):113–117, 2002.
- [113] J. Herman, R. Kučera, and J. Šimša. *Equations and Inequalities: Elementary Problems and Theorems in Number Theory*, volume 1. CMS Books, Springer-Verlag, Heidelberg, 2000.
- [114] Reuben Hersh. Fresh Breezes in the Philosophy of Mathematics. *American Mathematical Monthly*, 102:589–594, 1995.
- [115] Reuben Hersh. *What Is Mathematics, Really?* Oxford University Press, Oxford, 1999.
- [116] Arend Heyting. *Intuitionism: An Introduction*. North-Holland, Amsterdam, 1956.
- [117] Yozo Hida, Xiaoye S. Li, and David H. Bailey. Algorithms for Quad-Double Precision Floating Point Arithmetic. In *Proceedings of the Fifteenth Conference on Computer Arithmetic*, Washington, 2001. IEEE Computer Society.
- [118] T. P. Hill. The First-Digit Phenomenon. *American Scientist*, 86:358, 1998.



- [119] Georges Ifrah. *The Universal History of Numbers*. John Wiley and Sons, Hoboken, NJ, 2000.
- [120] A. Jaffe and F. Quinn. Theoretical Mathematics: Towards a Cultural Synthesis of Mathematics and Theoretical Physics. *Bulletin (New Series) of the American Mathematical Society*, 29:1–13, 1993.
- [121] A. Jaffe and F. Quinn. Response to Comments on “Theoretical Mathematics”. *Bulletin (New Series) of the American Mathematical Society*, 30:208–211, 1994.
- [122] D. J. Jeffrey, D. E. G. Hare, and R. M. Corless. Unwinding the Branches of the Lambert  $W$  Function. *Mathematical Sciences*, 21:1–7, 1996.
- [123] George Johnson. A World of Doughnuts and Spheres. *New York Times*, Apr. 20 2003.
- [124] A. P. Juskevic, E. Winter, and P. Hoffmann. Leonhard Euler; Christian Goldbach: Letters 1729-1764. *Notices of the German Academy of Sciences at Berlin*, 1965.
- [125] John Kadvany. *Imre Lakatos and the Guises of Reason*. Duke University Press, Durham, NC, 2001.
- [126] Y. Kanada, Y. Ushiro, and H. Kuroda. The Computation of Pi to 1,241,100,000,000 Decimal Digits. to appear.
- [127] A. Karatsuba and Yu Ofman. Multiplication of Many-Digital Numbers by Automatic Computers. *Doklady Akademii Nauk SSSR*, 145:293–294, 1962.
- [128] Kiran Kedlaya, Bjorn Poonen, Ravi Vakil, and Kiran S. Kedlaya. *The William Lowell Putnam Mathematical Competition 1985–2000: Problems, Solutions, and Commentary*. Mathematical Association of America, Washington, 2002.
- [129] John M. Keynes. Newton the Man. In James R. Newman, editor, *The World of Mathematics*, volume I. Simon and Schuster, New York, 1956.
- [130] E. Klarreich. Prime Time. *New Scientist*, Nov. 11 2000.
- [131] Donald E. Knuth. *The Art of Computer Programming*, volume 2. Addison-Wesley, Boston, 1998.
- [132] Gina Kolata. Does Gödel’s Theorem Matter to Mathematics? *Science*, 218:779–780, Nov. 19, 1982.

- [133] Charles Krauthammer. Paul Erdős, Sweet Genius. *Washington Post*, page A25, Sept. 27, 1996.
- [134] Thomas S. Kuhn. *The Structure of Scientific Revolutions*. University of Chicago Press, Chicago, 1970.
- [135] L. Kuipers and H. Niederreiter. *Uniform Distribution of Sequences*. Wiley-Interscience, Boston, 1974.
- [136] Gilbert Labelle. Sur l'Inversion et l'Itération Continue des Séries Formelles. *European Journal of Combinatorics*, 1:113–138, 1980.
- [137] Jeff Lagarias. The  $3x+1$  Problem and Its Generalizations. *American Mathematical Monthly*, 92:3–23, 1985.
- [138] Imre Lakatos. *Proofs and Refutations: The Logic of Mathematical Discovery*. Cambridge University Press, Cambridge, 1977.
- [139] Imre Lakatos. *The Methodology of Scientific Research Programmes: Philosophical Papers*, volume 1. Cambridge University Press, 1978.
- [140] Clement Lam. The Search for a Finite Projective Plane of Order 10. In *The Organic Mathematics Project Proceedings*, volume 20, Ottawa, 1997. Canadian Mathematical Society.
- [141] Leonard Lewin. *Polylogarithms and Associated Functions*. North-Holland, Amsterdam, 1981.
- [142] Shangzhi Li, Falai Chen, Yaohua Wu, and Yunhua Zhang. *Mathematics Experiments*. World Scientific Press, Singapore, 2003.
- [143] John E. Littlewood. *Littlewood's Miscellany*. Cambridge University Press, Cambridge, second edition, edited by Béla Bollobás edition, 1988.
- [144] L. Lovász, J. Pelikán, and K. Vesztergombi. *Discrete Mathematics, Elementary and Beyond*. Springer-Verlag, Heidelberg, 2003.
- [145] David T. Lykken. The Genetics of Genius. In Andrew Steptoe, editor, *Genius and the Mind: Studies of Creativity and Temperament*. Oxford University Press, Oxford, 1998.
- [146] Desmond MacHale. *Comic Sections*. Boole Press, Dublin, 1993.
- [147] D. Mackenzie. Homage to an Itinerant Master. *Science*, 275:759, 1997.

- [148] Mason Macklem. Multidimensional Modelling of Image Fidelity Measures. <http://www.cecm.sfu.ca/Preprints03/preprints03.html>, 2003.
- [149] Norman MacRae. *John von Neumann: The Scientific Genius Who Pioneered the Modern Computer, Game theory, Nuclear Deterrence, and Much More*. American Mathematical Society, Providence, 1992.
- [150] Trevor Major. Solomon's Bronze Basin and Pi. <http://www.apologeticspress.org/abdiscr/abdiscr33.html>, 2002.
- [151] Benoit B. Mandelbroit. *The Fractal Geometry of Nature*. W. H. Freeman, New York, 1988.
- [152] Yann Martel. *Life of Pi*. Harcourt, New York, 2002.
- [153] P. B. Medawar. *Advice to a Young Scientist*. Harper Colophon, New York, 1981.
- [154] Stephan Mertens. A Physicist's Approach to Number Partitioning. *Theoretical Computer Science*, 265(1-2):79–108, 2001.
- [155] Gordon E. Moore. Cramming More Components onto Integrated Circuits. *Electronics*, 38(8):114–117, 1965.
- [156] David Mumford. *Personal communication*, 2003.
- [157] Ivan Niven. A Simple Proof That Pi Is Irrational. *Bulletin of the American Mathematical Society*, 53:509, 1947.
- [158] Andrew Odlyzko. The  $10^{22}$ -nd Zero of the Riemann Zeta Function. In M. van Frankenhuysen and M. L. Lapidus, editors, *Dynamical, Spectral, and Arithmetic Zeta Functions*, pages 193–144. American Mathematical Society, Providence, 2001.
- [159] Joseph O'Donnell. *Avatars of the Word: From Papyrus to Cyberspace*. Harvard University Press, Cambridge, 1998.
- [160] C. D. Olds, Anneli Lax, and Giulian Davidoff. The Geometry of Numbers. In *Anneli Lax New Mathematics Library*, volume 41. Mathematical Association of America, Washington, 2000.
- [161] Atanasios Papoulis. *Signal Analysis*. McGraw-Hill, New York, 1977.

- [162] Roger Penrose. *The Emperor's New Mind*. American Philological Association, New York, 1989.
- [163] Max Planck and Frank Gaynor. *Scientific Autobiography and Other Papers*. Philosophical Library, New York, 1949.
- [164] George Polya. On Picture-Writing. *American Mathematical Monthly*, 63:689–697, 1956.
- [165] George Polya. *Mathematical Discovery: On Understanding, Learning and Teaching Problem Solving* (Combined Edition). John Wiley, Hoboken, NJ, 1981.
- [166] George Polya. *Mathematics and Plausible Reasoning, Volume I and II*. Princeton University Press, Princeton, 1990 (orig. published 1968).
- [167] Paul Preuss. Mathematics: The Statue in the Stone. *Berkeley Lab Research Review*, 23:12–13, Fall 2000.
- [168] S. D. Rabinowitz and S. Wagon. A Spigot Algorithm for Pi. *American Mathematical Monthly*, 102:195–203, 1995.
- [169] Ed Regis. *Who Got Einstein's Office?* Addison-Wesley, Boston, 1986.
- [170] Adrian Rice. What Makes a Great Mathematics Teacher? *American Mathematical Monthly*, 106:540–542, 1999.
- [171] Bernhard Riemann. Über die Anzahl der Primzahlen unter einer Gegebenen Grösse. In Hermann Weyl, editor, *Das Kontinuum und Andere Monographien*. Chelsea Publishing Co., New York, 1972.
- [172] Sara Robinson. Mathematician Fills In a Blank for a Fresh Insight on Art. *New York Times*, Jul. 30 2002.
- [173] K. Roth. Rational Approximations to Algebraic Numbers. *Mathematika*, 2:1–20, 1955.
- [174] Zeev Rudnick and Peter Sarnak. The  $n$ -Level Correlations of L-Functions. *Comptes Rendus de l'Académie Bulgare des Sciences*, 319:1027–1032, 1994.
- [175] Bertrand Russell. *The Study of Mathematics: Philosophical Essays*. London, 1910.
- [176] Carl Sagan. *Contact*. Doubleday, New York, 1986.

- [177] George Santayana. *Persons and Places*. Scribner, New York, 1944.
- [178] Peter Shor. Algorithms for Quantum Computation: Discrete Logarithms and Factoring. *Proceedings of the 35th Annual Symposium on Foundations of Computer Science*, pages 124–134, 1994.
- [179] Elaine Simmt and Brent Davis. Fractal Cards: A Space for Exploration in Geometry and Discrete Mathematics. *Mathematics Teacher*, 91(2):102–108, 1998.
- [180] Stephen G. Simpson. Unprovable Theorems and Fast-Growing Functions. *Contemporary Mathematics*, 65:359–394, 1987.
- [181] N. J. A. Sloane and Simon Plouffe. *The Encyclopedia of Integer Sequences*. Academic Press, New York, 1995.
- [182] Boris Stoicheff. *Gerhard Herzberg: An Illustrious Life in Science*. NRC Press, Ottawa, 2002.
- [183] R. Stoneham. A General Arithmetic Construction of Transcendental Non-Liouville Normal Numbers from Rational Fractions. *Acta Arithmetica*, 16:239–253, 1970.
- [184] Karl R. Stromberg. *An Introduction to Classical Real Analysis*. Wadsworth, Thompson Learning, Florence, KY, 1981.
- [185] George Szpiro. Does the Proof Stack Up? *Nature*, 424:12–13, Jul. 3, 2003.
- [186] William P. Thurston and Silvio Levy (editor). *Three-Dimensional Geometry and Topology*. Princeton University Press, Princeton, 1997.
- [187] John W. Tukey. The Future of Data Analysis. *Annals of Mathematical Statistics*, 33:1–67, 1962.
- [188] Herbert W. Turnbull. *The Great Mathematicians*. Methuen, Methuen, MA, 1929.
- [189] Charles Van Loan. *Computational Frameworks for the Fast Fourier Transform*. SIAM, 1992.
- [190] E. Roy Weintraub and Ted Gayer. Equilibrium Proofmaking. *Journal of the History of Economic Thought*, 23(4):421–442, 2001.
- [191] Asia Weiss. *Personal communication*, 2003.

- [192] Stephen Wolfram. Geometry of Binomial Coefficients. *American Mathematical Monthly*, 91:566–571, 1984.
- [193] Stephen Wolfram. *A New Kind of Science*. Wolfram Media, Inc., Champaign, IL, 2002.
- [194] Don Zagier. The First 50 Million Prime Numbers. *The Mathematical Intelligencer*, 0:7–19, 1977.
- [195] D. Zeilberger. Identities in Search of Identity. *Theoretical Computer Science*, 117:23–38, 1993.
- [196] D. Zeilberger. Theorems for a Price: Tomorrow's Semi-Rigorous Mathematical Culture. *Notices of the American Mathematical Society*, 40:978–981, 1993.
- [197] Peter Zvengrowski and Filip Saidak. On the Modulus of the Riemann Zeta Function in the Critical Strip. *Preprint, University of Calgary*, 2002.



# Index

- Abel, Niels Henrik, 11
- Achilles, Alf-Christian, 32
- Adamchik, Victor, iv
- aeronautical engineering, 2
- AGM inequality, 236
- Al-Kāshī, 121, 123
- Al-Khowarizmi, 123
- Amazon.com, 111
- Andrews, George, 239Q
- Apostol, Tom, 85, 301
- Archimedes, 48, 123
  - calculation of  $\pi$ , 118
- Aristotelian experiment, 288
- arithmetic-geometric mean (AGM),  
**15**, 225
  - calculation of  $\pi$ , 126–127
  - high-precision arithmetic, 256
- Arndt, Jörg, 163Q
- Arrow, Kenneth, 14Q
- Aryabhata, 123
- Au-Yeung, Enrico, 67, 278
  
- Baconian experiment, 288
- Bailey-Crandall hypothesis, 171
- Ballantine’s series, 154
- Banchoff, Thomas, 301
- Barnsley, Michael, 57
- Barwise, Jon, 2
- Bauschke, Heinz, iv
- BBP-type formulas, 138, 145–150
  - BBP sequence, 171
  - normal numbers, 169–174
- A Beautiful Mind*, 132
- Bellard, Fabrice, 142, 145
- Benford, Frank, 169
  - Benford’s principle, 169
- Berlinski, David, 1Q
- Berry, Michael, 85Q
- Bert, Paul, 22
- Besicovitch, Abram S., 94
- Bessel functions
  - Euler’s constant, 157
- beta function, **217**
  - normal numbers, 190
- Bianchii group, 101
- biology, 2
- Bohr, Harald, 110Q
  - Bohr-Mollerup theorem, 216, 230
- Bollobás, Béla, vi, 94, 110
- Bolzano, Bernhard, 11
- Borel, Emile
  - Borel’s normal number theorem, 186
  - Borel-Cantelli lemma, 174, 186–187
- Borrell, Lee, 34
- Borwein cubic/quartic iteration, *see*  $\pi$ ,  
algorithms for computing
- Borwein, David, 150, 302
- Borwein, Peter, iv, 127, 136, 138
- Bouyer, Martine, 130, 155
- Bradley, David, iv



- Brahmagupta, 123  
 Brent, Richard, 126, 155, 255  
     Brent–MacMillan algorithm, 157  
 Briggs, Henry, 87Q  
 Broadhurst, David, 9–10, 66, 69, 302  
 Brouwer, Luitzen Egbertus Jan, 127, 270Q  
 Buffon’s needle, 158  
  
 Cantor, Georg, 270  
 Carathéodory, Constantin, 205Q  
 Cardano, Girolamo, 205  
 Catalan, Eugène  
     Catalan numbers, 25  
     Catalan’s constant, 73, 145, 164, 267, 268  
 Cauchy, Augustin-Louis, 11  
     Cauchy-Goursat theorem, 200  
     Cauchy-Schwarz inequality, 213, 214  
 Cayley, Arthur  
     counting rooted trees, 25–26  
 cellular automata, 57  
 Cervone, David, 301  
 Césaro average, 92  
 Chaitin, Gregory, iv, 34, 283, 287Q  
     Chaitin’s constant, 164, 187, 193  
 Champernowne number, 168, 188–189  
 Chang Hong, 123  
 Chang, Jen, iv  
 Chang, Sun-Yung Alice, 45Q  
 chaos theory, *see* dynamical systems  
 Chretien, Jean, 197Q  
 Chu Shih-Chieh, 53  
 Chudnovsky, David and Gregory, iv, vi, 125–126, 130, 168, 303  
 civilization, 45Q  
 Clarke, Arthur C., 43Q  
 Clausen, Thomas, 124  
     Clausen functions, 102  
 Clay Mathematics Institute, 44, 67, 85  
 Cohen, Joel, 241  
 computer algebra systems, 2, 72  
     failures of, 202–206  
     proofs using, 4  
 constant recognition, 70, 75, 257  
     HJLS algorithm, 258  
     practical extensions, 262  
     PSLQ algorithm, 258–260  
     PSLQ, multipair, 260–262  
 constructivism, 87  
*Contact*, 132  
 continued fractions, 263  
     Euler’s constant, 269  
 convolutions, 242–244  
     cyclic convolution, **242**  
     linear convolution, **242**  
 Copeland-Erdős constant, **194**  
 Corless, Robert, iv, 202, 204  
 cosmology, 2  
 Costa, Celso, 302  
     Costa surface, 100, 113, 302  
 Coxeter, H. S. M. (Donald), 20, 97, 98Q, 99, 302, 303  
 Cramer, Harald, 78  
 Crandall, Richard, iv, 8, 33, 76, 91, 97, 170, 174, 178, 189, 207  
 cryptography, 38  
 Csicsery, George Paul, vi  
  
 Darwin, Charles, 13Q  
 Dase, Johan Zacharias, 124  
     calculation of  $\pi$ , 121  
 Davis, Martin, 270, 271Q  
 De Doelder, 68  
 de la Llave, Rafael, 277, 278Q  
 de la Vallée Poussin, 14  
 De Lagny, Thomas, 124

- de Morgan, Augustus, 18Q, 50
- de Smit, Bart, 34, 42
- Dedekind, Richard, 10, 37Q
- derangement, **232**
- Desmic configuration, 21
- Devlin, Keith, 2
- Dickson, Stewart, 34
- dilogarithm, 234
- Dirac, Paul, 21
- Dirichlet, Lejeune, 219
- discrete cosine transform (DCT), 269
- discrete Fourier transform (DFT), 242–244
- Duijvestijn's dissection, 94, 95
- dynamical systems, 60, 265–266
- Dyson, Freeman, 83–85
  
- economics, 2, 48
- Edison, Thomas, 51
- Einstein's brain, 98
- Eisenstein integers, 101
- elliptic integrals
  - complementary integrals, 224
  - complete, 133, 224
  - decimal expansion, 164
- Embree, Mark, 34
- ENIAC, 122, 153, 303
  - photograph, 153
- entanglement, 47
- Epstein, David, iiQ, 112, 277, 278Q
- equidistributed sequence, 171, **182**
- Erdős, Paul, 7, 20, 38, 81Q, 92, 194–195
  - Erdős Number Project, 195
  - photo, 195
- Escher, Maurits C., 42, 98
- Euclid
  - Elements, 22, 37
  - Euclidean algorithm, 257
  - prime numbers, 75
- Euler, Leonhard, 48, 119, 157
  - Euler numbers, 59
  - Euler sums, 67–69
    - quantum field theory, 69
  - Euler's constant, **157**, 164, 269
  - Euler-Maclaurin summation, 68, 211
  - Euler-Zagier sums, *see* Euler sums
  - gamma function, 219
  - Riemann zeta function, 113–115
- Experimental Mathematics* (journal), 2, 31, 113, 277, 279
- experimental mathematics, **3**, **298**
  
- factorization, 6, 47
- false proofs, 112
- Faraday, Michael, 43
- fast Fourier transform (FFT), 122, 131, 209–211, 242–244
  - high-precision arithmetic, 250–252
- Fateman, Richard, iv, 33, 92
- Fee, Greg, iv
- Feigenbaum, Mitchell, 62
  - Feigenbaum's constant, 62, 164
- Fenchel, Werner
  - Fenchel-Legendre conjugate, **214**
  - Fenchel-Young inequality, **213**
- Ferguson, Claire, 63
- Ferguson, D. F., 122, 130
- Ferguson, Helaman, iv, vi, 34, 41, 61, 63–67, 100, 258, 301, 302
  - sculptures, 34, 63–67, 100–106, 301, 302
- Ferguson, Samuel, 41
- Fermat's Last Theorem, 10, 112
- Feynman, Richard, 49, 51
  - Feynman diagrams, 69
- Fibonacci, Leonardo, 123

- Finch, Steven, iv, 34  
 finite attractor, 171, 182, 184  
 fluid dynamics, 2  
 Four Color Theorem, 8, 46, 112  
 fractal cards, 108  
 Frege, Gottlob, 11, 270  
 Fundamental Theorem of Algebra,  
     197–199  
  
 Galilean experiment, 288  
 Galois, Evariste, 195  
 Galway, William, iv, 150  
 Gamma function, **216**, 216–219  
     formulas, 218, 230  
     Psi function, 231  
     volumes of balls, 219  
 Gauss, Carl Friedrich, 14Q, 14–15, 48,  
     50, 75, 76Q, 109, 109Q, 131  
     arithmetic-geometric mean  
         (AGM), 14–15  
     Gauss-Kuz'min distribution, 264,  
         269  
     Prime Number Theorem, 75–76  
         prime number theorem, 14  
 Gauss, Eugene, 115  
 Gelfond-Schneider theorem, 87, 96  
 generating functions, 25  
 Gerriets, Marty, 131Q  
 Girgensohn, Roland, iv, 195, 278  
 Gödel, Kurt, 1Q, 270  
     Gödel statements, 39  
     Gödel's theorem, 8, 39, 41  
 Golay, Marcel, 7  
 Goldbach, Christian, 68, 113  
     Goldbach conjecture, 283  
 Goldston, Daniel, 81  
 Goodstein, Reuben  
     Goodstein sequences, 39  
     Goodstein's theorem, 40  
  
 Gosper, William, 125, 128, 130  
 Gourdon, Xavier, 34  
 Grabiner, Judith, 11  
 Graham, Ronald, iv, 74  
 Granville, Andrew, iv, 35  
 gravitational boosting, 43  
 Gray, Alfred, 100, 302  
 Gregory, James  
     Gregory's series, **57**, 59, 119  
     Gregory–Leibniz formula, 119,  
         158, 229  
 Griffeath, David, 35, 108Q  
 Grossman, Jerry, 35, 195  
 Guilloud, Jean, 130, 155  
  
 Hadamard, Jacques, 11Q, 14  
     prime number theorem, 12  
 Haenel, Christoph, iv, 163Q  
 Hales, Thomas, 35, 41  
 Hamming, Richard, 53Q  
 happy end problem, 92  
 Hardy, G. H., 12Q, 37Q, 48Q, 110, 270  
 Hayes, Brian, 46Q  
 Hepting, Darrell, 56  
 Hersh, Reuben, 46  
 Herzberg, Gerhard, 109  
 Heyting, Arend, 127  
 Hida, Yozo, iv  
 high-precision arithmetic, 245–257  
     advanced operations, 254–257  
     arbitrary precision, 247–249  
     double-double, 245–247  
     fast Fourier transform (FFT),  
         250–252  
     IEEE-754 standard, 245  
     Karatsuba Multiplication, 249–250  
     quad-double, 245–247  
*Hitchhiker's Guide to the Galaxy*, 133

- HJLS algorithm, *see* constant recognition
- Hoffman, David, 113, 302
- Hoffman, James, 302
- Hohmann transfer ellipses, 43
- Hölder's inequality, 213, 216, 218
- hot spot lemma, 176, 189
- hypergeometric function  
     calculation of  $\pi$ , 140  
     normal numbers, 189
- hypocycloid, **96**
- Indo-Arabic number system, 118
- integer relation algorithm, 60–61, 73, 257  
     calculation of  $\pi$ , 140  
     LLL, 61  
     polynomials, 60  
     PSLQ, 61
- intuitionism, 87
- Jaffe, Arthur, 289, 290Q
- Jeffrey, Dave, 202, 204, 206Q
- Jeffrey, David, iv
- Joint Photographic Experts Group (JPEG), 269
- The Joy of Pi*, 132
- Joyce, David, 35, 37
- Joyce, Jeff, iv
- JPEG, *see* Joint Photographics Experts Group
- Julia set, 303
- Julia sets, ii, 37, 266
- Takeya's problem, 94
- Kamata, Yoshikiyo, 124
- Kanada, Yasumasa, 8–9, 35, 74, 127–130, 132  
     frequency of digits of  $\pi$ , 164
- Kantian experiment, 287
- Keynes, John Maynard, 43Q, 49Q, 115Q
- Khintchine's constant, 164, 264
- Klein, Ester, 92
- Klein, Felix, 18, 19
- knots, 106, 303
- Knuth, Donald, 35, 138
- Kuhn, Thomas, 13Q
- Lakatos, Imre, 46, 47Q, 280Q, 282Q
- Lam, Clement, 4
- Lambert, Johann  
     Heinrich  
         Lambert  $W$  function, 303
- Lambert, Johann Heinrich, 122
- Land, Edwin, 51
- Landen, John, 234, 235
- Laplace transform, **220**
- Law and Order*, 133
- Legendre, Adrien-Marie, 122, 224  
     Legendre's relation, 133, **224**
- Lehmann, W., 124
- Lehmer's polynomial, 9
- Lehrer, Tom, 42Q
- Leibniz, Gottfried Wilhelm, 119, 270Q
- Lenstra, Hendrik, 34, 42
- Lenstra-Lenstra-Lovasz algorithm, *see* LLL
- Levy, Silvio, iiQ, 277, 278Q
- Lewin, Leonard, 102, 103, 235, 235Q
- Lewis, Adrian, iv
- Li, Xiaoye, iv
- Life of Pi*, 132, 162Q
- Liljedahl, Peter, 48
- Lindemann, Ferdinand von, 122
- Liouville, Joseph  
     Liouville numbers, 187  
     Liouville's theorem, 193, 200
- Lisoněk, Petr, iv

- Littlewood, John Edensor, 7, 76, 94,  
     96Q, 110  
 Liu Hui, 123  
 LLL, 61, 257  
 log sine integral, 102  
 logistic iteration, **60**, 60–62  
 Lucas numbers, 268  
 Luke, Russell, iv  
 Lykken, David, 50  
  
 Machin, John, 119, 124  
     Machin formula, 119, 155  
 Macklem, Mason, iv  
 Madelung's constant, 164  
 Madhava of Sangamagramma, 121, 123  
 Mandelbrot, Benoit, 55, 168, 303  
     Mandelbrot set, 37  
 Mann iteration, 92  
 Martel, Yann, *see* Life of Pi  
 Mathematics  
     humanist philosophy of, 46  
     standards of rigor in, 46  
*Matrix Reloaded*, 133  
 Matsunaga, Yoshisuke, 124  
 Mayer, Ernst, 8  
 mean iteration, 227  
     Archimedian iteration, 228  
     Gaussian iteration, 227  
     logarithmic mean, 227  
 Medawar, P.B., 287, 287, 288Q  
 Meeks, William, 302  
 Merit Factor problem, 7  
 Mersenne primes, 5  
 meteorology, 48  
 Millennium Prize, 5, 44  
 Milnor, John, 12Q  
 Minovitch, Michael, 43  
 Monte Carlo simulation, 160  
 Montgomery, Hugh  
     Riemann zeta function, 81–85  
 Moody, Robert, 98Q  
 Moore, Gordon, 4  
     Moore's Law, i, 2, 4  
 Morin, Mathew, iv  
 Mueller, Julia, 81  
 multivalued function, 204  
 Mumford, David, iv, 99Q  
  
 Napier, John, 87  
 Nemeth, Tomas, 100, 302  
 Newcomb, Simon, 169  
 Newton, Isaac, 48, 49, 51, 117Q, 119,  
     124  
     calculation of  $\pi$ , 119–121  
     Newton's method, 198, 254–256,  
     303  
 Niven, Ivan, 159  
 normal number, **163**, 164–186,  
     293–294  
     absolutely normal, **164**  
     addition, 188  
     multiplication, 187  
     topological properties, 187  
 North, Joseph Roy, 57  
  
 O'Donnell, James, 111Q  
 Odlyzko, Andrew, iv, vi, 83  
 Otho, Valentius, 123  
  
 P versus NP problem, 5–6  
 Papadopoulos, Jason, 8  
 parallel computer platforms, 4  
 Paris-Harrington theorem, 39  
 Parseval's equation, 67  
 Pascal, Blaise  
     Pascal's triangle, 53–57, 301  
     binomial coefficients, 55  
 Penrose, Roger, 127  
 perceptual dipper effect, 269

- Percival, Colin, 8, 142, 145  
 Perelman, Grigori, 44  
 periodic attractor, 182, 184  
 Peters, Klaus, iv, 113  
 physics, 2  
*Pi* (movie), 132  
 $\pi$   
   algorithms for calculating, 118  
     BBP algorithm, 136–143  
     Borwein quartic algorithm, 127, 133–136  
     early algorithms, 119–121  
     Salamin–Brent algorithm, 126–127, 255  
   algorithms for computing  
     spigot algorithm, 159  
   Biblical references, 117, 155  
   calculations of, 8–9  
   continued fraction, 125  
   early estimates, 117  
   formulas for calculating, 122, 125, 126, 128, 131, 137, 140, 145, 153, 154  
   frequency of digits, 164  
   irrationality, 121, 159  
   popular media, 132–133, 162  
   strings of digits, 127  
   transcendental, 121  
 Pick’s theorem, 19–20  
 Planck, Max, 13Q  
 platonism, 270  
 Plouffe, Simon, 59, 136, 138  
 Poincaré, Henri, 270Q  
   Poincaré Conjecture, 44, 112  
 Pollak, Henry, 74  
 Polya, George, iQ  
   pictorial representations, 22–26  
 polygamma function, *see* Psi function  
 polylogarithms  
   Euler sums, 69  
 Pomerance, Carl, 76, 178, 189  
 Pope John-Paul II, 111  
 primality checking, 6  
 prime counting function  $\pi(x)$ , 76, 81  
 prime number theorem, 14, 75–76  
 primitive prime factor, 151  
 principal-valued function, 204  
 probability, 48  
 projective plane, 4  
 proper power, 151  
 pseudorandom generators, 161, 176  
 Psi function, 231  
 PSLQ, 61, 63, 140, 145, 149, 258–260, 302  
   calculation of  $\pi$ , 140  
   Feynman diagrams, 69, 71  
   multipair, 260–262  
   statement of algorithm, 258  
 psychology, 2  
 Ptolemy, 123  
*Q*-linear Machin-type BBP arctangent  
   formula, 151  
 quantum computing, 6, 7, 47  
 quantum physics, 2, 13, 48, 83  
 qubit, 47  
 Quinn, Frank, 289, 290Q  
 Rademacher, Hans, 112  
 Ramanujan, Srinivasa, 48, 131, 195, 302  
   calculation of  $\pi$ , 125–126  
   Catalan’s constant, 267  
   lost notebook, 238  
 Ramsey’s theorem, 38–39  
 Record, Robert, 13Q  
 Reese, George, 35

- Riemann zeta function, **67**  
 Euler sums, 69  
 Feynman diagrams, 71  
 Hurwicz zeta function, 264  
 modulus, 109  
 normal numbers, 190  
 normality of, 164  
 prime numbers, 78, 81
- Riemann, Georg Bernhard, 11, 113  
 Prime Number Theorem, 78  
 Riemann Hypothesis, 7, 78–80, 83, 85, 110, 112, 133, 285, 287
- Robinson, John, 302
- Romanus, Adrianus, 123
- Roth's theorem, 175
- Rudolph, Daniel, iv
- Russell, Bertrand, 11, 37, 98, 98Q, 270
- Rutherford, W., 124
- Salamin, Eugene, 126, 155, 255
- Santayana, George, 116Q
- Sarnak, Peter, 44Q
- Scharein, Rob, iv, 19
- Schwalbe, Dan, 100, 302
- Schwarz inequality, 207, 208
- Sebah, Pascal, 34
- Seki Kowa, 124
- Sendov, Hristo, iv
- Serrano, Luis, iv
- Shanks, Daniel, 127, 130
- Shanks, William, 119, 122, 124
- Sharp, Abraham, 124
- Shor, Peter, 6, 47
- Siddhanta, 123
- Siegel, Carl Ludwig, 78
- Sierpinski, Waclaw, 187  
 Sierpinski cube, 36  
 Sierpinski triangle, 55, 57–58, 108  
 Sierpinski's number, 191
- simple connectedness, 44
- Simpson's paradox, 289
- The Simpsons*, 132, 134
- Sims, Brailey, vi
- singular value, 133
- Skewes number, 76
- Sloane, Neil, iv, 34, 59
- sociology, 2
- Square Root Bug, 202
- Star Trek*, 132
- Stephanos, 21
- Stirling, James, 14  
 Stirling's formula, **221**, 222–223
- Stoicheff, Boris, 109
- Stoneham, R. G., 177
- Störmer, F.C.W., 128
- strong law of large numbers, 186
- supernovas, 2
- Sylvester, James, 20–22  
 Sylvester's theorem, 19–21
- Szekeres, George, 92
- Takano, K., 128
- Takebe, Katahiro, 124
- tangent numbers, 59
- Tartaglia, Niccolo Fontana, 205
- Thurston, Bill, 113Q
- topology, experimental, 106
- trapezoidal formula, 222
- Trefethen, Nick, vi, 26, 34
- trilogarithm, 234
- Tsu Chung-Chih, 118, 121, 123
- Tukey, J. W., 241Q
- Turing, Alan, 1, 270, 271  
 Turing machines  
 halting problem for Turing machines, 193
- Ulam, Stanislaw, 160

- uncertainty principle, 206–213
- Van Ceulen, Ludolph, 123
- Vertesi, Janet, iv
- Vierling, Angela, 18, 36
- Viète, Francois, 123
  - Viète’s product, 236
- von Neumann, John, 1, 48, 270, 271
  - calculation of  $\pi$ , 122
- Von Vega, Georg F., 124
  
- Wagon, Stan, 100, 160, 302
- Wang Fau, 123
- Weeks, Jeff, 36, 106Q
- Weiss, Asia, iv, vi, 100Q
- Weisstein, Eric, 32
- Wiles, Andrew, 10
- Wilf-Zeilberger algorithm, 276, 282
- Will Roger’s phenomenon, 289
- Witten, Edward, 111
- Wolfram, Stephen, 56
- Wrench, John W., 130
- Wright, Tim, 34
  
- Z-linear Machin-type BBP arctangent formula, 151
- Zagier, Don, 76Q
- Zeilberger, Doron, 276, 283, 283Q, 285, 286Q
- Zucker, John, iv



## Preface

“Moreover a mathematical problem should be difficult in order to entice us, yet not completely inaccessible, lest it mock our efforts. It should be to us a guidepost on the mazy path to hidden truths, and ultimately a reminder of our pleasure in the successful solution. . . .

Besides it is an error to believe that rigor in the proof is the enemy of simplicity.”

David Hilbert, Paris International Congress, 1900 [210]

As we recounted in the first volume of this work, *Mathematics by Experiment: Plausible Reasoning in the 21st Century* [43], when we started our collaboration in 1985, relatively few mathematicians employed computations in serious research work. In fact, there appeared to be a widespread view in the field that “real mathematicians don’t compute.” In the ensuing years, computer hardware has skyrocketed in power and plummeted in cost, thanks to the remarkable phenomenon of Moore’s Law. In addition, numerous powerful mathematical software products, both commercial and noncommercial, have become available. But just as importantly, a new generation of mathematicians is eager to use these tools, and consequently numerous new results are being discovered.

The experimental methodology described in these books provides a compelling way to generate understanding and insight; to generate and confirm or confront conjectures; and generally to make mathematics more tangible, lively, and fun for both the professional researcher and the novice. Furthermore, the experimental approach helps broaden the interdisciplinary nature of mathematical research: a chemist, physicist, engineer, and a mathematician may not understand each others’ motivation or technical language, but they often share an underlying computational approach, usually to the benefit of all parties involved.

A typical scenario of using this experimental methodology is the following. Note the “dialogue” between human and computer, which is very typical of this approach to mathematical research:

1. Studying a mathematical problem to identify aspects that need to be better understood.

2. Using a computer to explore these aspects, by working out specific examples, generating plots, etc.
3. Noting patterns or other phenomena evident in the computer-based results that relate to the problem under study.
4. Using computer-based tools to identify or “explain” these patterns.
5. Formulating a chain of credible conjectures that, if true, would resolve the question under study.
6. Deciding if the potential result points in the desired direction and is worth a full-fledged attempt at formal proof.
7. Performing additional computer-based experiments to gain greater confidence in the key conjectures.
8. Confirming these conjectures by rigorous proof.
9. Using symbolic computing software to double-check analytical derivations.

Our goal in these books is to present a variety of *accessible* examples of modern mathematics where intelligent computing plays a significant role (along with a few examples showing the limitations of computing). We have concentrated primarily on examples from analysis and number theory, as this is where we have the most experience, but there are numerous excursions into other areas of mathematics as well (see the Table of Contents). For the most part, we have contented ourselves with outlining reasons and exploring phenomena, leaving a more detailed investigation to the reader. There is, however, a substantial amount of new material, including numerous specific results that have not yet appeared in the mathematical literature, as far as we are aware.

This work is divided into two volumes, each of which can stand by itself. The first volume, *Mathematics by Experiment: Plausible Reasoning in the 21st Century* [43], presents the rationale and historical context of experimental mathematics, and then presents a series of examples that exemplify the experimental methodology. We include in the first volume a reprint of an article co-authored by two of us that complements this material. This second volume, *Experimentation in Mathematics: Computational Paths to Discovery*, continues with several

chapters of additional examples. Both volumes include a chapter on numerical techniques relevant to experimental mathematics.

Each volume is targeted to a fairly broad cross-section of mathematically trained readers. Most of the first volume should be readable by anyone with solid undergraduate coursework in mathematics. Most of this volume should be readable by persons with upper-division undergraduate or graduate-level coursework. None of this material involves highly abstract or esoteric mathematics.

Some programming experience is valuable to address the material in this book. Readers with no computer programming experience are invited to try a few of our examples using commercial software such as *Mathematica* and *Maple*. Happily, much of the benefit of computational-experimental mathematics can be obtained on any modern laptop or desktop computer—a major investment in computing equipment and software is not required.

Each chapter concludes with a section of commentary and exercises. This permits us to include material that relates to the general topic of the chapter, but which does not fit nicely within the chapter exposition. This material is not necessarily sorted by topic nor graded by difficulty, although some hints, discussion and answers are given. This is because mathematics in the raw does not announce, “I am solved using such and such a technique.” In most cases, half the battle is to determine how to start and which tools to apply.

We are grateful to our colleagues Victor Adamchik, Heinz Bauschke, Peter Borwein, David Bradley, Gregory Chaitin, David and Gregory Chudnovsky, Robert Corless, Richard Crandall, Richard Fateman, Greg Fee, Helaman Ferguson, Steven Finch, Ronald Graham, Andrew Granville, Christoph Haenel, David Jeffrey, Jeff Joyce, Adrian Lewis, Petr Lisonek, Russell Luke, Mathew Morin, David Mumford, Andrew Odlyzko, Hristo Sendov, Luis Serrano, Neil Sloane, Daniel Rudolph, Asia Weiss, and John Zucker who were kind enough to help us prepare and review material for this book; to Mason Macklem, who helped with material, indexing (note that in the index definitions are marked in bold, and quotes with a suffix “Q”), and more; to Jen Chang and Rob Scharein, who helped with graphics; to Janet Vertesi who helped with bibliographic research; to Will Galway, Xiaoye Li, and Yozo Hida, who helped with computer programming; and to numerous others who have assisted in one way or another in this work. We owe a special debt of gratitude to Klaus Peters for urging us to write this book and for helping us nurse it into existence. Finally, we wish to acknowledge the assistance and the patience exhibited by our spouses and family members

during the course of this work.

Borwein's work is supported by the Canada Research Chair Program and the Natural Sciences and Engineering Council of Canada. Bailey's work is supported by the Director, Office of Computational and Technology Research, Division of Mathematical, Information, and Computational Sciences of the U.S. Department of Energy, under contract number DE-AC03-76SF00098.

Jonathan M. Borwein  
Simon Fraser University  
Burnaby, British Columbia V5A 1S6, Canada  
E-mail: [jborwein@cecm.sfu.ca](mailto:jborwein@cecm.sfu.ca)

David H. Bailey  
Lawrence Berkeley National Laboratory  
Berkeley, CA 94720, USA  
Email: [dhbailey@lbl.gov](mailto:dhbailey@lbl.gov)

Roland Girgensohn  
Technische Universität München  
85747 Garching bei München, Germany  
Email: [girgen@cecm.sfu.ca](mailto:girgen@cecm.sfu.ca)

August 2003

## Graphics

Several of the nicer graphics that appear in black-and-white in various chapters of this volume, as well as some from the first volume, have been rendered in the Color Supplement (Chapter 8) of the first volume.

## Experimental Mathematics Web Site

The authors have established a web site containing an updated collection of links to many of the URLs mentioned in the two volumes, plus errata, software, tools, and other web useful information on experimental mathematics. This can be found at the following URL:

<http://www.expmath.info>

# List of Figures

1.1	A pictorial proof of Archimedes' inequality . . . . .	3
1.2	Convergence of $s_n(1/x)$ to $g(x)$ . . . . .	9
2.1	The Gibbs phenomenon for Fourier series . . . . .	100
2.2	The Gibbs phenomenon for Fejér series . . . . .	100
2.3	25th iterate for $q = 1/\sqrt{2}$ . . . . .	107
2.4	25th iterate for $q = (\sqrt{5} - 1)/2$ . . . . .	107
2.5	25th iterate for $q = 2/3$ . . . . .	108
2.6	25th iterate for $q = 3/4$ . . . . .	108
2.7	Approximation $f_3$ to Lebesgue's function . . . . .	119
2.8	The oscillatory Fourier transforms $\widehat{f}_2, \widehat{f}_8, \widehat{f}_{16}$ . . . . .	135
3.1	How does one identify two knots? . . . . .	198
3.2	The knot $8_8$ and Reidemeister moves . . . . .	199
3.3	The knots $10_{161}$ and $10_{162}$ . . . . .	200
3.4	The knot equivalent to both $10_{161}$ and $10_{162}$ . . . . .	200
4.1	A Ferrer diagram . . . . .	202
4.2	Where $ \theta_2/\theta_3  > 1$ and where $ \theta_4/\theta_3  > 1$ (first quadrant) . . . . .	236
4.3	Approximations to $x^{x^{x^{\dots}}}$ . . . . .	240
4.4	Solutions to $x^{x^{x^{\dots}}}$ . . . . .	241
5.1	Zeros of zero-one polynomials . . . . .	276
5.2	A hyperbolicity cone . . . . .	280
5.3	A Young tableau along with all lattice paths of shape $(3,2)$ . . . . .	285
6.1	Trapping the zeros . . . . .	304

6.2	Real branches of Lambert's function that satisfy $W \exp W = x$ . . .	305
6.3	The Riemann surface for the Lambert function. . . . .	308
6.4	Which function $\phi_k$ is this? . . . . .	319
6.5	Which equation $\mathcal{E}_k$ generated this? . . . . .	320
6.6	Comparison of three direct search methods . . . . .	323
6.7	A nonconvex set with the fixed point property . . . . .	326
7.1	Newton-Julia set for $p(x) = x^3 - 1$ . . . . .	337
7.2	The pseudospectra of matrix $A(0.5)$ . . . . .	364
7.3	Daisy matrix pseudospectra . . . . .	365

# List of Tables

1.1	Error $ E_1(N, M, 0.9) $ for $N = 5 \times 10^k (1 \leq k \leq 4)$ and $0 \leq M \leq 6$	38
1.2	Error $ E_1(N, M, 1) $ for $N = 5 \times 10^k (1 \leq k \leq 6)$ and $0 \leq M \leq 6$	38
1.3	Error $ E_2(N + 1, M, 1) $ for $N = 5 \times 10^k (1 \leq k \leq 6)$ and $0 \leq M \leq 6$	38
1.4	Error $ E^*(N, M) $ for $N = 5 \times 10^k (1 \leq k \leq 2)$ and $0 \leq M \leq 6$	39
3.1	Values of $E_{n,k}$ for various $n$ and $k$ .	161
3.2	Values of $M_{n,k}$ for various $n$ and $k$ .	162
3.3	Values of $D_{n,k}$ for various $n$ and $k$ .	163





# Contents

<b>Preface</b>	<b>i</b>
<b>List of Figures</b>	<b>v</b>
<b>List of Tables</b>	<b>vii</b>
<b>1 Sequences, Series, Products and Integrals</b>	<b>1</b>
1.1 Pi is not $22/7$ . . . . .	2
1.2 Two Products . . . . .	5
1.3 A Recursive Sequence Problem . . . . .	8
1.4 High Precision Fraud . . . . .	12
1.5 Knuth's Series Problem . . . . .	17
1.6 Ahmed's Integral Problem . . . . .	19
1.7 Evaluation of Binomial Series . . . . .	23
1.7.1 The Case of Non-Negative $k$ . . . . .	25
1.7.2 Some Results for Negative $k$ . . . . .	30
1.8 Continued Fractions of Tails of Series . . . . .	32
1.8.1 Gregory's Series Reexamined . . . . .	33
1.8.2 Euler's Continued Fraction . . . . .	34
1.8.3 Gauss's Continued Fraction . . . . .	35
1.8.4 Perron's Continued Fraction . . . . .	40
1.9 Partial Fractions and Convexity . . . . .	41
1.10 Log-concavity of Poisson Moments . . . . .	46
1.11 Commentary and Additional Examples . . . . .	48
<b>2 Fourier Series and Integrals</b>	<b>77</b>
2.1 The Development of Fourier Analysis . . . . .	78

2.2	Basic Theorems of Fourier Analysis . . . . .	80
2.2.1	Fourier Series . . . . .	80
2.2.2	Fourier Transforms . . . . .	86
2.3	More Advanced Fourier Analysis . . . . .	89
2.3.1	The Poisson Summation Formula . . . . .	89
2.3.2	Convolution Theorems . . . . .	93
2.3.3	Summation Kernels . . . . .	94
2.4	Examples and Applications . . . . .	98
2.4.1	The Gibbs Phenomenon . . . . .	98
2.4.2	A Function with Given Integer Moments . . . . .	101
2.4.3	Bernoulli Convolutions . . . . .	102
2.5	Some Curious Sinc Integrals . . . . .	106
2.5.1	The Basic Sinc Integral . . . . .	109
2.5.2	Iterated Sinc Integrals . . . . .	111
2.6	Korovkin's Three Function Theorems . . . . .	113
2.7	Commentary and Additional Examples . . . . .	116
<b>3</b>	<b>Zeta Functions and Multi-Zeta Values</b>	<b>143</b>
3.1	Reflection and Continuation of Zeta . . . . .	144
3.1.1	The Riemann Hypothesis . . . . .	146
3.2	Special Values of the Zeta Function . . . . .	147
3.2.1	Zeta at Even Positive Integers . . . . .	147
3.2.2	Zeta at Odd Positive Integers . . . . .	149
3.2.3	A Taste of Ramanujan . . . . .	151
3.3	Other L-series . . . . .	153
3.4	Multi-Zeta Values . . . . .	155
3.4.1	Various Methods of Attack . . . . .	156
3.4.2	Reducibility and Dimensional Conjectures . . . . .	157
3.5	Double Euler Sums . . . . .	164
3.6	Duality Evaluations and Computations . . . . .	167
3.7	Proof of the Zagier Conjecture . . . . .	173
3.8	Extensions and Discoveries . . . . .	176
3.9	Multi-Clausen Values . . . . .	178
3.10	Commentary and Additional Examples . . . . .	180

<b>4</b>	<b>Partitions and Powers</b>	<b>201</b>
4.1	Partition Functions . . . . .	201
4.1.1	Euler’s Pentagonal Number Theorem . . . . .	203
4.1.2	Modular Properties of Partitions . . . . .	204
4.1.3	The “Exact” Formula for $p(n)$ . . . . .	206
4.2	Singular Values . . . . .	207
4.3	Crystal Sums and Madelung’s Constant . . . . .	213
4.3.1	Sums of Squares . . . . .	214
4.3.2	Multidimensional Sums . . . . .	216
4.3.3	Madelung’s Constant . . . . .	219
4.4	Some Fibonacci Sums . . . . .	221
4.5	A Characteristic Polynomial Triumph . . . . .	224
4.6	Commentary and Additional Examples . . . . .	227
<b>5</b>	<b>Primes and Polynomials</b>	<b>245</b>
5.1	Giuga’s Prime Number Conjecture . . . . .	245
5.1.1	Computation of Exclusion Bounds . . . . .	246
5.1.2	Giuga Sequences . . . . .	251
5.1.3	Lehmer’s Problem . . . . .	253
5.2	Disjoint Genera . . . . .	253
5.3	Gröbner Bases and Metric Invariants . . . . .	254
5.3.1	Formulation of the Polynomial System . . . . .	255
5.3.2	Gröbner Bases . . . . .	257
5.4	A Sextuple of Metric Invariants . . . . .	257
5.5	A Quintuple of Related Invariants . . . . .	260
5.5.1	Some Open Questions on Invariants . . . . .	262
5.6	Sloane’s Harmonic Designs . . . . .	262
5.7	Commentary and Additional Examples . . . . .	265
<b>6</b>	<b>The Power of Constructive Proofs II</b>	<b>289</b>
6.1	A More General AGM Iteration . . . . .	289
6.1.1	The Functional Equation for $A_N$ . . . . .	290
6.1.2	The Quadratic Case Recovered . . . . .	291
6.1.3	The Cubic Case Solved . . . . .	292
6.2	Variational Methods and Proofs . . . . .	295
6.3	Maximum Entropy Optimization . . . . .	299

6.4	A Magnetic Resonance Entropy . . . . .	301
6.5	Computational Complex Analysis . . . . .	302
6.6	The Lambert Function . . . . .	305
6.6.1	The Lambert Function and Stirling's Formula . . . . .	306
6.6.2	The Lambert Function and Riemann Surfaces . . . . .	307
6.6.3	The Lambert Function in Brief . . . . .	309
6.7	Commentary and Additional Examples . . . . .	311
<b>7</b>	<b>Numerical Techniques II</b>	<b>329</b>
7.1	The Wilf-Zeilberger Algorithm . . . . .	329
7.2	Prime Number Computations . . . . .	330
7.3	Roots of Polynomials . . . . .	334
7.4	Numerical Quadrature . . . . .	336
7.4.1	Gaussian Quadrature . . . . .	338
7.4.2	Error Function Quadrature . . . . .	340
7.4.3	Tanh-Sinh Quadrature . . . . .	343
7.4.4	Practical Considerations for Quadrature . . . . .	345
7.4.5	2-D and 3-D Quadrature . . . . .	346
7.5	Infinite Series Summation . . . . .	346
7.5.1	Computation of Multiple Zeta Constants . . . . .	349
7.6	Commentary and Additional Examples . . . . .	351
	<b>Bibliography</b>	<b>369</b>
	<b>Index</b>	<b>385</b>

# Chapter 1

## Sequences, Series, Products and Integrals

Several years ago I was invited to contemplate being marooned on the proverbial desert island. What book would I most wish to have there, in addition to the Bible and the complete works of Shakespeare? My immediate answer was: Abramowitz and Stegun's *Handbook of Mathematical Functions*. If I could substitute for the Bible, I would choose Gradsteyn and Ryzhik's *Table of Integrals, Series and Products*. Compounding the impiety, I would give up Shakespeare in favor of Prudnikov, Brychkov and Marichev's *Tables of Integrals and Series*. . . On the island, there would be much time to think about waves on the water that carve ridges on the sand beneath and focus sunlight there; shapes of clouds; subtle tints in the sky. . . With the arrogance that keeps us theorists going, I harbor the delusion that it would be not too difficult to guess the underlying physics and formulate the governing equations. It is when contemplating how to solve these equations—to convert formulations into explanations—that humility sets in. Then, compendia of formulas become indispensable.

Michael Berry, 2001 [23]

In the first volume, we presented numerous examples of experimental mathematics in action. In particular, we examined how a computational-experimental approach could be used to identify constants and sequences, evaluate definite

integrals and infinite series, discover new identities involving fundamental constants and functions of mathematics, provide a more intuitive approach to mathematical proofs, and formulate conjectures that can lead to important advances in the field. In this chapter, we introduce our discussion with a number of additional intriguing examples in the realm of sequences, series, products and integrals.

## 1.1 Pi is not 22/7

We first consider an example from the early history of  $\pi$ , as described in Section 2.1 of the first volume.

Even *Maple* or *Mathematica* “knows”  $\pi \neq 22/7$ , since

$$0 < \int_0^1 \frac{(1-x)^4 x^4}{1+x^2} dx = \frac{22}{7} - \pi, \quad (1.1.1)$$

though it would be prudent to ask “why” it can perform the evaluation and “whether” we should trust it?

Assume we trust it. Then the integrand is strictly positive on the interior of the interval of integration, and the answer in (1.1.1) is necessarily an area and thus strictly positive, despite millennia of claims that  $\pi$  is  $22/7$ . Of course,  $22/7$  is one of the early continued fraction approximations to  $\pi$ . The first four are  $3, 22/7, 333/106, 355/113$ .

In this case, computing the indefinite integral provides immediate reassurance. We obtain

$$\int_0^t \frac{x^4 (1-x)^4}{1+x^2} dx = \frac{1}{7} t^7 - \frac{2}{3} t^6 + t^5 - \frac{4}{3} t^3 + 4t - 4 \arctan(t). \quad (1.1.2)$$

This is easily confirmed by differentiation, and the Fundamental Theorem of Calculus substantiates (1.1.1).

In fact one can take this idea a bit further. We note that

$$\int_0^1 x^4 (1-x)^4 dx = \frac{1}{630}, \quad (1.1.3)$$

and we observe that

$$\frac{1}{2} \int_0^1 x^4 (1-x)^4 dx < \int_0^1 \frac{(1-x)^4 x^4}{1+x^2} dx < \int_0^1 x^4 (1-x)^4 dx. \quad (1.1.4)$$

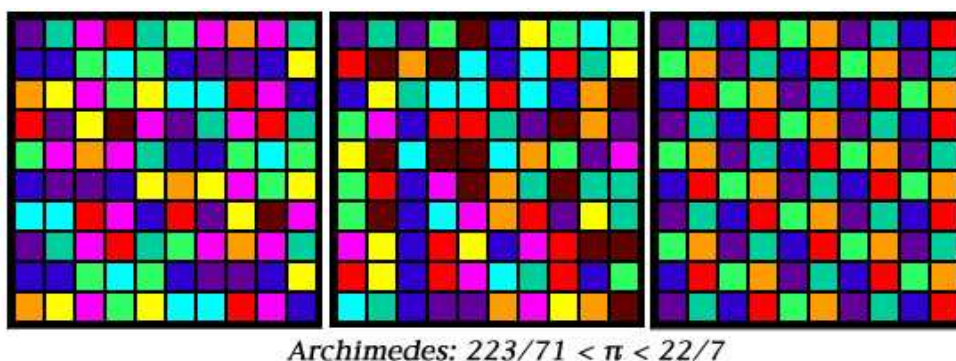


Figure 1.1: A pictorial proof of Archimedes' inequality

On combining this with (1.1.1) and (1.1.3), we straightforwardly derive  $223/71 < 22/7 - 1/630 < \pi < 22/7 - 1/1260 < 22/7$ , and so re-obtain Archimedes' famous computation

$$3\frac{10}{71} < \pi < 3\frac{10}{70} \quad (1.1.5)$$

(illustrating that it is sometimes better not to fully reduce a fraction to lowest terms).

This derivation of the estimate above seems first to have been written down in *Eureka*, the Cambridge student journal in 1971 [96]. The integral in (1.1.1) was apparently shown by Kurt Mahler to his students in the mid-1960s, and it had appeared in a mathematical examination at the University of Sydney in November, 1960. Figure 1.1 shows the estimate graphically illustrated. The three  $10 \times 10$  arrays color the digits of the first hundred digits of  $223/71$ ,  $\pi$ , and  $22/7$ . One sees a clear pattern on the right ( $22/7$ ), a more subtle structure on the left ( $223/71$ ), and a “random” coloring in the middle ( $\pi$ ).

It is tempting to ask if there is a clean general way to mimic (1.1.1) for more general rational approximations, or even continued fraction convergents. This is indeed possible to some degree as discussed by Beukers in [24]. The most satisfactory result is

$$a_n\pi - \frac{b_n}{c_n} = \int_0^1 \frac{t^{2n} (1-t^2)^{2n} ((1+it)^{3n+1} + (1-it)^{3n+1})}{(1+t^2)^{3n+1}} dt, \quad (1.1.6)$$

for  $n \geq 1$ , where the integers  $a_n, b_n$  and  $c_n$  are implicitly defined by the integral

in (1.1.6). The first three integrals evaluate to  $14\pi - 44$ ,  $968\pi - 45616/15$  and  $75920\pi - 1669568/7$ , so again we start with  $\pi - 22/7$ .

Unlike Beukers' preliminary attempts in [24], such as the seemingly promising

$$\int_0^1 \frac{t^n (1-t)^n}{(t^2+1)^{n+1}} dt,$$

this set of approximates actually produces an explicit if weak *irrationality estimate* [44, 24]: for large  $n$ ,

$$\left| \pi - \frac{p_n}{q_n} \right| \geq \frac{1}{q_n^{1.0499}}.$$

As Beukers sketches one consequence of this explicit sequence

$$\left| \pi - \frac{p}{q} \right| \geq \frac{1}{q^{21.04\dots}}$$

for all integers  $p, q$  with sufficiently large  $q$ . (Here  $21.04\dots = 1 + 1/0.0499$ . In fact, in 1993 Hata by different methods had improved the number 21.4 to 8.02.)

While it is easy to discover “natural” results like

$$\frac{1}{5} \int_0^1 \frac{x(1-x)^2}{(1+x)^3} dx = \frac{7}{10} - \log(2), \quad (1.1.7)$$

the fact that  $7/10$  is again a convergent to  $\log 2$  seems to be largely a happenstance. For example,

$$\begin{aligned} \int_0^1 \frac{x^{12}(1-x)^{12}}{16(1+x^2)} dx &= \frac{431302721}{137287920} - \pi \\ \int_0^1 \frac{x^{12}(1-x)^{12}}{16} dx &= \frac{1}{1081662400} \end{aligned}$$

leads to the true, if inelegant, estimate that  $5902037233/1878676800 < \pi < 224277414953/71389718400$ , where the interval is of size  $1.39 \cdot 10^{-9}$ .

In contrast to this easy symbolic success, *Maple* struggles with the following version of the *sophomore's dream*:

$$\int_0^1 \frac{1}{x^x} dx = \sum_{n=1}^{\infty} \frac{1}{n^n}. \quad (1.1.8)$$



When students are asked to confirm this, they most typically mistake numerical validation for symbolic proof:  $1.291285997 = 1.291285997$ . One seems to need to nurse a computer system, starting with integrating

$$x^{-x} = \exp(-x \log x) = \sum_{n=0}^{\infty} \frac{(-x \log x)^n}{n!},$$

term by term. See Exercise 3 at the end of this chapter.

## 1.2 Two Products

Consider the product

$$\prod_{n=2}^{\infty} \frac{n^3 - 1}{n^3 + 1} = \frac{2}{3}, \quad (1.2.9)$$

which has a rational value, and the seemingly simpler one (squares instead of cubes)

$$\prod_{n=2}^{\infty} \frac{n^2 - 1}{n^2 + 1} = \frac{\pi}{\sinh(\pi)}, \quad (1.2.10)$$

which evaluates to a transcendental number. *Mathematica* and *Maple* successfully evaluate such products, although not always in the same form. In this case, *Mathematica* produces expressions involving the gamma function, while *Maple* returns the values shown above. In either case, we learn little or nothing from the results, since the software typically cannot recreate the steps of validation. In such a situation, it often pays to ask our software to evaluate the finite products and then take limits. Note that in earlier versions of *Maple* or *Mathematica*, the infinite products would have been returned unevaluated, so that we may have been led directly to the finite products. Nowadays the system knows more, but we often learn less! To use a modern educational term, we are not led to “unpack” the concepts.

When asked to evaluate the finite products *Maple* returns expressions involving Gamma function values. For the first product (1.2.9), this expression can be simplified to

$$\prod_{n=2}^N \frac{n^3 - 1}{n^3 + 1} = \frac{2}{3} \frac{N^2 + N + 1}{N(N + 1)},$$

and from this, one may get the idea that the evaluation can be done by telescoping. This directly leads to the following proof, which just consists of filling in the intermediate steps *Maple* still does not care to tell us:

$$\begin{aligned} \prod_{n=2}^N \frac{n^3 - 1}{n^3 + 1} &= \prod_{n=2}^N \frac{(n-1)(n^2 + n + 1)}{(n+1)(n^2 - n + 1)} = \frac{\prod_{n=0}^{N-2} (n+1)}{\prod_{n=2}^N (n+1)} \cdot \frac{\prod_{n=2}^N (n^2 + n + 1)}{\prod_{n=1}^{N-1} (n^2 + n + 1)} \\ &= \frac{2}{N(N+1)} \cdot \frac{N^2 + N + 1}{3} \rightarrow \frac{2}{3}. \end{aligned}$$

The second finite product does not simplify in any helpful way; however, the Gamma function expression, together with the *Maple* evaluation of the infinite product, gives us the hint that the sin-product formula

$$\sin(\pi x) = \pi x \prod_{n=1}^{\infty} \left(1 - \frac{x^2}{n^2}\right), \quad (1.2.11)$$

which we met in Chapter 5 of the first volume, plays a role here. With this idea, the proof of the evaluation is simple: By complexification (and holomorphy), it follows from (1.2.11) that

$$\frac{\sinh(\pi)}{\pi} = 2 \prod_{n=2}^{\infty} \frac{n^2 + 1}{n^2},$$

and we get

$$\frac{\sinh(\pi)}{\pi} \cdot \prod_{n=2}^{\infty} \frac{n^2 - 1}{n^2 + 1} = 2 \prod_{n=2}^{\infty} \frac{n^2 - 1}{n^2} = 1,$$

since the final product is again telescoping.

Do these evaluations generalize in a useful manner? For example, does the product  $\prod_{n=2}^{\infty} (n^4 - 1)/(n^4 + 1)$  have an evaluation in terms of basic constants? *Maple* tells us that indeed

$$\prod_{n=2}^{\infty} \frac{n^4 - 1}{n^4 + 1} = \frac{\pi \sinh(\pi)}{\cosh(\sqrt{2}\pi) - \cos(\sqrt{2}\pi)}, \quad (1.2.12)$$

and it again produces a Gamma function expression for the finite product. For analogous products with fifth powers, *Maple* fails to return an evaluation. However, we now have enough hints to try our own hands at these products: Apparently we have to use properties of the Gamma function. In fact, setting  $\omega = \exp(\pi i/r)$  and using the relations  $\prod_{j=1}^{2r} (n - z\omega^j)^{(-1)^j} = (n^r - z^r)/(n^r + z^r)$ , as well as  $\sum_{j=1}^{2r} \omega^j (-1)^j = 0$  and  $\prod_{j=1}^{2r} (\omega^j)^{(-1)^j} = -1$ , it follows from the product representation of the Gamma function

$$\Gamma(x) = \lim_{n \rightarrow \infty} \frac{n! n^x}{x(x+1) \cdots (x+n)} \quad (1.2.13)$$

that, for  $r \in \mathbb{N}$ ,  $r > 1$ , and  $z \in \mathbb{C} \setminus \mathbb{N}$ ,

$$\prod_{n=1}^{\infty} \frac{n^r - z^r}{n^r + z^r} = \prod_{n=1}^{\infty} \prod_{j=1}^{2r} \left(1 - \frac{z\omega^j}{n}\right)^{(-1)^j} = - \prod_{j=1}^{2r} \Gamma(-z\omega^j)^{-(-1)^j}.$$

Hence, for  $m \in \mathbb{N}$ ,

$$\prod_{n=1, n \neq m}^{\infty} \frac{n^r - z^r}{n^r + z^r} = - \frac{m^r + z^r}{(m^r - z^r)\Gamma(-z)} \prod_{j=1}^{2r-1} \Gamma(-z\omega^j)^{-(-1)^j},$$

where as  $z \rightarrow m$ ,

$$(m^r - z^r)\Gamma(-z) = \frac{m^r - z^r}{(m-z)} \frac{\Gamma(m+1-z)}{(m-1-z) \cdots (1-z)(-z)} \rightarrow r m^{r-1} \frac{1}{m!(-1)^m}.$$

This gives the finite evaluation

$$\begin{aligned} P_r(m) &= \prod_{n=1, n \neq m}^{\infty} \frac{n^r - m^r}{n^r + m^r} \\ &= (-1)^{m+1} \frac{2m(m!)}{r} \prod_{j=1}^{2r-1} \Gamma(-m\omega^j)^{-(-1)^j}. \end{aligned} \quad (1.2.14)$$

When  $r = 2s$  is even, this can in a few steps be further reduced to

$$-(-1)^m \frac{2^\epsilon \pi m}{s} (\sinh \pi m)^{(-1)^s} \prod_{j=1}^{s-1} \left( \cosh \left( 2\pi m \sin \left( \frac{j\pi}{2s} \right) \right) - \cos \left( 2\pi m \cos \left( \frac{j\pi}{2s} \right) \right) \right)^{(-1)^j},$$

where  $\epsilon$  is 0 or 1 as  $s$  is respectively odd or even. From this, evaluations (1.2.10) and (1.2.12) immediately follow as special cases.

Interestingly, for odd  $r \geq 5$ , these products do not seem to have a closed form “nicer” than (1.2.14). In particular, they do not seem to be rational numbers like  $P_3(1)$  (and in fact  $P_3(m)$ ). The use of integer relation algorithms—to 400 digits—shows that  $P_5(1)$  satisfies no integer polynomial with degree less than 21 and Euclidean norm less than  $5 \cdot 10^{18}$ .

### 1.3 A Recursive Sequence Problem

The following problem on a recursively defined sequence appeared in *American Mathematical Monthly* (Problem 10901, [58]). We will describe here how the problem, which really is a problem about functional equations, can be solved via the experimental approach.

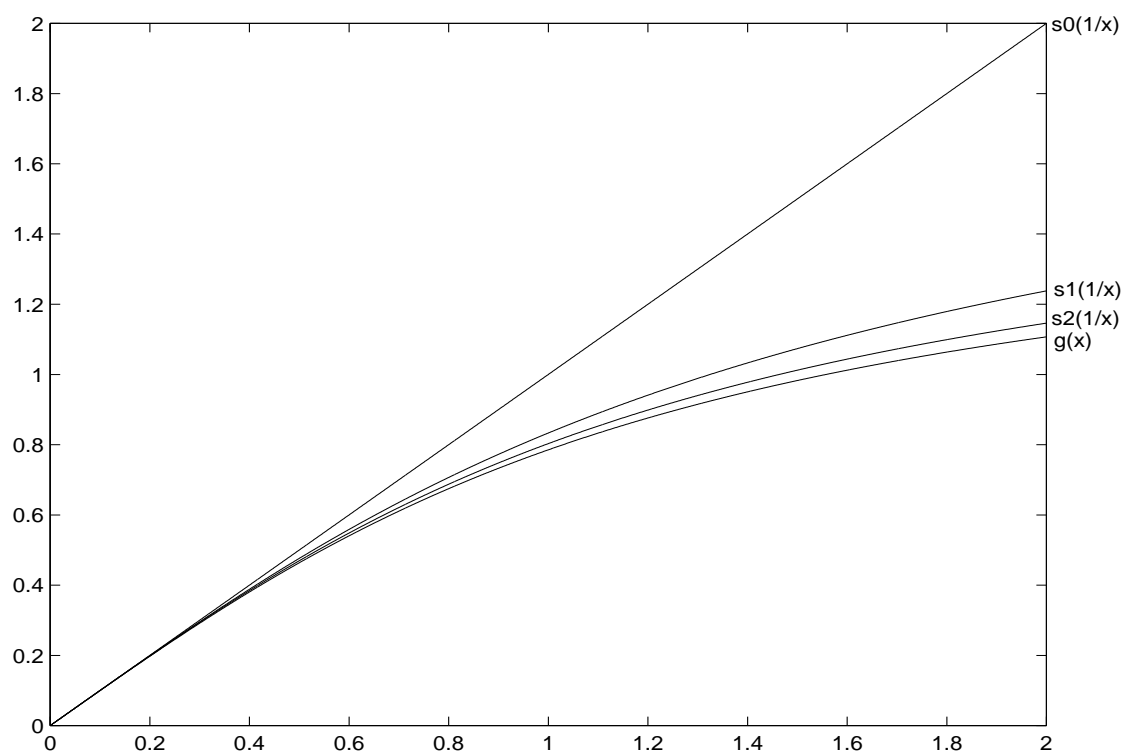
**Problem:** Let  $a_1 = 1$ ,

$$\begin{aligned} a_2 &= \frac{1}{2} + \frac{1}{3}, & a_3 &= \frac{1}{3} + \frac{1}{7} + \frac{1}{4} + \frac{1}{13}, \\ a_4 &= \frac{1}{4} + \frac{1}{13} + \frac{1}{8} + \frac{1}{57} + \frac{1}{5} + \frac{1}{21} + \frac{1}{14} + \frac{1}{183}, \end{aligned}$$

and continue the sequence, constructing  $a_{n+1}$  by replacing each fraction  $1/d$  in the expression for  $a_n$  with  $1/(d+1) + 1/(d^2+d+1)$ . Compute  $\lim_{n \rightarrow \infty} a_n$ .

**Solution:** We first observe that if  $s_0(x) = 1/x$  and  $s_{n+1}(x) = s_n(x+1) + s_n(x^2+x+1)$  for  $n \geq 0$ ,  $x > 0$ , then  $a_n = s_{n+1}(1)$ . What do these functions  $s_n(x)$  look like? Like  $s_0(x)$ , the plots of successive  $s_n(x)$  resemble reciprocal functions. If we instead examine the functions  $s_n(1/x)$ , we find that these are fairly well behaved, appearing to converge quickly to a smooth, monotone increasing function  $g(x)$  (see Figure 1.2). Indeed, we find fairly good convergence (to roughly four decimal places) for  $n = 25$ , by comparing  $s_{24}(1/x)$  with  $s_{25}(1/x)$ . What is this function  $g(x)$ ?

Examining the sequence of calculated numerical values used for plotting, we find that while  $g(x) = \lim_n s_n(1/x)$  is not defined at zero, it appears that  $\lim_{x \rightarrow 0} g(x) = 0$ . Further, it appears that  $g'(0) = 1$ ,  $g(1) \approx 0.7854$  and  $g'(1) = 1/2$ . Needless to say, the value 0.7854 is an approximation to  $\pi/4$ . These observations suggest that perhaps  $g(x) = \arctan x$ .

Figure 1.2: Convergence of  $s_n(1/x)$  to  $g(x)$

Let  $f(x) = \arctan(1/x)$  for  $x > 0$ . By applying the addition formula for the tangent, we note that

$$\begin{aligned} \tan[f(x+1) + f(x^2+x+1)] &= \frac{\frac{1}{x+1} + \frac{1}{x^2+x+1}}{1 - \frac{1}{x+1} \cdot \frac{1}{x^2+x+1}} \\ &= \frac{1}{x} = \tan[f(x)]. \end{aligned}$$

This means that  $f(x)$  satisfies  $f(x) = f(x+1) + f(x^2+x+1)$ , confirming that we are on the right track. In fact, we are finished if we can show that  $s_n(x)$  converges pointwise to  $f(x)$ .

To demonstrate this, we first verify that the function  $E(x) = 1/(xf(x))$  decreases strictly to 1 as  $x \rightarrow \infty$ . By differentiation, it suffices to show that  $-\arctan(x) + x/(x^2+1) < 0$ . But this follows since  $-\arctan(x) + x/(x^2+1)$  is strictly decreasing (its derivative is  $-2x^2/(x^2+1)^2$ ), and it starts at 0 for  $x = 0$ .

The second step is to show that for all  $x > 0$ , we have

$$f(x) \leq s_n(x) \leq f(x) \cdot E(x+n). \quad (1.3.15)$$

For  $n = 0$  this is merely the condition  $xf(x) \leq 1$ . Now if (1.3.15) holds for some  $n > 0$ , then we infer

$$f(x+1) \leq s_n(x+1) \leq f(x+1) \cdot E(x+n+1),$$

and (using the monotonicity of  $E$ )

$$\begin{aligned} f(x^2+x+1) &\leq s_n(x^2+x+1) \leq f(x^2+x+1) \cdot E(x^2+x+1+n) \\ &\leq f(x^2+x+1) \cdot E(x+n+1). \end{aligned}$$

Adding (and using the functional equation for  $f$ ), we obtain (1.3.15) for  $n+1$ . These facts together imply  $\lim_{n \rightarrow \infty} s_n(x) = f(x)$  for each  $x > 0$ .

Thus we have demonstrated here that  $a_n \rightarrow \pi/4$  [58].  $\square$

An extension of these methods leads to the following theorem:

**Theorem 1.3.1** *Let  $\mathcal{A} = \{s : \mathbb{R}_+ \rightarrow \mathbb{R} : \lim_{x \rightarrow \infty} xs(x) = 1\}$  and define a mapping  $T : \mathcal{A} \rightarrow \mathcal{A}$  by  $(Ts)(x) = s(x+1) + s(x^2+x+1)$  for  $s \in \mathcal{A}$ . Then*

the sequence  $(s_n)$  defined by the iteration  $s_{n+1} = Ts_n$  converges pointwise to  $f(x) = \arctan(1/x)$ , for every  $s_0 \in \mathcal{A}$ . Equivalently, every orbit of  $T$  converges pointwise to  $f$ , which is the unique fixed point of  $T$  in  $\mathcal{A}$ .

**Proof.** Define  $e(x) = \inf_{y \geq x} s_0(y)/f(y)$  and  $E(x) = \sup_{y \geq x} s_0(y)/f(y)$ . Then  $f(x)e(x) \leq s_0(x) \leq f(x)E(x)$  for all  $x > 0$ , while  $e(x)$  increases to 1 and  $E(x)$  decreases to 1 as  $x \rightarrow \infty$ . Now the same induction as in Step 2 above gives us

$$f(x)e(x+n) \leq s_n(x) \leq f(x)E(x+n) \quad \text{for all } n \geq 0, x > 0.$$

This implies  $s_n(x) \rightarrow f(x)$  for  $x > 0$ . This argument, slightly modified, also shows that  $f(x) = \arctan(1/x)$  is the unique fixed point of  $T$  in  $\mathcal{A}$ .  $\square$

The same procedure works for  $1/x \rightarrow 1/(x+y)+y/(x^2+xy+1)$ , for  $0 < y \leq 1$ . More generally, there are similar functional equations for other inverse functions. Thus, for  $l(x) = \log(1 + 1/x)$  the equation is

$$l(x) = l(2x + 1) + l(2x).$$

The corresponding iteration, for  $x = 1$ , starting with  $s_0(x) = 1/x$ , produces the classical result

$$\sum_{k=2^n}^{2^{n+1}-1} 1/k \rightarrow \log(2) \quad \text{as } n \rightarrow \infty.$$

Similarly, for  $\tau(x) = \operatorname{arctanh}(1/x) = \frac{1}{2} \log((1+x)/(1-x))$ , the functional equation is

$$\tau(x) = \tau(x+1) + \tau(x^2 + x - 1),$$

for  $x > 1$ . Likewise for  $\sigma = x \mapsto \operatorname{arcsinh}(1/x)$  we have

$$\sigma(x) = \sigma(\sqrt{x^2+1}) + \sigma\left(x\sqrt{x^2+1}\left(\sqrt{x^2+1} + \sqrt{x^2+2}\right)\right).$$

And, for  $\rho(x) = \arcsin(1/x)$ , we have

$$\rho(x) = \rho(\sqrt{x^2+1}) + \rho\left(x\sqrt{x^2+1}\left(x + \sqrt{x^2-1}\right)\right),$$

for all  $x \geq 1$ .

For these four functional equations, the result corresponding to Theorem 1.3.1 can be established. In fact, the basic inequality corresponding to Step 2 above for the last two functions would read

$$f(x) \cdot e(\sqrt{x^2+n}) \leq s_n(x) \leq f(x) \cdot E(\sqrt{x^2+n}),$$

for  $x \geq 1$ , where  $s_n$  is defined, as before, by  $s_0 \in \mathcal{A}$  and

$$s_{n+1}(x) = s_n(\sqrt{x^2+1}) + s_n\left(x\sqrt{x^2+1}\left(\sqrt{x^2+1} + \sqrt{x^2+2}\right)\right)$$

in the case of  $\sigma$ , and

$$s_{n+1}(x) = s_n(\sqrt{x^2+1}) + s_n\left(x\sqrt{x^2+1}\left(x + \sqrt{x^2-1}\right)\right)$$

in the case of  $\rho$ . By contrast,

$$\rho(x) = \rho\left(\frac{x^2}{(x-1)\sqrt{x^2-1} + \sqrt{2x-1}}\right) - \rho\left(\frac{x}{x-1}\right)$$

is another functional equation for  $\arcsin(1/x)$  which does not have convergent orbits.

## 1.4 High Precision Fraud

Consider the sums

$$\sum_{n=1}^{\infty} \frac{\lfloor n \tanh(\pi) \rfloor}{10^n} \stackrel{?}{=} \frac{1}{81},$$

an evaluation which is wrong but valid to 268 decimal places, and

$$\sum_{n=1}^{\infty} \frac{\lfloor n \tanh(\pi/2) \rfloor}{10^n} \stackrel{?}{=} \frac{1}{81},$$

which is valid to “only” 12 places. Both series actually evaluate to transcendental numbers.



What underlies these “fraudulent” evaluations? The “quick” reason is that  $\tanh(\pi)$  and  $\tanh(\pi/2)$  are almost integers, with, e.g.,  $0.99 < \tanh(\pi) < 1$ . Therefore,  $\lfloor n \tanh(\pi) \rfloor$  will be equal to  $n - 1$  for many  $n$ ; precisely for  $n = 1, \dots, 268$ . Since

$$\sum_{n=1}^{\infty} \frac{n-1}{10^n} = \frac{1}{81},$$

this explains the evaluations. Looking more closely at this argument, one is directly led to *continued fractions* as the deeper reason behind the frauds. For any irrational positive  $\alpha$ , we can write

$$\begin{aligned} \alpha &= [a_0, a_1, \dots, a_n, a_{n+1}, \dots] \\ &= a_0 + \frac{1}{a_1 + \frac{1}{a_2 + \frac{1}{a_3 + \dots}}} \end{aligned}$$

with integral  $a_n$  and  $a_0 \geq 0$ ,  $a_n \geq 1$  for  $n \geq 1$ . This is hard to compute by hand, but easy even on a small computer or calculator. For the parameters in our series, we get

$$\tanh(\pi) = [0, 1, \mathbf{267}, 4, 14, 1, 2, 1, 2, 2, 1, 2, 3, 8, 3, 1, \dots] \quad (1.4.16)$$

and

$$\tanh\left(\frac{\pi}{2}\right) = [0, 1, \mathbf{11}, 14, 4, 1, 1, 1, 3, 1, 295, 4, 4, 1, 5, 17, 7, \dots]. \quad (1.4.17)$$

It cannot be a coincidence that the integers 267 and 11 (each equal to the number of places of agreement with  $1/81$  in the respective formula) appear in these expansions! There must be a connection between series of the type  $\sum \lfloor n\alpha \rfloor z^n$  and the continued fraction expansion of an irrational  $\alpha$ . In fact, consider the infinite continued fraction approximations for  $\alpha$  generated by

$$\begin{aligned} p_{n+1} &= p_n a_{n+1} + p_{n-1}, & p_0 &= a_0 = \lfloor \alpha \rfloor, & p_{-1} &= 1, \\ q_{n+1} &= q_n a_{n+1} + q_{n-1}, & q_0 &= 1, & q_{-1} &= 0. \end{aligned}$$

Then for  $n \geq 0$ ,  $p_{2n}/q_{2n}$  increases to  $\alpha$ , while  $p_{2n+1}/q_{2n+1}$  decreases to  $\alpha$  and

$$\frac{1}{q_n(q_n + q_{n+1})} < \left| \alpha - \frac{p_n}{q_n} \right| < \frac{1}{q_n q_{n+1}}.$$

Let further  $\epsilon_n = q_n \alpha - p_n$ . Then from the above it follows that

$$|\epsilon_{n+1}| < \frac{1}{q_n + q_{n+1}} < |\epsilon_n| < \frac{1}{q_{n+1}} \leq 1.$$

All of this is standard and may be found in [132], [207], or [174]. Our aim now is to show a relationship between the above series and the continued fraction expansion of  $\alpha$ . A first key is the following lemma, which we will not prove here since it requires some knowledge about linear Diophantine equations (cf. [47], where this material is taken from).

**Lemma 1.4.1** *For any irrational  $\alpha > 0$  and  $n, N \in \mathbb{N}$ , we have*

$$\begin{aligned} \lfloor n\alpha + \epsilon_N \rfloor &= \lfloor n\alpha \rfloor && \text{for } n < q_{N+1}, \\ \lfloor n\alpha + \epsilon_N \rfloor &= \lfloor n\alpha \rfloor + (-1)^N && \text{for } n = q_{N+1}. \end{aligned}$$

**Theorem 1.4.2** *For irrational  $\alpha > 0$ ,*

$$\sum_{n=1}^{\infty} \lfloor n\alpha \rfloor z^n = \frac{p_0 z}{(1-z)^2} + \sum_{n=0}^{\infty} (-1)^n \frac{z^{q_n} z^{q_{n+1}}}{(1-z^{q_n})(1-z^{q_{n+1}})}.$$

**Proof.** Let

$$G_\alpha(z, w) = \sum_{n=1}^{\infty} z^n w^{\lfloor n\alpha \rfloor}, \quad (1.4.18)$$

for  $|z|, |w| < 1$ . Then for  $N > 0$ ,

$$\begin{aligned} (1 - z^{q_N} w^{p_N}) G_\alpha(z, w) &= \sum_{n=1}^{q_N} z^n w^{\lfloor n\alpha \rfloor} \\ &= \sum_{n=1}^{\infty} z^{n+q_N} (w^{\lfloor (n+q_N)\alpha \rfloor} - w^{\lfloor n\alpha \rfloor + p_N}) \\ &= \sum_{n=1}^{\infty} z^{n+q_N} w^{\lfloor n\alpha \rfloor + p_N} (w^{\lfloor n\alpha + \epsilon_N \rfloor - \lfloor n\alpha \rfloor} - 1) \\ &= z^{q_{N+1} + q_N} w^{\lfloor q_{N+1}\alpha \rfloor + p_N} (w^{(-1)^N} - 1) + O(z^{q_{N+1} + q_N + 1}) \\ &= z^{q_{N+1} + q_N} w^{p_{N+1} + p_N} (-1)^N \frac{w - 1}{w} + O(z^{q_{N+1} + q_N + 1}), \end{aligned} \quad (1.4.19)$$

since  $\lfloor q_{N+1}\alpha \rfloor = \lfloor \epsilon_{N+1} \rfloor + p_{N+1} = p_{N+1}$  if  $N$  is odd, and  $= p_{N+1} - 1$  if  $N$  is even.

Now write  $P_N = \sum_{n=1}^{q_N} z^n w^{\lfloor n\alpha \rfloor}$  and  $Q_N = 1 - z^{q_N} w^{p_N}$ . Then  $A_N = Q_N P_{N+1} - Q_{N+1} P_N$  is a polynomial of degree at most  $q_N + q_{N+1}$  in  $z$ , and therefore it follows from (1.4.19) that

$$\begin{aligned} A_N &= Q_{N+1}(Q_N G_\alpha - P_N) - Q_N(Q_{N+1} G_\alpha - P_{N+1}) \\ &= (-1)^N \frac{w-1}{w} z^{q_N} w^{p_N} z^{q_{N+1}} w^{p_{N+1}}. \end{aligned}$$

This in turn implies

$$\frac{P_{N+1}}{Q_{N+1}} - \frac{P_N}{Q_N} = \frac{A_N}{Q_N Q_{N+1}} = (-1)^N \frac{w-1}{w} \frac{z^{q_N} w^{p_N} z^{q_{N+1}} w^{p_{N+1}}}{Q_N Q_{N+1}}.$$

Next summing from zero to infinity, and noting that (1.4.19) implies that  $G_\alpha - P_N/Q_N$  tends to 0 as  $N$  tends to infinity, shows that

$$G_\alpha(z, w) = \frac{z w^{p_0}}{1 - z w^{p_0}} - \frac{1-w}{w} \sum_{n=0}^{\infty} (-1)^n \frac{z^{q_n} w^{p_n} z^{q_{n+1}} w^{p_{n+1}}}{(1 - z^{q_n} w^{p_n})(1 - z^{q_{n+1}} w^{p_{n+1}})}.$$

Now differentiating with respect to  $w$  and then letting  $w$  tend to 1 proves the assertion.  $\square$

This theorem was first proved (for  $\alpha \in (0, 1)$ ) by Mahler in [157].

**Example 1.4.3**  $\alpha = \tanh(\pi)$ .

In this case,  $q_n = 1, 1, 268, 1073, \dots$  for  $n = 0, 1, 2, 3, \dots$ , and thus

$$\sum_{n=1}^{\infty} \lfloor n \tanh(\pi) \rfloor z^n = \frac{z^2}{(1-z)^2} - \frac{z^{269}}{(1-z)(1-z^{268})} + \dots$$

Therefore,

$$\frac{1}{81} - 2 \cdot 10^{-269} \leq \sum_{n=1}^{\infty} \frac{\lfloor n \tanh(\pi) \rfloor}{10^n} \leq \frac{1}{81} + 2 \cdot 10^{-269},$$

and similarly for  $\alpha = \tanh(\frac{\pi}{2})$ .  $\square$

**Example 1.4.4**  $\alpha = e^{\pi\sqrt{163/9}}$ .

With one of our favorite transcendental numbers,  $\alpha = e^{\pi\sqrt{163/9}} = [640320, 1653264929, \dots]$ , we get the incorrect evaluation

$$\sum_{n=1}^{\infty} \frac{\lfloor ne^{\pi\sqrt{163/9}} \rfloor}{2^n} \stackrel{?}{=} 1280640,$$

which is, however, correct to at least half a billion digits.  $\square$

**Example 1.4.5**  $\alpha = \log_{10}(2)$ .

In this case,  $\lfloor n\alpha \rfloor + 1$  is the number of decimal digits of  $2^n$ . Then  $q_n = 1, 3, 10, 93, \dots$  for  $n = 0, 1, 2, 3, \dots$ , and as a result the transcendental number  $\sum \lfloor n \log_{10}(2) \rfloor / 2^n$  is equal to  $146/1023$  to 30 decimal digits. Interestingly, if  $e(n)$ , respectively  $o(n)$ , count the number of even, respectively odd, decimal digits of  $n$ , then

$$\sum_{n=1}^{\infty} \frac{o(2^n)}{2^n} = \frac{1}{9}$$

is rational, while

$$\sum_{n=1}^{\infty} \frac{e(2^n)}{2^n} = \sum_{n=1}^{\infty} \frac{\lfloor n \log_{10}(2) \rfloor + 1}{2^n} - \sum_{n=1}^{\infty} \frac{o(2^n)}{2^n}$$

is transcendental. We will not prove the transcendency result here, but the evaluation for the sum with  $o(2^n)$  follows in the next theorem.  $\square$

**Theorem 1.4.6** *If  $o(n)$  counts the odd decimal digits of  $n$ , then*

$$\sum_{n=1}^{\infty} \frac{o(2^n)}{2^n} = \frac{1}{9}.$$

**Proof.** Let  $0 < q < 1$  and  $m \in \mathbb{N}$ ,  $m > 1$ , and consider the base- $m$  expansion of  $q$ ,

$$q = \sum_{n=1}^{\infty} \frac{a_n}{m^n} \quad \text{with } 0 \leq a_n < m,$$

where when ambiguous we take the terminating expansion. Then  $a_n$  is the remainder of  $\lfloor m^n q \rfloor$  modulo  $m$ , and therefore we can just as well write

$$q = \sum_{n=1}^{\infty} \frac{\lfloor m^n q \rfloor \pmod{m}}{m^n}. \quad (1.4.20)$$

Now let  $F(q) = \sum_{k=1}^{\infty} c_k q^k$  be a power series with radius of convergence 1. Then for  $0 < q < 1$ , from (1.4.20) we get, by exchanging the order of summation (as is valid within the radius of convergence),

$$F(q) = \sum_{k=1}^{\infty} c_k q^k = \sum_{k=1}^{\infty} c_k \sum_{n=1}^{\infty} \frac{\lfloor m^n q^k \rfloor \pmod{m}}{m^n} = \sum_{n=1}^{\infty} \frac{f(n)}{m^n},$$

with  $f(n) = \sum_{k \geq 1} c_k (\lfloor m^n q^k \rfloor \pmod{m})$ . Now if  $q = 1/b$ , where  $b$  is an integer multiple of  $m$ , then  $\lfloor m^n/b^k \rfloor \pmod{m}$  is the  $k$ -th digit (mod  $m$ ) of the base- $b$  expansion of the integer  $m^n$ . (Here we start the numbering of the digits with 0, e.g., the 0-th digit of 1205 is 5.) Thus for  $F(q) = q/(1-q)$  and  $m = 2$  (and  $b$  even),  $f(n)$  counts the odd digits in the base- $b$  expansion of  $2^n$ . For  $b = 10$ , we have  $f(n) = o(2^n)$ , and we obtain

$$\frac{1}{9} = F\left(\frac{1}{10}\right) = \sum_{n=1}^{\infty} \frac{o(2^n)}{2^n}.$$

□

## 1.5 Knuth's Series Problem

We give an account here of the solution, by one of the present authors (Borwein) to a problem recently posed by Donald E. Knuth of Stanford University in the *American Mathematical Monthly* (Problem 10832, Nov. 2000):

**Problem:** Evaluate

$$S = \sum_{k=1}^{\infty} \left( \frac{k^k}{k!e^k} - \frac{1}{\sqrt{2\pi k}} \right).$$

**Solution:** We first attempted to obtain a numerical value for  $S$ . *Maple* produced the approximation

$$S \approx -0.08406950872765599646.$$

Based on this numerical value, the Inverse Symbolic Calculator, available at the URL <http://www.cecm.sfu.ca/projects/ISC>, with the “Smart Lookup” feature, yielded the result

$$S \approx -\frac{2}{3} - \frac{1}{\sqrt{2\pi}}\zeta\left(\frac{1}{2}\right). \quad (1.5.21)$$

Calculations to even higher precision (50 decimal digits) confirmed this approximation. Thus within a few minutes we “knew” the answer.

Why should such an identity hold? One clue was provided by the surprising speed with which *Maple* was able to calculate a high-precision value of this slowly convergent infinite sum. Evidently the *Maple* software knew something that we did not. Peering under the covers, we found that *Maple* was using the Lambert  $W$  function, which is the functional inverse of  $w(z) = ze^z$ .

Another clue was the appearance of  $\zeta(1/2)$  in the above experimental identity, together with an obvious allusion to Stirling’s formula in the original problem. This led us to conjecture the identity

$$\sum_{k=1}^{\infty} \left( \frac{1}{\sqrt{2\pi k}} - \frac{P(1/2, k-1)}{(k-1)!\sqrt{2}} \right) = \frac{1}{\sqrt{2\pi}}\zeta\left(\frac{1}{2}\right), \quad (1.5.22)$$

where  $P(x, n)$  denotes the Pochhammer function  $x(x+1)\cdots(x+n-1)$ , and where the binomial coefficients in the LHS of (1.5.22) are the same as those of the function  $1/\sqrt{2-2x}$ . *Maple* successfully evaluated this summation, as shown on the RHS. We now needed to establish that

$$\sum_{k=1}^{\infty} \left( \frac{k^k}{k!e^k} - \frac{P(1/2, k-1)}{(k-1)!\sqrt{2}} \right) = -\frac{2}{3}.$$

Guided by the presence of the Lambert  $W$  function

$$W(z) = \sum_{k=1}^{\infty} \frac{(-k)^{k-1} z^k}{k!},$$

an appeal to Abel's limit theorem suggested the conjectured identity

$$\lim_{z \rightarrow 1} \left( \frac{dW(-z/e)}{dz} + \frac{1}{2-2z} \right) = 2/3.$$

Here again, *Maple* was able to evaluate this summation and establish the identity.  $\square$

As can be seen from this account, the above manipulation took considerable human ingenuity, in addition to computer-based symbolic manipulation. We include this example to highlight a challenge for the next generation of mathematical computing software—these tools need to more completely automate this class of operations, so that similar derivations can be accomplished by a significantly broader segment of the mathematical community.

## 1.6 Ahmed's Integral Problem

The same comments apply to an “experimental” solution to a problem posed by Zafar Ahmed in the *American Mathematical Monthly* [3]:

**Problem:** Evaluate

$$F = \int_0^1 \frac{\arctan(\sqrt{x^2+2})}{\sqrt{x^2+2}(x^2+1)} dx.$$

**Solution:** Since presently available symbolic computing software is unable to produce a closed-form evaluation, we try to identify the integral via its numerical value,

$$F \approx 0.51404189589007076139762973957688287.$$

The Inverse Symbolic Calculator (with the integer relations algorithm clicked on) declares that this number matches  $5\pi^2/96$ . A test to higher precision confirms this evaluation.

It remains to prove the experimentally well-founded conjecture

$$F = \frac{5}{96} \pi^2.$$

An idea is needed, which, as always, takes more human insight than any computer algebra system (at present) has built in. However, such a system can help to quickly identify promising starting points—and it can then carry out the symbolic manipulations which, taken together, constitute a proof. One possible starting point for this problem is to generalize: Using *Maple*, set

```
> assume(x>0,p>0); interface(showassumed=0);
> g := arctan(p*sqrt(x^2+2))/sqrt(x^2+2)/(x^2+1);
```

$$g = \frac{\arctan(p \sqrt{x^2 + 2})}{\sqrt{x^2 + 2} (x^2 + 1)}$$

```
> G := Int(g(x,p),x=0..1);
```

$$G = \int_0^1 \frac{\arctan(p \sqrt{x^2 + 2})}{\sqrt{x^2 + 2} (x^2 + 1)} dx,$$

so that  $F = G(1)$ . On the other hand,

```
> diff(g,p);
```

$$\frac{1}{(1 + p^2 (x^2 + 2)) (x^2 + 1)}$$

is a rational function in  $x$  and  $p$ , so that it may pay to write

$$F = G(1) = \int_0^1 \int_0^1 \partial g(x, p) / \partial p dp dx = \int_0^1 \int_0^1 \partial g(x, p) / \partial p dx dp,$$

where the exchange of order of integration is justified by Fubini's (Tonelli's) theorem. Thus we are led to

```
> int(diff(g,p),x=0..1);
```



$$\frac{1}{4} \frac{-4p \arctan\left(\frac{p}{\sqrt{1+2p^2}}\right) + \pi \sqrt{1+2p^2}}{(p^2+1)\sqrt{1+2p^2}}$$

> map(int, expand(int(diff(g,p), x=0..1)), p=0..1);

$$\int_0^1 -\frac{p \arctan\left(\frac{p}{\sqrt{1+2p^2}}\right)}{(p^2+1)\sqrt{1+2p^2}} dp + \frac{1}{16} \pi^2$$

> G1 := map(int, expand(int(diff(g,p), x=0..1)), p=0..1)-Pi^2/16;

$$G1 = \int_0^1 -\frac{p \arctan\left(\frac{p}{\sqrt{1+2p^2}}\right)}{(p^2+1)\sqrt{1+2p^2}} dp$$

and  $F = G1 + \pi^2/16$ . Now using  $\arctan(y) + \arctan(1/y) = \pi/2$  for  $y > 0$  and then doing a change of variables  $x = 1/p$ , we obtain

> G2:=subs(arctan(p/sqrt(1+2\*p^2))=Pi/2-arctan(sqrt(1+2\*p^2)/p), G1);

$$G2 = \int_0^1 -\frac{p \left(\frac{1}{2} \pi - \arctan\left(\frac{\sqrt{1+2p^2}}{p}\right)\right)}{(p^2+1)\sqrt{1+2p^2}} dp$$

> G3:=map(int, expand(op(1, G2)), p=0..1);

$$G3 = -\frac{1}{24} \pi^2 + \int_0^1 \frac{p \arctan\left(\frac{\sqrt{1+2p^2}}{p}\right)}{(p^2+1)\sqrt{1+2p^2}} dp$$

> G4:=simplify(student[changevar](x=1/p, G3, x));

$$G4 = -\frac{1}{24}\pi^2 + \int_1^\infty \frac{\arctan(\sqrt{x^2+2})}{(x^2+1)\sqrt{x^2+2}} dx$$

> H:=G4+Pi^2/24;

$$H = \int_1^\infty \frac{\arctan(\sqrt{x^2+2})}{(x^2+1)\sqrt{x^2+2}} dx$$

so that  $F = G1 + \pi^2/16 = G4 + \pi^2/16 = H + \pi^2/48$ . This evaluates  $F - H = \pi^2/48$ . This suggests we also evaluate

$$F + H = \int_0^\infty \int_0^1 \partial g(x,p)/\partial p dp dx = \int_0^1 \int_0^\infty \partial g(x,p)/\partial p dx dp.$$

Indeed,

> int(diff(g,p),x=0..infinity);

$$-\frac{1}{2} \frac{\pi (p + \sqrt{1-2p^2})}{(p^2+1)\sqrt{1+2p^2}}$$

> FpH := int(int(diff(g,p),x=0..infinity),p=0..1);

$$FpH = \frac{1}{12} \pi^2.$$

Now the result is proved by

> solve({f-h=Pi^2/48,f+h=Pi^2/12},{f,h});

$$\{f = \frac{5}{96} \pi^2, h = \frac{1}{32} \pi^2\}.$$

□

**Generalization:** We note that in the same fashion we can evaluate

$$\begin{aligned} & \int_0^1 \frac{\arctan(\sqrt{x^2 + b^2})}{\sqrt{x^2 + b^2}(x^2 + 1)} dx - \int_1^\infty \frac{\arctan(\sqrt{x^2 + b^2})}{\sqrt{x^2 + b^2}(-1 + x^2 + b^2)} dx \\ &= \frac{3\pi \arctan(\sqrt{b^2 - 1})}{4\sqrt{b^2 - 1}} - \frac{1\pi \arctan(\sqrt{b^4 - 1})}{2\sqrt{b^2 - 1}}, \\ & \int_0^1 \frac{\arctan(\sqrt{x^2 + b^2})}{\sqrt{x^2 + b^2}(x^2 + 1)} dx + \int_1^\infty \frac{\arctan(\sqrt{x^2 + b^2})}{\sqrt{x^2 + b^2}(x^2 + 1)} dx \\ &= \frac{\pi \arctan(\sqrt{b^2 - 1})}{\sqrt{b^2 - 1}} - \frac{1\pi \arctan(\sqrt{b^4 - 1})}{2\sqrt{b^2 - 1}}, \end{aligned}$$

and for  $b = \sqrt{2}$  this yields the previous closed form. Moreover, for  $b = 1$ ,

$$\int_0^1 \frac{\arctan(\sqrt{x^2 + 1})}{(x^2 + 1)^{3/2}} dx = \left( \frac{1}{4} - \frac{\sqrt{2}}{2} \right) \pi + \frac{3}{2} \sqrt{2} \arctan(\sqrt{2}),$$

and for  $b = 0$ ,

$$\int_0^1 \frac{\arctan(x)}{x(x^2 + 1)} dx = \frac{G}{2} + \frac{1}{8} \pi \log(2),$$

where  $G$  is Catalan's constant.

## 1.7 Evaluation of Binomial Series

A classical binomial series, derived from the arctan series, and given in [44], is

$$\sum_{n \geq 1} \frac{-9n + 18}{\binom{2n}{n}} = 2 \frac{\pi}{\sqrt{3}}. \quad (1.7.23)$$

A more modern sum, due to Bill Gosper [124], is

$$\sum_{n \geq 0} \frac{50n - 6}{\binom{3n}{n} 2^n} = \pi. \quad (1.7.24)$$

In [7], whole classes of formulas for  $\pi$  of this type are proved, such as

$$\sum_{n \geq 0} \frac{S_k(n)}{\binom{8kn}{4kn} (-4)^{kn}} = \pi,$$

where  $S_k(n)$  is a polynomial in  $n$  of degree  $4k$  with rational coefficients, explicitly computable (for fixed  $k$ ).

Motivated by such results, we shall consider here the following two families of sums:

$$\begin{aligned} b_2(k) &= \sum_{n \geq 1} \frac{n^k}{\binom{2n}{n}}, \\ b_3(k) &= \sum_{n \geq 1} \frac{n^k}{\binom{3n}{n} 2^n}, \end{aligned}$$

for  $k \in \mathbb{Z}$ . We shall record closed forms for the sums  $b_2(k)$  and recursion formulas for the sums  $b_3(k)$ , both in the case of positive  $k$ . These were discovered with the help of integer relation and similar methods, described below. The case of negative  $k$  is more complex; we shall finish with some primarily experimental results. This material is taken from [54] (for  $b_2(k)$  with negative  $k$ ) and from [59].

The key observation is that the sums have integral representations involving the *polylogarithms*  $L_p(z) = \sum_{n > 0} z^n / n^p$ , see [61]. Using the following properties of the  $\beta$ -function (see Section 5.4 of the first volume):

$$\begin{aligned} \frac{1}{\binom{2n}{n}} &= (2n+1) \beta(n+1, n+1) = n \beta(n, n+1) \quad \text{and} \\ \frac{1}{\binom{3n}{n}} &= (3n+1) \beta(2n+1, n+1) = n \beta(2n+1, n) = 2n \beta(2n, n+1), \end{aligned}$$

we find that

$$\begin{aligned} b_2(k) &= \int_0^1 L_{-k}(x(1-x)) + 2 L_{-k-1}(x(1-x)) dx \quad (1.7.25) \\ &= \int_0^1 \frac{L_{-k-1}(x(1-x))}{x} dx, \end{aligned}$$

and

$$\begin{aligned} b_3(k) &= \int_0^1 L_{-k} \left( \frac{x^2(1-x)}{2} \right) + 3 L_{-k-1} \left( \frac{x^2(1-x)}{2} \right) dx \quad (1.7.26) \\ &= \int_0^1 \frac{L_{-k-1} \left( \frac{x^2(1-x)}{2} \right)}{1-x} dx = 2 \int_0^1 \frac{L_{-k-1} \left( \frac{x^2(1-x)}{2} \right)}{x} dx. \end{aligned}$$

For fixed  $k \geq 0$ , these integrals are easy to compute symbolically in *Maple* and (with some additional effort) in *Mathematica*.

### 1.7.1 The Case of Non-Negative $k$

For integer  $k \geq 0$ ,  $L_{-k}(x)$  is clearly a rational function, and it is useful to write it as a partial fraction

$$L_{-k}(x) = \sum_{j=1}^{k+1} \frac{c_j^k}{(x-1)^j}.$$

Since  $x dL_{-k}(x)/dx = L_{-k-1}(x)$ , we may obtain the recursion

$$c_j^k = - (j c_j^{k-1} + (j-1) c_{j-1}^{k-1}). \quad (1.7.27)$$

Let

$$\begin{aligned} M_2(k, x) &= L_{-k}(x) + 2 L_{-k-1}(x) \\ M_3(k, x) &= L_{-k}(x) + 3 L_{-k-1}(x). \end{aligned}$$

We may then easily verify that the coefficients of the partial fraction of  $M_2$  and  $M_3$  are governed by recursion (1.7.27) with initial conditions given by  $c_1^0(2) = 1$ ,  $c_2^0(2) = 2$ ,  $c_j^0(2) = 0$  otherwise and  $c_1^0(3) = 2$ ,  $c_2^0(3) = 3$ ,  $c_j^0(3) = 0$  otherwise, respectively.

This in turn is easily verified—by hand or in a computer algebra system—to yield

$$\begin{aligned} c_j^k(2) &= \frac{(-1)^{k+j}}{j} \sum_{m=1}^j (-1)^m (2m-1) m^{k+1} \binom{j}{m}, \\ c_j^k(3) &= \frac{(-1)^{k+j}}{j} \sum_{m=1}^j (-1)^m (3m-1) m^{k+1} \binom{j}{m}. \end{aligned} \quad (1.7.28)$$

As is often the case, this is somewhat easier to verify than to find.

Next we observe that the values of the integrals in (1.7.25) and (1.7.26) are of the form

$$b_2(k) = \sum_j c_j^k(2) \int_0^1 (1-x(1-x))^{-j} dx$$

and

$$b_3(k) = \sum_j c_j^k(3) \int_0^1 (1 - x^2(1-x)/2)^{-j} dx, \quad (1.7.29)$$

respectively. So we set ourselves the task of hunting for a recursion for

$$d_2(j) = \int_0^1 (1 - x(1-x))^{-j} dx$$

and for

$$d_3(j) = \int_0^1 (1 - x^2(1-x)/2)^{-j} dx.$$

By computing the first few cases, we determine that  $d_2(j)$  is a rational combination of 1 and  $\pi/\sqrt{3}$ , while  $d_3(j)$  is a rational combination of 1,  $\log 2$  and  $\pi$ . Thus it is reasonable to hunt for two-term recursions for  $d_2$  and three-term recursions for  $d_3$ .

Now integer relation computations come to the rescue. We look for relations between  $d_2(p)$ ,  $d_2(p+1)$ , and  $d_2(p+2)$ , say for  $0 \leq p \leq 4$ , and are rewarded by the relations  $[2, 2, -3]$ ,  $[-2, 9, -6]$ ,  $[6, -16, 9]$ ,  $[-10, 23, -12]$ ,  $[14, -30, 15]$ . By inspection we have  $d_2(0) = 1$ ,  $d_2(1) = 2\pi/(3\sqrt{3})$  and

$$(4p - 10) d_2(p - 2) - (7p - 12) d_2(p - 1) + (3p - 3) d_2(p) = 0 \quad (1.7.30)$$

for  $p \geq 2$ .

For  $d_3$ , we look for four-term relations between  $d_3(p)$ ,  $d_3(p+1)$ ,  $d_3(p+2)$  and  $d_3(p+3)$ , and we return  $[-3, -24, 78, -50]$ ,  $[-24, 183, -310, 150]$ ,  $[-105, 500, -696, 300]$ ,  $[240, -975, 1236, -500]$ ,  $[-429, 1608, -1930, 750]$ . A little more intense pattern matching leads to  $d_3(0) = 1$ ,  $d_3(1) = 3 \log(2)/5 + \pi/5$ ,  $d_3(2) = 9/25 + 48 \log(2)/125 + 37\pi/250$ , while  $d_3(3) = 627/1250 + 972 \log(2)/3125 + 843\pi/6250$ , as is predicted by

$$\begin{aligned} 3(3p - 10)(3p - 8) d_3(p - 3) - (79(p - 2)(p - 3) + 21 + p) d_3(p - 2) \\ + (77p - 153)(p - 2) d_3(p - 1) - 25(p - 1)(p - 2) d_3(p) = 0 \end{aligned} \quad (1.7.31)$$

for  $p \geq 3$ .

In each case once discovered one can prove the recursion by considering the indefinite integral from 0 to  $t$ , which *Mathematica* and *Maple* can perform. One may then verify that the integral has a zero at  $t = 1$ . We illustrate this for the case  $N = 2$ . We combine the integrals in (1.7.30) and consider

$$\int_0^t \frac{4p-10}{(1-x(1-x))^{p-2}} - \frac{7p-12}{(1-x(1-x))^{p-1}} + \frac{3p-3}{(1-x(1-x))^p} dx \quad (1.7.32)$$

$$= -\frac{(2t-1)(t-1)t}{(1-t+t^2)^{p-1}}.$$

If we differentiate this last expression back and simplify we recover the integrand as required. Since the right-hand side of (1.7.32) has a zero at  $t = 1$  we are done. Similarly for (1.7.31), and it is to assure the zero at  $t = 1$  that the factor of  $1/2^n$  is needed.

The quantities  $d_2(j)$  (and therefore also the  $b_2(k)$ ) can in fact be computed explicitly (but this realization for us came after having found the recursion). To find the explicit formula, substitute  $y = 2x - 1$  in the integral  $d_2(j)$  to get

$$d_2(j) = 4^j \int_0^1 \frac{1}{(3+x^2)^j} dx.$$

This satisfies the recursion  $d_2(1) = 2\pi/(3\sqrt{3})$  and

$$d_2(j+1) = \frac{2}{3j} (1 + (2j-1)d_2(j)).$$

This leads to the explicit representation

$$d_2(j) = \frac{1}{3^j} \binom{2j-2}{j-1} \cdot \left( \sum_{i=1}^{j-1} \frac{3^i}{(2i-1) \binom{2i-2}{i-1}} + \frac{2}{\sqrt{3}} \pi \right).$$

Putting this together with formula (1.7.28) and simplifying as much as possible gives

$$b_2(k-1) = \frac{(-1)^k}{2} \sum_{j=1}^k (-1)^j j! \mathcal{S}_k^{(j)} \frac{\binom{2j}{j}}{3^j} \left( \sum_{i=0}^{j-1} \frac{3^i}{(2i+1) \binom{2i}{i}} + \frac{2}{3\sqrt{3}} \pi \right),$$

for  $k \geq 1$ , where  $\mathcal{S}_k^{(j)}$  are the *Stirling numbers of the second kind*,

$$\mathcal{S}_k^{(j)} = \frac{(-1)^j}{j!} \sum_{m=0}^j (-1)^m m^k \binom{j}{m}.$$

For a similar explicit formula for  $b_3(k)$ , we would need to evaluate the integrals  $d_3(j)$  explicitly. This would involve doing the partial fraction decomposition of the integrand  $1/[(1+x)^j(x^2-2x+2)^j]$ , and while possible in principle, it leads to such unwieldy recursions that we have refrained from doing that.

To recapitulate, we have now proven that

$$b_2(k) = p_k + q_k \frac{\pi}{\sqrt{3}},$$

with explicitly given rationals  $p_k, q_k$ , and

$$b_3(k) = r_k + s_k \pi + t_k \log 2,$$

with certain rationals  $r_k, s_k, t_k$ , for which we have very efficient iterations. Explicitly,

$$\begin{aligned} \sum_{n \geq 1} \frac{1}{\binom{2n}{n}} &= \frac{1}{3} + \frac{2}{9} \frac{\pi}{\sqrt{3}}, \\ \sum_{n \geq 1} \frac{n}{\binom{2n}{n}} &= \frac{2}{3} + \frac{2}{9} \frac{\pi}{\sqrt{3}}, \\ \sum_{n \geq 1} \frac{n^2}{\binom{2n}{n}} &= \frac{4}{3} + \frac{10}{27} \frac{\pi}{\sqrt{3}}, \\ \sum_{n \geq 1} \frac{1}{\binom{3n}{n} 2^n} &= \frac{2}{25} - \frac{6}{125} \log(2) + \frac{11}{250} \pi, \\ \sum_{n \geq 1} \frac{n}{\binom{3n}{n} 2^n} &= \frac{81}{625} - \frac{18}{3125} \log(2) + \frac{79}{3125} \pi, \\ \sum_{n \geq 1} \frac{n^2}{\binom{3n}{n} 2^n} &= \frac{561}{3125} + \frac{42}{15625} \log(2) + \frac{673}{31250} \pi. \end{aligned}$$



In particular, we can deduce, by elimination, that (1.7.23) and (1.7.24) hold, and we can deduce the corresponding formulas

$$\begin{aligned}\sum_{n \geq 1} \frac{-150n^2 + 230n - 36}{\binom{3n}{n} 2^n} &= \pi, \\ \sum_{n \geq 1} \frac{575n^2 - 965n + 273}{\binom{3n}{n} 2^n} &= 6 \log 2.\end{aligned}$$

Moreover, the recursions derived for  $b_2(k)$  and  $b_3(k)$ , namely  $b_2(k) = \sum_j c_j^k(2)d_2(j)$  and  $b_3(k) = \sum_j c_j^k(3)d_3(j)$ , are sufficiently concise that we can compute symbolic values such as those of  $b_3(200)$  or of  $b_2(300)$  in a few seconds.

Without the factor of  $1/2^n$ , the sum  $b_3$  would not have such an evaluation in terms of simple constants. In general, the position of the poles of  $L_{-k}(ax^2(1-x))$ , i.e., of the zeros of  $f_a(x) = 1 - ax^2(1-x)$ , determines the evaluation of  $\sum a^n n^k / \binom{3n}{n}$ . If  $f_a$  has a sufficiently simple factorization, then we can expect an evaluation of the sum in terms of more basic constants. For example, with  $a = -1/4$ , we get (using *Maple* to evaluate the integral)

$$\begin{aligned}\sum_{n \geq 1} \frac{(-1)^n}{\binom{3n}{n} 4^n} &= -\frac{1}{28} - \frac{3}{32} \log(2) + \frac{13}{112} \frac{\arctan\left(\frac{\sqrt{7}}{5}\right)}{\sqrt{7}}, \\ \sum_{n \geq 1} \frac{(-1)^n n}{\binom{3n}{n} 4^n} &= -\frac{81}{1568} - \frac{9}{256} \log(2) + \frac{17}{6272} \frac{\arctan\left(\frac{\sqrt{7}}{5}\right)}{\sqrt{7}},\end{aligned}$$

while for  $a = 1$  *Maple* returns an expression which can only be simplified to

$$\sum_{n \geq 1} \frac{1}{\binom{3n}{n}} = \frac{4}{23} + \frac{2}{23} \sum_{23r^3 + 55r + 23 = 0} r \log(1987 - 598r + 621r^2).$$

Similarly, the sums  $\sum a^n n^k / \binom{2n}{n}$  lead to similar recursions; for example the classical

$$\sum_{n \geq 1} (-1)^n \frac{n^k}{\binom{2n}{n}} = r_k + s_k \frac{\operatorname{arctanh}(1/\sqrt{5})}{\sqrt{5}},$$

with appropriate rationals  $r_k, s_k$ . Another tractable example is the sum  $\sum n^k / \binom{4n}{2n}$ , which has

$$\sum_{n \geq 1} \frac{n^k}{\binom{4n}{2n}} = r_k + s_k \frac{\pi}{\sqrt{3}} + t_k \frac{\operatorname{arctanh}(1/\sqrt{5})}{\sqrt{5}},$$

again for appropriate rationals.

### 1.7.2 Some Results for Negative $k$

For  $k = -1$  and  $k = -2$ , the sums  $b_2(k)$  and  $b_3(k)$  can still be computed explicitly via the integrals (1.7.25) and (1.7.26):

$$\begin{aligned} \sum_{n \geq 1} \frac{1}{n \binom{2n}{n}} &= \frac{1}{3} \frac{\pi}{\sqrt{3}}, \\ \sum_{n \geq 1} \frac{1}{n^2 \binom{2n}{n}} &= \frac{1}{18} \pi^2 = \frac{\zeta(2)}{3}, \\ \sum_{n \geq 1} \frac{1}{n \binom{3n}{n} 2^n} &= \frac{1}{10} \pi - \frac{1}{5} \log(2), \\ \sum_{n \geq 1} \frac{1}{n^2 \binom{3n}{n} 2^n} &= \frac{1}{24} \pi^2 - \frac{1}{2} \log(2)^2. \end{aligned}$$

For  $k < -2$ , however, it seems that the integrals have no accessible antiderivative, so that a direct computation does not appear possible. One might conjecture that the sums  $b_3(k)$  then have no explicit reduction to simpler constants, but that conjecture would be wrong. One just has to expand the range of constants among which to hunt for a relation. In fact, it turns out that *multi-dimensional polylogarithms*

$$L_{a_1, \dots, a_m}(z) = \sum_{n_1 > \dots > n_m > 0} \frac{z^{n_1}}{n_1^{a_1} \dots n_m^{a_m}}$$

(with positive integers  $a_j$ ) will appear in the evaluations, for suitable  $z$ . For example, some of the relations proved in [54], as part of a more comprehensive

analysis, are the following (note that  $L_n(1) = \zeta(n)$ ):

$$\begin{aligned} \sum_{n \geq 1} \frac{1}{n^3} \binom{2n}{n} &= \frac{2}{3} \pi \operatorname{Im} (L_2(e^{i\pi/3})) - \frac{4}{3} \zeta(3), \\ \sum_{n \geq 1} \frac{1}{n^4} \binom{2n}{n} &= \frac{17}{36} \zeta(4), \\ \sum_{n \geq 1} \frac{1}{n^5} \binom{2n}{n} &= 2\pi \operatorname{Im} (L_4(e^{i\pi/3})) - \frac{19}{3} \zeta(5) + \frac{2}{3} \zeta(3) \zeta(2), \\ \sum_{n \geq 1} \frac{1}{n^6} \binom{2n}{n} &= -\frac{4}{3} \pi \operatorname{Im} (L_{4,1}(e^{i\pi/3})) + \frac{3341}{1296} \zeta(6) - \frac{4}{3} \zeta^2(3). \end{aligned}$$

In [54, 59], the terms such as  $\operatorname{Im} (L_{4,1}(e^{i\pi/3}))$  were termed *Clausen functions*.

Motivated by these results, one may conjecture that also the sums  $b_3(k)$  for negative  $k$  can be evaluated in terms of multi-dimensional polylogarithms, with suitable parameters and at suitable points  $z$ . To find the right parameters, we employ integer relation detection schemes between the sums  $b_3(k)$  and various polylogarithms. This of course involves a lot of trial and error. But in the end, the following evaluations are found by PSLQ.

None of these evaluations are proved yet. They are almost certainly true, since they have been verified to at least 100 digits. Only the evaluation for  $b_3(-3)$  can be proved rigorously, by laborious polylog manipulations—the interested

reader may feel challenged to try it.

$$\begin{aligned}
\sum_{n=1}^{\infty} \frac{1}{n^3 \binom{3n}{n} 2^n} &= -\frac{33}{16} \zeta(3) + \frac{1}{6} \log^3(2) - \frac{1}{24} \pi^2 \log(2) + \pi \operatorname{Im}(L_2(i)) \\
&= -\frac{1}{4} \zeta(3) + \frac{1}{6} \log^3(2) - \frac{1}{24} \pi^2 \log(2) - 4 \operatorname{Re}(L_{2,1}(i)) \\
&= -\frac{39}{16} \zeta(3) + \frac{1}{8} \pi^2 \log(2) - 4 \operatorname{Re}(L_{2,1}(\tfrac{1+i}{2})) + 4 \operatorname{Re}(L_3(\tfrac{1+i}{2})), \\
\sum_{n=1}^{\infty} \frac{1}{n^4 \binom{3n}{n} 2^n} &= -\frac{143}{16} \zeta(3) \log(2) + \frac{91}{640} \pi^4 - \frac{3}{8} \log^4(2) + \frac{3}{8} \pi^2 \log^2(2) \\
&\quad - 8 L_4(\tfrac{1}{2}) - 8 \operatorname{Re}(L_{3,1}(\tfrac{1+i}{2})) - 8 \operatorname{Re}(L_4(\tfrac{1+i}{2})), \\
\sum_{n=1}^{\infty} \frac{1}{n^5 \binom{3n}{n} 2^n} &= \frac{405}{32} \zeta(5) + \frac{21}{4} \zeta(3) \log^2(2) - \frac{1}{10} \pi^4 \log(2) - \frac{23}{144} \pi^2 \log^3(2) \\
&\quad - \frac{13}{8} \pi^2 \zeta(3) + \frac{1}{240} \log^5(2) - 13 L_5(\tfrac{1}{2}) - \frac{25}{2} L_{4,1}(\tfrac{1}{2}) \\
&\quad + 3 \pi \operatorname{Im}(L_4(i)) - 16 \operatorname{Re}(L_{4,1}(\tfrac{1+i}{2})) + 16 \operatorname{Re}(L_5(\tfrac{1+i}{2})).
\end{aligned}$$

Note that  $\operatorname{Im}(L_2(i)) = G$ , namely Catalan's constant.

Of course, there are still relations between the various polylogarithmic constants employed here (as is evidenced, for example, by the different evaluations for  $b_3(-3)$ ; note also that  $L_2(1/2) = \pi^2/12 - \log^2(2)/2$  and  $L_3(1/2) = 7\zeta(3)/8 - \pi^2 \log(2)/12 + \log^3(2)/6$ ). This means that evaluations other than the ones given above are possible. We have tried to find those evaluations with the smallest rational factors. Again, we must emphasize that the evaluations given here are still conjectural.

## 1.8 Continued Fractions of Tails of Series

We have already seen several examples of continued fractions in this chapter—see for example the formulas 1.4.16 and 1.4.17. In this section, we observe that the tails of the Taylor series for many standard functions, such as arctan and log, can be expressed as continued fractions in a variety of ways. A surprising side effect is that some of these continued fractions provide dramatic accelerations for the underlying power series. These investigations were motivated by a surprising observation about Gregory's series (see Sections 1.3 and 5.6.4 of the first volume).

### 1.8.1 Gregory's Series Reexamined

As discussed in Section 1.3 of the first volume, Gregory's series for  $\pi$ ,

$$\pi = 4 \sum_{k=1}^{\infty} \frac{(-1)^{k+1}}{2k-1} = 4(1 - 1/3 + 1/5 - 1/7 + \cdots), \quad (1.8.33)$$

when truncated to 5,000,000 terms, gives a value that differs strangely from the true value of  $\pi$ :

```
3.14159245358979323846464338327950278419716939938730582097494182230781640...
3.14159265358979323846264338327950288419716939937510582097494459230781640...
      2              -2              10              -122             2770
```

The series value differs, as one might expect from a series truncated to 5,000,000 terms, in the seventh decimal place—a “4” where there should be a “6”. But the next 13 digits are correct! Then, following another erroneous digit, the sequence is once again correct for an additional 12 digits. This pattern continues as shown. It is explained, *ex post facto*, by substituting  $N = 10^7$  in the result below:

**Theorem 1.8.1** *For integer  $N$  divisible by 4 the following asymptotic expansion holds:*

$$\begin{aligned} \frac{\pi}{2} - 2 \sum_{k=1}^{N/2} \frac{(-1)^{k-1}}{2k-1} &\sim \sum_{m=0}^{\infty} \frac{E_{2m}}{N^{2m+1}} \\ &= \frac{1}{N} - \frac{1}{N^3} + \frac{5}{N^5} - \frac{61}{N^7} + \cdots, \end{aligned} \quad (1.8.34)$$

where the coefficients are the even Euler numbers 1, -1, 5, -61, 1385, -50521...

The observation on the digits in the Gregory series arrived in the mail from Joseph Roy North in 1987. After verifying its truth numerically (which is much quicker today), it was an easy matter to generate a large number of the “errors” to high precision. The authors of [37] then recognized the sequence of errors above as the Euler numbers—with the help of Sloane's *Handbook of Integer Sequences*. The presumption that this sequence of errors is a form of Euler-Maclaurin summation is now formally verifiable for any fixed  $N$  in *Maple*. This

allowed them to determine that this phenomenon is equivalent to a set of identities between Bernoulli and Euler numbers, which could with considerable effort have been established. Secure in the knowledge that this observation holds, it is then easier, however, to use the *Boole summation formula*, which applies directly to alternating series and Euler numbers (see [37]). Because  $N$  was a power of ten, the asymptotic expansion was obvious on the computer screen.

This is a good example of a phenomenon that really does not become apparent without working to reasonably high precision (who recognizes 2, -2, 10?), and which highlights the role of pattern recognition and hypothesis validation in experimental mathematics.

It was an amusing additional exercise to compute  $\pi$  to 5,000 digits from the Gregory series. Indeed, with  $N = 200,000$  and correcting using the first thousand even Euler numbers, Borwein and Limber [62] obtained 5,263 digits of  $\pi$  (plus 12 guard digits). Thus, while the alternating Gregory series is very slowly convergent, the errors are highly predictable.

## 1.8.2 Euler's Continued Fraction

Identities such as

$$\begin{aligned} & a_0 + a_1 + a_1 a_2 + a_1 a_2 a_3 + a_1 a_2 a_3 a_4 & (1.8.35) \\ = a_0 + & \frac{a_1}{1 - \frac{a_2}{1 + a_2 - \frac{a_3}{1 + a_3 - \frac{a_4}{1 + a_4}}}} \end{aligned}$$

are easily verified symbolically. The general form can then be obtained by substituting  $a_N + a_N a_{N+1}$  for  $a_N$  and checking that the shape of the right hand side is preserved. This allows many series to be reexpressed as finite continued fractions. For example, with  $a_0 = 0, a_1 = x, a_2 = -x^2/3, a_3 = -3x^2/5, \dots$  we obtain, in the limit, the continued fraction for arctan due to Euler:

$$\arctan(x) = \frac{x}{1 + \frac{x^2}{3 - x^2 + \frac{9x^2}{5 - 3x^2 + \frac{25x^2}{7 - 5x^2 + \dots}}}}. \quad (1.8.36)$$

When  $x = 1$ , this becomes the first continued fraction for  $2/\pi$  given by Lord Brouncker (1620–1684):

$$\frac{2}{\pi} = \frac{1}{1 + \frac{9}{2 + \frac{25}{2 + \frac{49}{2 + \dots}}}}$$

If we let  $a_0 = \sum_1^N b_k$  be the initial segment of a similar series, we may use (1.8.35) to replace the next  $M$  terms, say, by a continued fraction. Applied to arctan this leads to:

$$\begin{aligned} \arctan(z) = \sum_{n=1}^N (-1)^{n-1} \frac{z^{2n-1}}{2n-1} + \frac{(-1)^N z^{2N+1}}{2N+1} + \frac{(2N+1)^2 z^2}{(2N+3) - (2N+1)z^2} + \\ \frac{(2N+3)^2 z^2}{(2N+5) - (2N+3)z^2} + \frac{(2N+5)^2 z^2}{(2N+7) - (2N+5)z^2} + \dots \end{aligned} \quad (1.8.37)$$

### 1.8.3 Gauss's Continued Fraction

An immediately richer vein lies in Gauss's continued fraction for the ratio of two hypergeometric functions  $\frac{F(a, b+1; c+1; z)}{F(a, b; c; z)}$ , see [207]. Recall that within its radius of convergence, the Gaussian hypergeometric function is defined by

$$\begin{aligned} F(a, b; c; z) = 1 + \frac{ab}{c} z + \frac{a(a+1)b(b+1)}{c(c+1)} z^2 \\ + \frac{a(a+1)(a+2)b(b+1)(b+2)}{c(c+1)(c+2)} z^3 + \dots \end{aligned} \quad (1.8.38)$$

The general continued fraction is developed by a reworking of the *contiguity relation*

$$F(a, b; c; z) = F(a, b+1; c+1; z) - \frac{a(c-b)}{c(c+1)} z F(a, b+1; c+2; z),$$

and formally, at least, this is quite easy to derive. Convergence and convergence estimates are more delicate. In the limit, for  $b = 0$ , this process yields

$$F(a, 1; c; z) = \frac{1}{1 - \frac{\frac{a}{c}z}{1 - \frac{\frac{(c-a)}{c(c+1)}z}{1 - \frac{\frac{c(a+1)}{(c+1)(c+2)}z}{1 - \frac{\frac{2(c-a+1)}{(c+2)(c+3)}z}{1 - \frac{\frac{(c+1)(a+2)}{(c+3)(c+4)}z}{1 - \dots}}}}}}}} \quad (1.8.39)$$

which is the case of present interest.

It is well known and easy to verify that  $\log(1+z) = zF(1, 1; 2; -z)$ . It is then a pleasant surprise to discover that  $\log(1+z) - z = \frac{1}{2}z^2F(2, 1; 3; -z)$ ,  $\log(1+z) - z + \frac{1}{2}z^2 = \frac{1}{3}z^3F(3, 1; 4; -z)$ , and to conjecture that

$$\log(1+z) + \sum_{n=1}^{N-1} \frac{(-1)^n z^n}{n} = \frac{z^N}{N} F(N, 1; N+1; -z). \quad (1.8.40)$$

This is easy to first verify for a few cases and then confirm rigorously. As always, a formula for  $\log$  leads correspondingly to one for  $\arctan$ :

$$\arctan(z) - \sum_{n=0}^{N-1} \frac{(-1)^n z^{2n+1}}{2n+1} = \frac{z^{2N+1}}{2N+1} F\left(N + \frac{1}{2}, 1; N + \frac{3}{2}; -z^2\right). \quad (1.8.41)$$

Happily, in both cases (1.8.39) is applicable—as it is for a variety of other functions, including for example  $\log[(1+z)/(1-z)]$ ,  $(1+z)^k$ , and  $\int_0^z (1+t^n)^{-1} dt = zF(1/n, 1; 1+1/n; -z^n)$ . Note that this last function recaptures  $\log(1+z)$  and  $\arctan(z)$  for  $n = 1$  and  $2$  respectively.

We give the explicit continued fractions for (1.8.40) and (1.8.41) in the conventional more compact form.

**Theorem 1.8.2** *Gauss's continued fractions for  $\log$  and  $\arctan$  are:*

$$\begin{aligned} & \log(1+z) + \sum_{n=1}^{N-1} \frac{(-1)^n z^n}{n} \\ &= \frac{(-1)^{N+1} z^N}{N} + \frac{N^2 z}{N+1} + \frac{1^2 z}{N+2} + \frac{(N+1)^2 z}{N+3} + \frac{2^2 z}{N+4} + \dots \end{aligned} \quad (1.8.42)$$



and

$$\begin{aligned} \arctan(z) &= \sum_{n=0}^{N-1} \frac{(-1)^n z^{2n+1}}{2n+1} \\ &= \frac{(-1)^N z^{2N+1}}{2N+1} + \frac{(2N+1)^2 z^2}{2N+3} + \frac{2^2 z^2}{2N+5} + \frac{(2N+3)^2 z^2}{2N+7} + \frac{4^2 z^2}{2N+9} + \cdots \end{aligned} \quad (1.8.43)$$

See [56] for details.

Suppose we return to Gregory's series, but add a few terms of the continued fraction for (1.8.41). One observes numerically that if the results are with  $N = 500,000$ , adding only five terms of the continued fraction has the effect of increasing the precision by more than 30 digits.

**Example 1.8.3** *Hypergeometric functions.*

Let

$$\begin{aligned} E_1(N, M, z) &= \log(1+z) \\ &\quad - \left( -\sum_{n=1}^N \frac{(-z)^n}{n} - \frac{(-z)^{N+1}}{N+1} F_M(N+1, 1; N+2; -z) \right) \\ E_2(N, M, z) &= \arctan(z) \\ &\quad - \left( -\sum_{n=0}^{N-1} \frac{(-1)^n z^{2n+1}}{2n+1} - \frac{(-1)^N z^{2N+1}}{2N+1} F_M\left(N + \frac{1}{2}, 1; N + \frac{3}{2}; -z^2\right) \right). \end{aligned} \quad (1.8.44)$$

Then  $E_1(N, M, z)$  and  $E_2(N, M, z)$  measure the precision of the approximations to  $\log(1+z)$  and  $\arctan(x)$  by the first  $N$  terms of Taylor series and then adding  $M$  terms of their continued fractions respectively. Let

$$E^*(N, M) = E_2(N, M, 1/2) + E_2(N, M, 1/5) + E_2(N, M, 1/8).$$

Tables 1.1, 1.2, 1.3, and 1.4 record the data for the approximations to the constants  $\log(1.9)$ ,  $\log(2)$ ,  $\arctan(1)$  and  $\arctan(1/2) + \arctan(1/5) + \arctan(1/8)$  respectively. Note that  $\arctan(1) = \arctan(1/2) + \arctan(1/5) + \arctan(1/8)$  is a Machin formula we saw in Chapter 3 of the first volume.  $\square$

		$5 \times 10$	$5 \times 10^2$	$5 \times 10^3$	$5 \times 10^4$
$M$	0	$0.48 \times 10^{-4}$	$0.13 \times 10^{-25}$	$0.15 \times 10^{-232}$	$0.13 \times 10^{-2292}$
	1	$0.43 \times 10^{-4}$	$0.11 \times 10^{-25}$	$0.14 \times 10^{-232}$	$0.11 \times 10^{-2292}$
	2	$0.40 \times 10^{-8}$	$0.11 \times 10^{-31}$	$0.14 \times 10^{-240}$	$0.11 \times 10^{-2302}$
	3	$0.34 \times 10^{-8}$	$1.00 \times 10^{-32}$	$0.12 \times 10^{-240}$	$0.10 \times 10^{-2302}$
	4	$0.12 \times 10^{-11}$	$0.40 \times 10^{-37}$	$0.50 \times 10^{-248}$	$0.41 \times 10^{-2312}$
	5	$0.10 \times 10^{-11}$	$0.35 \times 10^{-37}$	$0.45 \times 10^{-248}$	$0.37 \times 10^{-2312}$
	6	$0.78 \times 10^{-15}$	$0.31 \times 10^{-42}$	$0.40 \times 10^{-255}$	$0.33 \times 10^{-2321}$

Table 1.1: Error  $|E_1(N, M, 0.9)|$  for  $N = 5 \times 10^k (1 \leq k \leq 4)$  and  $0 \leq M \leq 6$ 

		$5 \times 10$	$5 \times 10^2$	$5 \times 10^3$	$5 \times 10^4$	$5 \times 10^5$	$5 \times 10^6$
$M$	0	$0.99 \times 10^{-2}$	$1.00 \times 10^{-3}$	$1.00 \times 10^{-4}$	$1.00 \times 10^{-5}$	$1.00 \times 10^{-6}$	$1.00 \times 10^{-7}$
	1	$0.97 \times 10^{-2}$	$1.00 \times 10^{-3}$	$1.00 \times 10^{-4}$	$1.00 \times 10^{-5}$	$1.00 \times 10^{-6}$	$1.00 \times 10^{-7}$
	2	$0.91 \times 10^{-6}$	$1.00 \times 10^{-9}$	$1.00 \times 10^{-12}$	$1.00 \times 10^{-15}$	$1.00 \times 10^{-18}$	$1.00 \times 10^{-21}$
	3	$0.86 \times 10^{-6}$	$1.00 \times 10^{-9}$	$1.00 \times 10^{-12}$	$1.00 \times 10^{-15}$	$1.00 \times 10^{-18}$	$1.00 \times 10^{-21}$
	4	$0.31 \times 10^{-9}$	$0.39 \times 10^{-14}$	$0.40 \times 10^{-19}$	$0.40 \times 10^{-24}$	$0.40 \times 10^{-29}$	$0.40 \times 10^{-34}$
	5	$0.28 \times 10^{-9}$	$0.39 \times 10^{-14}$	$0.40 \times 10^{-19}$	$0.40 \times 10^{-24}$	$0.40 \times 10^{-29}$	$0.40 \times 10^{-34}$
	6	$0.22 \times 10^{-12}$	$0.34 \times 10^{-19}$	$0.36 \times 10^{-26}$	$0.36 \times 10^{-33}$	$0.36 \times 10^{-40}$	$0.36 \times 10^{-47}$

Table 1.2: Error  $|E_1(N, M, 1)|$  for  $N = 5 \times 10^k (1 \leq k \leq 6)$  and  $0 \leq M \leq 6$ 

		$5 \times 10$	$5 \times 10^2$	$5 \times 10^3$	$5 \times 10^4$	$5 \times 10^5$	$5 \times 10^6$
$M$	0	$0.50 \times 10^{-2}$	$0.50 \times 10^{-3}$	$0.50 \times 10^{-4}$	$0.50 \times 10^{-5}$	$0.50 \times 10^{-6}$	$0.50 \times 10^{-7}$
	1	$0.48 \times 10^{-2}$	$0.50 \times 10^{-3}$	$0.50 \times 10^{-4}$	$0.50 \times 10^{-5}$	$0.50 \times 10^{-6}$	$0.50 \times 10^{-7}$
	2	$0.44 \times 10^{-6}$	$0.49 \times 10^{-9}$	$0.50 \times 10^{-12}$	$0.50 \times 10^{-15}$	$0.50 \times 10^{-18}$	$0.50 \times 10^{-21}$
	3	$0.42 \times 10^{-6}$	$0.49 \times 10^{-9}$	$0.50 \times 10^{-12}$	$0.50 \times 10^{-15}$	$0.50 \times 10^{-18}$	$0.50 \times 10^{-21}$
	4	$0.15 \times 10^{-9}$	$0.19 \times 10^{-14}$	$0.20 \times 10^{-19}$	$0.20 \times 10^{-24}$	$0.20 \times 10^{-29}$	$0.20 \times 10^{-34}$
	5	$0.14 \times 10^{-9}$	$0.19 \times 10^{-14}$	$0.20 \times 10^{-19}$	$0.20 \times 10^{-24}$	$0.20 \times 10^{-29}$	$0.20 \times 10^{-34}$
	6	$0.10 \times 10^{-12}$	$0.17 \times 10^{-19}$	$0.18 \times 10^{-26}$	$0.18 \times 10^{-33}$	$0.18 \times 10^{-40}$	$0.18 \times 10^{-47}$

Table 1.3: Error  $|E_2(N + 1, M, 1)|$  for  $N = 5 \times 10^k (1 \leq k \leq 6)$  and  $0 \leq M \leq 6$

		$5 \times 10$	$5 \times 10^2$
$M$	0	$0.31 \times 10^{-32}$	$0.37 \times 10^{-304}$
	1	$0.19 \times 10^{-33}$	$0.23 \times 10^{-305}$
	2	$0.11 \times 10^{-37}$	$0.15 \times 10^{-311}$
	3	$0.26 \times 10^{-38}$	$0.37 \times 10^{-312}$
	4	$0.56 \times 10^{-42}$	$0.92 \times 10^{-318}$
	5	$0.13 \times 10^{-42}$	$0.23 \times 10^{-318}$
	6	$0.59 \times 10^{-46}$	$0.13 \times 10^{-323}$

Table 1.4: Error  $|E^*(N, M)|$  for  $N = 5 \times 10^k (1 \leq k \leq 2)$  and  $0 \leq M \leq 6$ 

After some further numerical experimentation, it is clear that for large  $a, c$  the continued fraction  $F(a, 1, c; z)$  is rapidly convergent. And indeed the rough rate is apparent. This is part of the content of the next theorem:

**Theorem 1.8.4** ([56]) *Suppose  $-1 \leq z < 0$ , with  $a \geq 2$  and  $a + 1 \leq c \leq 2a$ . Then the following error estimate holds for all  $M \geq 2$ :*

$$\begin{aligned} & |F(a, 1, c; z) - F_M(a, 1; c; z)| \\ & \leq \frac{\Gamma(n+1)(n+a)\Gamma(n+c-a)\Gamma(a)\Gamma(c)}{\Gamma(n+a)\Gamma(n+c)a\Gamma(c-a)} \left( \frac{2a}{(c-2)\left(1-\frac{2}{z}\right) + (2a-c)} \right)^M, \end{aligned}$$

where  $n = \lfloor M/2 \rfloor$  and  $F_M(a, 1; c; z)$  is the  $M$ -th convergent of the continued fraction to  $F(a, 1, c; z)$ .

We leave it as an exercise to compare the estimates in Theorem 1.8.4 with the computed errors in Tables 1.1 and 1.2 (using  $a = N$  and  $c = N + 1$ ) and Table 1.3 (using  $a = N + 1/2$  and  $c = N + 3/2$ ). The results are very good.

In [207], one can find listed many explicit continued fractions, which can be derived from Gauss's continued fraction or various of its limiting cases. These include  $\exp$ ,  $\tanh$ ,  $\tan$  and various less elementary functions. One especially attractive fraction is that for  $J_{n-1}(z)/J_n(z)$  and  $I_{n-1}(z)/I_n(z)$ , where  $J$  and  $I$  are

*Bessel functions of the first kind.* In particular,

$$\frac{J_{n-1}(2z)}{J_n(2z)} = \frac{n}{z} - \frac{\frac{z}{(n+1)}}{1 - \frac{\frac{z^2}{(n+1)(n+2)}}{1 - \frac{\frac{z^2}{(n+2)(n+3)}}{1 - \dots}}}. \quad (1.8.45)$$

Setting  $z = i$  and  $n = 1$  leads to the very beautiful continued fraction  $I_1(2)/I_0(2) = [1, 2, 3, 4, \dots]$ . In general, arithmetic simple continued fractions correspond to such ratios.

An example of a more complicated situation is:

$$\frac{(2z')^{2N+1} \mathbf{F}\left(N + \frac{1}{2}, \frac{1}{2}; N + \frac{3}{2}; z^2\right)}{(N+1) \binom{2N+2}{N+1} \mathbf{F}\left(-\frac{1}{2}; \cdot; z^2\right)} = \frac{\arcsin(z)}{\sqrt{1-z^2}} - \sigma_{2N}(z), \quad (1.8.46)$$

where  $\sigma_{2N}$  is the  $2N$ -th Taylor polynomial for  $(\arcsin z)/\sqrt{1-z^2}$ . Only for  $N = 0$  is this precisely of the form of Gauss's continued fraction.

### 1.8.4 Perron's Continued Fraction

Another continued fraction expansion is based on Stieltjes' work on the moment problem (see Perron [174]) and leads to similar acceleration. In volume 2, page 18 of [174], one finds a beautiful continued fraction for

$$\int_0^z \frac{t^\mu}{1+t} dt = \frac{z}{(\mu+1) + \frac{z}{(\mu+2) - (\mu+1)z + \frac{(\mu+2)^2 z}{(\mu+3) - (\mu+2)z + \dots}}}, \quad (1.8.47)$$

valid for  $\mu > -1, -1 < z \leq 1$ . One can observe that this can be proved by Euler's continued fraction if we write

$$\frac{1}{z^\mu} \int_0^z \frac{t^\mu}{1+t} dt = \frac{z}{\mu+1} - \frac{z^2}{\mu+2} + \frac{z^3}{\mu+3} - \frac{z^4}{\mu+4} + \dots$$

and observe that (1.8.47) follows from (1.8.35) in the limit.

Since

$$\frac{z^{\mu+1}}{\mu+1} F(\mu+1, 1; \mu+2; -z) = \int_0^z \frac{t^\mu}{1+t} dt, \quad (1.8.48)$$

$$\frac{z^{2\mu+1}}{2\mu+1} F\left(\mu+\frac{1}{2}, 1; \mu+\frac{3}{2}; -z^2\right) = \int_0^z \frac{t^{2\mu}}{1+t^2} dt, \quad (1.8.49)$$

for  $\mu > 0$ , on examining (1.8.40) and (1.8.41) this is immediately applicable to provide Euler continued fractions for the tail of the log and arctan series. Explicitly, we obtain:

**Theorem 1.8.5** *Perron's continued fractions for (1.8.40) and (1.8.41) are:*

$$\begin{aligned} & \log(1+z) + \sum_{n=1}^{N-1} \frac{(-1)^n z^n}{n} \\ &= \frac{(-1)^{N+1} z^N}{N} + \frac{N^2 z}{(N+1) - Nz} + \frac{(N+1)^2 z}{(N+2) - (N+1)z} + \dots \end{aligned} \quad (1.8.50)$$

and

$$\begin{aligned} & \arctan z - \sum_{n=0}^{N-1} \frac{(-1)^n z^{2n+1}}{2n+1} \\ &= \frac{(-1)^N z^{2N+1}}{2N+1} + \frac{(2N+1)^2 z^2}{(2N+3) - (2N+1)z^2} + \frac{(2N+3)^2 z^2}{(2N+5) - (2N+3)z^2} + \dots \end{aligned} \quad (1.8.51)$$

Moreover, while the Gauss and Euler/Perron continued fractions obtained are quite distinct, the convergence behavior is very similar to that of the previous section. Note also the coincidence of (1.8.51) and (1.8.37). Indeed as we have seen Theorem 1.8.5 coincides with a special case of (1.8.35).

## 1.9 Partial Fractions and Convexity

We consider a network *objective function*  $p_N$  given by

$$p_N(q) = \sum_{\sigma \in S_N} \left( \prod_{i=1}^N \frac{q_{\sigma(i)}}{\sum_{j=i}^N q_{\sigma(j)}} \right) \left( \sum_{i=1}^N \frac{1}{\sum_{j=i}^N q_{\sigma(j)}} \right)$$

summed over *all*  $N!$  permutations; so a typical term is

$$\left( \prod_{i=1}^N \frac{q_i}{\sum_{j=i}^N q_j} \right) \left( \sum_{i=1}^N \frac{1}{\sum_{j=i}^N q_j} \right).$$

For example, with  $N = 3$  this is

$$q_1 q_2 q_3 \left( \frac{1}{q_1 + q_2 + q_3} \right) \left( \frac{1}{q_2 + q_3} \right) \left( \frac{1}{q_3} \right) \left( \frac{1}{q_1 + q_2 + q_3} + \frac{1}{q_2 + q_3} + \frac{1}{q_3} \right).$$

This arose as the objective function in research into coupon collection and the researcher, Ian Affleck, wished to show  $p_N$  was *convex* on the positive orthant.

First we try to simplify the expression for  $p_N$ . The *partial fraction decomposition* gives:

$$\begin{aligned} p_1(x_1) &= \frac{1}{x_1}, \\ p_2(x_1, x_2) &= \frac{1}{x_1} + \frac{1}{x_2} - \frac{1}{x_1 + x_2}, \\ p_3(x_1, x_2, x_3) &= \frac{1}{x_1} + \frac{1}{x_2} + \frac{1}{x_3} - \frac{1}{x_1 + x_2} - \frac{1}{x_2 + x_3} - \frac{1}{x_1 + x_3} \\ &\quad + \frac{1}{x_1 + x_2 + x_3}. \end{aligned} \tag{1.9.52}$$

Partial fraction decompositions are another arena in which computer algebra systems are hugely useful. The reader is invited to try performing the third case in (1.9.52) by hand. It is tempting to predict the “same” pattern will hold for  $N = 4$ . This is easy to confirm (by computer if not by hand) and so we are led to:

**Conjecture 1.9.1** *For each  $N \in \mathbb{N}$  the function*

$$p_N(x_1, \dots, x_N) = \int_0^1 \left( 1 - \prod_{i=1}^N (1 - t^{x_i}) \right) \frac{dt}{t} \tag{1.9.53}$$

*is convex, indeed  $1/p_N$  is concave.*

One may check symbolically that this is true for  $N < 5$  via a large Hessian computation. But this is impractical for larger  $N$ . That said, it is easy to numerically sample the Hessian for much larger  $N$ , and it is always positive definite.

Unfortunately while the integral is convex, the integrand is not, or we would be done. Nonetheless, the process was already a success, as the researcher was able to rederive his objective function in the form of (1.9.53).

A year later, Omar Hjab suggested re-expressing (1.9.53) as the *joint expectation* of Poisson distributions. See “Convex,” *SIAM Electronic Problems and Solutions* at the URL <http://www.siam.org/journals/problems/99-002.htm>. Explicitly this leads to:

**Lemma 1.9.2** *If  $x = (x_1, \dots, x_n)$  is a point in the positive orthant  $\mathbb{R}_+^n$ , then*

$$\int_0^\infty \left( 1 - \prod_{i=1}^n (1 - e^{-tx_i}) \right) dt = \left( \prod_{i=1}^n x_i \right) \int_{\mathbb{R}_+^n} e^{-\langle x, y \rangle} \max(y_1, \dots, y_n) dy, \quad (1.9.54)$$

where  $\langle x, y \rangle = x_1 y_1 + \dots + x_n y_n$  is the Euclidean inner product.

**Proof.** Let us denote the left-hand side of (1.9.54) by  $f$ . Since

$$1 - e^{-tx_i} = x_i \int_0^t e^{-x_i y_i} dy_i,$$

it follows that

$$1 - \prod_{i=1}^n (1 - e^{-tx_i}) = \left( \prod_{i=1}^n x_i \right) \left( \int_{\mathbb{R}_+^n} e^{-\langle x, y \rangle} dy - \int_{S_t^n} e^{-\langle x, y \rangle} dy \right),$$

where

$$S_t^n = \{y \in \mathbb{R}_+^n \mid 0 < y_i \leq t \text{ for } i = 1, \dots, n\}.$$

Hence

$$f(x) = \left( \prod_{i=1}^n x_i \right) \int_0^\infty dt \int_{\mathbb{R}_+^n \setminus S_t^n} e^{-\langle x, y \rangle} dy = \left( \prod_{i=1}^n x_i \right) \int_0^\infty dt \int_{\mathbb{R}_+^n} e^{-\langle x, y \rangle} \chi_t(y) dy,$$

where

$$\chi_t(y) = \begin{cases} 1 & \text{if } \max(y_1, \dots, y_n) > t, \\ 0 & \text{otherwise.} \end{cases}$$

Therefore

$$\begin{aligned} f(x) &= \left( \prod_{i=1}^n x_i \right) \int_{\mathbb{R}_+^n} e^{-\langle x, y \rangle} dy \int_0^\infty \chi_t(y) dt \\ &= \left( \prod_{i=1}^n x_i \right) \int_{\mathbb{R}_+^n} e^{-\langle x, y \rangle} \max(y_1, \dots, y_n) dy. \end{aligned}$$

□

It follows from the lemma that

$$p_N(x) = \int_{\mathbb{R}_+^N} e^{-(y_1 + \dots + y_N)} \max\left(\frac{y_1}{x_1}, \dots, \frac{y_N}{x_N}\right) dy,$$

and hence that  $p_N$  is positive, decreasing and convex, as is the integrand. To derive the stronger result that  $1/p_N$  is concave we proceed as follows. Let

$$h(a, b) = \frac{2ab}{a+b}.$$

Then  $h$  is concave and concavity of  $1/p_N$  is equivalent to

$$p_N\left(\frac{x+x'}{2}\right) \leq h(p_N(x), p_N(x')) \text{ for all } x, x' \in \mathbb{R}_+^N. \quad (1.9.55)$$

To establish this, define

$$m(x, y) = \min\left(\frac{x_1}{y_1}, \dots, \frac{x_n}{y_n}\right) \text{ for } x, y \in \mathbb{R}_+^N.$$

Then, since

$$m(x, y) + m(x', y) \leq 2m\left(\frac{x+x'}{2}, y\right),$$

we have

$$\begin{aligned} p_N\left(\frac{x+x'}{2}\right) &= \int_{\mathbb{R}_+^n} \frac{e^{-(y_1 + \dots + y_n)}}{m\left(\frac{x+x'}{2}, y\right)} dy \leq \int_{\mathbb{R}_+^n} \frac{2e^{-(y_1 + \dots + y_n)}}{m(x, y) + m(x', y)} dy \\ &= \int_{\mathbb{R}_+^n} e^{-(y_1 + \dots + y_n)} h\left(\frac{1}{m(x, y)}, \frac{1}{m(x', y)}\right) dy \leq h(p_N(x), p_N(x')). \end{aligned}$$



Where we leave it to the reader to confirm that the final assertion follows since  $h$  is concave and  $\int_{\mathbb{R}_+^N} e^{-(y_1+\dots+y_N)} dy = 1$ . This is a form of *Jensen's inequality*.  $\square$

Observe that since  $h(a, b) \leq \sqrt{ab} \leq (a + b)/2$ , it follows from (1.9.55) that  $p_N$  is log-convex (and convex). A little more analysis of the integrand shows  $p_N$  is strictly convex. There is still no truly direct proof of the convexity of  $p_N$ .

An amusing related example, which cries out for generalization, is that for  $a > b > c > d > 0$ , the function

$$f(x) = \frac{a^x - b^x}{c^x - d^x}$$

is convex on the real line, but  $\log f(x)$  is convex on the real line only when  $ad \geq bc$ , and is concave on the real line when  $ad < bc$ . These assertions are fairly easy to deduce from:

**Proposition 1.9.3** *Let  $g_\mu(x) = (e^{2\mu x} - 1)/(e^{2x} - 1)$ ,  $\ell_\mu(x) = \log g_\mu(x)$ , and  $\ell_{\mu,\nu}(x) = \log(g_\mu(x) - g_\nu(x))$ . Then, for  $\mu > 1$ ,  $a > b > 1$ , and all real  $x$ ,*

1.  $g'_\mu(x) \geq 0$ ,  $\ell'_\mu(x) \geq 0$ ,  $\ell''_\mu(x) \geq 0$ ,  $g''_\mu(x) \geq 0$ ,
2.  $\ell''_a(x) - \ell''_b(x) \geq 0$ ,  $g''_a(x) - g''_b(x) \geq 0$ ,
3.  $\ell''_{a,b}(x) \geq 0$  when  $a - b \geq 1$ , and  $\ell''_{a,b}(x) < 0$  when  $a - b < 1$ .

Note that Item 2 says  $\ell_\mu$  becomes more convex as the parameter  $\mu$  increases, and similarly for  $g_\mu$ . Note also that

$$\log \left( \frac{\sinh(\mu x)}{\sinh(x)} \right) = \log \left( \frac{e^{2\mu x} - 1}{e^{2x} - 1} \right) - 2x.$$

As an example, with some care, the convex conjugate of the function  $f : x \mapsto \log(\sinh(3x)/\sinh(x))$  can be symbolically nursed to  $g : y \mapsto y/2 \cdot \log \left[ (y + \sqrt{-3y^2 + 16})/(-2y + 4) \right] + \log \left[ (-2 + \sqrt{-3y^2 + 16})/2 \right]$ . Since in turn the conjugate of  $g$  is much more easily computed to be  $f$ , this produces a symbolic computational proof that  $f$  and  $g$  are convex and are mutually conjugate.

## 1.10 Log-concavity of Poisson Moments

Recall that a sequence  $\{a_n\}$  is *log-convex* if  $a_{n+1}a_{n-1} \geq a_n^2$ , for  $n \geq 1$  and is *log concave* when the sign is reversed. Consider the *unsolved* Problem 10738 posed by Radu Theodorescu in the 1999 *American Mathematical Monthly* [200]:

**Problem:** For  $t > 0$  let

$$m_n(t) = \sum_{k=0}^{\infty} k^n \exp(-t) \frac{t^k}{k!}$$

be the  $n$ th moment of a *Poisson distribution* with parameter  $t$ . Let  $c_n(t) = m_n(t)/n!$ . Show

- a)  $\{m_n(t)\}_{n=0}^{\infty}$  is log-convex for all  $t > 0$ .
- b)  $\{c_n(t)\}_{n=0}^{\infty}$  is not log-concave for  $t < 1$ .
- c\*)  $\{c_n(t)\}_{n=0}^{\infty}$  is log-concave for  $t \geq 1$ .

**Solution.** (a) Neglecting the factor of  $\exp(-t)$  as we may, this reduces to

$$\sum_{k,j \geq 0} \frac{(jk)^{n+1} t^{k+j}}{k!j!} \leq \sum_{k,j \geq 0} \frac{(jk)^n t^{k+j}}{k!j!} k^2 = \sum_{k,j \geq 0} \frac{(jk)^n t^{k+j}}{k!j!} \frac{k^2 + j^2}{2},$$

and this now follows from  $2jk \leq k^2 + j^2$ .

(b) As

$$m_{n+1}(t) = t \sum_{k=0}^{\infty} (k+1)^n \exp(-t) \frac{t^k}{k!},$$

on applying the binomial theorem to  $(k+1)^n$ , we see that  $m_n(t)$  satisfies the recurrence

$$m_{n+1}(t) = t \sum_{k=0}^n \binom{n}{k} m_k(t), \quad m_0(t) = 1.$$

In particular for  $t = 1$ , this produces the sequence

$$1, 1, 2, 4, 9, 21, 51, 127, 323, 835, 2188, 5798, 15511, 41835, \dots$$

These are the *Bell numbers*, which can be discovered from consulting Sloane's on-line integer sequence recognition tool that we have mentioned above. This tool can also tell us that for  $t = 2$ , we have obtained generalized Bell numbers, and can give us the exponential generating functions. The Bell numbers were known earlier to Ramanujan.

Now an explicit computation shows that

$$t \frac{1+t}{2} = c_0(t) c_2(t) \leq c_1(t)^2 = t^2$$

exactly if  $t \geq 1$ . Also, preparatory to the next part, a simple calculation shows that

$$\sum_{n \geq 0} c_n u^n = \exp(t(e^u - 1)). \quad (1.10.56)$$

(c\*) (The “\*” indicates this was the unsolved component.) We appeal to a recent theorem due to E. Rodney Canfield [74]. A search in 2001 on *MathSciNet* for “Bell numbers” since 1995 turned up 18 items. This article showed up as paper #10. Later, *Google* found the paper immediately! Canfield proves the lovely and quite difficult result below.

**Theorem 1.10.1** *If a sequence  $1, b_1, b_2, \dots$  is non-negative and log-concave, then so is the sequence  $1, c_1, c_2, \dots$  determined by the generating function equation*

$$\sum_{n \geq 0} c_n u^n = \exp\left(\sum_{j \geq 1} b_j \frac{u^j}{j}\right).$$

Using equation (1.10.56) above, we apply this to the sequence  $b_j = t/(j-1)!$  which is log-concave exactly for  $t \geq 1$ .  $\square$

Indeed, symbolic computation—facilitated by the recursion above—strongly suggests the *only* violation of log-concavity of the sequence  $\{c_n(t)\}_{n=0}^{\infty}$  for  $t > 0$ , occurs as illustrated in b). We have not been able to prove this conjecture. It seems to require a significant strengthening of Theorem 1.10.1 to cover the case when the first term of  $\{1, b_n\}$  is replaced by  $b_0 \neq 1$ .

It transpired that the given solution to (c) was the only one received by the *Monthly*. This is quite unusual. The reason might well be that it relied on the following sequence of steps:

Question  $\Rightarrow$  Computer Algebra System  $\Rightarrow$  Web Interface $\Rightarrow$  Search Engine  $\Rightarrow$  Digital Library $\Rightarrow$  Hard New Paper  $\Rightarrow$  Answer

Now if only we could automate this!

## 1.11 Commentary and Additional Examples

1. **Dictionaries are like timepieces.** Samuel Johnson observed that dictionaries are like clocks: The best do not run true, and the worst are better than none. The same is true of tables and databases. We quoted Michael Berry as saying “Compounding the impiety, I would give up Shakespeare in favor of Prudnikov, Brychkov and Marichev” [177]. That excellent compendium contains

$$\sum_{k=1}^{\infty} \sum_{l=1}^{\infty} \frac{1}{k^2 (k^2 - kl + l^2)} = \frac{\pi^{\alpha} \sqrt{3}}{30}, \quad (1.11.57)$$

where the “ $\alpha$ ” is probably “4” [177, vol. 1, entry 9, pg. 750]. Integer relation methods suggest that no reasonable value of  $\alpha$  works. What is intended in formula (1.11.57)? Note that

(a)

$$\begin{aligned} & \sum_{n=1}^{\infty} \sum_{m=1}^{\infty} \frac{2}{n^2 (n^2 - mn + m^2)} + \sum_{n=1}^{\infty} \sum_{m=1}^{\infty} \frac{2}{n^2 (n^2 + mn + m^2)} = \\ & \sum_{m,n \in \mathbb{Z}} \sum_{mn \neq 0} \frac{1}{n^2 (n^2 + mn + m^2)} = 6 \zeta(4) \end{aligned}$$

(b)

$$\sum_{n=1}^{\infty} \sum_{m=1}^{n-1} \frac{1}{nm (n^2 + mn + m^2)} + \sum_{n=1}^{\infty} \sum_{m=1}^{\infty} \frac{1}{n^2 (n^2 + mn + m^2)} = \frac{13}{12} \zeta(4)$$

(c)

$$\begin{aligned}
\sum_{n=1}^{\infty} \sum_{m=1}^{\infty} \frac{1}{n^2(n^2 + mn + m^2)} &= \frac{2}{\sqrt{3}} \operatorname{Im} \sum_{n=1}^{\infty} \frac{\Psi\left(1 + n \frac{-1+i\sqrt{3}}{2}\right)}{n^3} \\
&= \frac{2}{\sqrt{3}} \operatorname{Im} \int_0^{\infty} \frac{\operatorname{Li}_3\left(e^{\left(\frac{-1+i\sqrt{3}}{2}\right)t}\right)}{e^t - 1} dt \\
&= 1.00445719820157402755414025 \dots
\end{aligned}$$

2. **A series for pi with first term 22/7.** In [96] an estimate of  $\pi - 355/113$  is also given.

(a) Let  $P(t) = 4 - 4t^2 + 5t^4 - 4t^5 + t^6$  and  $Q(t) = t(1-t)$ . Show that

$$\frac{4}{1+t^2} = \frac{4 - 4t^2 + 5t^4 - 4t^5 + t^6}{1 + t^4(1-t)^4/4}$$

and so

$$\pi = \int_0^1 \frac{P(t)}{1 + Q^4(t)/4} dt = \frac{1}{2} \int_0^1 \frac{P(t) + P(1-t)}{1 + Q^4(t)/4} dt.$$

(b) Observe that  $P(t) + P(1-t) = 6 + 2Q(t) - Q^2(t) - 2Q^3(t)$  and deduce that

$$\pi = \frac{1}{2} \sum_{n=0}^{\infty} \left(-\frac{1}{4}\right)^n \int_0^1 \frac{6 + 2Q(t) - Q^2(t) - 2Q^3(t)}{1 + Q^4(t)/4} dt = \sum_{n=0}^{\infty} a_n$$

where

$$a_n = \left(-\frac{1}{4}\right)^n \left\{ \frac{3(4n)!^2}{(8n+1)!} + \frac{(4n+1)!^2}{(8n+3)!} - \frac{(4n+2)!^2}{2(8n+5)!} - \frac{(4n+3)!^2}{(8n+7)!} \right\}.$$

(c) This series, which can be neatly written as a sum of four  ${}_5F_4$  functions, gains roughly three digits per term.

Check that the series has constant term 22/7 and continues

$$\frac{22}{7} - \frac{76}{15015} \left(-\frac{1}{4}\right) + \frac{543}{37182145} \left(-\frac{1}{4}\right)^2 - \frac{308}{6511704225} \left(-\frac{1}{4}\right)^3 + \dots$$

(d) Find an economical way to show

$$\frac{355}{113} - \frac{33}{10^8} < \pi < \frac{355}{113} + \frac{24}{10^8}.$$

3. **A sophomore's dream.** Show that

$$\int_0^1 x^x dx = \sum_{n=1}^{\infty} \frac{(-1)^{n+1}}{n^n},$$

and that

$$\int_0^1 x^{-x} dx = \sum_{n=1}^{\infty} \frac{1}{n^n}.$$

Give meaning to and evaluate  $\int_0^1 x^{x^x} dx$ . In each case, it helps to start by writing the integrand as a series and to justify integrating term by term.

4. **Some binary digit algorithms.** Whenever a function satisfies a suitable addition formula, its inverse admits algorithms to compute binary (or other base) representations [57]. We begin with an introductory example. Let  $x \geq 0$ . Set

$$a_0 = x \quad \text{and} \quad a_{n+1} = \frac{2a_n}{1 - a_n^2} \quad (\text{with } a_{n+1} = -\infty \text{ if } a_n = \pm 1).$$

Then

$$\sum_{\substack{a_n < 0 \\ n \geq 0}} \frac{1}{2^{n+1}} = \frac{\arctan x}{\pi}.$$

Let an interval  $I \subseteq \mathbb{R}$  and subsets  $D_0, D_1 \subseteq I$  with  $D_0 \cup D_1 = I$  and  $D_0 \cap D_1 = \emptyset$  be given, as well as functions  $r_0 : D_0 \rightarrow I$ ,  $r_1 : D_1 \rightarrow I$ . Then consider the system (S) of the following two functional equations for an unknown function  $f : I \rightarrow [0, 1]$ .

$$2f(x) = f(r_0(x)) \quad \text{if } x \in D_0, \quad (\text{S}_0)$$

$$2f(x) - 1 = f(r_1(x)) \quad \text{if } x \in D_1. \quad (\text{S}_1)$$

Such a system always leads to an iteration:

$$a_0 = x \quad \text{and} \quad a_{n+1} = \begin{cases} r_0(a_n) & a_n \in D_0, \\ r_1(a_n) & a_n \in D_1. \end{cases} \quad (1.11.58)$$

Then:

$$f(a_0) = f(x) = \sum_{\substack{a_n \in D_1 \\ n \geq 0}} \frac{1}{2^{n+1}}. \quad (1.11.59)$$

**Several examples.** Here are some elementary transcendental functions  $f$  that satisfy a system of type (S) with algebraic  $r_0, r_1$ .

(a)  $f(x) = \log x / \log 2$ ,  $I = [1, 2]$ ,  $D_0 = [1, \sqrt{2}]$ ,  $D_1 = [\sqrt{2}, 2]$ ,  
 $r_0(x) = x^2$ ,  $r_1(x) = x^2/2$ .

Of course, the recursion (1.11.58) and (1.11.59) can be used to compute the binary expansion of the function  $f$ . Take for example  $x = a_0 = 3/2$ . Then  $a_1 = 9/8$ ,  $a_2 = 81/64$ ,  $a_3 = 6561/4096$ ,  $a_4 = 3^{16}/2^{25}$ , and so on. The first 20 binary digits of  $\log 3 / \log 2$  are  $(1.10010101110000000001)_2$ .

(b)  $f(x) = \arccos(x)/\pi$ ,  $I = [-1, 1]$ ,  $D_0 = (0, 1]$ ,  $D_1 = [-1, 0]$ ,  
 $r_0(x) = 2x^2 - 1$ ,  $r_1(x) = 1 - 2x^2$ .

Note that  $r_0^k = r_0 \circ \cdots \circ r_0$  is the Chebyshev polynomial of the first kind  $T_{2^k}$  for  $[-1, 1]$ .

(c)  $f(x) = 2 \arcsin(x)/\pi$ ,  $I = [0, 1]$ ,  $D_0 = [0, 1/\sqrt{2}]$ ,  $D_1 = [1/\sqrt{2}, 1]$ ,  
 $r_0(x) = 2x\sqrt{1-x^2}$ ,  $r_1(x) = 2x^2 - 1$ .

(d)  $f(x) = \begin{cases} \arctan(x)/\pi & x \in [0, \infty) \\ 1 + \arctan(x)/\pi & x \in [-\infty, 0) \end{cases}$ ,  $D_0 = [0, \infty)$ ,  $D_1 = [-\infty, 0)$ ,  
 $r_0(x) = \frac{2x}{1-x^2}$ ,  $r_0(1) = -\infty$ ,  $r_1(x) = \frac{2x}{1-x^2}$ ,  $r_1(-1) = -\infty$ .

(e)  $f(x) = \operatorname{arccot}(x)/\pi$ ,  $I = \mathbb{R} \cup \{-\infty\}$ ,  $D_0 = [0, \infty)$ ,  $D_1 = [-\infty, 0)$ ,  
 $r_0(x) = \frac{x^2-1}{2x}$ ,  $r_0(0) = -\infty$ ,  $r_1(x) = \frac{x^2-1}{2x}$ .

(f)  $f(x) = \operatorname{arsinh}(x)/\log 2$ ,  $I = [0, 3/4]$ ,  $D_0 = [0, 1/2\sqrt{2}]$ ,  $D_1 = [1/2\sqrt{2}, 3/4]$ ,  
 $r_0(x) = 2x\sqrt{1+x^2}$ ,  $r_1(x) = 5/2x\sqrt{1+x^2} - 3/2x^2 - 3/4$ .

(g)  $f(x) = \arccos(x)/\pi$  satisfies

$$\begin{aligned} 3f(x) &= f(4x^3 - 3x) && \text{if } x \in (1/2, 1], \\ 3f(x) - 1 &= f(-4x^3 + 3x) && \text{if } x \in (-1/2, 1/2], \\ 3f(x) - 2 &= f(4x^3 - 3x) && \text{if } x \in [-1, -1/2]. \end{aligned}$$

That means that ternary representations of  $f$  can be computed by the following recursion:

$$\text{Set } a_0 = x, \quad a_{n+1} = \begin{cases} 4a_n^3 - 3a_n & a_n \in (1/2, 1] \cup [-1, -1/2] \\ -4a_n^3 + 3a_n & a_n \in (-1/2, 1/2]. \end{cases}$$

Then

$$\frac{\arccos(x)}{\pi} = \sum_{a_n \in (-1/2, 1/2]} \frac{1}{3^{n+1}} + \sum_{a_n \in [-1, -1/2]} \frac{2}{3^{n+1}}.$$

Full details are given in [57].

5. **A two-term recursion.** Problem: Determine the behavior of the iteration  $u_{n+1} = |u_n| - u_{n-1}$  for arbitrary real starting points  $u_0 = x$  and  $u_1 = y$ . Then attempt to generalize this behavior. (Taken from [127].)

Solution: Numerical testing shows the iteration has period nine. One proof is to explicitly compute, (preferably using a symbolic math program) the messy looking algebraic function this determines. It can be seen that each case returns  $[x, y]$ . More explicitly

$$A = \begin{bmatrix} 0 & 1 \\ -1 & 1 \end{bmatrix} \quad B = \begin{bmatrix} 0 & 1 \\ -1 & -1 \end{bmatrix}$$

can be used to represent the iteration. Then  $B^3 = I = -A^3$  and analysis of the cases above shows that it always devolves to  $A^3 B A^3 B^2$  which is the identity.

A nice generalization is that for  $M, K > 1$  and integer, the iteration

$$u_{k+1} = \cos\left(\frac{\pi}{K}\right)(|u_k| + u_k) + \cos\left(\frac{\pi}{M}\right)(|u_k| - u_k) - u_{k-1}$$

has period  $KM + K - M$ . This can again be checked symbolically for many small  $M$  and  $K$ . One may prove this by considering

$$A = \begin{bmatrix} 0 & 1 \\ -1 & 2 \cos(\pi/K) \end{bmatrix} \quad B = \begin{bmatrix} 0 & 1 \\ -1 & 2 \cos(\pi/M) \end{bmatrix}$$

and reducing the iteration to  $(A^K B)^{M-2} A^K B^2$ .



6. **A rational function recursion.** Problem: Solve the recursion  $u_0 = 2$  and

$$u_{n+1} = \frac{2u_n + 1}{u_n + 2}.$$

(Taken from [127]). Solution: The first few numerators are

$$2, 5, 14, 41, 122, 365, 1094, 3281, 9842, 29525, 88574$$

and clearly satisfy  $n_{k+1} = 3n_k - 1$ . Sloane's online sequence recognition tool announces that the numerators are  $(3^n + 1)/2$ . The denominators are one smaller and so

$$u_n = \frac{3^{n+1} + 1}{3^{n+1} - 1}.$$

This can now easily be proved by induction.

7. **A sequence with nines.** Problem: Determine the number of nines in the tail of the sequence

$$u_{n+1} = 3u_n^4 + 4u_n^3,$$

with  $u_0 = 9$ . (Taken from [127]).

Solution: The first six or so cases show there is a 5 followed by  $2^k$  occurrences of 9. Define  $u_k = 6 \cdot 10^{2^k} - 1$  and show inductively that this is so.

8. **Continued fraction of Champernowne's number.** Compute the first ten or twenty terms of the continued fraction for Champernowne's number, namely the decimal constant  $0.123456789101112131415\dots$ . Explain the phenomenon observed.
9. **A sequence involving square roots.** Problem: Consider the iteration

$$c_{n+1} = c_n + r - \frac{c_n}{\sqrt{1 + c_n^2}}, \quad c_0 \geq 1,$$

where  $r$  is a positive constant. For which  $r$  does the sequence  $\{c_n\}$  converge? In case of convergence to  $c \neq c_0$ , prove that  $\lim(c_{n+1} - c)/(c_n - c)$  exists and determine its value. In case of divergence, find a precise asymptotic expression for  $c_n$ .

Solution: Justify the following assertions: When  $0 < r < 1$ , the sequence converges to  $c = r/\sqrt{1-r^2}$  and either  $c_n = c$  for every integer  $n \geq 0$ , or  $\lim(c_{n+1} - c)/(c_n - c) = 1 - (1 - r^2)^{3/2}$ . When  $r = 1$ ,  $c_n \sim (3n/2)^{1/3}$ ; and when  $r > 1$ ,  $c_n \sim (r - 1)n$ .

10. **A sequence involving exponentials.** Problem: Define a sequence  $\{t_k\}$  by setting

$$t_1 = 1, \quad t_{k+1} = t_k \exp(-t_k), \quad k = 1, 2, \dots$$

Determine the behavior of the sequence.

Solution: Note that  $t_k$  tends monotonically to a limit  $\ell$  which must necessarily be zero. Hence  $t_{k+1}^{-1} - t_k^{-1} = t_k^{-1}(\exp t_k - 1)$ , which tends to  $\exp'(0) = 1$  as  $k$  tends to infinity. Whence, since Cesàro averaging preserves limits,

$$\frac{1}{mt_m} = \frac{1}{m} \sum_{k=1}^{m-1} \frac{e^{t_k} - 1}{t_k} + \frac{1}{mt_1}$$

also tends to 1, and

$$\lim_{m \rightarrow \infty} mt_m = 1.$$

The reader is invited to perform a similar analysis for a more general  $g : [0, 1] \mapsto [0, 1]$ .

11. **Some double integrals.** Problem: Evaluate the following integrals:

(a)

$$\int_0^1 \int_0^1 \frac{1}{(1 - x^2 y^2)} dx dy \quad \left( = \frac{\pi^2}{8} \right)$$

(b)

$$\int_0^1 \int_0^1 \frac{1}{(1 - xy)} dx dy \quad \left( = \frac{\pi^2}{6} \right)$$

(c)

$$\int_{-1}^1 \int_{-1}^1 \frac{1}{\sqrt{1 + x^2 + y^2}} dx dy \quad \left( = 4 \log(2 + \sqrt{3}) - \frac{2\pi}{3} \right)$$

(d)

$$\frac{1}{4} \int_0^1 \int_0^1 \frac{1}{(x+y) \sqrt{(1-x)(1-y)}} dx dy \quad (= G)$$

(e)

$$\int_0^1 \int_0^1 \frac{1-x}{(1-xy)|\log(xy)|} dx dy \quad (= \gamma).$$

12. **A double summation.** Evaluate

$$\sum_{n=1}^{\infty} \sum_{m=1}^{\infty} \frac{1}{(m^2n + mn^2 + 2mn)} \quad (= 7/4).$$

13. **Infinite series and dilogarithms.** Prove that for positive integers  $a$  and  $b$ ,

$$\sum_{m=0}^{\infty} \frac{z^m}{(am+b)^2} = \frac{z^{-b/a}}{a} \sum_{k=0}^{a-1} e^{-2\pi i bk/a} \text{Li}_2(z^{1/a} e^{2\pi i k/a}),$$

and note the finitude of the sum. If one believes in Li evaluations as “fundamental,” then this leads to a finite expression for the Broadhurst  $V$  constant (mentioned in Chapter 2 of the first volume), as well as some other BBP sums.

14. **An infinite product evaluation.** Evaluate

$$\frac{z-1}{w-1} \prod_{n=1}^{\infty} \frac{w^{1/2^n} + 1}{z^{1/2^n} + 1}.$$

for  $w, z > 0$ . Hint: Examine the case  $w = 2$ .

15. **Quasi-elliptic integrals.** Show that

$$\int \frac{6x}{\sqrt{x^4 + 4x^3 - 6x^2 + 4x + 1}} dx = \log \left( a(x) + b(x) \sqrt{D(x)} \right) \quad (1.11.60)$$

where

$$\begin{aligned} a(x) &= x^6 + 12x^5 + 45x^4 + 44x^3 - 33x^2 + 43 \\ b(x) &= x^4 + 10x^3 + 30x^2 + 22x - 11 \\ D(x) &= x^4 + 4x^3 - 6x^2 + 4x + 1. \end{aligned}$$

Hint: Confirm that in (1.11.60)  $a' = 6x, /b$  and hence that (1.11.60) is a pseudo-elliptic integral in the sense described in [206]. The following is taken from the abstract of [206]:

In this report we detail the following story. Several centuries ago, Abel noticed that the well-known elementary integral

$$\int \frac{dx}{\sqrt{x^2 + 2bx + c}} = \log \left( x + b + \sqrt{x^2 + 2bx + c} \right)$$

is just a presage of more surprising integrals of the form

$$\int \frac{f(x)dx}{\sqrt{D(x)}} = \log \left( p(x) + q(x)\sqrt{D(x)} \right).$$

Here  $f$  is a polynomial of degree  $g$  and the  $D$  are certain polynomials of degree  $\deg D(x) = 2g+2$ . Specifically,  $f(x) = p'(x)/q(x)$  (so  $q$  divides  $p'$ ). Note that, morally, one expects such integrals to produce inverse elliptic functions and worse, rather than an innocent logarithm of an algebraic function.

Abel went on to study abelian integrals, and it was Chebyshev who explained—using continued fractions—what is going on with these “quasi-elliptic” integrals. Recently, the second author computed all the polynomials  $D$  over the rationals of degree 4 that have an  $f$  as above. We explain various contexts in which the present issues arise. These contexts include symbolic integration of algebraic functions, the study of units in function fields and, given a suitable polynomial  $g$ , the consideration of the period length of the continued fraction expansion of the numbers  $\sqrt{g(n)}$  as  $n$  varies over the integers. But the major content of this survey is an introduction to period continued fractions in hyperelliptic—thus quadratic—function fields.

16. **Clausen’s product.** Prove that

$$\begin{aligned} & {}_2F_1^2(a, b, a + b + 1/2, z) \\ = & {}_3F_2(2a, a + b, 2b, a + b + 1/2, 2a + 2b, z) \end{aligned} \quad (1.11.61)$$

Hint: Show both sides satisfy the following differential equation

$$\begin{aligned} 0 &= x^2(x-1)y^{(3)}(x) - 3x(a+b+1/2 - (a+b+1)x)y^{(2)}(x) \\ &+ [(2(a^2+b^2+4ab) + 3(a+b)+1)x - (a+b)(2(a+b)+1)]y'(x) \\ &+ 4ab(a+b)y(x), \end{aligned}$$

are analytic at zero and have appropriate initial values. Use Clausen's product to deduce that

$$\arcsin^2(x) = \frac{1}{2} \sum_{n=1}^{\infty} \frac{(2x)^{2n}}{n^2 \binom{2n}{n}}.$$

Hint:  $\arcsin(x) = x \cdot {}_2F_1(1/2, 1/2, 3/2; x^2)$ .

17. **An application of Clausen's product.** Consider the hypergeometric function

$$G_a : x \mapsto F\left(a, a + \frac{1}{2}, 2a + 1; x\right)$$

and show that

$$G_{ab} = G_a^b \quad \text{for all } a, b \in \mathbb{R}.$$

By considering  $G_{-1/2}$  determine the closed form for  $G_a$ .

18. **Putnam problem 1987-A6.** Let  $n$  be a positive integer and let  $a_3(n)$  be the number of zeroes in the ternary expansion of  $n$ . Determine for which positive  $x$  the series  $\sum_{n=1}^{\infty} x^{a_3(n)}/n^3$  converges. Answer: For  $x < 25$ . In the  $b$ -ary analogue  $\sum_{n=1}^{\infty} x^{a_b(n)}/n^b$  converges if and only if  $x < b^b - b + 1$ .
19. **Putnam problem 1987-B2.** For  $r, s$  nonnegative with  $r+s \leq t$ , evaluate

$$\sum_{k=0}^s \frac{\binom{s}{k}}{\binom{t}{r+k}} \quad \left( = \frac{t+1}{(t+1-s)\binom{t-s}{r}} \right).$$

20. **Putnam problem 1987-B4.** Let  $x_0 = 4/5$  and  $y_0 = 3/5$  and consider the dynamical system

$$x_{n+1} \leftarrow x_n \cos(y_n) - y_n \sin(y_n) \quad y_{n+1} \leftarrow x_n \sin(y_n) + y_n \cos(y_n).$$

Does the system converge and if so to what? Hint: It may help to consider  $z = x + iy$ .

21. **Putnam problem 1989–A2.** Evaluate

$$\int_0^a \int_0^b e^{\max(a^2 y^2, b^2 x^2)} dy dx.$$

Hint:

$$\frac{2}{ab} \int_0^{ab} \int_0^z e^{z^2} dw dz = \frac{e^{a^2 b^2} - 1}{ab}.$$

22. **Putnam problem 1990–A1.** Let  $T_0 = 2$ ,  $T_1 = 3$ , and  $T_2 = 6$ . Find a simple formula for  $T_n$  where

$$T_n = (n + 4) T_{n-1} - 4n T_{n-2} + (4n - 8) T_{n-3} \quad (= n! + 2^n).$$

23. **Putnam problem 1990–B5.** Is there an infinite sequence of non zero reals  $\{a_n\}$  so that

$$p(x) = \sum_{k=0}^n a_k x^k$$

has  $n$  distinct real roots for all  $n$ ? Hint: Let  $a_n = (-10)^{-n^2}$  and evaluate  $p$  at  $10^{2k}$ .

24. **Putnam problem 1993–A2.** Suppose that a sequence  $\{x_n\}$  of nonzero real numbers satisfies  $x_n^2 = 1 + x_{n+1} x_{n-1}$  for all  $n$ . Show that for some real  $a$ , the sequence satisfies  $x_{n+1} = a x_n - x_{n-1}$ . Hint: Examine values of  $(x_{n+1} + x_{n-1})/x_n$  for various  $n$ .

25. **Putnam problem 1997–A3.** Evaluate

$$\mathcal{E} = \int_0^\infty \sum_{k=0}^\infty \frac{(-1)^k x^{2k+1}}{2^k k!} \sum_{k=0}^\infty \frac{x^{2k}}{4^k (k!)^2} dx.$$

Answer:  $1.6487212707001281\dots = \sqrt{e}$ . Solution: The integrand is the expression  $x \exp(-x^2/2) I_0(x)$ , which *Maple* can identify and integrate. Alternately, one can interchange the integral and the second sum to obtain

$$\mathcal{E} = \sum_{k=0}^\infty \frac{\int_0^\infty x^{2k+1} e^{-1/2 x^2} dx}{4^k (k!)^2} = \sum_{k=0}^\infty \frac{1}{2^k k!}.$$

26. **Putnam problem 1999–A3.** Consider the power series expansion

$$\frac{1}{1 - 2x - x^2} = \sum_{n \geq 0} a_n x^n.$$

Prove that for each integer  $n \geq 0$ , there is an integer  $m$  such that

$$a_n^2 + a_{n+1}^2 = a_m.$$

Answer: It transpires that

$$a_n^2 + a_{n+1}^2 = a_{2n+1}, \quad (1.11.62)$$

which remains to be proven. Hint: The first 15 coefficients are

1, 2, 5, 12, 29, 70, 169, 408, 985, 2378, 5741, 13860, 33461, 80782, 195025,

and the desired squares are

5, 29, 169, 985, 5741, 33461, 195025,

which is more than enough to spot the pattern. To prove this either explicitly use the closed form for

$$a_n = \frac{1}{2\sqrt{2}} \left( (1 + \sqrt{2})^{n+1} - (-\sqrt{2} + 1)^{n+1} \right),$$

or show that both sides of (1.11.62) satisfy the same recursion (and initial conditions).

27. **Log-concavity.** An easy criterion for log-concavity of a sequence is Newton's lemma that *if  $\sum_{k=0}^n a_k x^k$  is a real polynomial with only real roots then its coefficients are log-concave, as are  $a_k / \binom{n}{k}$  ( $0 < k < n$ ).*

- (a) Use Rolle's theorem to prove Newton's lemma.
- (b) Deduce that the binomial coefficients, row by row, as the Stirling numbers of the first and second kind are log-concave. In particular, they are unimodal.
- (c) The Catalan numbers are log-convex.

- (d) The Motzkin numbers,  $M_n$  (see Item 41), are log concave. This was first shown in 1998 and may be established analytically, as in [104], starting with the recursion

$$(n+2)M_n = (2n+1)M_{n-1} + 3(n-1)M_{n-2},$$

with ordinary generating function

$$\frac{(1-x) - \sqrt{1-2x-3x^2}}{2x^2}.$$

- (e) Prove that the sequence  $x_n = M_n/M_{n-1}$  is increasing (equivalent to log-concavity) by considering that the function  $f : [2, \infty) \rightarrow \mathbb{R}$  by  $f(x) = 2$  on  $[2, 3]$  and by

$$f(x) = \frac{2x+3}{x+2} + \frac{3(x-1)}{x+2} \frac{1}{f(x-1)}$$

thereafter. Thus  $f(n) = x_n$ . Show that  $f$  is continuous, increasing and piece-wise smooth with  $\lim_{x \rightarrow \infty} f(x) = 3$ .

28. **Putnam problem 1999–A4.** Sum the series

$$\mathcal{S} = \sum_{m=1}^{\infty} \sum_{n=1}^{\infty} \frac{m^2 n}{3^m(n 3^m + m 3^n)} \quad \left( = \frac{9}{32} \right). \quad (1.11.63)$$

Hint: Interchange  $m$  and  $n$  and average to obtain  $\mathcal{S} = (\sum_{n \geq 1} n/3^n)^2$ .

29. **Putnam problem 1999–A6.** Consider the sequence defined by  $a_1 = 1$ ,  $a_2 = 2$ ,  $a_3 = 24$ , and, for  $n > 3$ ,

$$a_n = \frac{6a_{n-1}^2 a_{n-3} - 8a_{n-1} a_{n-2}^2}{a_{n-2} a_{n-3}}. \quad (1.11.64)$$

Show that  $n$  divides  $a_n$  for all  $n$ . Hint: Consider the much simpler linear recursion satisfied by  $b_n = a_n/a_{n-1}$ , which is solved by  $2^{n-1}(2^{n-1} - 1)$  so that

$$a_n = 2^{n(n-1)/2} \prod_{k=1}^{n-1} (2^k - 1).$$

Write  $n$  as  $2^a b$  where  $b$  is odd and observe that  $a \leq n(n-1)/2$  and  $b$  divides  $2^{\phi(b)} - 1$ , where  $\phi$  is Euler's totient function.



30. **Berkeley problem 1.3.3.** Establish the limit of the recursion  $x_0 = 1$  and

$$x_{n+1} = \frac{3 + 2x_n}{3 + x_n},$$

for  $n > 0$ . Does the initial value matter? Answer:  $\ell = \frac{\sqrt{13}-1}{2}$ . Note: The *Maple* code

```
iter1:=proc(y,n) local x,k; x:=y; for k to n do
x:=(3+2*x)/(3+x);od;solve(Minpoly(x,2))[1];end:
```

answers this symbolically.

31. **Berkeley problem 1.3.4.** Similarly for

$$x_{n+1} = \frac{1}{2 + x_n}.$$

Answer:  $\ell = \sqrt{2} - 1$ .

32. **Continued fraction errors.** Compare the estimates in Theorem 1.8.4 with the computed errors in Table 1.1 using  $a = N$  and  $c = N + 1$ , and in Table 1.3 using  $a = N + 1/2$  and  $c = N + 3/2$ .

33. **Berkeley problems 5.7.1, 5.7.2, and 5.7.3.** Evaluate

$$\int_0^{2\pi} e^{it} dt \quad (= 2\pi),$$

$$\int_0^{2\pi} e^{it-it} dt \quad (= 2\pi),$$

and, for  $a > b > 0$ , show

$$\frac{1}{2\pi} \int_0^{2\pi} \frac{1}{|ae^{it} - b|^4} dt = \frac{b^2 + a^2}{(a^2 - b^2)^3}.$$

Hint: Proofs use the Cauchy integral formula for derivatives.

34. **A double integral—series equivalence.** For nonnegative integer  $m$ ,  
 (a) show that

$$-i \int_0^\infty \int_0^\infty \frac{t^m e^{(ix-1)t}}{1+x^2} dt dx = \frac{m!}{2^{m+1}} \left( \sum_{n=0}^m \frac{2^n}{n+1} - i \frac{\pi}{2} \right). \quad (1.11.65)$$

- (b) Hence, establish that the moment evaluations:

$$\int_0^\infty t^m \int_0^\infty \frac{\cos(tx)}{1+x^2} dx dt = \frac{m!}{2^{m+1}} \frac{\pi}{2} \quad (1.11.66)$$

and

$$\int_0^\infty t^m \int_0^\infty \frac{\sin(tx)}{1+x^2} dx dt = \frac{m!}{2^{m+1}} \sum_{n=0}^m \frac{2^n}{n+1}. \quad (1.11.67)$$

In particular, (c) show that (1.11.67) also equals

$$\frac{m!}{2(m+1)} \sum_{k=0}^m \frac{1}{\binom{m}{k}}.$$

Hint: The left-hand side is

$$-i m! \int_0^\infty \frac{(1-ix)^{-1-m}}{1+x^2} dx.$$

Both sides then of (1.11.65) satisfy the recursion

$$2(m+1)r(m) - m(m+1)r(m-1) = m!,$$

with initial value  $r(0) = \frac{1}{2} - i \frac{\pi}{4}$ . Now justify exchanging the order of integration.

35. **An open summation problem.** Does

$$\sum_{n=1}^{\infty} \frac{\left(\frac{2}{3} + \frac{1}{3} \sin(n)\right)^n}{n}$$

converge? This is an open problem.

36. **Hadamard inequality.** Let  $A = \{a_{ij}\}$  be a real  $n \times n$  positive semidefinite Hermitian matrix.

(a) Show

$$\det(A) \leq \prod_{i=1}^n a_{ii},$$

with equality if and only if  $A$  is a diagonal matrix or if some  $a_{ii}$  is zero.

(b) By applying the result of (a) to  $AA^*$  obtain the inequality

$$|\det(A)| \leq \left( \prod_{i=1}^n \sum_{j=1}^n |a_{ij}|^2 \right)^{1/2},$$

for arbitrary square matrices. Hint: Apply the arithmetic-geometric mean inequality to the diagonalization of  $A$ .

37. **A tangent series.** (From [136, pg. 83]). For each positive  $n$  evaluate

$$\mathcal{T}_n = \sum_{k=0}^{n-1} \tan^2 \left( \frac{\pi (2k+1)}{4n} \right),$$

with proof. Answer:  $\mathcal{T}_n = n(2n-1)$ .

38. **Putnam problem 1989–B3.** Let  $f : [0, \infty) \rightarrow [0, \infty)$  be differentiable and satisfy  $f'(x) = 6f(2x) - 3f(x)$ . Assume that  $|f(x)| \leq \exp(-\sqrt{x})$ . Determine a formula for the moments

$$\mu_n = \int_0^\infty x^n f(x) dx$$

of  $f$  for  $n = 1, 2, 3, \dots$ . Deduce that  $\{3^n \mu_n / n!\}$  converges and that the limit is only zero if  $\mu_0$  is 0. Hint: Use integration by parts in the formula for  $\mu_n$ , which then involves  $f'(x)$ ; a recursion for  $\mu_n$  will then be apparent.

39. **Putnam problem 1992–A4.** Consider an infinitely differentiable real function  $f$  with

$$f\left(\frac{1}{n}\right) = \frac{n^2}{1+n^2}.$$

for positive integers  $n$ . Determine the Maclaurin series of  $f$ . Hint: Consider the series of  $f(x) - 1/(1 + x^2)$ , and recall that if a  $C^\infty$  function is zero on a sequence converging to zero, then all of its derivatives are zero (although this doesn't necessarily imply that the function is analytic in a neighborhood of zero).

40. **Putnam problem 2000–A4.** Show that the improper integral

$$\mathcal{I} = \lim_{M \rightarrow \infty} \int_0^M \sin(x) \sin(x^2) dx \quad (1.11.68)$$

exists. Hint: Numerical experimentation shows that a limit of approximately 0.4917 is reached. The existence of the limit can be rigorously established in two ways: (a) Since the integrand equals  $(\cos(x^2 - x) - \cos(x^2 + x))/2$ , it suffices to show that  $\lim_{M \rightarrow \infty} \int_0^M \cos(x + x^2) dx$  exists. After a change of variables, it suffices to consider

$$\sum_{k=0}^{n-1} \int_{(k-1/2)\pi}^{(k+1/2)\pi} \frac{\cos(u)}{\sqrt{1+4u}} du.$$

This converges by the alternating series test. (b) Use Cauchy's theorem to integrate the entire functions  $\exp(ix^2 \pm ix)$  over a triangular path with vertices at 0,  $M$  and  $(1+i)M$ . Easy estimates show that the integrals over the vertical and the diagonal edges converge.

41. **Several classic sequences.** In each case try to find a generating function or rule for the sequence below:

- (a) 6, 28, 496, 8128, 33550336, 8589869056, 137438691328, 2305843008139952128,  $\dots$
- (b) 1, 1, 2, 5, 15, 52, 203, 877, 4140, 21147, 115975, 678570, 4213597, 27644437,  $\dots$
- (c) 1, 1, 2, 4, 9, 21, 51, 127, 323, 835, 2188, 5798, 15511, 41835,  $\dots$
- (d) 1, 2, 6, 22, 94, 454, 2430, 14214, 89918, 610182, 4412798,  $\dots$
- (e) 1, 4, 11, 16, 24, 29, 33, 35, 39, 45, 47, 51, 56, 58, 62, 64,  $\dots$
- (f) 1, 20, 400, 8902, 197281, 4865617,  $\dots$

Answers: a) The first few perfect numbers.

b) The *Motzkin numbers*. Among other interpretations they count the number of ways to join  $n$  points on a circle by nonintersecting chords, and the number of length  $n$  paths from  $(0, 0)$  to  $(n, 0)$  that do not go below the horizontal axis and are made up of steps  $(1, 1)$ ,  $(1, -1)$  and  $(1, 0)$ . The generating function (see Item 27) is  $(1 - x - \sqrt{1 - 2x - 3x^2})/(2x^2)$ .

c) The *Bell numbers*, whose exponential generating function is  $\exp(e^x - 1)$ .

d) Values of *Bell polynomials*, in this case counting ways of placing  $n$  labeled balls into  $n$  unlabelled (but 2-colored) boxes. (See Section 1.10.)

e) *Aronson's sequence*, whose definition is: "t is the first, fourth, eleventh, ... letter of this sentence."

f) The number of possible chess games after  $n$  moves.

42. **Duality for Mahler's generating function.** Let  $\alpha > 0$  be irrational and consider  $G_\alpha(z, w) = \sum_{n=1}^{\infty} z^n w^{\lfloor n\alpha \rfloor}$  as in (1.4.18). Define  $F_\alpha(z, w) = \sum_{n=1}^{\infty} z^n \sum_{m=1}^{\lfloor n\alpha \rfloor} w^m$ . Show

(a)

$$F_\alpha(z, w) = \frac{z}{1-z} G_{\alpha-1}(w, z)$$

(b)

$$F_\alpha(z, w) + F_{\alpha-1}(w, z) = \frac{z}{1-z} \frac{w}{1-w}.$$

43. **A continued fraction form of Mahler's generating function.** Let  $\alpha > 0$  be irrational. Show that

$$\frac{1-w}{w} \sum_{n=1}^{\infty} z^n w^{\lfloor n\alpha \rfloor} = \frac{1}{c_0 + \frac{1}{c_1 + \frac{1}{c_2 + \dots}}}, \quad (1.11.69)$$

where

$$c_0 = \frac{z^{-1}w^{-a_0} - 1}{w^{-1} - 1}$$

and for  $n \geq 1$

$$c_n = \frac{z^{-q_{n-2}} w^{-p_{n-2}} [z^{-a_n q_{n-1}} w^{-a_n p_{n-1}} - 1]}{z^{-q_{n-1}} w^{-p_{n-1}} - 1} = \frac{z^{q_n} w^{p_n} - z^{-q_{n-2}} w^{-p_{n-2}}}{z^{-q_{n-1}} w^{-p_{n-1}} - 1}.$$

Here  $\{a_n\}$  are the convergents, and  $p_n/q_n$  are the partial quotients of the continued fraction for  $\alpha$ . In particular, sums like

$$\sum_{n=1}^{\infty} \frac{1}{3^n 2^{\lfloor n\alpha \rfloor}} \quad \text{and} \quad \sum_{n=1}^{\infty} \frac{1}{2^n 3^{\lfloor n\alpha \rfloor}}$$

are non-quadratic irrationals since their continued fractions are clearly unbounded.

- (a) Apply (1.11.69) for  $\alpha = (1 + \sqrt{5})/2$  so that  $a_n = 1$  and  $p_n$  and  $q_n$  are Fibonacci numbers. Apply this also to  $\sqrt{2} \pm 1$ .
- (b) Using Exercise 42 deduce a continued fraction for  $\sum_{n=1}^{\infty} \lfloor n\alpha \rfloor z^n$ .

More details and quite broad extensions—found experimentally—may be found in [48].

44. **Beatty's Theorem.** Let irrational numbers  $\sigma, \tau > 0$  be given. Use the Mahler continued fraction to show that the sets of integer parts

$$\mathcal{S} = \{\lfloor n\sigma \rfloor : 0 < n \in \mathbb{N}\} \quad , \quad \mathcal{T} = \{\lfloor n\tau \rfloor : 0 < n \in \mathbb{N}\}$$

partition  $\mathbb{N} \setminus \{0\}$  if and only if  $\sigma > 1$  and

$$\frac{1}{\sigma} + \frac{1}{\tau} = 1.$$

What happens in the case that  $\sigma$  is rational?

[Beatty's Theorem, as often the case was rediscovered by Beatty. It was known to Lord Raleigh and others earlier. This is a good example of Stigler's Law of Eponymy, namely that "No scientific law is named after its original discoverer" [137, page 60]. Stigler's law is named after Stephen Stigler, the son of the 1982 Nobel prize winning economist George Stigler. Neither Stigler was the "discoverer" of this principle.]

Hint: Observe first that  $\sigma, \tau > 1$  is necessary. Set  $\alpha = \sigma - 1$  and  $\beta = \tau - 1$ . Note that  $\mathcal{S}$  and  $\mathcal{T}$  partition the positive integers if and only if

$$G_\alpha(z, z) + G_\beta(z, z) = \sum_{n=1}^{\infty} z^n = \frac{z}{1-z}.$$

Use Exercise 42 to show this happens if and only if

$$G_\beta(z, z) = G_{\alpha^{-1}}(z, z),$$

which in turn happens exactly when  $\lfloor n\beta \rfloor = \lfloor n(\alpha^{-1}) \rfloor$  for all  $n \geq 1$ . This last equivalence holds if and only if  $\beta\alpha = 1$ , and this is the same as

$$\frac{1}{\sigma} + \frac{1}{\tau} = 1.$$

(A direct proof of the “if” is quite easy.)

45. **Wilker’s inequalities.** Show that for  $0 < x < \pi/2$  one has

$$2 + \frac{2}{45}x^3 \tan(x) > \frac{\sin^2(x)}{x^2} + \frac{\tan(x)}{x} > 2 + \frac{16}{\pi^4}x^3 \tan(x),$$

and that the constants  $2/45$  and  $16/\pi^4$  are the best possible.

46. **A Gamma integral.** Show that

$$\int_0^{\infty} e^{iy} y^{a-1} dy = i^a \Gamma(a)$$

for  $0 < a < 1$ . Hence evaluate  $\int_0^{\infty} \cos(x^b/b) dx$  for  $b > 1$ . Hint: Use Cauchy’s theorem on a contour that goes from 0 to  $R$  and then on a quarter circle to  $iR$  and back on the vertical axis to 0.

47. **The Airy integral.** For real  $x$  the Airy integral is defined by

$$\text{Ai}(x) = \frac{1}{\pi} \int_0^{\infty} \cos\left(\frac{1}{3}t^3 + xt\right) dt.$$

Integrate by parts to show that the integral is well defined. Then obtain the value of  $\text{Ai}(0)$ . Finally, show that  $\text{Ai}^2(z)$  satisfies  $w'' - 4zw' - 2w = 0$ . See also [139]. Some difficult Airy-related integrals are derived and discussed in connection with the quantum bouncer in [88].

48. **Zeros of the Airy function.** Let  $N(\text{Ai})$  denote the real zeros of the Airy function  $\text{Ai}$ . Then, for  $n = 2, 3, \dots$  we define the Airy zeta function

$$Z(n) = \sum_{a \in N(\text{Ai})} \frac{1}{a^n} = \frac{\pi T_{n-1}(0)}{\Gamma(n)}$$

where

$$\begin{aligned} T_n(z) &= C^{(n)}(z) \int_0^\infty \text{Ai}^2(u) du \\ &\quad - \sum_{j=1}^n \binom{n}{j} C^{(n-j)}(z) \frac{d^{j-1}}{dz^{j-1}} \text{Ai}^2(z) + \frac{d^{n-1}}{dz^{n-1}} (\text{Ai}(z)\text{Bi}(z)), \end{aligned}$$

with  $C(z) = \text{Bi}(z)/\text{Ai}(z)$ , [88]. Here  $\text{Ai}$  and  $\text{Bi}$  are the fundamental solutions to  $w'' = zw$  with  $\text{Ai}(0) = \sqrt[3]{3}/(3\Gamma(2/3))$  and  $\text{Bi}(0) = 3^{5/6}/(3\Gamma(2/3))$ , and  $\text{Ai}$  is the Airy integral.

49. **Fun with Airy functions.** Let  $\alpha_n = \text{Ai}^{(n)}(0)$  and  $\beta_n = \text{Bi}^{(n)}(0)$ .

- (a) Show that  $\alpha_{n+3} = (n+1)\alpha_n$ , with a similar recursion for  $\beta_n$ .  
 (b) Show that  $\alpha_n^* = (\text{Ai}^2)^{(n)}(0)$  and  $\beta_n^* = (\text{Bi}^2)^{(n)}(0)$  both satisfy  $\delta_{n+3} = (4n+2)\delta_n$ .  
 (c) Consider  $\gamma_n = (\text{Ai}/\text{Bi})^{(n)}(0)$  and  $\delta_n = (\text{Ai Bi})^{(n)}(0)$ . Show that  $\delta_{n+3} = (4n+2)\delta_n$ , and obtain a recursion for  $\gamma_n$  via the convolution

$$\sum_{j=0}^n \binom{n}{j} \delta_j \gamma_{n-j} = \beta_n^*.$$

Note the analogy with the Bernoulli numbers.

50. **More on the Airy zeta function.** Express  $Z(n)$  of Exercise 48 explicitly as a polynomial in

$$X = \frac{3^{5/6}}{2\pi} \Gamma^2(2/3) = \frac{1}{2\pi \text{Ai}(0)\text{Bi}(0)}.$$



For example,

$$\begin{aligned} Z(2) &= X^2 = 3 \frac{3^{2/3} \Gamma^4(2/3)}{4 \pi^2}, \\ Z(7) &= X^7 - \frac{7}{12} X^4 + \frac{13}{180} X, \\ \text{and} \\ Z(10) &= X^{10} - \frac{5}{6} X^7 + \frac{209}{1008} X^4 - \frac{17}{1296} X. \end{aligned}$$

Note that

$$\int_0^\infty \text{Ai}^2(t) dt = \text{Ai}'(0)^2 = \frac{\sqrt[3]{3} \Gamma^2(2/3)}{4 \pi^2}.$$

Exercise 48 may help find the desired polynomials.

51. **Gosper's continued fraction for  $\pi$ .** During his record 1985 computation of the simple continued fraction for  $\pi$ , Gosper found some large convergents, but no strong evidence to suggest that  $\pi$  has a bounded or unbounded continued fraction. Gosper describes how continued fractions allow you to “see” what a number is. “[I]t's completely astounding ... it looks like you are cheating God somehow” [4, page 112]. He goes on to talk about how this sense of surprise has driven him to extensive work with continued fractions.

52. **A positivity problem.** Show that  $f$  defined by

$$f(x) = \sum_{j=0}^{\infty} \frac{(-2)^j x^{2^j}}{\prod_{i=1}^j (2^i - 1)}$$

is strictly positive for  $0 < x < 1$ . Hint:  $g = x \mapsto f(x)/x$  satisfies  $g'(x) + 2g(x^2) = 0$  and  $g(0) = 1, g(1) = 0$ .

53. **Ramanujan's AGM continued fraction.** The assertions below are extensions of those in Entry 12 of Chapter 18 of Ramanujan's Second Notebook. Let

$$\mathcal{R}_\eta(a, b) = \frac{a}{\eta + \frac{b^2}{\eta + \frac{4a^2}{\eta + \frac{9b^2}{\eta + \dots}}}}$$

for  $a, b, \eta > 0$ .

(a) Then show the marvelous fact that

$$\mathcal{R}_\eta \left( \frac{a+b}{2}, \sqrt{ab} \right) = \frac{\mathcal{R}_\eta(a, b) + \mathcal{R}_\eta(b, a)}{2}.$$

This relies on knowing for  $y > 0$  that

$$\begin{aligned} \frac{2}{\eta} \sum_{k \geq 0} \frac{\operatorname{sech}((2k+1)\pi y/2)}{1 + \{(2k+1)/\eta\}^2} &= \mathcal{R}_\eta(\theta_2^2(q), \theta_3^2(q)) \\ \frac{1}{\eta} \sum_{k \in \mathbb{Z}} \frac{\operatorname{sech}(k\pi y)}{1 + \{(2k)/\eta\}^2} &= \mathcal{R}_\eta(\theta_3^2(q), \theta_2^2(q)), \end{aligned}$$

where  $q = \exp(-\pi y)$ , and observing how  $y \mapsto \frac{y}{2}$  interacts with these two identities. While these two series are hard to derive, they are easy to verify numerically.

- (b) The continued fraction is hard to compute directly for  $a = b$ . We shall find a way to use the last hyperbolic series to compute it.
- (c) Determine the relationship between  $\mathcal{R}_\eta$  and  $\mathcal{R}_1$ . Hence, for  $0 < b < a$ , show that

$$\mathcal{R}_1(a, b) = \frac{\pi}{2} \sum_{n \in \mathbb{Z}} \frac{a K(k)}{K^2(k) + a^2 n^2 \pi^2} \operatorname{sech} \left( n\pi \frac{K(k')}{K(k)} \right), \quad (1.11.70)$$

where  $k = b/a = \theta_2^2/\theta_3^2$ .

(d) For  $y = 1$ , this evaluates

$$\mathcal{R}_1 \left( 1, \frac{1}{\sqrt{2}} \right) = \frac{\pi}{2} \sum_{n \in \mathbb{Z}} \frac{K(1/\sqrt{2}) \operatorname{sech}(n\pi)}{K^2(1/\sqrt{2}) + n^2 \pi^2},$$

with similar evaluations for  $\mathcal{R}_1(1, k_N)$  at the  $N$ -th singular value discussed in Section 4.2.

(e) Denote  $\mathcal{R}(a) = \mathcal{R}_1(a, a)$ . Write (1.11.70) as a Riemann sum to deduce that

$$\mathcal{R}(a) = \int_0^\infty \frac{\operatorname{sech} \left( \frac{\pi x}{2a} \right)}{1 + x^2} dx = 2a \sum_{k=1}^\infty \frac{(-1)^{k+1}}{1 + (2k-1)a}, \quad (1.11.71)$$

where the final equality comes from the Cauchy-Lindelöf Theorem (Theorem 6.5.4). The final sum provides an analytic continuation of  $\mathcal{R}$ , and can be written in the computationally efficient form  $\mathcal{R}(a) = 1/2 \Psi(3/4 + 1/4 a^{-1}) - 1/2 \Psi(1/4 + 1/4 a^{-1})$ . (Here  $\Psi = \Gamma'/\Gamma$  is the digamma function.) Observe that

$$\begin{aligned} \mathcal{R}(a) &= \frac{2a}{1+a} F\left(\frac{1}{2a} + \frac{1}{2}, 1; \frac{1}{2a} + \frac{3}{2}; -1\right) \\ &= 2 \int_0^1 t^{1/a} (1+t^2)^{-1} dt = a \int_0^\infty \operatorname{sech}(ax) e^{-x} dx \end{aligned}$$

is now of the form for Gauss's continued fraction of Section 7.8 of the first volume.

(f) Conclude that  $\mathcal{R}(1) = \log 2$  and  $\mathcal{R}(1/2) = 2 - \pi/2$  with similar evaluations for all Egyptian fractions  $(1/n)$ .

(g) Prove that

$$\mathcal{R}(2) = \sqrt{2} \left\{ \frac{\pi}{2} - \log(1 + \sqrt{2}) \right\}.$$

Evaluate also  $\mathcal{R}(3/2)$  and  $\mathcal{R}(5)$ .

(h) One can derive the rapidly convergent  $\zeta$ -series

$$\mathcal{R}(a) = \frac{\pi}{2} \sec\left(\frac{\pi}{2a}\right) + \frac{2a^2(1+8a-106a^2+280a^3+9a^4)}{(a^2-1)(9a^2-1)(5a-1)(7a-1)} + \mathcal{C}(a)$$

where

$$\mathcal{C}(a) = \frac{1}{2} \sum_{n \geq 1} \{\zeta(2n+1) - 1\} \frac{(3a-1)^{2n} - (a-1)^{2n}}{(4a)^{2n}}.$$

54. **A fast series.** It is possible to deduce from the previous exercise, using Poisson summation (see Theorem 2.3.1), that

(a) For  $0 < b \leq a$ ,  $k = b/a$ ,  $K = K(k)$ , and  $K' = K(k')$  we have

$$\begin{aligned} \mathcal{R}_1(a, b) &= \mathcal{R}\left(\frac{\pi a}{2K'}\right) + \frac{\pi}{\cos(K'/a)} \frac{1}{e^{2K/a} - 1} \\ &+ \frac{2\pi a}{K'} \sum_{d \in O^+} \frac{(-1)^{(d-1)/2}}{1 - \pi^2 d^2 a^2 / (4K'^2)} \frac{1}{e^{\pi d K / K'} - 1}. \end{aligned}$$

This allows a highly effective computation of the continued fraction when  $a$  is close to  $b$ , and so the series (1.11.70) is not effective. Determine the proper interpretation when  $a = b$ .

- (b) The AGM relationship,  $\mathcal{R}_1(b, a) = 2 \mathcal{R}_1((a+b)/2, \sqrt{ab}) - \mathcal{R}_1(a, b)$ , then allows one to compute  $\mathcal{R}_1(a, b)$  for  $a < b$  equally effectively. (See [39] for details.)
- (c) Denoting by  $O^+$  the positive odd integers, one may establish

$$\int_{-\infty}^{\infty} \frac{\operatorname{sech}(bx)}{1+x^2} e^{iax} dx = \frac{\pi}{\cosh b} e^{-a} + \frac{2\pi}{b} \sum_{d \in O^+} \frac{(-1)^{(d-1)/2} e^{-\pi da/(2b)}}{1 - \pi^2 d^2/(4b^2)}.$$

Interpret this relation for the possibility that  $b$  is an odd integer times  $\pi/2$ . Compare the next exercise.

### 55. A cosh integral.

- (a) Show that, for  $|a| \leq b$ , one has

$$\begin{aligned} & \int_0^{\infty} \frac{\cosh(at)}{\cosh(bt)(1+t^2)} dt \\ &= \frac{\pi}{2} \frac{1}{b \sin(b)} \left( \sin(b) \int_0^a \frac{\sin(a-t)}{\cos\left(\frac{1}{2} \frac{\pi t}{b}\right)} dt - \sin(a) \int_0^b \frac{\sin(b-t)}{\cos\left(\frac{1}{2} \frac{\pi t}{b}\right)} dt \right) \\ &+ \frac{\pi}{2} \frac{\sin(a)}{\sin(b)} - \frac{\sin(a-b)}{\sin(b)} \int_0^{\infty} \frac{1}{\cosh(bt)(1+t^2)} dt. \end{aligned}$$

Hint: Show that both sides satisfy the same second-order differential equation in  $a$  with the same values for  $a = 0$  (namely the value  $\int_0^{\infty} \operatorname{sech}(bx)/(1+x^2) dx = \mathcal{R}(\pi/2/b)$ ) and for  $a = b$  ( $\pi/2$ ).

- (b) Deduce that

$$\begin{aligned} & \int_0^{\infty} \frac{\operatorname{sech}(bt)}{1+t^2} \cos(at) dt \\ &= \frac{\cosh(a)}{2} \left\{ \Psi\left(\frac{3}{4} + \frac{1}{2} \frac{b}{\pi}\right) - \Psi\left(\frac{1}{4} + \frac{1}{2} \frac{b}{\pi}\right) \right\} \\ &- \frac{\pi}{2b} \int_0^a \sinh(a-t) \operatorname{sech}\left(\frac{\pi t}{2b}\right) dt \end{aligned}$$

for all  $a, b \geq 0$ .

- (c) Note that as  $b \rightarrow 0$ , the final integrand is convergent to zero but not uniformly, and the prior evaluation becomes

$$\int_0^\infty \frac{\cos(at)}{1+t^2} dt = \frac{\pi}{2} e^{-a}.$$

The final integral in (b) equals  $\int_0^a \cosh(a-t) \arctan\left(\sinh\left(\frac{\pi t}{2b}\right)\right) dt$ .

- (d) Show for  $|a| < b$  that

$$\int_0^\infty \frac{\operatorname{cosech}(bt)}{1+t^2} \sinh(at) dt = \pi \sum_{k=1}^\infty \frac{\sin(k\pi(1-a/b))}{b+k\pi}.$$

56. **Infinitely differentiable functions.** A lovely result of E. Borel [197, pg. 191] shows that for every real sequence  $\{a_n\}$  there is an infinitely differentiable function on  $\mathbb{R}$  such that  $f^{(n)}(0) = a_n$ .

A remarkable explicit example occurs via Ramanujan's continued fraction. Use equation (1.11.71) to show that the function  $a \mapsto \mathcal{R}(a)$  has Maclaurin series  $\sum_{n \geq 0} E_{2n} a^{2n+1}$ , with zero radius of convergence. Hint: Show that  $\int_0^\infty \operatorname{sech}(\pi x/2) x^{2n} dx = E_{2n}$  for each positive integer  $n$ . Then estimate

$$\left| \mathcal{R}(a) - \sum_{n=0}^{N-1} a^{2n+1} E_{2n} \right| \leq a^{2N+1} \int_0^\infty \operatorname{sech}\left(\frac{\pi x}{2}\right) x^{2N} dx = a^{2N+1} |E_{2N}|,$$

where  $E_{2N}$  denote the Euler numbers.

57. **Two expected distances.** These results originate with James D. Klein.

- (a) The expected distance between two random points on different sides of the unit square:

$$\begin{aligned} & \frac{2}{3} \int_0^1 \int_0^1 \sqrt{x^2 + y^2} dx dy + \frac{1}{3} \int_0^1 \int_0^1 \sqrt{1 + (y-u)^2} du dy \\ &= 0.869009055274534463884970594345406624856719 \dots \\ &= \frac{2}{9} + \frac{1}{9} \sqrt{2} + \frac{5}{9} \log(1 + \sqrt{2}). \end{aligned}$$

(b) The expected distance between two random points on different faces of the unit cube:

$$\begin{aligned} & \frac{4}{5} \int_0^1 \int_0^1 \int_0^1 \int_0^1 \sqrt{x^2 + y^2 + (z - w)^2} dw dx dy dz \\ & + \frac{1}{5} \int_0^1 \int_0^1 \int_0^1 \int_0^1 \sqrt{1 + (y - u)^2 + (z - w)^2} du dw dy dz \\ & = 0.92639005517404672921816358654777901444496019010734 \dots \\ & = \frac{4}{75} + \frac{17}{75} \sqrt{2} - \frac{2}{25} \sqrt{3} - \frac{7}{75} \pi \\ & \quad + \frac{7}{25} \log(1 + \sqrt{2}) + \frac{7}{25} \log(7 + 4\sqrt{3}). \end{aligned}$$

(c) Show that the first term in (b) is

$$\begin{aligned} & \frac{\sqrt{2\pi}}{5} \sum_{n=2}^{\infty} \frac{F(1/2, -n+2; 3/2; 1/2)}{(2n+1) \Gamma(n+2) \Gamma(5/2-n)} \\ & \quad + \frac{4}{15} \sqrt{2} + \frac{2}{5} \log(\sqrt{2} + 1) - \frac{1}{75} \pi \end{aligned}$$

and the second term is

$$\begin{aligned} & \frac{\sqrt{\pi}}{10} \sum_{n=0}^{\infty} \frac{F(1, 1/2, -1/2-n, -n-1; 2, 1/2-n, 3/2; -1)}{(2n+1) \Gamma(n+2) \Gamma(3/2-n)} \\ & \quad - \frac{2}{25} + \frac{\sqrt{2}}{50} + \frac{1}{10} \log(\sqrt{2} + 1). \end{aligned}$$

This allows one to numerically compute the expectation to high precision and to express both of the individual integrals in terms of the same set of constants. These expectations have actually been checked by computer simulations. Hint: Reduce the first integral to a three dimensional one and use the binomial theorem on both.

58. **Euler: The Master of Us All.** Laplace once wrote, “Read Euler, read Euler. He is the master of us all.” This quote is used for the title of Dunham’s lovely book about Euler [107]. It contains a concise biography and

emphasizes his breadth and mastery of all the topics of his period. Dunham's discussions on Euler and logarithms, and on Euler and infinite series, are especially worth reading in conjunction with this chapter. The former emphasizes the centrality of logarithms and their generalizations to Euler's work, and the latter discusses in illuminating detail Euler's conquest of the *Basel problem* of evaluating  $\zeta(2)$ .





## Chapter 2

# Fourier Series and Integrals

Having contested the various results [Biot and Poisson] now recognise that they are exact but they protest that they have invented another method of expounding them and that this method is excellent and the true one. If they had illuminated this branch of physics by important and general views and had greatly perfected the analysis of partial differential equations, if they had established a principal element of the theory of heat by fine experiments . . . they would have the right to judge my work and to correct it. I would submit with much pleasure . . . But one does not extend the bounds of science by presenting, in a form said to be different, results which one has not found oneself and, above all, by forestalling the true author in publication.

Joseph Fourier, c. 1825 [135]

It is often useful to decompose a given function into its components, analyze those and then put the function together again, possibly in a different way. One of the most classical (and mathematically interesting) methods is to use trigonometric functions. This is the basis for the theory of Fourier analysis. One can think of a sound (a certain tone played on the violin, say) as consisting of countably many oscillations with different discrete frequencies, which together define the pitch and the specific timbre of the tone. These component frequencies can be identified via Fourier analysis—mathematically, by computing the Fourier series of a periodic function. Of course, in reality no oscillation is precisely periodic and a sound will consist of a continuum of frequencies—mathematically, this is analyzed by taking the continuous Fourier transform. Thus, the Fourier

transform arises from Fourier series by taking more and more frequencies into account. This process is described by the Poisson summation formula. Finally, the question arises if a function thus analyzed can be reconstructed from its Fourier series. It turns out that even for a continuous function, the Fourier series may not be everywhere convergent. Thus one is led to consider special summation methods, or “kernels.”

These are the topics covered in the next few sections. The interested reader can find more details than we have room to give here in the introductory exposition in Chapter 8 of Stromberg’s book [197], or in the exhaustive classical treatment of Zygmund [213], or in the more modern books by Katznelson [144] and Butzer/Nessel [72].

Fourier methods are always a good tool for the evaluation of sums and integrals. Once such an object is identified as a Fourier series or integral, that knowledge can be used for the evaluation, or methods such as the Parseval equation or Poisson summation can be brought to bear. We will see some examples of this in the present chapter, and we have already seen and will see more of that in the first volume and throughout this second volume.

## 2.1 The Development of Fourier Analysis

We start with some historical background here, which we have adapted in part from the MacTutor website <http://www-gap.dcs.st-and.ac.uk/~history>, and also from R. Bhatia’s monograph on Fourier series [25].

Joseph Fourier was one of the more colorful figures of mathematical history. Originally intending to be a Catholic priest, Fourier declined to take his vows when he realized that he could not extinguish his interest in mathematics. Shortly afterwards he became involved in the movement that led to the French Revolution in 1793, but fortunately for modern mathematics he was spared the guillotine, and was able to study mathematics at the Ecole Normale in Paris under the tutelage of Lagrange. A few years later he was appointed as a scientific adviser for Napoleon’s expedition to Egypt. When Napoleon’s army was defeated by Nelson at the battle of the Nile, Fourier and the other French advisers insisted that they be able to retain some of the artifacts they had found there. The British refused, but at least permitted the French to make a catalog of what they felt were the more important items. Fourier was given this task by the French commanders. The eighth item on his catalog was the Rosetta

stone, which had been recognized by the French scientists on the expedition as a possible key to understanding of the Egyptian language. Later in Europe, when published copies of the inscriptions were made available, Champollion, a student who had been inspired by Fourier himself to study Egyptology, succeeded in the first translation.

Fourier's principal contributions to mathematics, namely Fourier analysis and Fourier series, paralleled and even stimulated the development of the entire field of real analysis. Fourier analysis had its origin in the 1700s, when d'Alembert derived the wave equation that describes the motion of a vibrating string, starting with an initial "function," which at the time was restricted to an analytic expression. In 1755 Daniel Bernoulli gave another solution for the problem in terms of *standing waves*, namely waves associated with the  $n + 1$  points  $0, 1/n, 2/n, \dots, (n - 1)/n, 1$  on the string that remain fixed. The motion for  $n = 1, 2, \dots$  is the *first harmonic*, the *second harmonic* and so on. Bernoulli asserted that every solution to the problem of the plucked string is merely a sum of these harmonics.

Beginning in 1804, Joseph Fourier began to analyze the conduction of heat in solids. He not only discovered the basic equations governing heat conduction, but he developed methods to solve them, and in the process developed and extended Fourier analysis to a much broader range of scientific problems. He described his work in his book *The Analytical Theory of Heat*, which is regarded as one of the most important books in the history of physics.

Like Bernoulli, Fourier asserted that any continuous function can be written as

$$f(x) = \sum_{n=-\infty}^{\infty} A_n e^{int}. \quad (2.1.1)$$

But Fourier claimed that his representation is valid not only for  $f$  given by a single analytical formula, but for  $f$  given by any *graph*, which at the time was a more general object, encompassing, for example, a piecewise combination of different analytic expressions. Fourier was not able to prove his assertions, at least not to the satisfaction of the mathematical community at the time (and certainly not to the standards required today). But other mathematicians were intrigued, and pursued these questions with renewed determination.

Dirichlet was the first to find a rigorous proof. He defined a real function as we now understand the term, namely as a general mapping from one set of reals to another, thus decoupling analysis from geometry. He then was able

to prove that for every “piecewise smooth” function  $f$ , the Fourier series of  $f$  converges to  $f(x)$  at any point  $x$  where  $f$  is continuous, and to the average value  $(f(x+) + f(x-))/2$  if  $f$  has a jump discontinuity at  $x$ . This was the first major convergence result for Fourier series.

Mathematicians realized that to handle functions that have infinitely many discontinuities, it was necessary to generalize the notion of an integral beyond the intuitive idea of the area under a curve. Riemann succeeded in developing a theory of integration that could handle such functions, and using this theory, he was able to exhibit an example of a function that did not satisfy Dirichlet’s piecewise continuous condition, yet still possessed a pointwise convergent Fourier series. Cantor observed that changing a function  $f$  at a few points does not change its Fourier coefficients. In the course of asking how many points can be changed while preserving Fourier coefficients, he was led to the notion of countably infinite and uncountably infinite sets. Ultimately Lebesgue extended Dirichlet’s, Riemann’s, and Cantor’s results into what we now know as measure theory, where sets of measure zero, almost everywhere equality of functions, and almost everywhere convergence of functions supersede the simple concepts that prevailed in the 1700s.

In summary, it is not an exaggeration to say that all of modern real and complex analysis has its roots in Fourier series and Fourier analysis.

## 2.2 Basic Theorems of Fourier Analysis

### 2.2.1 Fourier Series

We will consider  $2\pi$ -periodic functions  $f : \mathbb{R} \rightarrow \mathbb{C}$ . For  $p > 0$ , we write  $f \in L^p(\mathbb{T})$  if such an  $f$  is Lebesgue measurable and satisfies

$$\|f\|_p = \left[ \int_{-\pi}^{\pi} |f(t)|^p dt \right]^{1/p} < \infty$$

( $\mathbb{T}$  stands for *Torus*). Note that  $\|\cdot\|_p$  is not a norm on  $L^p$ , since any function  $f$  which is 0 almost everywhere will have  $\|f\|_p = 0$ . (Later we will identify functions which are equal a.e.) In what follows, we will be mainly interested in the spaces  $L^1(\mathbb{T})$  and  $L^2(\mathbb{T})$ . Note that  $\|\cdot\|_1 \leq \|\cdot\|_2$  and  $L^2(\mathbb{T}) \subsetneq L^1(\mathbb{T})$ . If  $f$  is  $2\pi$ -periodic and continuous on  $\mathbb{R}$ , then we write  $f \in C(\mathbb{T})$  and equip this space with the uniform norm.

For a function  $f \in L^1(\mathbb{T})$ , define the  $n$ -th Fourier coefficient ( $n \in \mathbb{Z}$ ) by

$$\widehat{f}_n = \int_{-\pi}^{\pi} f(t) e^{-int} dt.$$

This is motivated by the insight that if we write

$$s_n(t) = \frac{1}{2\pi} \sum_{k=-n}^n c_k e^{ikt} \quad (2.2.2)$$

for some sequence  $c_k$  and assume  $L^1(\mathbb{T})$ -convergence of  $(s_n)$  to some  $f \in L^1(\mathbb{T})$ , then  $\widehat{f}_n = c_n$ . Thus for arbitrary  $f \in L^1(\mathbb{T})$  we will write, formally and suggestively,

$$f(t) \sim \frac{1}{2\pi} \sum_{n=-\infty}^{\infty} \widehat{f}_n e^{int},$$

where in general no assertion about convergence of this series is implied. Any convergence statement is to be read in the sense of symmetric limits, i.e. of

$$s_n(f, t) = \frac{1}{2\pi} \sum_{k=-n}^n \widehat{f}_k e^{ikt}.$$

Fourier coefficients usually are complex numbers, even when  $f$  is a real-valued function. Sometimes it is desirable to have a real-valued series for  $f$ . Then the Fourier series can be equivalently written as

$$f(t) \sim \frac{1}{2\pi} a_0 + \frac{1}{\pi} \sum_{n=1}^{\infty} (a_n \cos(nt) + b_n \sin(nt))$$

where

$$\begin{aligned} a_n &= \int_{-\pi}^{\pi} f(t) \cos(nt) dt \quad \text{and} \\ b_n &= \int_{-\pi}^{\pi} f(t) \sin(nt) dt. \end{aligned}$$

Note that

$$s_n(f, t) = \frac{1}{2\pi} a_0 + \frac{1}{\pi} \sum_{k=1}^n (a_k \cos(kt) + b_k \sin(kt)).$$

**Example 2.2.1** *Fourier series of a symmetric function.*

Define  $f \in L^1(\mathbb{T})$  by  $f(t) = (\pi - t)/2$  for  $t \in [0, 2\pi)$ . Then  $\widehat{f}_n = -i\pi/n$  for  $n \neq 0$  and  $\widehat{f}_0 = 0$ , and its Fourier series is given by

$$f(t) \sim \frac{-i}{2} \sum_{\substack{n=-\infty \\ n \neq 0}}^{\infty} \frac{1}{n} e^{int}, \quad (2.2.3)$$

or, equivalently,

$$f(t) \sim \sum_{n=1}^{\infty} \frac{\sin(nt)}{n}. \quad (2.2.4)$$

The right-hand side of (2.2.4) equals 0 at  $t = 0$ , while the left-hand side by definition equals  $\pi/2$ . Thus, equality cannot hold pointwise here. The situation would improve if we were to define  $f(0) = 0$ . In fact, it follows from

$$\sum_{n=1}^{\infty} \frac{z^n}{n} = -\text{Ln}(1 - z) \quad \text{for } |z| \leq 1, z \neq 1, \quad (2.2.5)$$

by setting  $z = e^{it}$  and taking real and imaginary parts, that

$$\sum_{n=1}^{\infty} \frac{\sin(nt)}{n} = \frac{\pi - t}{2} \quad \text{and} \quad (2.2.6)$$

$$\sum_{n=1}^{\infty} \frac{\cos(nt)}{n} = -\ln |2 \sin(t/2)|, \quad (2.2.7)$$

for all  $t \in (0, 2\pi)$ , and even with uniform convergence on every closed subinterval of  $(0, 2\pi)$ .  $\square$

The question now is under what condition the Fourier series of a function converges to that function. The answer depends on the definition of “convergence” and is most interesting in the cases of pointwise,  $L^1(\mathbb{T})$ - and  $L^2(\mathbb{T})$ -convergence.

**Pointwise and uniform convergence.** As we have seen in the above example (and as is clear from the computation of Fourier series),  $L^1(\mathbb{T})$ -functions which are equal a.e. will have the same Fourier series. By the *uniqueness theorem for Fourier series*, the converse is also true: Functions with the same Fourier series

are equal a.e. It is *not* true that the Fourier series of any continuous function is pointwise convergent to that function. An example due to Lebesgue is given in Item 6 at the end this chapter. Such functions must have a complicated structure, because they cannot be of bounded variation. A function  $f : \mathbb{R} \rightarrow \mathbb{C}$  is of bounded variation on  $(a, b)$  if there is an  $M > 0$  with

$$\sum_{k=0}^{n-1} |f(t_{k+1}) - f(t_k)| \leq M \quad \text{for all } a < t_0 < t_1 < \cdots < t_n < b.$$

The infimum of all such  $M$  is the total variation of  $f$ , thus

$$V_a^b(f) = \sup \left\{ \sum_{k=0}^{n-1} |f(t_{k+1}) - f(t_k)| : a < t_0 < t_1 < \cdots < t_n < b, n \in \mathbb{N} \right\}.$$

The set of all functions of bounded variation on  $(a, b)$  is denoted  $BV(a, b)$ . An equivalent characterization (due to Lebesgue) is that  $f$  can be written as the difference of two bounded increasing functions. Thus any  $BV$ -function  $f$  is differentiable a.e., and the one-sided limits  $f(t+)$  and  $f(t-)$  exist in  $(a, b)$ .

**Theorem 2.2.2 (Jordan test).** *For  $f \in L^1(\mathbb{T}) \cap BV(a, b)$  we have*

$$\lim_{n \rightarrow \infty} s_n(f, t) = \frac{f(t-) + f(t+)}{2} \quad \text{for every } t \in (a, b)$$

*and uniformly on every compact subinterval of  $(a, b)$  where  $f$  is continuous. If  $f \in C(\mathbb{T})$ , then the convergence is uniform on  $\mathbb{R}$ .*

Note that this theorem is proved via Césaro summation, and thus via the Fejér kernel, which we will discuss in Section 2.3.3.

The Jordan test is another explanation of the convergence properties of the Fourier series for the function  $f$  from the previous example. If another function  $f \in C(\mathbb{T})$  is defined by  $f(t) = t^2$  on  $[-\pi, \pi]$ , then

$$f(t) \sim \frac{1}{3}\pi^2 + 4 \sum_{n=1}^{\infty} \frac{(-1)^n}{n^2} \cos(nt).$$

Since  $f$  is continuous on  $\mathbb{R}$ , the Fourier series converges uniformly to  $f$ , by the Jordan test. However, since  $s_n(f, t)$  is uniformly convergent, this also follows directly from the uniqueness theorem for Fourier series.

For  $t = 0$  we get from this the identity

$$\sum_{n=1}^{\infty} \frac{(-1)^{n-1}}{n^2} = \frac{\pi^2}{12}.$$

By separating even and odd parts, this proves that

$$\zeta(2) = \sum_{n=1}^{\infty} \frac{1}{n^2} = \frac{\pi^2}{6}. \quad \square$$

To conclude the section on pointwise convergence, we note that the Fourier series of an  $f \in L^1(\mathbb{T})$  may be divergent everywhere (by an example due to Kolmogorov in 1926). If, however,  $f \in L^p(\mathbb{T})$  for  $p > 1$ , then  $s_n(f, t)$  converges to  $f(t)$  a.e. (Carleson/Hunt 1966).

**Convergence in  $L^1(\mathbb{T})$ .** From now on it makes sense to identify functions which are equal almost everywhere, so that  $\|\cdot\|_1$  and  $\|\cdot\|_2$  are now norms in the respective spaces. Thus we will deal, strictly speaking, with equivalence classes of functions instead of pointwise defined functions, although this will not be denoted explicitly. For example, the statement that such a function is continuous will mean that in its equivalence class we can find a continuous function, then denoted by the same symbol. This notation is unusual but convenient, and it corresponds to how one deals with these objects informally. It can lead to dangerous pitfalls, though, as we will see later. In general, the Fourier series of an  $f \in L^1(\mathbb{T})$  need not be convergent to  $f$  with respect to the  $L^1(\mathbb{T})$ -norm. Of the many restrictions on  $f$  which imply convergence, we mention here only one which is particularly simple (and has a convincing analog in the case of Fourier transforms).

**Theorem 2.2.3** *Let  $f \in L^1(\mathbb{T})$ . If  $\sum_{n=-\infty}^{\infty} |\widehat{f}_n| < \infty$ , then  $f \in C(\mathbb{T})$  and*

$$f(t) = \frac{1}{2\pi} \sum_{n=-\infty}^{\infty} \widehat{f}_n e^{int},$$

*with convergence in the  $L^1(\mathbb{T})$ -norm as well as uniformly on  $\mathbb{T}$ .*

There are functions in  $L^1(\mathbb{T})$  (even continuous ones) for which the Fourier coefficients are not absolutely summable. It is a difficult and in general open



problem to characterize those sequences which are Fourier sequences of an  $L^1(\mathbb{T})$ -function. A weak necessary condition follows from the Riemann-Lebesgue lemma (see next subsection and Exercise 14): The Fourier coefficients of any  $f \in L^1(\mathbb{T})$  satisfy  $\lim_{|n| \rightarrow \infty} \widehat{f}_n = 0$ .

**Convergence in  $L^2(\mathbb{T})$ .** In contrast to the  $L^1$ -case, for  $p > 1$  the Fourier series of any  $f \in L^p(\mathbb{T})$  converges to  $f$  in the  $L^p(\mathbb{T})$ -norm. For  $p = 2$  this follows directly from the usual Hilbert space theory: The trigonometric functions constitute an orthogonal basis for  $L^2(\mathbb{T})$ . This implies that the Fourier coefficients of an  $f \in L^2(\mathbb{T})$  are square-summable and that every square-summable sequence is the sequence of Fourier coefficients of an  $f \in L^2(\mathbb{T})$ . (That is the Riesz-Fischer theorem.) Another consequence of Hilbert space theory is the Parseval equation.

**Theorem 2.2.4 (Parseval's formula.)** *For any  $f, g \in L^2(\mathbb{T})$ , the identity*

$$\frac{1}{2\pi} \sum_{n=-\infty}^{\infty} \widehat{f}_n \overline{\widehat{g}_n} = \int_{-\pi}^{\pi} f(t) \overline{g(t)} dt \quad (2.2.8)$$

*holds. In particular, for  $f = g$  we get  $\frac{1}{2\pi} \sum_{n=-\infty}^{\infty} |\widehat{f}_n|^2 = \int_{-\pi}^{\pi} |f(t)|^2 dt$ .*

**Example 2.2.5** *Parseval's formula and the zeta function.*

Applying the Parseval equation to the function  $f(t) = (\pi - t)/2$  on  $(0, 2\pi)$  from the first example in this section again gives the identity, after simplifying,

$$\zeta(2) = \sum_{n=1}^{\infty} \frac{1}{n^2} = \frac{\pi^2}{6}.$$

This is the same result as in the previous example, but the Parseval equation is conceptually simpler than the Jordan test. Applying the Parseval equation to  $f(t) = t^2/4 - \pi t/2 + \pi^2/6$  gives

$$\zeta(4) = \sum_{n=1}^{\infty} \frac{1}{n^4} = \frac{\pi^4}{90},$$

and this can be continued to give  $\zeta(2n)$  as a rational multiple of  $\pi^{2n}$  for any  $n \in \mathbb{N}$ . The general formula is given in the next chapter, by a different method. Similar formulas for  $\zeta(2n+1)$  are unknown (and highly unlikely to exist; see the next chapter for more information).  $\square$

**Example 2.2.6** *Parseval's formula and Euler sums.*

Multiplying the evaluations (2.2.6) and (2.2.7), using the Cauchy product, simplifying and doing a partial fraction decomposition gives

$$-\ln |2 \sin(t/2)| \cdot \frac{\pi - t}{2} = \sum_{n=1}^{\infty} \frac{1}{n} \sum_{k=1}^{n-1} \frac{1}{k} \sin(nt) \quad \text{on } (0, 2\pi). \quad (2.2.9)$$

Now Parseval proves the Euler sum formula that we first mentioned in the first volume:

$$\frac{1}{4} \int_0^{2\pi} (\pi - t)^2 \ln^2(2 \sin(t/2)) dt = \pi \sum_{n=1}^{\infty} \frac{(\sum_{k=1}^{n-1} k^{-1})^2}{n^2}.$$

□

## 2.2.2 Fourier Transforms

We now consider functions  $f : \mathbb{R} \rightarrow \mathbb{C}$ . For  $p > 0$ , we write  $f \in L^p(\mathbb{R})$  if such an  $f$  is Lebesgue measurable and satisfies

$$\|f\|_p = \left[ \int_{-\infty}^{\infty} |f(t)|^p dt \right]^{1/p} < \infty.$$

As before, functions which are equal a.e. will be identified, so that  $\|\cdot\|_p$  is a norm. In what follows, we will be mainly interested in the spaces  $L^1(\mathbb{R})$  and  $L^2(\mathbb{R})$ . In contrast to the periodic case, there is now no inclusion relation between these spaces. If  $f : \mathbb{R} \rightarrow \mathbb{C}$  is continuous, we write  $f \in C(\mathbb{R})$  and equip this space with the uniform norm.

The Fourier transform on  $L^1(\mathbb{R})$  is now a direct analog of the Fourier coefficients on  $L^1(\mathbb{T})$ . By further analogy to the previous subsection, a Fourier transform on  $L^2(\mathbb{R})$  would also be of interest. However, the definition of such an  $L^2(\mathbb{R})$ -transform is not as straightforward as before, since these spaces are not contained in each other. There is, however, a meaningful transform on  $L^2(\mathbb{R})$ , and we will discuss this after giving the properties of the  $L^1(\mathbb{R})$ -transform.

**Fourier Transform on  $L^1(\mathbb{R})$ .** For a function  $f \in L^1(\mathbb{R})$ , define the Fourier transform of  $f$  to be the function  $\widehat{f} : \mathbb{R} \rightarrow \mathbb{C}$  given by

$$\widehat{f}(x) = \int_{-\infty}^{\infty} f(t) e^{-ixt} dt.$$

As the example  $f = I_{(-\pi, \pi)}$  (characteristic function) with  $\widehat{f}(x) = 2 \sin(\pi x)/\pi$  shows, the Fourier transform of an  $f \in L^1(\mathbb{R})$  need not be in  $L^1(\mathbb{R})$ . It is not difficult to show, however, that such an  $\widehat{f}$  is always continuous, with  $\|\widehat{f}\|_{\infty} \leq \|f\|_1$ . The *Riemann-Lebesgue lemma* says that, additionally,  $\lim_{|x| \rightarrow \infty} \widehat{f}(x) = 0$ . It is proved by approximating  $f$  by step functions.

Under what conditions can an  $f \in L^1(\mathbb{R})$  be reconstructed from its Fourier transform? In principle, this is always possible: By the *uniqueness theorem for Fourier transforms*, functions with the same Fourier transform must be equal a.e. In practice, one would like a simple formula for this inversion. Such a formula is given in the inversion theorem below, whose proof is not easy: It depends on constructing and investigating a suitable summation kernel for the Fourier transform (often the Gauss or the Fejér kernel are used).

**Theorem 2.2.7** *If  $f \in L^1(\mathbb{R})$  is such that  $\widehat{f} \in L^1(\mathbb{R})$ , then  $f, \widehat{f} \in C(\mathbb{R})$  and*

$$f(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \widehat{f}(x) e^{ixt} dx \quad (2.2.10)$$

for all  $t \in \mathbb{R}$ .

Conditions which are good for  $\widehat{f} \in L^1(\mathbb{R})$  are given in Item 10 at the end of this chapter.

**Example 2.2.8** *Sinc integrals.*

For  $f(t) = \max\{1 - |t|, 0\}$  we compute  $\widehat{f}(x) = \text{sinc}^2(x/2) \in L^1(\mathbb{R})$ , where  $\text{sinc } x = (\sin x)/x$ . Thus by Theorem 2.2.7, we immediately get

$$\int_{-\infty}^{\infty} \text{sinc}^2(x/2) e^{ixt} dx = 2\pi \max\{1 - |t|, 0\} \quad (2.2.11)$$

for all  $t$ , and especially  $\int_{-\infty}^{\infty} \text{sinc}^2(x/2) dx = 2\pi$ . Maple's built-in integral transform package knows this integral:

```
inttrans[fourier]((sin(x/2)/(x/2))^2,x,t); returns 2 Pi (-2 t
Heaviside(t) + t Heaviside(t - 1) - Heaviside(t - 1)
+ t Heaviside(t + 1) + Heaviside(t + 1))
```

This is an inelegant form of the answer we gave. Note that newer versions of Maple can evaluate the integral directly as well, while in previous versions this was only possible with the use of the Fourier transform, so that researchers had to have more knowledge about what they were doing. Thus this is another example of being able to watch technology in progress!  $\square$

**Fourier Transform on  $L^2(\mathbb{R})$ .** Since  $L^2(\mathbb{R})$  is not a subset of  $L^1(\mathbb{R})$ , in contrast to the periodic case, the definition of the Fourier transform cannot be directly transferred onto this space: The function  $f(t)e^{-ixt}$  may not be integrable! For this reason, the Fourier transform on  $L^2(\mathbb{R})$  is usually defined as the continuation of the Fourier transform on  $L^1(\mathbb{R}) \cap L^2(\mathbb{R})$  which is dense in  $L^2(\mathbb{R})$ . An equivalent (and more practical) definition is the following. First of all, note that for  $f \in L^2(\mathbb{R})$ ,  $f_A = f \cdot \chi_{(-A,A)} \in L^1(\mathbb{R})$  for any  $A > 0$ . It can be proved, with some effort, that then  $\widehat{f}_A \in L^2(\mathbb{R})$ .

**Definition 2.2.9** *Let  $f \in L^2(\mathbb{R})$ . If the  $L^2(\mathbb{R})$ -limit of the functions  $\widehat{f}_A$  for  $A \rightarrow \infty$  exists in  $L^2(\mathbb{R})$ , then it is called the Fourier (or Plancherel) transform of  $f$  and is again denoted by  $\widehat{f}$ .*

By definition, the Fourier transform of an  $f \in L^2(\mathbb{R})$  is always in  $L^2(\mathbb{R})$ . It need not be continuous, nor does the Riemann/Lebesgue lemma hold. It can be proved that  $\|f\|_2 = \|\widehat{f}\|_2$  (Parseval equation), and that every function in  $L^2(\mathbb{R})$  is the Fourier transform of an  $f \in L^2(\mathbb{R})$ . An  $f \in L^2(\mathbb{R})$  is reconstructible from its Fourier transform by the same process as in Theorem 2.2.7.

**Theorem 2.2.10** *For any  $f \in L^2(\mathbb{R})$ ,*

$$f(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \widehat{f}(x) e^{ixt} dx = \lim_{A \rightarrow \infty} \frac{1}{2\pi} \int_{-A}^A \widehat{f}(x) e^{ixt} dx \quad (2.2.12)$$

*in  $L^2(\mathbb{R})$ .*

**Example 2.2.11** *Fourier transform of sine-exponential.*

We have already computed  $\widehat{f}(x) = 2 \sin(\pi x)/x$  for  $f(t) = \chi_{(-\pi, \pi)}(t)$ . This  $\widehat{f}$  is in  $L^2(\mathbb{R})$ , but not in  $L^1(\mathbb{R})$ . Theorem 2.2.10 now says that

$$\int_{-\infty}^{\infty} \frac{\sin(\pi x)}{\pi x} e^{ixt} dx = \chi_{(-\pi, \pi)}(t) \quad \text{a.e.} \quad \square$$

It would be interesting to replace this a.e.-statement with a pointwise statement, since a.e.-statements are more difficult to handle when the goal is exact evaluation of a series or integral, and so they are less useful for experimental mathematics. In fact, a standard theorem of Fourier analysis (Jordan's theorem, see [72, p. 205]) is: *If  $f \in L^1(\mathbb{R})$  is of bounded variation in an interval including the point  $t$ , then*

$$\lim_{a \rightarrow \infty} \frac{1}{2\pi} \int_{-a}^a \widehat{f}(x) e^{ixt} dx = \frac{1}{2} (f(t+) + f(t-)).$$

Applied to the above example, this gives

$$\lim_{a \rightarrow \infty} \int_{-a}^a \frac{\sin(\pi x)}{\pi x} e^{ixt} dx = \begin{cases} 0 & \text{for } |t| > \pi, \\ 1 & \text{for } |t| < \pi, \\ \frac{1}{2} & \text{for } |t| = \pi. \end{cases} \quad (2.2.13)$$

## 2.3 More Advanced Fourier Analysis

### 2.3.1 The Poisson Summation Formula

There are obvious similarities between Fourier series and Fourier transforms. Thus one would think that there are connections between the two concepts. That is indeed so, and the link is provided by the Poisson summation formula.

As a first example, take a function  $F \in L^1(\mathbb{T})$  and note that  $\widehat{F}_n = \widehat{f}(n)$  if  $f \in L^1(\mathbb{R})$  is defined to equal  $F$  on  $(-\pi, \pi)$  and to vanish everywhere else. This already is a simple special case of the Poisson summation formula.

A second approach to Poisson summation can be motivated by the following question. Take a function  $g \in C(\mathbb{R})$ , and assume that the sequence  $(g(\frac{n}{w}))$  is absolutely summable for each  $w > 0$ . Consider  $F_w(t) = \sum_{n=-\infty}^{\infty} g(\frac{n}{w}) e^{int}$  for

$w > 0$ . Such functions  $F_w$  are often investigated in connection with summation procedures for Fourier series, and we will do so in the next subsection. A first graphical observation is that the restrictions to  $(-\pi, \pi)$  of these functions  $F_w$  tend to be concentrated more and more around 0 for increasing  $w$ . If, however, the functions  $\frac{1}{w}F_w(t/w)$  are plotted, then there seems to be convergence towards a limit function which depends on  $g$ . What is this limit function, and how can we prove convergence? The answer is again given by the Poisson summation formula.

In general, the Poisson formula links the finite and the infinite transforms via the so-called *periodization*, an operation which associates  $L^1(\mathbb{R})$ -functions in a natural way with  $2\pi$ -periodic functions: For  $f \in L^1(\mathbb{R})$ , set  $F(t) = \sum_{j \in \mathbb{Z}} f(t + 2\pi j)$  for all  $t$  for which the limit exists. The next theorem is (one version of) the classical Poisson formula, linking the Fourier series of  $F$  with the Fourier transform of  $f$ .

**Theorem 2.3.1 (Poisson summation formula).** *Let  $f \in L^1(\mathbb{R})$ .*

a) *The periodization  $F$  exists for almost every  $t \in \mathbb{T}$ , and we have  $F \in L^1(\mathbb{T})$  and  $\|F\|_1 \leq \|f\|_1$ .*

b) *The Fourier series of  $F$  is*

$$\sum_{j=-\infty}^{\infty} f(t + 2\pi j) \sim \frac{1}{2\pi} \sum_{n=-\infty}^{\infty} \widehat{f}(n) e^{int}, \quad (2.3.14)$$

*in other words, we have  $\widehat{F}_n = \widehat{f}(n)$  for all  $n \in \mathbb{Z}$ .*

**Proof.** a) Obviously  $f(t + 2\pi j)$  is integrable over  $[-\pi, \pi]$ , and we have

$$\begin{aligned} \sum_{j=-\infty}^{\infty} \int_{-\pi}^{\pi} |f(t + 2\pi j)| dt &= \sum_{j=-\infty}^{\infty} \int_{-\pi+2\pi j}^{\pi+2\pi j} |f(t)| dt \\ &= \int_{-\infty}^{\infty} |f(t)| dt = \|f\|_{L^1(\mathbb{R})} < \infty. \end{aligned}$$

By B. Levi's theorem, the series  $F$  is absolutely convergent a.e., we have  $F \in L^1(\mathbb{T})$ , and summation and integration can be exchanged. Now we also get  $\|F\|_{L^1(\mathbb{T})} \leq \|f\|_{L^1(\mathbb{R})}$  by using the triangle inequality, doing the exchange and then using the above computation.

b) Now applying B. Levi's theorem to the functions  $f(t + 2\pi j) \cdot e^{-nt}$  for fixed  $n$ , and using the fact that the function  $e^{-int}$  is  $2\pi$ -periodic, we get with the same summation trick as before,

$$\widehat{F}_n = \int_{-\pi}^{\pi} F(t) e^{-int} dt = \int_{-\infty}^{\infty} f(t) e^{-int} dt = \widehat{f}(n).$$

□

Of course, it is interesting to ask when the identity holds pointwise instead of just in the sense of Fourier series. From the Jordan test it can be deduced that if  $f \in L^1(\mathbb{R}) \cap BV(\mathbb{R})$  and  $f(t) = \frac{1}{2}(f(t+) + f(t-))$  everywhere, then equality holds for all  $t$ .

**Example 2.3.2** *Fourier series of hyperbolic functions.*

Choose  $y > 0$  and define

$$f(t) = \begin{cases} e^{-yt} & \text{for } t > 0, \\ 0 & \text{for } t < 0, \end{cases}$$

and  $f(0) = 1/2$ . Then  $\widehat{f}(x) = (y + ix)^{-1}$ , and Poisson says that

$$\sum_{j=-\infty}^{\infty} f(t + 2\pi j) = \frac{1}{2\pi} \sum_{n=-\infty}^{\infty} \frac{1}{y + in} e^{int},$$

which is equivalent to

$$\frac{1}{2\pi} \left( \frac{1}{y} + 2 \sum_{n=1}^{\infty} \frac{y \cos(nt) + n \sin(nt)}{y^2 + n^2} \right) = \sum_{j>t/(2\pi)} e^{-y(t+2\pi j)} + \begin{cases} \frac{1}{2}, & t \in 2\pi\mathbb{Z}, \\ 0, & \text{otherwise.} \end{cases}$$

Setting  $t = 0$ , we get

$$\pi \coth(\pi y) = \frac{1}{y} + 2y \sum_{n=1}^{\infty} \frac{1}{y^2 + n^2}.$$

Setting  $t = 1/2$ , we get

$$\pi \operatorname{cosech}(\pi y) = \frac{1}{y} + 2y \sum_{n=1}^{\infty} \frac{(-1)^n}{x^2 + n^2}.$$

□

**Example 2.3.3** *A quadratic exponential identity.*

Now choose  $s > 0$  and set  $f(t) = e^{-st^2}$ . Then  $\widehat{f}(x) = \sqrt{\frac{\pi}{s}} e^{-x^2/(4s)}$  (see Section 5.6 of the first volume), and Poisson says that

$$\sum_{j=-\infty}^{\infty} e^{-s(t+2\pi j)^2} = \frac{1}{2} \sqrt{\frac{1}{\pi s}} \sum_{n=-\infty}^{\infty} e^{-n^2/(4s)} e^{int}. \quad (2.3.15)$$

By analyticity, this can be extended to hold for all  $\operatorname{Re}(s) > 0$ . We will meet this formula again in Section 3.1.  $\square$

**Example 2.3.4** *A series identity that does not hold pointwise.*

Take an even real-valued continuous  $g \in L^1(\mathbb{R})$  such that  $(g(\frac{n}{w}))$  is absolutely summable for each  $w > 0$  and assume that  $\widehat{g} \in L^1(\mathbb{R})$ . Then using Poisson and the inversion theorem 2.2.7 it follows that

$$\sum_{n=-\infty}^{\infty} g\left(\frac{n}{w}\right) e^{int} = w \sum_{j=-\infty}^{\infty} \widehat{g}(w(t + 2\pi j)), \quad (2.3.16)$$

where equality is meant in  $L^1(\mathbb{T})$ . This equality does not hold pointwise! In Katznelson's book [144] an example is given of a function  $f \in L^1(\mathbb{R})$  with  $\widehat{f} \in L^1(\mathbb{R})$  for which equality does not hold pointwise in Poisson's formula (2.3.14), because the periodization does not converge uniformly. This is the "dangerous pitfall" that we mentioned earlier: Even though the left-hand side of (2.3.16) is a continuous function, the right-hand side is not necessarily continuous; it is only equal to a continuous function a.e. To deduce pointwise equality from this would be wrong! This is a noteworthy difference to the situation in Theorem 2.2.7.

In any case, this gives an answer to the question at the beginning of the present subsection: The limit function (in the  $L^1(\mathbb{R})$ -sense) of the  $F_w$ 's, restricted to  $(-\pi, \pi)$  and then suitably rescaled, is the Fourier transform of  $g$ . This follows from (2.3.16), since

$$\begin{aligned} \left\| \frac{1}{w} F_w\left(\frac{t}{w}\right) \cdot \chi_{(-\pi w, \pi w)} - \widehat{g}(t) \right\|_1 &= \int_{-\pi w}^{\pi w} \left| \sum_{\substack{j=-\infty \\ j \neq 0}}^{\infty} \widehat{g}(t + 2\pi j w) \right| dt + \int_{|t| > \pi w} |\widehat{g}(t)| dt \\ &\leq 2 \int_{|t| > w} |\widehat{g}(t)| dt \rightarrow 0 \quad (w \rightarrow \infty) \end{aligned}$$



if  $\widehat{g} \in L^1(\mathbb{R})$ . Often enough (but not always) this convergence is even uniform on  $\mathbb{R}$ .

This is a very visual theorem; it can be discovered and explored on the computer. In this sense Formula (2.3.16) is the experimental version of Poisson's summation formula!  $\square$

### 2.3.2 Convolution Theorems

As we have seen, the Fourier series even of a continuous function need not converge back to the function, neither in the  $L^1$ -sense nor pointwise. Often, convergence properties of such a series can be improved by putting additional factors into the series to "force convergence." For example, it can be proved (and will be in the next subsection) that for  $f \in C(\mathbb{T})$  the series

$$\sigma_n(f, t) = \frac{1}{2\pi} \sum_{k=-n}^n \left(1 - \frac{|k|}{n+1}\right) \widehat{f}_k e^{ikt}$$

converges to  $f$  uniformly as well as in the  $L^1$ -sense; this is Fejér's famous theorem.

How do these convergence factors work? If we set

$$F_n(t) = \sum_{k=-n}^n \left(1 - \frac{|k|}{n+1}\right) e^{ikt},$$

then the Fourier coefficients of  $\sigma_n(f, t)$  are the product of those of  $f$  and those of  $F_n$ . Thus it seems reasonable that convergence properties of  $\sigma_n$  can be deduced from suitable properties of  $F_n$ . An important relation between these objects is given by the following theorem.

**Theorem 2.3.5 (Convolution theorem for  $L^1(\mathbb{T})$ ).** *For  $f, g \in L^1(\mathbb{T})$  define*

$$h(t) = (f * g)(t) = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(t-u) g(u) du.$$

1. *Then the integral exists for a.e.  $t \in \mathbb{T}$ , and we have  $h \in L^1(\mathbb{T})$  and  $\|h\|_1 \leq \frac{1}{2\pi} \|f\|_1 \|g\|_1$ .*

2. Moreover,  $\widehat{h}_n = \frac{1}{2\pi} \widehat{f}_n \cdot \widehat{g}_n$  holds for the Fourier coefficients.

Applied to the question above, this convolution theorem says that  $\sigma_n(f, t) = (F_n * f)(t)$ . In the next section, this will lead to a proof of Fejér's theorem. We will in fact be able to treat much more general summation kernels.

In the following sections, we will also need a convolution theorem in  $L^1(\mathbb{R})$ .

**Theorem 2.3.6 (Convolution theorem for  $L^1(\mathbb{R})$ ).** For  $f, g \in L^1(\mathbb{R})$  define

$$h(t) = (f * g)(t) = \int_{-\infty}^{\infty} f(t-u)g(u) du.$$

1. Then the integral exists for a.e.  $t \in \mathbb{R}$ , and we have  $h \in L^1(\mathbb{R})$  and  $\|h\|_1 \leq \|f\|_1 \|g\|_1$ .

2. Moreover,  $\widehat{h}(x) = \widehat{f}(x) \cdot \widehat{g}(x)$  holds for the Fourier transforms.

The convolution in  $L^1(\mathbb{R})$  tends to make functions smoother but less localized. If  $g$  is, for example, the characteristic function of an interval, then  $h = f * g$  will be absolutely continuous for every  $L^1$ -function  $f$ . If, on the other hand, both  $f$  and  $g$  are  $L^1$ -functions with bounded supports, equal to, say,  $[-a, a]$  and  $[-b, b]$ , then the support of  $h = f * g$  will be equal to  $[-(a+b), a+b]$ .

Convolution theorems are quite important in computational mathematics! To compute a convolution one usually has to perform many multiplications, so that it is expensive in terms of time and memory. On the other hand, a single multiplication often is cheap. The convolution theorems (which have many analogs for different types of convolutions) say that convolutions can be transformed into multiplications and thus may be much easier to compute than appears on first glance. Techniques for computing convolutions using the fast Fourier transform (FFT) are presented in Chapter 6 of the first volume. The speed of convolutions computed using FFTs is a principal reason that Fourier theory in general, and the FFT in particular, are so important in computational science.

### 2.3.3 Summation Kernels

With regard to Fejér's sum  $\sigma_n$ , the convolution theorem says that  $\sigma_n(f, t) = (F_n * f)(t)$ , and we are interested in the question whether  $F_n * f \rightarrow f$  for  $n \rightarrow \infty$

in a suitable norm ( $L^1$  or uniformly). This question can be generalized: Under what conditions on a family of functions  $(K_w) \subseteq L^1(\mathbb{T})$  is there convergence  $\|K_w * f - f\|_1 \rightarrow 0$  ( $w \rightarrow \infty$ ) for all  $f \in L^1(\mathbb{T})$ ? Under what conditions is there uniform convergence for continuous  $f$ ? Any family of type  $(K_w)$  is called a *kernel*. If there is suitable convergence then  $(K_w)$  is also called an *approximate identity*, and the question is now open to systematic experimentation. Which conditions of a kernel make it an approximate identity?

It is proved in Katznelson's book [144] that the following conditions imply that some family  $(K_w) \subseteq L^1(\mathbb{T})$  is an approximate identity for  $L^p(\mathbb{T})$  ( $1 \leq p < \infty$ ) and for  $C(\mathbb{T})$ :

$$\begin{aligned} \text{(S1)} \quad & \frac{1}{2\pi} \int_{-\pi}^{\pi} K_w(t) dt = 1 \quad \text{for all } w, \\ \text{(S2)} \quad & \int_{-\pi}^{\pi} |K_w(t)| dt \leq M \quad \text{uniformly in } w, \\ \text{(S3)} \quad & \lim_{w \rightarrow \infty} \int_{\delta \leq |t| \leq \pi} |K_w(t)| dt = 0 \quad \text{for all } 0 < \delta < \pi. \end{aligned}$$

Of particular interest here are kernels of the form  $K_w(t) = \sum_{k=-\infty}^{\infty} g(\frac{k}{w}) e^{ikt}$  for suitable functions  $g$ , since many classical kernels are of this form. Now our direction should be clear: the "experimental" version (2.3.16) of the Poisson formula will be of use.

**Theorem 2.3.7** *Let  $K_w(t) = \sum_{k=-\infty}^{\infty} g(\frac{k}{w}) e^{ikt}$  and assume that  $g \in C(\mathbb{R}) \cap L^1(\mathbb{R})$ , that  $(g(\frac{n}{w}))$  is absolutely summable for each  $w > 0$ , and that  $g(0) = 1$  and  $\widehat{g} \in L^1(\mathbb{R})$ . Then  $(K_w)$  is an approximate identity.*

**Proof.** We have to check (S1)–(S3) above. Condition (S1) is a condition on the middle Fourier coefficient of  $K_w$  and follows from  $g(0) = 1$ . Regarding (S2), we have, using (2.3.16) and the theorem of B. Levi,

$$\begin{aligned} \int_{-\pi}^{\pi} |K_w(t)| dt &= \frac{1}{2\pi} \int_{-\pi}^{\pi} \left| w \sum_{j=-\infty}^{\infty} \widehat{g}(w(t + 2\pi j)) \right| dt \\ &\leq w \sum_{j=-\infty}^{\infty} \int_{-\pi}^{\pi} |\widehat{g}(w(t + 2\pi j))| dt \\ &= w \int_{-\infty}^{\infty} |\widehat{g}(wt)| dt = \|\widehat{g}\|_1, \end{aligned}$$

which is a uniform bound. Finally, we get by similar computations as before that

$$\int_{\delta \leq |t| \leq \pi} |K_w(t)| dt \leq \int_{|t| > w\delta} |\widehat{g}(t)| dt,$$

which tends to 0 for  $w \rightarrow \infty$  since  $\widehat{g} \in L^1(\mathbb{R})$ .  $\square$

All of the conditions in Theorem 2.3.7 are easily checked symbolically. Moreover, the methods employed here allow more detailed investigations of the approximation properties of the kernels by direct computation of the Fourier transform. For example, often the *Lebesgue constants*, defined as  $\|K_w\|_1$ , determine the rate of convergence of  $K_w * f$  to  $f$ , or, if  $K_w$  is not an approximate identity, the growth rate of  $K_w * f$ . The Lebesgue constants, as computed in the proof of (S2), satisfy  $\|K_w\|_1 \leq \|\widehat{g}\|_1$ . This bound is precise, since the same Poisson methods also gives

$$\|K_w\|_1 \geq \|\widehat{g}\|_1 - 2 \int_{|t| > \pi w} |\widehat{g}(t)| dt.$$

**Example 2.3.8** *The Fejér kernel.*

The kernel  $F_n$  as defined above comes from the Cesàro summation method applied to Fourier series; thus  $\sigma_n(f, t) = (F_n * f)(t)$ . By geometric summation,  $F_n$  can also be written as

$$F_n(t) = \frac{\sin^2(\frac{n+1}{2}t)}{(n+1)\sin^2(\frac{t}{2})}.$$

If we set  $g(x) = \max\{1 - |x|, 0\}$ , then  $F_n = K_{n+1}$ . Since  $\widehat{g}(t) = \text{sinc}^2(t/2) \in L^1(\mathbb{R})$ , the conditions of Theorem 2.3.7 are satisfied. Thus we deduce directly that  $F_n$  is an approximate identity in  $L^p(\mathbb{T})$  and in  $C(\mathbb{T})$ . Moreover, since  $\widehat{g}$  is non-negative, for the Lebesgue constants we get  $\|F_n\|_1 = \|\widehat{g}\|_1 = 2\pi$  (compare with the example in Section 2.2.2).  $\square$

**Example 2.3.9** *The Poisson kernel.*

Another important summation method is Abel summation. Applied to Fourier series, it leads to the Poisson kernel, defined as

$$P_r(t) = \sum_{k=-\infty}^{\infty} r^{|k|} e^{ikt} = \frac{1-r^2}{1-2r\cos(t)+r^2} \quad \text{for } 0 \leq r < 1.$$

The question is whether  $P_r * f \rightarrow f$  in  $L^p(\mathbb{T})$  or in  $C(\mathbb{T})$  for  $r \rightarrow 1$ . By setting  $w = -1/\ln(r)$  and  $g(x) = e^{-|x|}$ , we have  $P_r(t) = K_w(t)$ . Since  $\widehat{g}(t) = 2/(1+t^2) \in L^1(\mathbb{R})$ , Theorem 2.3.7 produces convergence of  $P_r * f$  to  $f$ . Similarly to the Fejér kernel,  $P_r$  as well as  $g$  and  $\widehat{g}$  are nonnegative, so that we again have  $\|P_r\|_1 = \|K_w\|_1 = \|\widehat{g}\|_1 = 2\pi$ .  $\square$

**Example 2.3.10** *The Dirichlet kernel.*

The same methods also explain why the usual summation of Fourier series does not always give convergence. This summation corresponds to the Dirichlet kernel

$$D_n(t) = \sum_{k=-n}^n e^{ikt} = \frac{\sin\left(\left(n + \frac{1}{2}\right)t\right)}{\sin\left(\frac{t}{2}\right)}$$

via  $s_n(f, t) = (D_n * f)(t)$ . The Dirichlet kernel is of the form  $K_n$  with  $g = \chi_{[-1,1]}$ . This  $g$  is neither continuous, nor does it have a Fourier transform in  $L^1(\mathbb{R})$  (its Fourier transform is  $\widehat{g}(t) = 2\text{sinc}(t)$ .) Thus Theorem 2.3.7 is not applicable. But the experimental Poisson formula (2.3.16) can still be used to estimate the Lebesgue constants and gives  $\|D_n\|_1 = \int_{-n\pi}^{n\pi} |2\text{sinc}(t)| dt + O(1) = \frac{8}{\pi} \ln(n) + O(1)$ . These are unbounded, and so the Dirichlet kernel can be expected to have worse summation properties than the Fejér and Poisson kernels above.  $\square$

In summary, the methods described here are very useful to gauge norm-convergence properties of kernels of the special form  $K_w(t) = \sum g\left(\frac{k}{w}\right) e^{ikt}$  in a direct, computational way, and they open the door to further experimentation. Our description has been adapted from [118], but reportedly these methods go back at least to Korovkin.

Of course, norm-convergence does not imply pointwise convergence. As described in [144], instead of (S1)–(S3) the following properties of a kernel can

be used to prove pointwise convergence of  $K_w * f$  to  $f$ :  $K_w$  satisfies (S1), is non-negative and even, and satisfies

$$\lim_{w \rightarrow \infty} \left( \sup_{\delta \leq |t| \leq \pi} K_w(t) \right) = 0 \quad \text{for all } 0 < \delta < \pi.$$

This allows one to decide, for given  $f \in L^1(\mathbb{T})$  and  $t_0 \in \mathbb{R}$ , whether  $(K_w * f)(t_0)$  converges to  $f(t_0)$ .

For the Fejér kernel, it leads to Lebesgue's condition: *If there exists a value  $\check{f}(t_0)$  such that*

$$\lim_{h \rightarrow 0} \frac{1}{h} \int_0^h \left| \frac{f(t_0 + t) + f(t_0 - t)}{2} - \check{f}(t_0) \right| dt = 0,$$

*then  $\sigma_n(f, t) \rightarrow \check{f}(t_0)$  for  $n \rightarrow \infty$ . In particular,  $\sigma_n(f, t) \rightarrow f(t)$  a.e.*

For the Poisson kernel, it leads to Fatou's condition: *If there exists a value  $\check{f}(t_0)$  such that*

$$\lim_{h \rightarrow 0} \frac{1}{h} \int_0^h \left( \frac{f(t_0 + t) + f(t_0 - t)}{2} - \check{f}(t_0) \right) dt = 0,$$

*then  $(P_r * f)(t_0) \rightarrow \check{f}(t_0)$  for  $r \rightarrow 1-$ . In particular,  $(P_r * f)(t) \rightarrow f(t)$  a.e.*

The convergence is uniform on closed subintervals where  $f$  is continuous.

## 2.4 Examples and Applications

### 2.4.1 The Gibbs Phenomenon

If a function  $f \in L^1(\mathbb{T})$  is of bounded variation, it may have jump discontinuities. The Jordan test says that the Fourier series of  $f$  converges to the center of the gap at such a point. Directly to the left and right of the jump, the series converges pointwise, but not uniformly on any interval containing the discontinuity, to the function. The function  $f(t) = (\pi - t)/2 = \sum n^{-1} \sin(nt)$  on  $[0, 2\pi]$  is a good example for this behavior, see Figure 2.1, where the series for  $(\pi - t)/2$  is evaluated to 20 terms.

One notices that the cut-off Fourier series “overshoots” the function at the discontinuity. These oscillations do not diminish when more terms are added;

they just move closer to the discontinuity. When the experimental physicist A. Michelson (famous for the Michelson-Morley experiment, which led to Einstein's special relativity) had built a machine to calculate Fourier series and fed it a discontinuous function, he noticed this phenomenon. It was unexpected for him, but after hand calculations confirmed this behavior, he wrote a letter to *Nature* in 1898, expressing his doubts that "a real discontinuity can replace a sum of continuous curves" (cited after Bhatia [25]). J. Willard Gibbs, one of the founders of modern thermodynamics, replied to this letter and clarified the matter. Thus here we have another example of a mathematical theorem which was experimentally discovered (by an experimental physicist)!

Now what is the explanation for the Gibbs phenomenon? Inspection of the picture shows that the largest overshoot seems to occur around the point  $\pi/N$  if  $N$  terms of the Fourier series are added. Thus we compute

$$s_N\left(f, \frac{\pi}{N}\right) = \sum_{n=1}^N \frac{\sin\left(\frac{n\pi}{N}\right)}{n} = \frac{\pi}{N} \sum_{n=1}^N \frac{\sin\left(\frac{n\pi}{N}\right)}{\frac{n\pi}{N}},$$

where the last sum is a Riemann sum for the integral

$$I = \int_0^{\pi} \frac{\sin(t)}{t} dt.$$

Therefore,  $s_N(f, \frac{\pi}{N}) \rightarrow I$  for  $N \rightarrow \infty$ . Since  $I/f(0+) = I/(\pi/2) \approx 1.178979744$ , this explains why the overshoot does not go away for large  $N$ . This overshoot of roughly 18% is not dependent on the function  $f$  used here as an example, but can be observed (and proved) for any jump discontinuity.

Does the Gibbs phenomenon vanish when we use Fejér's series instead of the Fourier series? Figure 2.2 shows the Fejér approximation to  $f$ , again to 20 terms. The oscillation is now replaced by a pronounced "undershoot" to the right of 0 (this can be explained by the positivity of the Fejér kernel); again, the undershoot can be observed to move closer to the discontinuity but not vanish altogether when more terms are added. In fact, we have to pay for the increased smoothness of the approximation by its reduced willingness to snuggle up to the limit function.

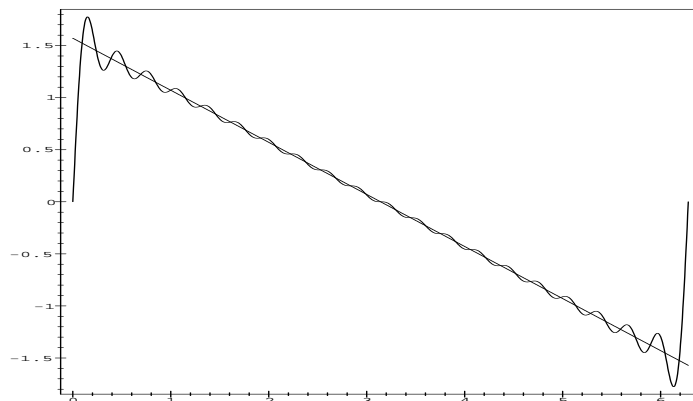


Figure 2.1: The Gibbs phenomenon for Fourier series

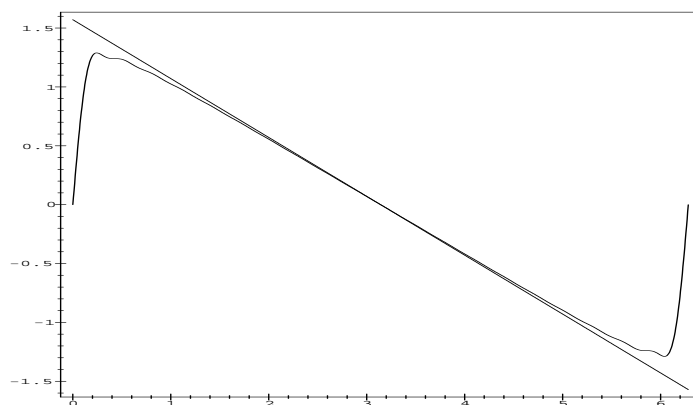


Figure 2.2: The Gibbs phenomenon for Fejér series



### 2.4.2 A Function with Given Integer Moments

The  $k$ th moment of a function  $f \in L^1(\mathbb{R})$  is defined as

$$\mu_k(f) = \int_{-\infty}^{\infty} f(t) t^k dt,$$

provided that  $t \mapsto f(t) t^k \in L^1(\mathbb{R})$ . The *Hamburger moment problem* is the problem to find a function  $f$  with a given sequence of moments  $(\mu_k)$ . This problem is underdetermined: There can be non-vanishing functions whose every moment is 0. This is easily seen by the following argument. Assume that  $f$  is  $k$  times differentiable with every derivative in  $L^1(\mathbb{R})$ . Then by partial integration,

$$\widehat{f^{(k)}}(x) = \int_{-\infty}^{\infty} f^{(k)}(t) e^{-ixt} dt = (ix)^k \widehat{f}(x),$$

and by the inversion theorem,  $\int_{-\infty}^{\infty} \widehat{f}(x) x^k dx = 2\pi f^{(k)}(0)$ . Thus if  $f \in L^1(\mathbb{R})$  is infinitely differentiable with every derivative in  $L^1(\mathbb{R})$  and satisfies  $f^{(k)}(0) = 0$  for all  $k$ , then all moments of  $\widehat{f}$  vanish. Of course, such non-trivial functions  $f$  exist, even with compact support.

Obviously, for an even function every odd moment vanishes. To generalize this, it is quite easy to find, for given  $n \in \mathbb{N}$ , a function whose  $k$ th moment is non-zero precisely when  $k \bmod n = 0$ . Just choose an infinitely differentiable function  $f \in L^1(\mathbb{R})$  with all derivatives in  $L^1(\mathbb{R})$ , and such that all its derivatives are non-zero at 0. Then set  $g(t) = f(t^n)$ , and by the chain and product rule of differentiation,  $g^{(k)}(0)$  is non-zero precisely for  $k \bmod n = 0$ . Thus,  $\widehat{g}$  satisfies the moment condition. If  $f$  is analytic, say  $f(t) = \sum_{j=0}^{\infty} a_j t^j$ , then  $g^{(k)}(0) = (qn)! a_q$  if  $k = qn$  and  $g^{(k)}(0) = 0$  otherwise. An example for such a function  $f$  is  $f(t) = (1+t) e^{-t^2/2}$ .

When this was first investigated some years ago, numerical fast Fourier transforms were used—as a test—to calculate the moments for  $g(t) = f(t^n)$ , where  $f(t) = (1+t) e^{-t^2/2}$  as above. The scheme for doing this is presented in Section 6.1 of the first volume. When this was done, it was noticed that the resulting moment values were extremely accurate, far more than one would expect based on what amounts to a simple step-function approximation to the Fourier integrals. Readers may recall that we also encountered this phenomenon in Section 5.2 of the first volume. This phenomenon, which is rooted in the Euler-Maclaurin

summation formula, is the foundation of some extremely efficient and highly accurate numerical quadrature schemes that we shall see in Sections 7.4.2 and 7.4.3.

### 2.4.3 Bernoulli Convolutions

Consider the discrete probability density on the real line with measure  $\frac{1}{2}$  at each of the two points  $\pm 1$ . The corresponding measure is the so-called Bernoulli measure, denoted  $b(X)$ . For every  $0 < q < 1$ , the infinite convolution of measures

$$\mu_q(X) = b(X) * b(X/q) * b(X/q^2) * \cdots \quad (2.4.17)$$

exists as a weak limit of the finite convolutions. The most basic theorem about these infinite Bernoulli convolutions is due to Jessen and Wintner ([140]). They proved that  $\mu_q$  is always continuous and that it is either absolutely continuous or purely singular. This statement follows from a more general theorem on infinite convolutions of purely discontinuous measures (Theorem 35 in [140]); however, it is not difficult to prove the statement directly with the use of Kolmogoroff's 0-1-law (which can be found, e.g., in [26]). The question about these measures is to decide for which values of the parameter  $q$  they are singular, and for which  $q$  they are absolutely continuous.

This question can be recast in a more real-analytic way by defining the distribution function  $F_q$  of  $\mu_q$  as

$$F_q(t) = \mu_q(-\infty, t], \quad (2.4.18)$$

and to ask for which  $q$  this continuous, increasing function  $F_q$  is singular, and for which it is absolutely continuous. Note that  $F_q$  satisfies  $F_q(t) = 0$  for  $t < -1/(1-q)$  and  $F_q(t) = 1$  for  $t > 1/(1-q)$ .

Another way to define the distribution function  $F_q$  is by functional equations:  $F_q$  is the only bounded solution of the functional equation

$$F(t) = \frac{1}{2}F\left(\frac{t-1}{q}\right) + \frac{1}{2}F\left(\frac{t+1}{q}\right) \quad (2.4.19)$$

with the above restrictions. Moreover, if  $F_q$  is absolutely continuous and thus has a density  $f_q \in L^1(\mathbb{R})$ , then  $f_q$  satisfies the functional equation

$$2q f(t) = f\left(\frac{t-1}{q}\right) + f\left(\frac{t+1}{q}\right) \quad (2.4.20)$$

almost everywhere. This is a special case of a much more general class of equations, namely two-scale difference equations. Those are functional equations of the type

$$f(t) = \sum_{n=0}^N c_n f(\alpha t - \beta_n) \quad (t \in \mathbb{R}) \quad (2.4.21)$$

with  $c_n \in \mathbb{C}$ ,  $\beta_n \in \mathbb{R}$  and  $\alpha > 1$ . They were first discussed by Ingrid Daubechies and Jeffrey C. Lagarias, who proved existence and uniqueness theorems and derived some properties of  $L^1$ -solutions [98, 99]. One of their theorems, which we state here in part for the general equation (2.4.21) and in part for the specific case (2.4.20), is the following:

**Theorem 2.4.1** 1. *If  $\alpha^{-1}(c_0 + \cdots + c_N) = 1$ , then the vector space of  $L^1(\mathbb{R})$ -solutions of (2.4.21) is at most 1-dimensional.*

2. *If, for given  $q \in (0, 1)$ , equation (2.4.20) has a non-trivial  $L^1$ -solution  $f_q$ , then its Fourier transform satisfies  $\widehat{f}_q(0) \neq 0$ , and is given by*

$$\widehat{f}_q(x) = \widehat{f}_q(0) \prod_{n=0}^{\infty} \cos(q^n x). \quad (2.4.22)$$

*In particular, for normalization we can assume  $\widehat{f}_q(0) = 1$ .*

3. *On the other hand, if the right-hand side of (2.4.22) is the Fourier transform of an  $L^1$ -function  $f_q$ , then  $f_q$  is a solution of (2.4.20).*

4. *Any non-trivial  $L^1$ -solution of (2.4.21) is finitely supported. In the case of (2.4.20), the support of  $f_q$  is contained in  $[-1/(1-q), 1/(1-q)]$ .*

This implies in particular that the question whether the infinite Bernoulli convolution (2.4.17) is absolutely continuous is equivalent to the question whether (2.4.20) has a non-trivial  $L^1$ -solution. Now what is known about these questions?

It is relatively easy to see that in the case  $0 < q < 1/2$ , the solution of (2.4.20) is singular; it is in fact a Cantor function, meaning that it is constant on a dense set of intervals. This was first proved by R. Kershner and A. Wintner [146]. (An example of a Cantor function is depicted in Figure 6.1 of the first volume.)

It is also easy to see that in the case  $q = 1/2$  there is an  $L^1$ -solution of (2.4.19), namely  $f_{1/2}(t) = \frac{1}{4} \chi_{[-2,2]}(t)$ . Moreover, this can be used to construct a solution for every  $q = 2^{-1/p}$  where  $p$  is an integer, namely

$$f_q(t) = f_{1/2}(t) * f_{1/2}(qt) * \cdots * f_{1/2}(q^{p-1}t). \quad (2.4.23)$$

This was first noted by Wintner via the Fourier transform [209]. Explicitly, we have

$$\begin{aligned} \widehat{f_{2^{-1/p}}}(x) &= \prod_{n=0}^{\infty} \cos(2^{-n/p}x) = \prod_{m=0}^{\infty} \prod_{k=0}^{p-1} \cos(2^{-(m+k/p)}x) \\ &= \widehat{f_{1/2}}(x) \cdot \widehat{f_{1/2}}(q^{1/p}x) \cdots \widehat{f_{1/2}}(q^{(p-1)/p}x), \end{aligned}$$

which is equivalent to (2.4.23) by the convolution theorem.

Note that the regularity of these solutions  $f_{2^{-1/p}}$  increases when  $p$  and thus  $q = 2^{-1/p}$  increases:  $f_{2^{-1/p}} \in C^{p-2}(\mathbb{R})$ . From the results given so far, one might therefore surmise that (2.4.20) would have a non-trivial  $L^1$ -solution for every  $q \geq 1/2$  with increasing regularity when  $q$  increases. This supposition, however, would be wrong, and it came as a surprise when Erdős proved in 1939 [111] that there are some values of  $1/2 < q < 1$  for which (2.4.20) does not have an  $L^1$ -solution, namely, the inverses of Pisot numbers. A *Pisot number* (discussed further in Exercise 13 of Chapter 7) is defined to be an algebraic integer greater than 1 all of whose algebraic conjugates lie inside the unit disk. The best known example of a Pisot number is the golden mean  $\varphi = (\sqrt{5}+1)/2$ . The characteristic property of Pisot numbers is that their powers quickly approach integers: If  $a$  is a Pisot number then there exists a  $\theta$ ,  $0 < \theta < 1$ , such that

$$\text{dist}(a^n, \mathbb{Z}) \leq \theta^n \quad \text{for all } n \in \mathbb{N}. \quad (2.4.24)$$

Erdős used this property to prove that if  $q = 1/a$  for a Pisot number  $a$ , then  $\limsup_{x \rightarrow \infty} |\widehat{f_q}(x)| > 0$ . Thus in these cases,  $f_q$  cannot be in  $L^1(\mathbb{R})$ , since that would contradict the Riemann-Lebesgue lemma. Erdős's proof uses the Fourier transform  $\widehat{f_q}$ : Consider, for  $N \in \mathbb{N}$ ,

$$\left| \widehat{f_q}(q^{-N}\pi) \right| = \prod_{n=1}^{\infty} |\cos(q^n\pi)| \cdot \prod_{n=0}^{N-1} |\cos(q^{-n}\pi)| =: C \cdot p_N,$$

where  $C > 0$ . Moreover, choose  $\theta \neq 1/2$  according to (2.4.24) and note that

$$\begin{aligned} p_N &= \prod_{\substack{n=0 \\ \theta^n \leq 1/2}}^{N-1} |\cos(q^{-n}\pi)| \cdot \prod_{\substack{n=0 \\ \theta^n > 1/2}}^{N-1} |\cos(q^{-n}\pi)| \\ &\geq \prod_{\substack{n=0 \\ \theta^n \leq 1/2}}^{N-1} \cos(\theta^n\pi) \cdot \prod_{\substack{n=0 \\ \theta^n > 1/2}}^{N-1} |\cos(q^{-n}\pi)| \\ &\geq \prod_{\substack{n=0 \\ \theta^n \leq 1/2}}^{\infty} \cos(\theta^n\pi) \cdot \prod_{\substack{n=0 \\ \theta^n > 1/2}}^{\infty} |\cos(q^{-n}\pi)| = C' > 0, \end{aligned}$$

independently of  $N$ .

In 1944, Raphaël Salem [186] showed that the reciprocals of Pisot numbers are the only values of  $q$  where  $\widehat{f}_q(x)$  does not tend to 0 for  $x \rightarrow \infty$ . In fact, no other  $q > \frac{1}{2}$  are known at all where  $F_q$  is singular. Moreover, the set of explicitly given  $q$  with absolutely continuous  $F_q$  is also not very big: The largest such set known to date was found by Adriano Garsia in 1962 [115]. It contains reciprocals of certain algebraic numbers (such as roots of the polynomials  $x^{n+p} - x^n - 2$  for  $\max\{p, n\} \geq 2$ ) besides the roots of  $\frac{1}{2}$ .

Matters remained in this state for more than 30 years; the question remained settled only for countably many  $q \in [\frac{1}{2}, 1)$ . The most recent significant progress then was made in 1995 by Boris Solomyak [195], who developed exciting new methods in geometric measure theory to prove that  $F_q$  is in fact absolutely continuous for almost every  $q \in [\frac{1}{2}, 1)$ . (See also [172] for a simplified proof and [171] for a survey and some newer results.)

This, however, yields no explicit result; the set of  $q$ 's for which the behavior of  $F_q$  is known explicitly is the same as before. Here we now suggest an experimental approach to at least identify  $q$ -values for which the behavior of  $F_q$  can be guessed. In fact, define a map  $T_q$ , mapping the set of  $L^1$ -functions with support in  $[-1/(1-q), 1/(1-q)]$  and with  $\widehat{f}_q(0) = 1$  into itself, by

$$(T_q f)(t) = \frac{1}{2q} \left( f\left(\frac{t-1}{q}\right) + f\left(\frac{t+1}{q}\right) \right) \quad \text{for } t \in \mathbb{R}.$$

Then note that the fixed points of  $T_q$  are the solutions of (2.4.20) and that  $T_q$  is nonexpansive (cf. Chapter 6). Therefore one may have hope that by iterating

the operator it may be possible to approximate the fixed point. In fact, if a sequence of iterates  $T_q^n f$  converges in  $L^1(\mathbb{R})$  for some initial function  $f$ , then the limit will be a fixed point of  $T_q$ , since  $T_q$  is continuous. It is, however, not easy to prove convergence; no convergence proof is known. It is, on the other hand, possible to prove a weaker result, namely convergence in the mean, provided that a fixed point exists: *If a solution  $f_q \in L^1(\mathbb{R})$  of (2.4.20) exists, then for every initial function  $f \in L^1(\mathbb{R})$  with support in  $[-1/(1-q), 1/(1-q)]$ ,*

$$\lim_{n \rightarrow \infty} \left\| \frac{1}{n} \sum_{k=0}^{n-1} T_q^k f - f_q \right\|_1 = 0.$$

This theorem follows from properties of Markov operators [150] and from a result by Mauldin and Simon [159], showing that if an  $L^1$ -density  $f_q$  exists, then it must be positive a.e. on its support.

In practice, we observe that the iterates usually seem to converge directly, even without the means. Plotting them, we hope to infer existence and regularity of  $L^1$ -solutions by visual inspection. The figures on the next pages show the 25th iterate for  $f = \chi_{-1/(1-q), 1/(1-q)}$  as initial function. Figure 2.3 shows convergence to  $f_{1/2} * f_{1/2}$  for  $q = 1/\sqrt{2}$ ; Figure 2.4 shows that for  $q = (\sqrt{5} - 1)/2$ , the iterates do not converge to a meaningful function. It is not known if there is a density for any rational  $q \in (1/2, 1)$ . Figures 2.5 and 2.6 show that there seems to be a continuous limit in both cases shown; moreover, regularity seems to increase when  $q$  increases.

## 2.5 Some Curious Sinc Integrals

Define

$$I_n = \int_0^\infty \operatorname{sinc}(x) \cdot \operatorname{sinc}\left(\frac{x}{3}\right) \cdots \operatorname{sinc}\left(\frac{x}{2n+1}\right) dx.$$

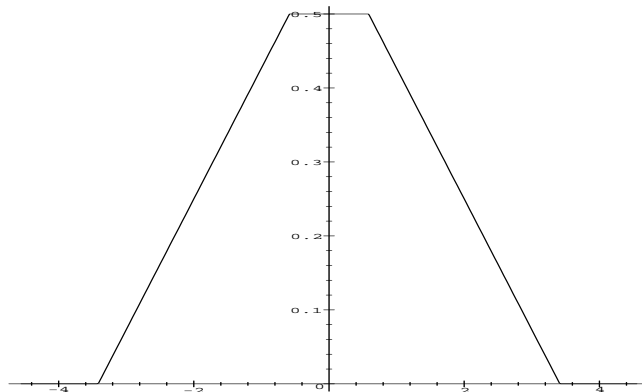
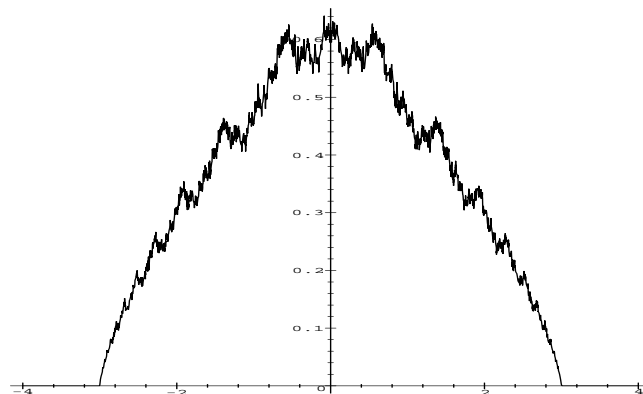
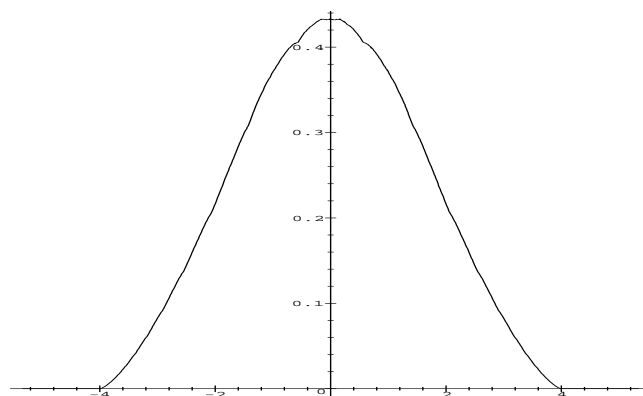


Figure 2.3: 25th iterate for  $q = 1/\sqrt{2}$



Figure 2.4: 25th iterate for  $q = (\sqrt{5} - 1)/2$

Figure 2.5: 25th iterate for  $q = 2/3$ Figure 2.6: 25th iterate for  $q = 3/4$



Then *Maple* or *Mathematica* evaluate

$$\begin{aligned}
 I_0 &= \int_0^\infty \operatorname{sinc}(x) dx = \frac{\pi}{2}, \\
 I_1 &= \int_0^\infty \operatorname{sinc}(x) \operatorname{sinc}\left(\frac{x}{3}\right) dx = \frac{\pi}{2}, \\
 &\vdots \\
 I_6 &= \int_0^\infty \operatorname{sinc}(x) \cdot \operatorname{sinc}\left(\frac{x}{3}\right) \cdots \operatorname{sinc}\left(\frac{x}{13}\right) dx = \frac{\pi}{2}, \quad \text{but} \\
 I_7 &= \int_0^\infty \operatorname{sinc}(x) \cdot \operatorname{sinc}\left(\frac{x}{3}\right) \cdots \operatorname{sinc}\left(\frac{x}{15}\right) dx = \\
 &\qquad\qquad\qquad \frac{467807924713440738696537864469}{935615849440640907310521750000} \pi,
 \end{aligned}$$

where the fraction is approximately 0.499999999992646 . . .

When this fact was recently verified by a researcher using a computer algebra package, he concluded that there must be a “bug” in the software. This conclusion may be too hasty, but it does raise the question: How far can we trust our computer algebra system? Or as computer scientists often ask, “Is it a bug or a feature?”

In this section, we will derive general formulas for this type of sinc integrals, thereby proving that all of the above evaluations are in fact correct. Thus, this is a somewhat cautionary example for too enthusiastically inferring patterns from symbolic or numerical computations. The material comes from [33], and additional information can also be found in [34].

### 2.5.1 The Basic Sinc Integral

It will turn out that the general multi-sinc integral can be reduced to the integral  $I_0$ , so that it makes sense to first evaluate this integral. Note that the function  $\operatorname{sinc}(x)$  is not an element of  $L^1(\mathbb{R})$ ! Thus, Lebesgue theory cannot be applied here directly, and in fact the integral has to be interpreted correctly. Here we use the usual interpretation

$$I_0 = \int_0^\infty \operatorname{sinc}(x) dx = \lim_{a \rightarrow \infty} \int_0^a \operatorname{sinc}(x) dx.$$

Thus we interpret it as an improper Riemann integral, or at best as the limit of Lebesgue integrals.

Of course, *Maple* or *Mathematica* directly evaluate  $I_0 = \frac{\pi}{2}$ , but this is not helpful for those who demand understanding or a proof. Where does this evaluation come from? Peering behind the covers, we find that *Maple* knows

$$\int_0^a \operatorname{sinc}(x) dx = \operatorname{Si}(a) \rightarrow \frac{\pi}{2} \quad \text{for } a \rightarrow \infty.$$

However, this just shifts the problem to another level, since Si equals the integral by definition. Here we will now give several proofs of this identity: One will be short (and incomplete), one will be wrong, and one will then be constructive!

For the first proof we just remember the Jordan theorem in Section 2.2.2, which directly implies that

$$\lim_{a \rightarrow \infty} \int_{-a}^a \operatorname{sinc}(\pi x) e^{ixt} dx = \frac{1}{2} (\chi_{(-\pi, \pi)}(t+) + \chi_{(-\pi, \pi)}(t-)),$$

so that  $t = 0$  gives the desired evaluation. However, this is only a proof modulo the Jordan theorem. A direct proof would still be preferable.

The second “proof” is not a proof, just an idea: Write the sinc function as an inner integral and then use Fubini. Writing  $1/x = \int_0^\infty e^{-tx} dt$ , we would have to use Fubini on the function  $g(t, x) = e^{-tx} \sin(x)$  on  $\mathbb{R} \times \mathbb{R}$ . The double integral that results from exchanging the integration order does, in fact, give

$$\int_0^\infty \int_0^\infty e^{-tx} \sin(x) dx dt = \int_0^\infty \frac{1}{1+t^2} dt = \frac{\pi}{2}.$$

However, this exchange is not allowed, since the function  $g$  is not in  $L^1(\mathbb{R}^2)$ .

But this idea can now be made into a proof which is valid and constructive. If  $g$  is not  $L^1$  on  $\mathbb{R}^2$ , then we just have to restrict the domain of  $g$  at first. Now Fubini is applicable in

$$\begin{aligned} \int_0^a \operatorname{sinc}(x) dx &= \int_0^a \int_0^\infty e^{-xt} \sin(x) dt dx \\ &= \int_0^\infty \int_0^a e^{-xt} \sin(x) dx dt \\ &= \int_0^\infty \frac{1}{1+t^2} [1 - e^{-at}(t \sin(a) + \cos(a))] dt, \\ &= \frac{\pi}{2} - \int_0^\infty \frac{e^{-at}}{1+t^2} (t \sin(a) + \cos(a)) dt, \end{aligned}$$

and the final integral goes to 0 for  $a \rightarrow \infty$  as follows by elementary estimates (see [26]).

Another constructive method to evaluate the sinc integral is given in Item 26 in the Exercises at the end of this chapter.

## 2.5.2 Iterated Sinc Integrals

Now let  $n \geq 1$  and  $a_0, a_1, \dots, a_n$  be positive reals. Our goal is to find inequalities (explicit formulas will be given in the exercises) for the integral

$$\tau = \int_0^\infty \prod_{k=0}^n \operatorname{sinc}(a_k x) dx$$

which in particular explain the behavior of the integrals  $I_n$ . For simplicity and without loss, we can assume that  $a_0 = 1$ .

**Theorem 2.5.1** *Let  $s = \sum_{k=1}^n a_k$ . If  $s \leq 1$ , then  $\tau = \frac{\pi}{2}$ ; if  $s > 1$ , then  $\tau < \frac{\pi}{2}$ .*

**Proof.** Let  $\tau(x) = \prod_{k=0}^n \operatorname{sinc}(a_k x)$ . Note that  $\operatorname{sinc}(a_k x) = \widehat{f_k}(x)$  with  $f_k = \frac{1}{2a_k} \chi_{[-a_k, a_k]}$ . Thus by the convolution theorem,  $\tau(x) = (f_0 * \dots * f_n)^\wedge(x)$ , and by the inversion theorem,

$$\int_{-\infty}^\infty \tau(x) dx = 2\pi (f_0 * \dots * f_n)(0) = 2\pi \frac{1}{2} \int_{-1}^1 (f_1 * \dots * f_n)(u) du. \quad (2.5.25)$$

Now since the support of  $f_k$  equals  $[-a_k, a_k]$ , the support of  $(f_1 * \dots * f_n)$  equals  $[-s, s]$ . If  $s \leq 1$ , then

$$\begin{aligned} \int_{-1}^1 (f_1 * \dots * f_n)(u) du &= \int_{-\infty}^\infty (f_1 * \dots * f_n)(u) du \\ &= (f_1 * \dots * f_n)^\wedge(0) = \prod_{k=1}^n \operatorname{sinc}(a_k 0) = 1, \end{aligned}$$

and  $\int_{-\infty}^\infty \tau(x) dx = \pi$  follows. If, on the other hand,  $s > 1$ , then the interval  $[-1, 1]$  is strictly inside the support of  $(f_1 * \dots * f_n)$ . Since  $(f_1 * \dots * f_n)$  is strictly positive in the interior of its support, we get

$$\int_{-1}^1 (f_1 * \dots * f_n)(x) dx < \int_{-\infty}^\infty (f_1 * \dots * f_n)(x) dx = 1,$$

and  $\int_{-\infty}^{\infty} \tau(x) dx < \pi$  follows.  $\square$

This theorem explains why the values of  $I_n$  suddenly drop below  $\frac{\pi}{2}$  at  $n = 7$  and not before: We have  $\frac{1}{3} + \frac{1}{5} + \cdots + \frac{1}{13} < 1$  but  $\frac{1}{3} + \frac{1}{5} + \cdots + \frac{1}{13} + \frac{1}{15} > 1$ .

A geometric interpretation of this behavior can also be given. Consider the polyhedra

$$P = \{(x_1, \dots, x_n) : -1 \leq \sum_{k=1}^n x_k \leq 1, -a_k \leq x_k \leq a_k \text{ for } k = 1, \dots, n\},$$

$$Q = \{(x_1, \dots, x_n) : -1 \leq \sum_{k=1}^n a_k x_k \leq 1, -1 \leq x_k \leq 1 \text{ for } k = 1, \dots, n\},$$

$$H = \{(x_1, \dots, x_n) : -1 \leq x_k \leq 1 \text{ for } k = 1, \dots, n\}.$$

Then by formula (2.5.25),

$$\begin{aligned} \tau &= \frac{\pi}{2^n a_1 \cdots a_n} \int_0^{\min(1, s)} (\chi_{[-a_1, a_1]} * \cdots * \chi_{[-a_n, a_n]})(x) dx \\ &= \frac{\pi}{2} \frac{\text{Vol}(P)}{2^n a_1 \cdots a_n} = \frac{\pi}{2} \frac{\text{Vol}(Q)}{\text{Vol}(H)}. \end{aligned}$$

Thus the value of  $\tau$  drops below  $\frac{\pi}{2}$  precisely when the constraint  $-1 \leq \sum a_k x_k \leq 1$  becomes active and “bites” into the hypercube  $H$ .

Of course, the same methods will also work for infinite products. Consider the function

$$C(x) = \prod_{n=1}^{\infty} \cos\left(\frac{x}{n}\right)$$

which is continuous since the product is absolutely convergent. We are interested in the integral  $\mu = \int_0^{\infty} C(x) dx$ . High precision numerical evaluation of this highly oscillatory integral is by no means straightforward but possible. We get

$$\int_0^{\infty} C(x) dx \approx 0.785380557298632873492583011467332524761,$$

while  $\frac{\pi}{4} \approx 0.785398$  only differs in the fifth significant place. Can this numerical evaluation  $\mu < \frac{\pi}{4}$  be confirmed symbolically? Indeed it can, by reduction to a sinc integral of the above type, only this time with an infinite product. Recall

the sine product (1.2.11) and note that a corresponding product for the cosine can be derived:

$$\sin(x) = x \prod_{n=1}^{\infty} \left(1 - \frac{x^2}{\pi^2 n^2}\right), \quad \cos(x) = \prod_{k=0}^{\infty} \left(1 - \frac{4x^2}{\pi^2 (2k+1)^2}\right).$$

Using this we get, by exchanging the order of multiplication,

$$C(x) = \prod_{n=1}^{\infty} \prod_{k=0}^{\infty} \left(1 - \frac{4x^2}{\pi^2 n^2 (2k+1)^2}\right) = \prod_{k=0}^{\infty} \operatorname{sinc}\left(\frac{2x}{2k+1}\right).$$

Now apply the theorem to get that

$$\mu = \int_0^{\infty} C(x) dx = \lim_{N \rightarrow \infty} \int_0^{\infty} \prod_{k=0}^N \operatorname{sinc}\left(\frac{2x}{2k+1}\right) dx < \frac{\pi}{4}.$$

This remarkable observation was made by Bernard Mares, then 17, and led to the entire development that we have given of the iterated sinc integrals. More examples are given in the Exercises. There is an interesting connection with random harmonic series in [187].

## 2.6 Korovkin's Three Function Theorems

In 1953, Pavel Korovkin [148] provided an approach to uniform approximation results that is especially well suited to computational assistance and discovery. While the result can be given much more generally, we limit ourselves to the two most basic cases.

Below we let  $\iota$  denote the identity function  $t \mapsto t$ , we let  $C[0,1]$  denote the continuous functions on the unit interval, and  $\rightrightarrows$  denotes uniform convergence (i.e., in the supremum norm). Also,  $[0,1]$  can easily be replaced by any compact interval  $[a,b]$ .

Recall that an operator between continuous function spaces is positive if it maps nonnegative functions to nonnegative functions (when linear this is necessarily a monotone and bounded linear operator). The motivating example is

**Example 2.6.1** *Bernstein operators.*

For  $n$  in  $\mathbb{N}$ , let  $\mathcal{B}_n(f)$  be defined by

$$\mathcal{B}_n(f)(t) = \sum_{k=0}^n f\left(\frac{k}{n}\right) \binom{n}{k} t^k (1-t)^{n-k}. \quad (2.6.26)$$

It is clear that the Bernstein operators are linear and positive and indeed take values which are polynomials.  $\square$

**Theorem 2.6.2 (First Korovkin three function theorem).** *Let  $L_n$  be a sequence of positive linear operators from  $C[0, 1]$  to  $C[0, 1]$ . Suppose that*

$$L_n(1) \rightrightarrows_n 1, \quad L_n(\iota) \rightrightarrows_n \iota, \quad L_n(\iota^2) \rightrightarrows_n \iota^2.$$

Then

$$L_n(f) \rightrightarrows_n f$$

as  $n \rightarrow \infty$  for all  $f$  in  $C[0, 1]$ .

**Proof.** The hypotheses imply that  $L_n(q) \rightrightarrows_n q$  for all quadratic  $q$ . Fix  $f$  in  $C[0, 1]$ ,  $x$  in  $[0, 1]$ , and  $\varepsilon > 0$ . We claim that one can find a quadratic  $q_x^\varepsilon$  with  $f \leq q_x^\varepsilon$  and  $f(x) + \varepsilon \geq q_x^\varepsilon$ . Thus

$$L_n(f) \leq L_n(q_x^\varepsilon) \rightrightarrows_n q_x^\varepsilon \leq f(x) + \varepsilon.$$

A compactness argument completes the proof. The details are left for the reader as Exercise 33.  $\square$

**Corollary 2.6.3 (Stone-Weierstrass)** *The Bernstein polynomials are uniformly dense in  $C[0, 1]$ .*

**Proof.** We check by hand or in a computer algebra system that  $\mathcal{B}_n(1) = 1$ ,  $\mathcal{B}_n(\iota) = \iota$ , and slightly more elaborately  $\mathcal{B}_n(\iota^2) = \iota^2 + \frac{1}{n}(\iota - \iota^2) \rightrightarrows_n \iota^2$ .  $\square$

In the periodic case, the role of  $t$  and  $t^2$  is taken by  $\sin$  and  $\cos$ , as the second Korovkin theorem shows.

**Theorem 2.6.4 (Second Korovkin three function theorem).** *Let  $L_n$  be a sequence of positive linear operators from  $C(\mathbb{T})$  to  $C(\mathbb{T})$ . Suppose that*

$$L_n(1) \rightrightarrows_n 1, \quad L_n(\sin) \rightrightarrows_n \sin, \quad L_n(\cos) \rightrightarrows_n \cos.$$

Then

$$L_n(f) \rightrightarrows_n f$$

as  $n \rightarrow \infty$  for all  $f$  in  $C(\mathbb{T})$ .

The great virtue of the Korovkin approach is that it provides us with a well formed program. We illustrate with the second theorem. For any kernel ( $K_n$ ), we may induce a sequence of linear operators  $\mathcal{K}_n(f) = K_n * f$  and must answer two questions: (i) Is each  $\mathcal{K}_n$  positive? (ii) Does  $\mathcal{K}_n(f) \rightrightarrows_n f$  for the three functions  $f = 1, \sin, \cos$ ? The first is usually easy to answer; the second frequently is a direct computation.

**Example 2.6.5** *Dirichlet and Fejér Operators.*

We revisit the uniform convergence properties of the Dirichlet and Fejér kernels from Section 2.3.3.

1. The Dirichlet kernel induces the operator

$$\mathcal{D}_n(f) = D_n * f$$

where  $D_n = \sin((n+1/2)t) / \sin(t/2)$ . This is quite easily seen not to be positive. (A good thing since we know that  $\mathcal{D}_n(f)$  need not converge uniformly to  $f$  for  $f$  in  $C(\mathbb{T})$ .)

2. The Fejér kernel induces the operator

$$\mathcal{F}_n(f) = F_n * f$$

where  $F_n = \sin^2((n+1)/2)t / [(n+1)\sin^2(t/2)] \geq 0$ . Thus, to recover Fejér's theorem on the uniform convergence of the Cesàro averages it suffices to compute

$$\mathcal{F}_n(1) = 1, \quad \mathcal{F}_n(\sin) = \frac{n}{n+1} \sin, \quad \mathcal{F}_n(\cos) = \frac{n}{n+1} \cos. \quad \square$$

## 2.7 Commentary and Additional Examples

1. **An error function evaluation.** (AMM Problem 11000, Mar. 2003) [149].

(a) Work out the ordinary generating function of  $\binom{2n}{n}$  and so evaluate

$$\sum_{n+m=N} \binom{2n}{n} \binom{2m}{m}.$$

(b) Evaluate

$$\int_0^{\frac{\pi}{2}} \cos^{2n}(x) \sin^{2m}(x) dx.$$

(c) Recall the *error function*,  $\operatorname{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x \exp(-t^2) dt$ , and show for  $a > 0$  that

$$a \int_0^{\frac{\pi}{2}} \operatorname{erf}(\sqrt{a} \cos x) \operatorname{erf}(\sqrt{a} \sin x) \sin(2x) dx = e^{-a} + a - 1.$$

(d) The previous evaluation can be viewed as an inner product of the functions  $\operatorname{erf}(\sqrt{a} \sin x) \sin x$  and  $\operatorname{erf}(\sqrt{a} \cos x) \cos x$ . Determine that

$$\begin{aligned} & \int_0^{\frac{\pi}{2}} \operatorname{erf}^2(\sqrt{a} \cos x) \cos^2(x) dx \\ &= \sum_{N=0}^{\infty} \frac{\left(\frac{-a}{4}\right)^{N+1} (8N+12) \binom{2N}{N}}{(N+2)!} F\left(\frac{1}{2}, -N, -N - \frac{1}{2}; \frac{3}{2}, -N + \frac{1}{2}; -1\right) \\ &= \frac{1}{2\pi} - 2 \int_0^1 \frac{e^{-1/2 a(1+x^2)}}{1+x^2} \left\{ I_0\left(\frac{1}{2} a(1+x^2)\right) - I_1\left(\frac{1}{2} a(1+x^2)\right) \right\} dx. \end{aligned}$$

2. **Failure of Fubini.** Evaluate these integrals:

(a)

$$\int_0^1 \int_0^1 \frac{x^2 - y^2}{(x^2 + y^2)^2} dx dy = -\frac{\pi}{4}$$

and

$$\int_0^1 \int_0^1 \frac{x^2 - y^2}{(x^2 + y^2)^2} dy dx = \frac{\pi}{4}$$



(b)

$$\int_0^1 \int_1^\infty (e^{-xy} - 2e^{-2xy}) dy dx - \int_1^\infty \int_0^1 (e^{-xy} - 2e^{-2xy}) dx dy = \ln(2)$$

(c)

$$\int_0^\infty \int_0^\infty \frac{4xy - x^2 - y^2}{(x+y)^4} dy dx = \int_0^\infty \int_0^\infty \frac{4xy - x^2 - y^2}{(x+y)^4} dx dy = 0$$

but, for all  $m, c > 0$ 

$$\int_0^{mc} \int_0^c \frac{4xy - x^2 - y^2}{(x+y)^4} dx dy = \frac{m}{(1+m)^2} \neq 0,$$

as  $c \rightarrow \infty$ .

In each case, explain why they differ without violating any known theorem.

**3. Failure of l'Hôpital's rule.** Evaluate these limits:Let  $f(x) = x + \cos(x) \sin(x)$  and  $g(x) = e^{\sin(x)} (x + \cos(x) \sin(x))$ .Then  $\lim_{x \rightarrow \infty} \frac{f(x)}{g(x)}$  does not exist although  $\lim_{x \rightarrow \infty} \frac{f'(x)}{g'(x)} = 0$ . This is a caution against carelessly dividing by zero!**4. Various Fourier series evaluations.**

- Compute the Fourier series of  $t/2$ ,  $|t|$ ,  $t^2$  and  $(t^3 - \pi^2 t)/3$  on  $[-\pi, \pi]$ .
- Plot the 6th and 12th Fourier polynomials against the function in each case.
- Compute enough Fourier coefficients of  $\sin(x^3)$  on  $[-\pi, \pi]$  to be convinced of Parseval's equation.
- Compute the Fourier series of  $t^2$  and  $(t^3 - \pi^2 t)/3$  on  $[0, 2\pi]$ .
- Use Parseval's equation with  $(t^3 - \pi^2 t)/3$  to evaluate  $\zeta(6)$ . Then apply Parseval to  $t^4/4$ .
- Show that

$$\int_0^{\pi/2} \log(2 \sin(t/2)) dt = -G,$$

where  $G$  is Catalan's constant.

(g) Show that for  $a > 0$ ,

$$\begin{aligned} \cos(ax) &= \frac{\sin(\pi a)}{\pi a} - 2 \frac{\sin(\pi a) a \cos(x)}{(a^2 - 1)\pi} + 2 \frac{\sin(\pi a) a \cos(2x)}{(a^2 - 4)\pi} \\ &\quad - 2 \frac{\sin(\pi a) a \cos(3x)}{(a^2 - 9)\pi} + \cdots \end{aligned} \quad (2.7.27)$$

Similarly evaluate the Fourier series for  $\exp(ax)$ .

(h) Substitute  $x = \pi$  in (2.7.27) to obtain the partial fraction expansion for  $\cot$  (compare the first example in Section 2.3.1) and integrate to recover the product formula for  $\sin$  (justifying all steps).

(i) Evaluate  $\sum_{n \geq 0} 1/(4n^2 - 1)$ .

5. **Two applications of Parseval's equation.** Use Parseval's equation in  $L^2(\mathbb{R})$  to evaluate

$$\int_{-\infty}^{\infty} \frac{\sin^2(t)}{t^2} dt = \pi \quad \text{and} \quad \int_{-\infty}^{\infty} \frac{\sin^4(t)}{t^4} dt = \frac{2\pi}{3}.$$

See also Exercise 28.

6. **Lebesgue's example** of a continuous function with divergent Fourier series.

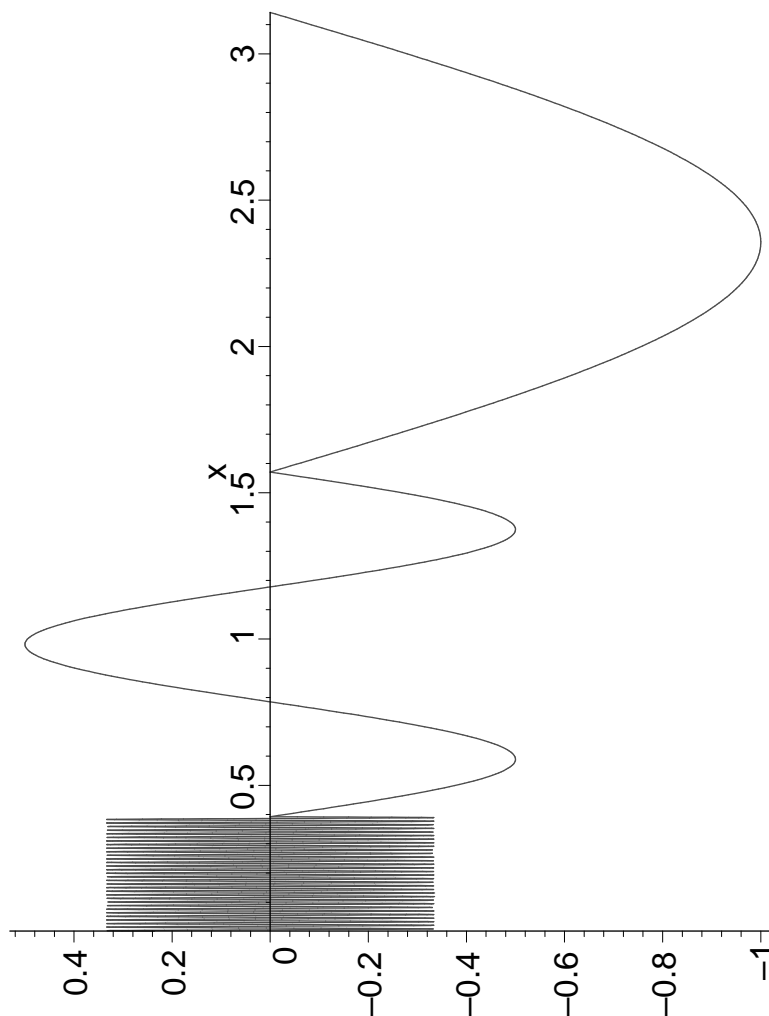
**Construction.** We follow Stromberg page 557 and let  $a_k = 2^{\sum_{j=1}^k j!}$  for  $k \geq 0$  and define

$$f_n(x) = \sum_{k=1}^n \frac{\sin(a_k |x|)}{k} \chi_k(|x|)$$

on  $[-\pi, \pi]$ , where  $\chi_k$  is the characteristic function of  $[\pi/a_k, \pi/a_{k-1}]$ , and extend  $f$  by  $2\pi$ -periodicity onto  $\mathbb{R}$ . Then  $f(x) = \lim_{n \rightarrow \infty} f_n(x)$  is continuous and the Fourier series is uniformly convergent on  $[\delta, 2\pi - \delta]$  for  $\delta > 0$ , but  $s_{a_k}(f, 0) \rightarrow \infty$  for  $k \rightarrow \infty$ .

**Convergence on  $[\delta, 2\pi - \delta]$ .** This is easy since  $f = f_n$  on this interval for large  $n$ , so that the "Riemann localization principle" for Fourier series can be used: *If  $f_1(t) = f_2(t)$  for every  $t$  in some nonvoid open interval  $I$ , then  $|s_n(f_1, t) - s_n(f_2, t)| \rightarrow 0$  for every  $t \in I$ .*

**Divergence at 0.** The divergence estimate comes as follows:

Figure 2.7: Approximation  $f_3$  to Lebesgue's function

- (a) We start with Dirichlet's kernel: it can be proved that the  $n$ -th partial sum behaves like

$$s_n(f, 0) = \frac{2}{\pi} \int_0^\pi f(t) \frac{\sin(nt)}{t} dt + \varepsilon_n$$

where  $\varepsilon_n \rightarrow 0$ .

(b) We estimate the first part of the integral as

$$\left| \int_0^{\pi/a_k} f(t) \frac{\sin(a_k t)}{t} dt \right| \leq \frac{\pi}{k+1}$$

since  $|\frac{\sin(x)}{x}| \leq 1$  and  $|f(t)| \leq \frac{1}{k+1}$ .

(c) We estimate the second part of the integral via

$$2 \int_{\pi/a_k}^{\pi/a_{k-1}} f(t) \frac{\sin(a_k t)}{t} dt + \frac{1}{k} \int_{\pi/a_k}^{\pi/a_{k-1}} \frac{\cos(2 a_k t)}{t} dt = k! \frac{\ln(2)}{k}, \quad (2.7.28)$$

and the second term on the left, say  $I_k$ , is no bigger than  $\frac{1}{2\pi k}$  on using the Bonnet second mean value theorem to write

$$|I_k| = \left| \frac{a_k}{k\pi} \int_{\pi/a_k}^{\psi} \cos(2 a_k t) dt \right| \leq \frac{1}{2k\pi},$$

for some  $\frac{a_{k-1}}{\pi} \leq \psi \leq \frac{a_k}{\pi}$ .

(d) We estimate the third part of the integral as

$$\begin{aligned} \left| \int_{\pi/a_{k-1}}^{\pi} f(t) \frac{\sin(a_k t)}{t} dt \right| &\leq \frac{a_{k-1}}{\pi} \left| \int_{\pi/a_{k-1}}^{\psi} f(t) \sin(a_k t) dt \right| \\ &\leq \frac{a_{k-1}}{\pi} \left( \left| f(t) \frac{\cos(a_k t)}{a_k} \right|_{\pi/a_{k-1}}^{\psi} + \left| \int_{\psi}^{\pi} f'(t) \frac{\cos(a_k t)}{a_k} dt \right| \right) \\ &\leq \frac{a_{k-1}}{a_k} \left( \frac{1}{\pi} + a_{k-1} \right) \rightarrow 0, \end{aligned}$$

on using the mean value theorem again, and then applying integration by parts with the estimates that  $|f(t)| \leq 1$ ,  $|f'(t)| \leq a_{k-1}$ . Thus, the dominant term is the first integral in (2.7.28) and  $s_{a_k}(f, 0) \rightarrow \infty$ .

7. **Non-uniqueness of Fourier series.** Can a trigonometric series converge a.e. on  $\mathbb{R}$  to a function  $\varphi \in L^1(T)$  and yet not be the Fourier series of  $\varphi$ ? This question was first answered in the affirmative with  $\varphi = 0$  in 1916 by the Russian analyst D. E. Menshov. His counterexample involves the Cantor set. For more details, see [197], from which the above text was cited.

8. **Nowhere differentiable continuous functions.** The first famous example of a continuous nowhere differentiable function was constructed by K. Weierstrass in 1872. (Tradition has it that Bolzano and Riemann constructed such examples before Weierstrass, but their examples did not become widely known.) Weierstrass's example was given in the form of a trigonometric series. We state it here as a series on  $[0, 1]$ , not on  $[-\pi, \pi]$  or  $[0, 2\pi]$  as before, because this will simplify matters later; we analogously write  $f \in L^1(0, 1)$ , and formulas and theorems on Fourier series are easily converted to this case. Weierstrass's example was

$$C_{a,b}(t) = \sum_{n=0}^{\infty} a^n \cos(b^n 2\pi t)$$

with  $|a| < 1$  and integral  $b > 1$ . Weierstrass proved that  $C_{a,b}$  is nowhere differentiable when  $b \in 2\mathbb{N} + 1$  and  $ab > 1 + \frac{3}{2}\pi$ . This once and for all settled the question if such functions could exist (at the beginning of the 19th century, Ampère “proved” that every continuous function must be differentiable at some point). Some questions were left open, however: It is clear that  $C_{a,b}$  is differentiable when  $|a|b < 1$ , since the series is then termwise differentiable. But what happens for  $|a|b$  between 1 and  $1 + \frac{3}{2}\pi$ ? This question gave several mathematicians a headache, until in 1916 G. H. Hardy proved that both  $C_{a,b}$  and the corresponding sine series

$$S_{a,b}(t) = \sum_{n=0}^{\infty} a^n \sin(b^n 2\pi t)$$

are nowhere differentiable whenever  $b$  is a real greater than 1, and  $ab > 1$ . In his paper, Hardy first treated the case when  $b \in \mathbb{N}$ , i.e., when the functions are given by their Fourier series, and only afterwards treated the general case of arbitrary real  $b$ . Hardy's methods were not easy, not even in the Fourier case (where he used the Poisson kernel, among other things). In the ensuing years, several other, simpler proofs have been published. In the middle of the 20th century, G. Freud and J.-P. Kahane gave conditions for the differentiability of lacunary Fourier series (where non-zero Fourier coefficients are spaced far apart), from which the non-differentiability of Weierstrass's function follows. Another approach to the Weierstrass functions uses functional equations.

Prove:

- (a) The Weierstrass sine series  $f = S_{a,2}$  with  $b = 2$  satisfies the system of two functional equations

$$f\left(\frac{t}{2}\right) = af(t) + \sin(t), \quad f\left(\frac{t+1}{2}\right) = af(t) - \sin(t) \quad (2.7.29)$$

for every  $t \in [0, 1]$ . The cosine series  $S_{a,2}$  satisfies an analogous system.

- (b) The Weierstrass function is the only bounded solution of the respective system on  $[0, 1]$ .

Hint for (b): Use Banach's fixed point theorem. (Note that it not only proves uniqueness, but can also be used to give an explicit approximation procedure. It is instructive to investigate how this procedure approaches the Weierstrass function for different initial guesses.)

9. **Replicative functions.** Let  $D$  be an interval containing  $(0, 1)$ . A function  $f : D \rightarrow \mathbb{C}$  is called *replicative* (on  $D$ ) if it satisfies the functional equation

$$\frac{1}{p} \sum_{k=0}^{p-1} f\left(\frac{t+k}{p}\right) = u(p)f(t) \quad \text{for all } t \in D \text{ and } p \in \mathbb{N}, \quad (2.7.30)$$

with a  $u : \mathbb{N} \rightarrow \mathbb{C}$  (which turns out to be unique if  $f \not\equiv 0$ ). This notion was introduced (with more generality) by D. E. Knuth in [147]. Examples are the cotangent ( $\cot(\pi t)$  is replicative on  $(0, 1)$  with  $u(p) = 1$ ), the Bernoulli polynomials ( $B_m(t)$  is replicative on  $\mathbb{R}$  with  $u(p) = 1/p^m$ ) and derivatives of the Psi function (the  $m$ -th derivative of  $\Psi = \Gamma'/\Gamma$  is replicative on  $\mathbb{R}_+$  with  $u(p) = p^m$ ). Functions which are replicative and 1-periodic have multiplicative Fourier coefficients.

**Theorem 2.7.1** (a) Let  $f : D \rightarrow \mathbb{C}$ ,  $f \not\equiv 0$ , be replicative on  $D$  with  $u(p)$ . Then  $u$  is necessarily multiplicative, i.e.,  $u(mn) = u(m) \cdot u(n)$  for all  $m, n \in \mathbb{N}$ .

(b) Let  $f \in L^1(0, 1)$  be replicative on  $(0, 1)$  with a sequence  $u(p)$ . Then  $\widehat{f}_{mn} = u(n)\widehat{f}_m$ .

This implies: If  $u \equiv 1$ , then  $f(t) = \widehat{f}_0$ . If  $u \neq 1$ , then

$$f(t) \sim \widehat{f}_{-1} \sum_{n=-\infty}^{-1} u(-n) e^{2\pi int} + \widehat{f}_1 \sum_{n=1}^{\infty} u(n) e^{2\pi int}.$$

(c) Let  $u$  be multiplicative and assume that  $f(t) = {}^L \sum_{n=1}^{\infty} u(n) e^{2\pi int}$  is pointwise convergent on  $[0, 1]$ , where  $L$  is a linear summation method. Then  $f$  is replicative on  $[0, 1]$ .

Part (c) of this theorem makes it easy to construct many different examples of replicative functions on  $[0, 1]$ . Verify the following Fourier series:

- (a)  $\sum n^{-2} \sin(2\pi nt) = -2\pi \int_0^t \ln(2 \sin(\pi x)) dx$  on  $[0, 1]$ ,  
 $\sum n^{-2} \cos(2\pi nt) = B_2(t)$  on  $[0, 1]$ .
- (b)  $\sum n^{-1} \sin(2\pi nt) = B_1(t)$  on  $(0, 1)$  and  $= 0$  on  $0, 1$ ,  
 $\sum n^{-1} \cos(2\pi nt) = -\ln(2 \sin(\pi t))$  on  $(0, 1)$  and  $= \infty$  on  $0, 1$ .
- (c)  $C_1 \sum \sin(2\pi nt) = \frac{1}{2} \cot(\pi t)$  on  $(0, 1)$  and  $= 0$  on  $0, 1$ ,  
 $C_1 \sum \cos(2\pi nt) = -\frac{1}{2}$  on  $(0, 1)$  and  $= \infty$  on  $0, 1$ ,  
 where  $C_1$  stands for Césaro summation.
- (d)  $A \sum n \sin(2\pi nt) = 0$  on  $[0, 1]$ ,  
 $A \sum n \cos(2\pi nt) = -1/(4 \sin^2(\pi t))$  on  $(0, 1)$  and  $= \infty$  on  $0, 1$ ,  
 where  $A$  stands for Abel summation.
- (e)  $\sum_{n=0}^{\infty} a^n \sin(p^n 2\pi t) = S_{a,p}(t)$  on  $[0, 1]$ ,  
 $\sum_{n=0}^{\infty} a^n \cos(p^n 2\pi t) = C_{a,p}(t)$  on  $[0, 1]$ ,  
 for  $p$  a prime, i.e., the nowhere differentiable Weierstrass functions can be replicative.

#### 10. Conditions for $\widehat{f} \in L^1(\mathbb{R})$ .

- (a) It is often useful to decide whether  $\widehat{f} \in L^1(\mathbb{R})$  for a given  $f \in L^1(\mathbb{R})$ , without having to explicitly compute  $\widehat{f}$ . The usual conditions assume differentiability properties of  $f$ , since smoothness of  $f$  translates into shrinkage of  $\widehat{f}$ . Thus,  $f \in C^2(\mathbb{R})$  is sufficient for  $\widehat{f} \in L^1(\mathbb{R})$ . However,

this condition does not cover the Fejér kernel, for example. A stronger condition, which is good for functions with bounded support, is given in the next theorem.

**Theorem 2.7.2** *Let  $f$  be an absolutely continuous function on the real line with compact support and let  $f'$  be of bounded total variation on  $\mathbb{R}$ , i.e.,  $V(f') < \infty$ . Then  $\widehat{f} \in L^1(\mathbb{R})$  and*

$$\|\widehat{f}\|_1 \leq 4\sqrt{V(f')\|f\|_1}. \quad (2.7.31)$$

This theorem presents another experimental challenge: Is the constant “4” appearing there best possible? The answer is not known. Non-systematic experimentation has found no value for the constant greater than  $\pi$ , which is achieved for the Fejér kernel. It is also not known if the “compact support” condition in the theorem is really needed.

Perform a systematic experiment on Theorem 2.7.2, in analogy to the experimentation for the uncertainty principle described in Section 5.2 of the first volume.

- (b) A quite different condition is due to Chandrasekharan: *If  $f \in L^1(\mathbb{R})$ , continuous at 0, and satisfies  $\widehat{f} \geq 0$  on  $\mathbb{R}$ , then  $\widehat{f} \in L^1(\mathbb{R})$ .* The disadvantage of this condition is that it uses  $\widehat{f}$  explicitly. It is applicable, however, to both the Fejér and the Poisson kernel.

11. **More kernels.** For each of the following kernels, decide whether (resp. for which parameters) it is an approximate identity in  $L^1(\mathbb{R})$ . Note that sometimes a version of Theorem 2.3.7 with weakened assumptions (allowing more variety in the kernels) is needed.

- (a) The *de la Vallée-Poussin kernel*  $V_m^n$ , depending on two integer parameters  $m, n$  with  $n > m$ , is defined by  $V_m^n(t) = \sum_{k=-(n+m)}^{n+m} a_{m,k}^n e^{ikt}$  where

$$a_{m,k}^n = \begin{cases} 1, & \text{if } |k| \leq n - m, \\ \frac{n+m+1-|k|}{2m+1}, & \text{if } n - m \leq |k| \leq n + m, \\ 0, & \text{otherwise.} \end{cases}$$

Hint: Let  $m, n$  tend to infinity such that  $n/m \rightarrow \lambda$ .



- (b) For  $\alpha > 0$ , the  $(C, \alpha)$ -kernel  $F_n^{(\alpha)}$  is defined as  $F_n^{(\alpha)}(t) = \sum_{k=-n}^n a_{n,k}^{(\alpha)} e^{ikt}$  where

$$a_{n,k}^{(\alpha)} = \begin{cases} \frac{\Gamma(n-|k|+\alpha+1)\Gamma(n+1)}{\Gamma(n-|k|+1)\Gamma(n+\alpha+1)}, & \text{if } |k| \leq n+1, \\ 0, & \text{otherwise.} \end{cases}$$

- (c) For parameters  $\alpha_0, \dots, \alpha_p \in \mathbb{R}$  with  $\alpha_0 + \dots + \alpha_p = 1$ , the *Blackman kernel*  $H_n^{(\alpha_0, \dots, \alpha_p)}$  is defined by  $H_n^{(\alpha_0, \dots, \alpha_p)}(t) = \sum_{k=-n}^n h_{n,k}^{(\alpha_0, \dots, \alpha_p)} e^{ikt}$  where

$$h_{n,k}^{(\alpha_0, \dots, \alpha_p)} = \sum_{j=0}^p \alpha_j \cos(jkt_n),$$

with  $t_n = \frac{2\pi}{2n+1}$ .

- (d) The **Fejér-Korovkin kernel**  $FK_n$  is defined as

$$FK_n(t) = \begin{cases} \frac{2 \sin^2(\pi/(n+2))}{n+2} \left[ \frac{\cos((n+2)t/2)}{\cos(\pi/(n+2)) - \cos t} \right]^2, \\ (n+2)/2, \end{cases}$$

depending on whether  $t \neq \pm\pi/(n+2) + 2j\pi$  or  $t = \pm\pi/(n+2) + 2j\pi$ , respectively. It can be written in the form  $FK_n(t) = \sum_{k=-n}^n a_{n,k} e^{ikt}$  where

$$a_{n,k} = \frac{(n-|k|+3) \sin \frac{|k|+1}{n+2} \pi - (n-|k|+1) \sin \frac{|k|-1}{n+2} \pi}{2(n+2) \sin(\pi/(n+2))}.$$

- (e) Finally, the *Jackson kernel*  $J_n$  is a rescaled version of the square of the Fejér kernel, namely

$$J_n(t) = \frac{3}{n(2n^2+1)} \left[ \frac{\sin(nt/2)}{\sin(t/2)} \right]^4.$$

12. **The Haar basis.** As we mentioned in the text, the trigonometric functions  $e^{int}$  constitute an orthogonal basis for the space  $L^2(\mathbb{T})$ , so that  $L^2$ -statements follow from general Hilbert space theory. Bases other than the trigonometric are of course conceivable and are in fact used in practice for the analysis of  $L^2$ -functions. Since about 15 years ago, certain

bases of  $L^2(\mathbb{R})$ , called *wavelet bases*, have found widespread use in signal analysis. Such bases are constructed as follows. Take a  $\psi \in L^2(\mathbb{R})$  and define  $\psi_{j,n}(t) = 2^{n/2} \psi(2^n t - j)$ . Then  $\psi$  is called an *orthogonal wavelet* if  $\{\psi_{j,n} : j, n \in \mathbb{Z}\}$  is an orthonormal basis of  $L^2(\mathbb{R})$ .

Show:  $\psi = \chi_{[0,1/2)} - \chi_{[1/2,1)}$  is an orthogonal wavelet. The associated basis  $\{\psi_{j,n}\}$  is called the *Haar basis* of  $L^2(\mathbb{R})$ .

13. **The Schauder basis.** The foundation of the theory of bases in Banach spaces was laid by J. Schauder in the 1930's. A sequence  $(x_n)$  in a Banach space  $B$  is called a *basis of  $B$*  if for every  $x \in B$  there exists a unique sequence of scalars  $(\alpha_n)$  with

$$x = \sum_{n=1}^{\infty} \alpha_n x_n \quad \text{in } B.$$

The trigonometric functions are not a basis for  $L^1(\mathbb{T})$  or for  $C(\mathbb{T})$ , although their span is dense in these spaces. The standard example of a basis for the space  $C[0, 1]$  is also due to Schauder (although G. Faber had used the same basis before Schauder, in a different analytical guise). This Faber-Schauder basis is the system of continuous functions  $\{\sigma_{0,0}, \sigma_{1,0}\} \cup \{\sigma_{i,n} : n \in \mathbb{N}, i = 0, \dots, 2^{n-1} - 1\}$ , where  $\sigma_{0,0}(t) = 1 - t$ ,  $\sigma_{1,0}(t) = t$ , and the function  $\sigma_{i,n}$  is the linear interpolation of the points

$$(0, 0), \quad \left(\frac{i}{2^{n-1}}, 0\right), \quad \left(\frac{2i+1}{2^n}, 1\right), \quad \left(\frac{i+1}{2^{n-1}}, 0\right), \quad (1, 0).$$

This system is a basis of the space  $C[0, 1]$ , more precisely: *Every continuous function  $f : [0, 1] \rightarrow \mathbb{R}$  has a unique, uniformly convergent expansion of the form*

$$f(x) = \gamma_{0,0}(f) \sigma_{0,0}(x) + \gamma_{1,0}(f) \sigma_{1,0}(x) + \sum_{n=1}^{\infty} \sum_{i=0}^{2^{n-1}-1} \gamma_{i,n}(f) \sigma_{i,n}(x),$$

where the coefficients  $\gamma_{i,n}(f)$  are given by  $\gamma_{0,0}(f) = f(0)$ ,  $\gamma_{1,0}(f) = f(1)$ , and

$$\gamma_{i,n}(f) = f\left(\frac{2i+1}{2^n}\right) - \frac{1}{2} f\left(\frac{i}{2^{n-1}}\right) - \frac{1}{2} f\left(\frac{i+1}{2^{n-1}}\right).$$

Knowing the Schauder basis expansion of a continuous function  $f$  can be useful in the analysis of  $f$ . For example, Faber proved in 1910 a criterion for differentiability of  $f$  in terms of its Schauder coefficients: *If  $f'(x_0) \in \mathbb{R}$  exists for some  $x_0 \in [0, 1]$ , then*

$$\lim_{n \rightarrow \infty} 2^n \cdot \min\{|\gamma_{i,n}(f)| : i = 0, \dots, 2^{n-1} - 1\} = 0. \quad (2.7.32)$$

Interestingly, this condition can be used to prove non-differentiability of the Weierstrass functions in an elementary way. Prove:

- (a) The Schauder coefficients of the Weierstrass sine series  $f = S_{a,2}$  satisfy the recursion

$$\gamma_{0,1}(f) = 0,$$

$$\gamma_{i,n+1}(f) = a\gamma_{i,n}(f) + \gamma_{i,n}(\sin) \text{ for } n \in \mathbb{N}, i = 0, \dots, 2^{n-1} - 1,$$

$$\gamma_{i,n+1}(f) = a\gamma_{i-2^{n-1},n}(f) - \gamma_{i-2^{n-1},n}(\sin) \text{ for } n \in \mathbb{N}, i = 2^{n-1}, \dots, 2^n - 1.$$

Hint: Use the functional equations (2.7.29).

- (b) Faber's condition (2.7.32) is not satisfied for  $f = S_{a,2}$ . Thus this function is nowhere differentiable.

It is instructive to experiment with the recursion in (a): to plot the Schauder coefficients and Schauder approximations for the Weierstrass and for other functions which satisfy similar functional equations. Details of this method can be found in [116] and [117].

14. **Riemann-Lebesgue lemma.** Deduce the following from the Riemann-Lebesgue lemma for every Lebesgue integrable function  $f$ .

- (a) For any real  $\sigma(t)$

$$\lim_{t \rightarrow \infty} \int_{\mathbb{R}} f(x) \cos^2(tx + \sigma(t)) dx = \frac{1}{2} \int_{\mathbb{R}} f(x) dx.$$

- (b) The coefficients  $\hat{f}(n) \rightarrow 0$  as  $n \rightarrow \infty$ .

Conclude that the trigonometric series  $\sum_{n>1} \sin(nt) / \log(n)$  is not the Fourier series of any integrable function.

When  $a_n$  is convex decreasing with limit zero and with  $\sum_{n>0} a_n/n = \infty$ , it is in fact the case that  $\sum_{n>0} a_n \cos(nt)$  is a Fourier series of an integrable function, but  $\sum_{n>0} a_n \sin(nt)$  is not [197, Chapter 8].

15. **A few Fourier transforms.** We have already seen many examples of Fourier transforms and their Laplace transform variants. The specialization to the Mellin transform is explored in the next chapter.

- (a) Show that the Fourier transform of  $\cos^2(ax)$  is the function  $\cos[y^2/(4a) - \pi/4]/\sqrt{2a}$ . Find the transform of  $\sin^2(ax)$ .
- (b) Show that for  $a > 0$ , the Fourier transform of  $|x| \exp(-a|x|)$  is  $a(2/\pi)^{1/2} (a^2 - y^2) / (a^2 + y^2)^2$ .
- (c) Find the transform of  $1/x^n$  and of  $1/(a^2 + x^2)$ .
- (d) Find all square-integrable solutions to  $\hat{f}/\sqrt{2\pi} = f$  (the fixed points of the normalized Fourier transform). Then experiment with the orbit of  $f \mapsto \hat{f}/\sqrt{2\pi}$  for various choices  $f_0$ .

16. **The isoperimetric inequality.** The ancient Greek geometers knew already that a circle with given perimeter encloses a larger area than any polygon with the same perimeter. In 1841 Steiner extended this result to simple closed plane curves. Here we will sketch a Fourier series proof (due to Hurwitz), for simplicity restricted to (piecewise)  $C^1$ -curves. Thus, assume that we have a simple, closed  $C^1$ -curve  $(x(t), y(t))$  in  $\mathbb{R}^2$  of length  $\int_{-\pi}^{\pi} (x'(s)^2 + y'(s)^2)^{1/2} ds = 2\pi$ . Without loss of generality we can assume that  $x'(s)^2 + y'(s)^2 = 1$  for all  $s$ . We wish to minimize the area inside the curve, given by

$$A = \int_{-\pi}^{\pi} x(s) y'(s) ds.$$

Show:  $A \geq \pi$  with equality if and only if  $(x(t) - \hat{x}_0)^2 + (y(t) - \hat{y}_0)^2 = 1$ . This is the *isoperimetric inequality*. Hint: Substitute Fourier series, transfer the formulas for derivatives from Section 2.4.2 to the  $L^1(\mathbb{T})$ -case and use Parseval's equation.

17. **The maximum principle.** In like fashion, employ Poisson's kernel to heuristically deduce that the maximum principle discussed briefly in Section 6.5 applies.

18. **The heat equation.** The one-dimensional heat equation

$$\frac{\partial \phi}{\partial t}(x, t) = \frac{\pi}{4i} \frac{\partial^2 \phi}{\partial x^2}(x, t),$$

is solved by the general theta function  $\sum_{n \in \mathbb{Z}} x^n \exp(-\pi i t n^2)$ . More usefully, when  $G : \mathbb{R} \mapsto \mathbb{C}$  is continuous and bounded we may solve the equation

$$\frac{\partial \phi}{\partial t}(x, t) = K \frac{\partial^2 \phi}{\partial x^2}(x, t),$$

with boundary condition  $\phi(x, t) \rightarrow G(x)$  as  $t \rightarrow 0^+$ , by the infinitely differentiable function

$$G * E_{1/\sqrt{2Kt}}(x) = \frac{1}{2\sqrt{\pi Kt}} \int_{\mathbb{R}} G(x-y) \exp(-y^2/2Kt) dy,$$

for  $x$  in  $\mathbb{R}$  and  $t > 0$ .

19. **The easiest three-dimensional Watson integral.** We start with the easiest integral to evaluate. Let

$$W_3(w) = \int_0^\pi \int_0^\pi \int_0^\pi \frac{1}{1 - w \cos(x) \cos(y) \cos(z)} dx dy dz,$$

for suitable  $w > 0$ .

(a) Prove that

$$\begin{aligned} W_3(1) &= \int_0^\pi \int_0^\pi \int_0^\pi \frac{1}{1 - \cos(x) \cos(y) \cos(z)} dx dy dz \\ &= \frac{1}{4} \Gamma^4\left(\frac{1}{4}\right) = 4\pi K\left(\frac{1}{\sqrt{2}}\right) \end{aligned}$$

via the binomial expansion and [44, Exercise 14, page 188].

(b) More generally

$$W_3((2kk')^2) = \pi^3 F\left(1/2, 1/2, 1/2; 1, 1; 4k^2(1-k^2)\right) = 4\pi K^2(k).$$

20. **The harder three-dimensional Watson integrals.** We now describe results largely in Joyce and Zucker [141, 142], where more background can also be found. The following integral arises in Gaussian and spherical models of ferromagnetism and in the theory of random walks.

- (a) One of the most impressive closed-form evaluations of a multiple integral is Watson's

$$\begin{aligned} W_1 &= \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} \frac{1}{3 - \cos(x) - \cos(y) - \cos(z)} dx dy dz \\ &= \frac{1}{96} (\sqrt{3} - 1) \Gamma^2\left(\frac{1}{24}\right) \Gamma^2\left(\frac{11}{24}\right) \\ &= 4\pi \left(18 + 12\sqrt{2} - 10\sqrt{3} - 7\sqrt{6}\right) K^2(k_6) \end{aligned} \quad (2.7.33)$$

where  $k_6 = (2 - \sqrt{3})(\sqrt{3} - \sqrt{2})$  is the sixth singular value of Section 4.2.

Note that  $W_1 = \pi^3 \int_0^\infty \exp(-3t) I_0^3(t) dt$  allows for efficient computation [141] where the Bessel function  $I_0(t)$  has been written as  $(1/\pi) \int_0^\pi \exp(t \cos(\theta)) d\theta$ . The evaluation (2.7.33) in its original form is due to Watson and is really a tour de force. In the next Exercise we describe a refined and simplified evaluation due to Joyce and Zucker [142].

- (b) Similarly, the integral

$$\begin{aligned} W_2 &= \int_0^\pi \int_0^\pi \int_0^\pi \frac{dx dy dz}{3 - \cos(x) \cos(y) - \cos(y) \cos(z) - \cos(z) \cos(x)} \\ &= \sqrt{3} \pi K^2\left(\sin\left(\frac{\pi}{12}\right)\right) = \frac{2^{1/3}}{8\pi} \beta^2\left(\frac{1}{3}, \frac{1}{3}\right) \end{aligned} \quad (2.7.34)$$

where  $\sin(\pi/12) = k_3$  is the third singular value, again as in Section 4.2. Indeed, as we shall see in Exercise 21, (2.7.34) is easier and can be derived on the way to (2.7.33).

- (c) The evaluation (2.7.34) then implies that

$$\frac{1}{\pi} W_2 = \int_0^\pi \int_0^\pi \frac{dy dz}{\sqrt{9 - 8 \cos(y) \cos(z) - \cos^2(y) - \cos^2(z) + \cos^2(y) \cos^2(z)}}$$

on performing the innermost integration carefully.

- (d) The expression inside the square-root factors as  $(\cos x \cos y + \cos x + \cos y - 3)(\cos x \cos y - \cos x - \cos y - 3)$ . Upon substituting  $s = x/2$ , and  $t = y/2$ , one obtains

$$\begin{aligned} & \int_0^{\frac{\pi}{2}} \int_0^{\frac{\pi}{2}} \frac{dy dx}{\sqrt{(1 - \sin^2(x) \sin^2(y)) (1 - \cos^2(x) \cos^2(y))}} \\ &= \frac{1}{4\pi} \sum_{m=0}^{\infty} \sum_{n=0}^{\infty} \beta^3 \left( n + \frac{1}{2}, m + \frac{1}{2} \right) \binom{m+n}{n} = \sqrt{3} K^2 \left( \sin \left( \frac{\pi}{12} \right) \right). \end{aligned}$$

## 21. More about the Watson integrals.

- (a) For  $a > 1, b > 1$  show that

$$\begin{aligned} & \frac{1}{2} \int_0^{\pi} \frac{1}{\sqrt{(a + \cos(y)) (b - \cos(y))}} dy = \frac{K \left( \sqrt{\frac{2(b+a)}{(1+b)(1+a)}} \right)}{\sqrt{(1+b)(1+a)}} \\ &= \int_0^{1/2\pi} \frac{1}{\sqrt{(1+b)(1+a) \cos^2(t) + (1-a)(1-b) \sin^2(t)}} dt \end{aligned}$$

- (b) A beautiful but harder to establish identity is that

$$\int_0^{\frac{\pi}{2}} K \left( \sqrt{c^2 \cos^2(s) + \sin^2(s)} \right) ds = K \left( \sqrt{\frac{1-c}{2}} \right) K \left( \sqrt{\frac{1+c}{2}} \right) \quad (2.7.35)$$

or equivalently that

$$\int_0^{\frac{\pi}{2}} K \left( \sqrt{1 - (2kk')^2 \cos^2(\theta)} \right) d\theta = K(k) K(k')$$

with  $k' = \sqrt{1 - k^2}$ . Hence

$$\int_0^{\frac{\pi}{2}} K \left( \sqrt{1 - (2k_N k'_N)^2 \cos^2(\theta)} \right) d\theta = \sqrt{N} K^2(k_N)$$

where  $k_N$  is the  $N$ -th singular value. This is especially pretty for  $N = 1, 3, 7$  so that  $2k_N k'_N = 1, 1/2, 1/8$  respectively.

- (c) Deduce that face centered cubic (FCC) lattice for the Green's function evaluates as

$$\frac{1}{\pi} W_2 = \int_0^{\frac{\pi}{2}} K \left( \sqrt{\frac{3}{4} \cos^2(s) + \sin^2(s)} \right) ds = \sqrt{3} K^2(k_3).$$

- (d) Correspondingly, Watson's evaluation for the simple cubic (SC) lattice relied on deriving

$$W_1 = \sqrt{2} \pi \int_0^{\pi} K \left( \frac{\cos(x) - 5}{2} \right) dx,$$

and the following extension of (2.7.35):

$$\begin{aligned} \int_0^{\frac{\pi}{2}} K \left( \sqrt{c^2 \cos^2(s) + d^2 \sin^2(s)} \right) ds = \\ K \left( \sqrt{\frac{1 - cd - \sqrt{(d^2 - 1)(c^2 - 1)}}{2}} \right) \\ \times K \left( \sqrt{\frac{1 + cd - \sqrt{(d^2 - 1)(c^2 - 1)}}{2}} \right). \end{aligned}$$

- (e) **The Generalized Watson integrals.** Let

$$\begin{aligned} W_1(w_1) &= \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} \frac{1}{3 - w_1 (\cos(x) - \cos(y) - \cos(z))} dx dy dz \\ W_2(w_2) &= \int_0^{\pi} \int_0^{\pi} \int_0^{\pi} \frac{dx dy dz}{3 - w_2 (\cos(x) \cos(y) - \cos(y) \cos(z) - \cos(z) \cos(x))}. \end{aligned}$$

In a beautiful study, Joyce and Zucker [142], using the sort of elliptic and hypergeometric transformations we have explored, now are able to show fairly directly that

$$W_2(-w_1(3 + zw_1)/(1 - w_1)) = (1 - w_1)^{1/2} W_1(w_1).$$

Verify that, with  $w_1 = -1$ , this leads to a quite direct evaluation of (2.7.33) from (2.7.34).



(f) It is also true that

$$W_1 = \sqrt{2} \pi \int_0^{\frac{\pi}{2}} K \left( \frac{1}{2} + \frac{1}{2} \sin^2(t) \right) dt.$$

(g) **A more symmetric form.** Show that

$$\begin{aligned} & \int_0^{\frac{\pi}{2}} K \left( \sqrt{1 - 4k^2(1 - k^2) \cos^2(x)} \right) dx \\ &= \int_0^{\frac{\pi}{2}} \int_0^{\frac{\pi}{2}} \frac{dt dx}{\sqrt{\cos^2(t) + 4k^2(1 - k^2) \cos^2(x) \sin^2(t)}} \end{aligned}$$

for  $0 < k < 1$ . Hint: For (21d) consider  $N = 3$  ( $c^2 = 3/4$ ) in (21c), and let  $a$  and  $b$  be defined as  $a = (3 - \cos x) / (1 + \cos x)$  and  $b = (3 + \cos x) / (1 - \cos x)$  in (21a).

22. **Watson integral and Burg entropy.** Consider the perturbed *Burg entropy* maximization problem

$$\begin{aligned} v(\alpha) &= \sup_{p \geq 0} \{ \log(p(x_1, x_2, x_3)) \mid \int_0^1 \int_0^1 \int_0^1 p(x_1, x_2, x_3) dx_1 dx_2 dx_3 = 1, \\ & \text{and for } k = 1, 2, 3, \int_0^1 \int_0^1 \int_0^1 p(x_1, x_2, x_3) \cos(2\pi x_k) dx_1 dx_2 dx_3 = \alpha \}, \end{aligned}$$

maximizing the log of a density  $p$  with given mean, and with the first three cosine moments fixed at a parameter value  $0 \leq \alpha < 1$ . It transpires that there is a parameter value  $\bar{\alpha}$  such that below and at that value  $v(\alpha)$  is attained, while above it is finite but unattained. This is interesting severally:

- (a) The general method—maximizing  $\int_T \log(p(t)) dt$  subject to a finite number of trigonometric moments—is frequently used. In one or two dimensions such spectral problems are always attained when feasible.
- (b) There is no easy way to see this problem qualitatively changes at  $\bar{\alpha}$ , but we can get an idea by considering

$$\bar{p}(x_1, x_2, x_3) = \frac{1/W_1}{3 - \sum_1^3 \cos(2\pi x_i)},$$

and checking that this is feasible for

$$\bar{\alpha} = 1 - 1/(3W_1) \approx 0.340537329550999142833$$

in terms of the first Watson integral,  $W_1$ .

- (c) By using Fenchel duality [61] one can show that this  $\bar{p}$  is optimal.  
 (d) Indeed, for all  $\alpha \geq 0$  the only possible optimal solution is of the form

$$\bar{p}_\alpha(x_1, x_2, x_3) = \frac{1}{\lambda_\alpha^0 - \sum_1^3 \lambda_\alpha^i \cos(2\pi x_i)},$$

for some real numbers  $\lambda_\alpha^i$ . Note that we have four coefficients to determine; using the four constraints we can solve for them. For  $0 \leq \alpha \leq \bar{\alpha}$ , the precise form is parameterized by the generalized Watson integral:

$$\bar{p}_\alpha(x_1, x_2, x_3) = \frac{1/W_1(w)}{3 - \sum_1^3 w \cos(2\pi x_i)},$$

and  $\alpha = 1 - 1/(3W_1(w))$ , as  $w$  ranges from zero to one. Note also that  $W_1(w) = \pi^3 \int_0^\infty I_0^3(wt) e^{-3t} dt$ , allows one to quickly obtain  $w$  from  $\alpha$  numerically. For  $\alpha > \bar{\alpha}$ , no feasible reciprocal polynomial can stay positive. Full details are given in [60].

23. **A “momentary” recursion.** Choose  $p \in \mathbb{N}$  and define polynomials  $q_k(x)$  recursively by  $q_0(x) = -1$  and  $q_{n+1}(x) = q'_n(x) - x^{p-1} q_n(x)$ . Give an explicit formula for  $q_n(0)$  (and for  $q_n(x)$ ?).
24. **The limit of certain Fourier transforms.** For  $p \in 2\mathbb{N}$  let  $f_p(t) = e^{-t^p/p}$ . In Figure 2.8, the functions  $\widehat{f}_p(x)$  are shown for  $p = 2, 8, 16$ . The figure suggests that there may be a limit function as  $p \rightarrow \infty$ . Identify this limit function!
25. **The Schilling equation.** The Schilling equation is the functional equation

$$4q f(qt) = f(t+1) + 2f(t) + f(t-1) \quad \text{for } t \in \mathbb{R}$$

with a parameter  $q \in (0, 1)$ . It has its origin in Physics, and although it has been studied intensively in recent years, there are still many open questions

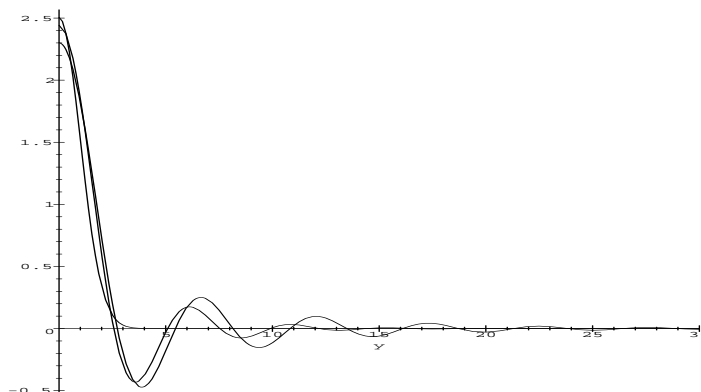


Figure 2.8: The oscillatory Fourier transforms  $\widehat{f}_2, \widehat{f}_8, \widehat{f}_{16}$

connected with it. The main question is to find values of  $q$  for which the Schilling equation has a non-trivial  $L^1$ -solution. Discuss this question! Hint: If an  $L^1$ -function  $f$  satisfies (2.4.20), then a rescaled version of  $f * f$  satisfies the Schilling equation.

26. **Another way to evaluate the sinc integral.** The evaluation of the integral  $\int_0^\infty \sin y/y dy = \pi/2$  also follows on taking the limit, via Binet's mean value theorem [197, p. 328], of the absolutely convergent integral

$$\int_0^\infty \frac{\sin y}{y^{1+\varepsilon}} dy = \frac{\pi \sec(\frac{\pi}{2}\varepsilon)}{2 \Gamma(1 + \varepsilon)}.$$

*Maple* happily provides the second integral in a form which simplifies to that we have given. A conventional Mellin transform based proof follows.

- (a) For  $0 < \varepsilon < 1$ , use the  $\Gamma$ -function to write

$$\int_0^\infty \frac{\sin y}{y^{1+\varepsilon}} dy = \frac{1}{\Gamma(\varepsilon + 1)} \int_0^\infty dx \int_0^\infty \sin(x) \exp(-xt)t^\varepsilon dt.$$

- (b) Interchange variables and evaluate the inner integral to  $t^\varepsilon/(t^2 + 1)$ .  
 (c) Then use the  $\beta$ -function to prove

$$\int_0^\infty \frac{t^\varepsilon}{t^2 + 1} dt = \beta\left(\frac{1}{2} - \frac{1}{\varepsilon}, \frac{1}{2} - \frac{1}{\varepsilon}\right) = \frac{\pi}{2} \sec\left(\frac{\pi}{2}\varepsilon\right).$$

Note:

$$\int_0^\infty \frac{\log^{2n}(s)}{s^2 + 1} ds = (-1)^n \left(\frac{\pi}{2}\right)^{2n+1} E_{2n}.$$

27. **An explicit formula for the sinc integrals.** Assume that  $n \geq 1$  and  $a_0, a_1, \dots, a_n > 0$ . For  $\gamma = (\gamma_1, \dots, \gamma_n) \in \{-1, 1\}^n$  define

$$b_\gamma = a_0 + \sum_{k=1}^n \gamma_k a_k \quad \text{and} \quad \epsilon_\gamma = \prod_{k=1}^n \gamma_k.$$

Show:

$$(a) \quad \sum_{\gamma \in \{-1, 1\}^n} \epsilon_\gamma b_\gamma^r = \begin{cases} 0, & \text{for } r = 0, 1, \dots, n-1, \\ 2^n n! \prod_{k=1}^n a_k, & \text{for } r = n, \end{cases}$$

where  $b_\gamma^0 = 1$  even if  $b_\gamma = 0$ . Hint: Expand both sides of  $e^{a_0 t} \prod_{k=1}^n (e^{a_k t} - e^{-a_k t}) = \sum_{\gamma \in \{-1, 1\}^n} \epsilon_\gamma e^{b_\gamma t}$  into a power series in  $t$  and compare coefficients.

$$(b) \quad \prod_{k=0}^n \sin(a_k x) = \frac{1}{2^n} \sum_{\gamma \in \{-1, 1\}^n} \epsilon_\gamma \cos(b_\gamma x - \frac{\pi}{2}(n+1)).$$

$$(c) \quad \int_0^\infty \prod_{k=0}^n \frac{\sin(a_k x)}{x} dx = \frac{\pi}{2} \frac{1}{2^n n!} \sum_{\gamma \in \{-1, 1\}^n} \epsilon_\gamma b_\gamma^n \text{sign}(b_\gamma).$$

$$(d) \quad \int_0^\infty \prod_{k=0}^n \text{sinc}(a_k x) dx = \frac{\pi}{2} \frac{1}{a_0} \left( 1 - \frac{1}{2^{n-1} n! a_1 \cdots a_n} \sum_{b_\gamma < 0} \epsilon_\gamma b_\gamma^n \right).$$

- (e) **The first “bite.”** If  $\sum_{k=1}^{n-1} a_k \leq a_0 < \sum_{k=1}^n a_k$ , then

$$\int_0^\infty \prod_{k=0}^n \text{sinc}(a_k x) dx = \frac{\pi}{2} \frac{1}{a_0} \left( 1 - \frac{(a_1 + \cdots + a_n - a_0)^n}{2^{n-1} n! a_1 \cdots a_n} \right).$$

28. **A special sinc integral.** Evaluate

$$\begin{aligned} \int_0^\infty \operatorname{sinc}^n(x) dx &= \frac{\pi}{2} \left( 1 + \frac{1}{2^{n-2}} \sum_{1 \leq r \leq \frac{n}{2}} \frac{(-1)^r (n-2r)^{n-1}}{(r-1)! (n-r)!} \right) \\ &= \frac{\pi}{2^n (n-1)!} \sum_{0 \leq r \leq \frac{n}{2}} (-1)^r \binom{n}{r} (n-2r)^{n-1}. \end{aligned}$$

Thus confirm the results of Exercise 5.

29. **A strange cosine integral.** Let  $C^*(x) = \cos(2x) \prod_{n=1}^\infty \cos(x/n)$ . Show symbolically that  $\int_0^\infty C^*(x) dx < \pi/8$ , and show numerically that

$$0 < \frac{\pi}{8} - \int_0^\infty C^*(x) dx < 10^{-41}.$$

This is thus hard to distinguish numerically from  $\pi/8$ ; compare Exercise 39.

30. **Multi-variable sinc integrals.** For  $x, y \in \mathbb{R}^m$  we write  $x \cdot y$  to denote the dot product. Define the *sinc space*  $\mathcal{S}^{m,n}$  to be the set of  $m \times (m+n)$  matrices  $S = (s_1 \ s_2 \ \cdots \ s_{m+n})$  of column vectors in  $\mathbb{R}^m$  such that

$$\int_{\mathbb{R}^m} \left| \prod_{k=1}^{m+n} \operatorname{sinc}(s_k \cdot y) \right| dy < \infty,$$

and a function  $\sigma : \mathcal{S}^{m,n} \rightarrow \mathbb{R}$  by

$$\sigma(S) = \int_{\mathbb{R}^m} \prod_{k=1}^{m+n} \operatorname{sinc}(s_k \cdot y) dy.$$

Correspondingly, define the *polyhedron space*  $\mathcal{P}^{m,n}$  to be the complete set of  $m \times (m+n)$  matrices  $P = (p_1 \ p_2 \ \cdots \ p_{m+n})$  and a function  $\nu : \mathcal{P}^{m,n} \rightarrow \mathbb{R}$  by

$$\nu(P) = \operatorname{Vol}\{x \in \mathbb{R}^n : |p_k \cdot x| \leq 1 \text{ for } k = 1, 2, \dots, m+n\}.$$

(a) Note that by change of basis, for  $S \in \mathcal{S}^{m,n}$  and  $P \in \mathcal{P}^{m,n}$  we have

$$\sigma(S) = |\det(M)| \sigma(MS) \quad \text{and} \quad \nu(P) = |\det(N)| \nu(NP)$$

for non-singular matrices  $M$  ( $m \times m$ ) and  $N$  ( $n \times n$ ).

- (b) The following correspondence between multidimensional sinc integrals and volumes of polyhedra can be proved with some effort (see [34]): If  $n \geq m$ , if  $A$  is a non-singular  $(m \times m)$ -Matrix, and if  $B$  is any  $(m \times n)$ -matrix having  $m$  of its columns linearly independent, then

$$\sigma(A|B) = \frac{\sigma(I^m|A^{-1}B)}{|\det(A)|} = \frac{\pi^m \nu(I^n|(A^{-1}B)^T)}{2^n |\det(A)|}.$$

Similarly, if  $n \geq m$ , if  $C$  is a non-singular  $(n \times n)$ -matrix, and if  $D$  is any  $(n \times m)$ -matrix such that  $C^{-1}D$  has  $m$  linearly independent rows, then

$$\nu(C|D) = \frac{\nu(I^n|C^{-1}D)}{|\det(C)|} = \frac{2^n \sigma(I^m|(C^{-1}D)^T)}{\pi^m |\det(C)|}.$$

- (c) Use the theorem from (b) to determine (with the use of symbolic integration) the volume of  $\{x \in \mathbb{R}^6 : |p_k \cdot x| \leq 1, k = 1, \dots, 11\}$ , where  $p_i$  is the  $i$ -th column of the matrix

$$P = \begin{pmatrix} 10 & 0 & 0 & 0 & 0 & 0 & 9 & 10 & -1 & -3 & 7 \\ 0 & 10 & 0 & 0 & 0 & 0 & -2 & -1 & -8 & 2 & -6 \\ 0 & 0 & 10 & 0 & 0 & 0 & -9 & 7 & -5 & 5 & 1 \\ 0 & 0 & 0 & 10 & 0 & 0 & 5 & -2 & -9 & -8 & -9 \\ 0 & 0 & 0 & 0 & 10 & 0 & -10 & -2 & -3 & 6 & -4 \\ 0 & 0 & 0 & 0 & 0 & 10 & -8 & 9 & 2 & 7 & -10 \end{pmatrix}.$$

Hint:  $\nu(P) = (32/(5\pi^5)) \int_{\mathbb{R}^5} \prod_{k=1}^{11} \text{sinc}(s_i \cdot y) dy$ , where

$$S = \begin{pmatrix} 10 & 0 & 0 & 0 & 0 & 9 & -2 & -9 & 5 & -10 & -8 \\ 0 & 10 & 0 & 0 & 0 & 10 & -1 & 7 & -2 & -2 & 9 \\ 0 & 0 & 10 & 0 & 0 & -1 & -8 & -5 & -9 & -3 & 2 \\ 0 & 0 & 0 & 10 & 0 & -3 & 2 & 5 & -8 & 6 & 7 \\ 0 & 0 & 0 & 0 & 10 & 7 & -6 & 1 & -9 & -4 & -10 \end{pmatrix}.$$

31. **Another iterated sinc integral.** Problem: For positive constants  $(a_i)$ , evaluate

$$\int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} \frac{\sin(a_1 x_1)}{x_1} \dots \frac{\sin(a_n x_n)}{x_n} \frac{\sin(a_1 x_1 + \dots + a_n x_n)}{x_1 + \dots + x_n} dx_1 \dots dx_n.$$

Answer:  $\pi^n \min(a_1, \dots, a_n)$ .

32. **Infinite series and Clausen's product.** For  $x$  and  $t$  appropriately restricted:

(a) Use Clausen's product to obtain

$$\sum_{n=0}^{\infty} \frac{(t)_n (-t)_n}{(2n)!} (2x)^{2n} = \cos(2t \arcsin(x))$$

and

$$-\frac{1}{2} \sum_{n=1}^{\infty} \frac{(t)_n (-t)_n}{(2n)!} (4 \sin^2 x)^n = \sin^2(tx).$$

(b) Obtain the Taylor series

$$\arcsin^2(x) = \frac{1}{2} \sum_{n \geq 1} \frac{(2x)^{2n}}{n^2 \binom{2n}{n}}$$

on taking an appropriate limit as  $t \rightarrow 0$  (see also Problem 16 of Chapter 1). Hence, show

$$\sum_{n \geq 1} \frac{1}{n^2 \binom{2n}{n}} = \frac{\pi^2}{18} \quad \text{and} \quad \sum_{n \geq 1} \frac{(-1)^n}{n^2 \binom{2n}{n}} = -2 \log^2 \left( \frac{1 + \sqrt{5}}{2} \right).$$

Evaluate  $\sum_{n \geq 1} 3^n / \binom{2n}{n}$  and both of  $\sum_{n \geq 1} (\pm 1)^n / \binom{2n}{n}$ .

33. **Proof of the Korovkin theorems.** Prove Theorems 2.6.2 and 2.6.4.

34. **Korovkin by inequalities.** An interesting recent approach to the Korovkin theorems is given in [205]. Recall that a subset of a continuous function space is a *subalgebra* if it is closed under pointwise multiplication. Therein, the following elegant lemma is proven:

**Lemma 2.7.3** *Suppose that  $\mathcal{A}$  is a norm-closed subalgebra of  $C[a, b]$  that contains 1. Let  $T$  be a positive linear operator on  $\mathcal{A}$  such that  $T(1) \leq 1$ . Then*

$$(a) \quad \mathcal{E}(h) = T(h)^2 - T(h^2) \geq 0,$$

- (b)  $|T(fg) - T(f)T(g)| \leq \mathcal{E}(f)\mathcal{E}(g)$ ,  
 (c)  $\|T(fg) - T(f)T(g)\|^2 \leq \|\mathcal{E}(f)\| \|\mathcal{E}(g)\|$ ,  
 (d)  $\|T(fg) - T(f)T(g)\|^2 \leq \|\mathcal{E}(f)\| \|\mathcal{E}(g) + \mathcal{E}(k)\|$ ,

for all elements  $f, g, h$  and  $k$  in the algebra.

**Proof.** (a) is established by observing that  $T(h + t1)^2 \geq 0$  for all real  $t$ . Then (b) follows with  $h$  replaced by  $f + tg$ , and (c) and (d) are easy consequences.  $\square$

It is now a nice problem to show that the first and second Korovkin theorems follow—if one knows that the polynomials are dense in  $C[a, b]$ . Moreover, the same approach will yield:

**Theorem 2.7.4 (Complex Korovkin theorem.)** *Let  $D = \{z \in \mathbb{C} : |z| \leq 1\}$ . Let  $T_n$  be positive linear operators on  $C(D)$  such that  $T_n(h) \Rightarrow h$  for  $h = 1, z$  and  $|z|^2$ . Then this holds for all  $h$  in  $C(D)$ .*

To prove this it helps to observe that positive operators preserve conjugates:  $T(\bar{h}) = \overline{T(h)}$  for all  $h$  in  $C(D)$ .

35. **Bézier curves.** The *Bézier curve* of degree  $n$  defined by  $n + 1$  points  $b_0, b_1, \dots, b_n$  is exactly the Bernstein polynomial interpolating the values at  $k/n$

$$\sum_{k=0}^n b_k \binom{n}{k} t^k (1-t)^{n-k}. \quad (2.7.36)$$

Typically, parametric cubic Bézier curves in the plane such as

$$\begin{aligned} x(t) &= -(1-t)^3 - t(1-t)^2 + \frac{3}{2}t^2(1-t) + t^3 \\ y(t) &= \frac{1}{2}(1-t)^3 + t(1-t)^2 + \frac{3}{4}t^2(1-t) + \frac{1}{2}t^3 \end{aligned} \quad (2.7.37)$$

are fitted together for smoothing purposes. To compute the values it is useful to observe Castlejau's algorithm that the basis functions  $B_{n,k} = t \mapsto \binom{n}{k} t^k (1-t)^{n-k}$  satisfy the recursion  $B_{n,-1} = B_{n-1,n} = 0$  and

$$B_{n,k}(t) = (1-t)B_{n-1,k}(t) + tB_{n-1,k-1}(t),$$



for  $0 \leq k \leq n$  and all real  $t$ .

36. **Bernstein polynomials.** Determine the appropriate Bernstein polynomials on  $[-1, 1]$ .
37. **Rate of approximation.** As we have seen, the rate of approximation is tied to the smoothness of the underlying function. In Lebesgue's proof of the Stone-Weierstrass Theorem, the main work is in showing that  $|\cdot|$  can be uniformly approximated by polynomials on  $[-1, 1]$ . Plot the first few Bernstein polynomials and observe that the approximation is worst at zero, where  $|t|$  is not differentiable.
38. **Korovkin kernels.** Apply the Korovkin theorems to the Poisson, Fejér-Korovkin and Jackson kernels respectively.
39. **Contriving coincidences.**

- (a) A consequence of the theta transform, (2.3.15), in the form  $s\theta_3^2(e^{-\pi s}) = \theta_3^2(e^{-\pi/s})$ , is that

$$\sum_{n \geq 1} e^{-(n/10)^2} \approx 5\Gamma\left(\frac{1}{2}\right) - \frac{1}{2}$$

and they agree through 427 digits, with similar more baroque estimates for higher powers of ten.

- (b) The fact that  $\alpha = \exp(\pi\sqrt{163}/3) \approx 640320$  lies deeper and relates to the fact that the only imaginary quadratic fields with unique factorization are  $\mathbf{Q}(\sqrt{-d})$  are with  $d = 1, 2, 4, 7, 11, 19, 43, 67$  and the largest 163.
- (c) This leads to a spectacular “billion-digit” fraud

$$\sum_{n=1}^{\infty} \frac{[n\alpha]}{2^n} \approx 1280640.$$

As we saw this is explained by Theorem 1.4.2 and the fact that as a continued fraction,

$$\alpha = [640320, 1653264929, 30, 1, 321, 2, 1, 1, 1, 4, 3, 4, 2, \dots].$$

(d) Determine the integers  $N_d$  such that

$$\sum_{n=1}^{\infty} \frac{[n\alpha_d]}{2^n} \approx N_d,$$

for  $\alpha_d = \exp(\pi\sqrt{d}/3)$  with  $d = 19, 43, 67, 163$ , and determine the error in each case.

These examples signal the danger of inferring a symbolic identity from tools like PSLQ without knowing the context. That said, we know of nearly no cases where such spectacular deception has occurred without contrivance.

## Chapter 3

# Zeta Functions and Multi-Zeta Values

I see some parallels between the shifts of fashion in mathematics and in music. In music, the popular new styles of jazz and rock became fashionable a little earlier than the new mathematical styles of chaos and complexity theory. Jazz and rock were long despised by classical musicians, but have emerged as art-forms more accessible than classical music to a wide section of the public. Jazz and rock are no longer to be despised as passing fads. Neither are chaos and complexity theory. But still, classical music and classical mathematics are not dead. Mozart lives, and so does Euler. When the wheel of fashion turns once more, quantum mechanics and hard analysis will once again be in style.

Freeman Dyson, 1996 [108]

The Riemann zeta function has already appeared in various contexts in earlier chapters. We start this chapter with gathering up its basic properties in one place. After some further discussion of special values of the function, we complete this chapter with a more detailed exploration of multiple zeta values (Euler sums), as introduced in Chapter 2 of the first volume.

The *zeta-function (of Riemann)* is defined by the following series

$$\zeta(s) = \sum_{n=1}^{\infty} \frac{1}{n^s}, \quad (3.0.1)$$

for  $\operatorname{Re}(s) > 1$ . The estimate

$$\frac{1}{n+1} + \frac{1}{n+2} + \cdots + \frac{1}{2n} > \frac{n}{2n} = \frac{1}{2},$$

and the comparison test for series, show that  $\zeta$  has a pole at  $s = 1$ . To go further we introduce the *alternating zeta function*

$$\alpha(s) = \sum_{n=1}^{\infty} \frac{(-1)^{n+1}}{n^s}, \quad (3.0.2)$$

and note that  $\alpha(1) = \log(2)$ . This series clearly converges (and is analytic as a uniform limit) for  $\operatorname{Re}(s) > 0$ . Moreover regrouping the terms shows that

$$\begin{aligned} \alpha(s) &= -\sum_{n=1}^{\infty} \frac{1}{(2n)^s} + \sum_{n=1}^{\infty} \frac{1}{(2n-1)^s} = 2^{-s}\zeta(s) - \sum_{n=1}^{\infty} \frac{1}{(2n-1)^s} \quad (3.0.3) \\ \zeta(s) &= \sum_{n=1}^{\infty} \frac{1}{(2n)^s} + \sum_{n=1}^{\infty} \frac{1}{(2n-1)^s} = 2^{-s}\zeta(s) + \sum_{n=1}^{\infty} \frac{1}{(2n-1)^s}. \end{aligned}$$

Then (3.0.3) shows  $\sum_{n=1}^{\infty} 1/(2n-1)^s = (1-2^{-s})\zeta(s)$  and that  $\alpha(s) = (1-2^{1-s})\zeta(s)$ . Thus

$$\zeta(s) = \frac{\alpha(s)}{(1-2^{1-s})} \quad (3.0.4)$$

for  $\operatorname{Re}(s) > 0$  which provides an analytic continuation of  $\zeta$  in the right halfplane, with  $\zeta(\frac{1}{2}) = 0.6048986430\dots$ . We also note that  $\alpha(2) = \frac{1}{2}\zeta(2) = \pi^2/12$ .

### 3.1 Reflection and Continuation of Zeta

There are various routes to extend  $\zeta$  into the left halfplane. We choose to start with the function  $\tau(t) = [\theta_3(e^{-\pi t}) - 1]/2$ , where  $\theta_3(q) = \sum_{n=-\infty}^{\infty} q^{n^2}$  as is discussed in Section 4.2. For  $\operatorname{Re}(s) > 1/2$  we choose to use the Mellin transform

$$M_s(\tau) = \int_0^{\infty} \tau(x)x^{s-1} dx$$

and write

$$M_s(\tau) = \sum_{n=1}^{\infty} n^{-2s} \pi^{-s} \int_0^{\infty} e^{-t} t^{s-1} dt = \frac{\Gamma(s)}{\pi^s} \zeta(2s). \quad (3.1.5)$$

Hence

$$\Gamma\left(\frac{s}{2}\right) \zeta(s) \pi^{-s/2} = \int_0^{\infty} t^{s/2-1} \tau(t) dt \quad (3.1.6)$$

$$\begin{aligned} &= \int_1^{\infty} t^{s/2-1} \tau(t) dt + \int_0^1 t^{-1/2} \tau\left(\frac{1}{t}\right) t^{s/2-1} dt \\ &+ \frac{1}{2} \int_0^1 (t^{-1/2} - 1) t^{s/2-1} dt. \end{aligned} \quad (3.1.7)$$

Here we have used the *theta transform* (2.3.15) to replace  $\tau(t)$  by  $\tau(1/t)$  on  $[1, \infty)$ . We deduce that

$$\Gamma\left(\frac{s}{2}\right) \zeta(s) \pi^{-s/2} = -\left(\frac{1}{s} + \frac{1}{1-s}\right) + \int_1^{\infty} \frac{t^{s/2} + t^{(1-s)/2}}{t} \tau(t) dt, \quad (3.1.8)$$

as we see on evaluating the final integral and sending  $t \rightarrow 1/t$  in the second integral in (3.1.6). Because  $\tau(t) = O(e^{-\pi t})$  as  $t \rightarrow \infty$ , the integral in (3.1.8) is an entire function of  $s$ , and as  $\Gamma$  has a simple pole at zero, we see that (3.1.8) extends  $\zeta$  analytically with a single simple pole at  $s = 1$ .

Most beautifully, we note that (3.1.8) is left unchanged by the substitution  $s \rightarrow 1 - s$  and so we obtain the famous functional equation or *reflection formula* for the Riemann zeta function:

$$\Gamma\left(\frac{s}{2}\right) \zeta(s) \pi^{-s/2} = \Gamma\left(\frac{1-s}{2}\right) \zeta(1-s) \pi^{(1-s)/2}. \quad (3.1.9)$$

Symmetry of  $\zeta$  around the line  $\operatorname{Re}(s) = \frac{1}{2}$  is now apparent. We also note that (3.1.9) shows that  $\zeta(-2n)$  is zero for even negative integers (because  $\Gamma$  has poles at those values). These are called “trivial” zeros. As we shall see in the next section  $\zeta(-2n + 1)$  is a Bernoulli number and so rational.

The reflection formula must represent one of the most beautiful findings in mathematics. The British analyst G. N. Watson, discussing his response to equally beautiful formulae of the wonderful Indian mathematical genius Ramanujan (1887–1920), such as those in Section 3.2.3, describes

a thrill which is indistinguishable from the thrill I feel when I enter the Sagrestia Nuovo of the Capella Medici and see before me the austere beauty of the four statues representing “Day,” “Night,” “Evening,” and “Dawn” which Michelangelo has set over the tomb of Giuliano de’Medici and Lorenzo de’Medici. [208]

### 3.1.1 The Riemann Hypothesis

The mathematical centrality of the zeta function can hardly be overestimated. It figures as Problem 8 (of 23) in Hilbert’s famous 1900 lecture and as Problem 5 (of 8) in the Millennium Problems posed by the Clay Foundation 100 years later. Central to the study of the zeta function is the *Riemann hypothesis*:

The only non-trivial zeroes of  $\zeta(s)$  for complex numbers  $s = \sigma + i\gamma$  lie on  $\sigma = 1/2$ .

The importance and present status of the Riemann Hypothesis in prime number theory has already been discussed in Chapter 2 of the first volume, and is developed further in Exercise 2. A related reason for the role of  $\zeta$  in number theory comes from the following:

**Lemma 3.1.1 (Euler Product.)** For  $\sigma = \operatorname{Re}(s) > 1$

$$\zeta(s) = \prod_p \left(1 - \frac{1}{p^s}\right)^{-1}, \quad (3.1.10)$$

where  $p$  runs over the primes.

**Proof.** This can be seen by expanding the finite product and using unique factorization:

$$\prod_{p \leq X} \left(1 - \frac{1}{p^s}\right) = \prod_{p \leq X} \left(1 + \frac{1}{p^s} + \frac{1}{p^{2s}} + \cdots\right) = \sum_{n \leq X} \frac{1}{n^s} + \mathcal{E}(s, X) \quad (3.1.11)$$

where  $\mathcal{E}(s, X) \leq \sum_{n > X} 1/n^\sigma \rightarrow_X 0$ . □

From this we may derive again that for  $\operatorname{Re}(s) \geq 1$ , the function  $\zeta(s)$  has a pole only at  $s = 1$ .

## 3.2 Special Values of the Zeta Function

### 3.2.1 Zeta at Even Positive Integers

We have already evaluated  $\zeta(2)$  in various ways. There are also several ways to find the closed form for  $\zeta(2n)$ . Let us start with the intuitive path followed by Euler. Euler intuited his product formula for  $\pi$  (1.2.11) from the analogy with the Fundamental Theorem of Algebra, and on writing down the Taylor series for  $\sin(x)/x$ , one is left to compare

$$\sum_{n=0}^{\infty} \frac{(-1)^n x^{2n}}{(2n+1)!} = \prod_{n=1}^{\infty} \left(1 - \frac{x^2}{\pi^2 n^2}\right). \quad (3.2.12)$$

Thus, one has

$$\zeta(2) = \sum_{n>0} \frac{1}{n^2} = \frac{\pi^2}{6}.$$

Considering the next term we have

$$\sum_{\substack{m,n>0 \\ m<n}} \frac{1}{n^2 m^2} = \frac{\pi^4}{120}.$$

Now

$$\frac{\pi^4}{60} + \zeta(4) = 2 \sum_{\substack{m,n>0 \\ m \neq n}} \frac{1}{n^2 m^2} + \sum_{m>0} \frac{1}{m^4} = \sum_{m,n>0} \frac{1}{n^2 m^2} = \zeta(2)^2,$$

and

$$\zeta(4) = \frac{\pi^4}{36} - \frac{\pi^4}{60} = \frac{\pi^4}{90}.$$

One can continue in like fashion—by hand or in a computer algebra system—and obtain  $\zeta(6) = \pi^6/945$ ,  $\zeta(8) = \pi^8/9450$ ,  $\zeta(10) = \pi^{10}/93555$ , and we discover that  $\zeta(2n) = q_n \pi^{2n}$  for some rational  $q_n$ , and it is clearly time to be more organized.

To do this we introduce the even *Bernoulli numbers* via the generating function

$$\frac{z}{e^z - 1} + \frac{z}{2} = \sum_{m=0}^{\infty} B_{2m} \frac{z^{2m}}{(2m)!}, \quad |z| \leq 2\pi, \quad (3.2.13)$$

and define  $B_1 = -\frac{1}{2}$ ,  $B_{2n+1} = 0$  for  $n > 0$ . It is easy to discover, equivalently, that

$$\sum_{k=0}^n \binom{n+1}{k} B_k = 0, \quad (3.2.14)$$

and that

$$\pi z \cot(\pi z) = \sum_{m=0}^{\infty} (-1)^m B_{2m} \frac{(2\pi z)^{2m}}{(2m)!}. \quad (3.2.15)$$

Returning to (1.2.11) and differentiating logarithmically, we obtain

$$\pi \cot(\pi z) = \frac{1}{z} - \sum_{n=1}^{\infty} \frac{2z}{n^2 - z^2} = \sum_{n=1}^{\infty} \zeta(2n) z^{2n-1}, \quad (3.2.16)$$

where the second identity comes from repeated use of the geometric series. Comparing coefficients in (3.2.15) and (3.2.16) yields

$$\zeta(2m) = (-1)^m B_{2m} \frac{(2\pi)^{2m}}{2(2m)!}. \quad (3.2.17)$$

Using the reflection formula (3.1.9) we deduce that for nonnegative  $n$

$$\zeta(-2n+1) = -\frac{B_{2n}}{2n}.$$

As the coefficient of  $B_{2n}$  in equation (3.2.14) is non-zero ( $n+1$ ), equation (3.2.14) is a practical formula for generating Bernoulli numbers. The first few Bernoulli numbers are

$$\frac{1}{6}, -\frac{1}{30}, \frac{1}{42}, -\frac{1}{30}, \frac{5}{66}, -\frac{691}{2730}.$$

Note that if one did too little computation, one might come away with the impression that the numerator is always “1.” Actually, the numerator is as hard to compute as the number, but a lovely theorem of Karl von Staudt and Thomas Clausen proves that the fractional part of  $B_{2n}$  and  $\sigma_{2n} = \sum_{p-1|2n} \frac{1}{p}$  agree. Thus  $B_{2n}$  and  $\sigma_{2n}$  have the same denominators, and the later can be very quickly computed, even for large  $n$ . The first 15 even values are

$$1, 6, 30, 42, 30, 66, 2730, 6, 510, 798, 330, 138, 2730, 6, 870, 14322$$

and the mystery as to why terms reoccur is explained by von Staudt’s result.



### 3.2.2 Zeta at Odd Positive Integers

It was only in 1976 that Roger Apéry proved that  $\zeta(3)$  is irrational. As of the end of 2002, it is known that one of the next four odd zeta values is irrational and that infinitely many are, but we can not prove that  $\zeta(5)$  is. They certainly are not simple rational multiples of powers of  $\pi$ , a fact that can be determined by integer relation computations.

Thanks to Apéry, who used the series for  $A_3$  below in his work, it is now well known that

$$\begin{aligned}\zeta(2) &= 3 \sum_{k=1}^{\infty} \frac{1}{k^2 \binom{2k}{k}} \\ \zeta(3) &= \frac{5}{2} \sum_{k=1}^{\infty} \frac{(-1)^{k-1}}{k^3 \binom{2k}{k}} \\ \zeta(4) &= \frac{36}{17} \sum_{k=1}^{\infty} \frac{1}{k^4 \binom{2k}{k}}\end{aligned}$$

(See [54].)

These results make it tempting to conjecture that

$$Z_5 = \zeta(5) / \sum_{k=1}^{\infty} \frac{(-1)^{k-1}}{k^5 \binom{2k}{k}}$$

is a simple rational or algebraic number. Sadly, or happily, we may use PSLQ to determine that *If  $Z_5$  satisfies a polynomial of degree  $\leq 25$  the Euclidean norm of coefficients exceeds  $2 \times 10^{37}$ .* And the order and norm can be extended *ad libidem*. Thus, any relatively prime integers  $p$  and  $q$  such that

$$\zeta(5) \stackrel{?}{=} \frac{p}{q} \sum_{k=1}^{\infty} \frac{(-1)^{k+1}}{k^5 \binom{2k}{k}}$$

have astronomically large  $q$ .

But a positive use of PSLQ yields in terms of *the first polylogarithms*:

$$\begin{aligned}\sum_{k=1}^{\infty} \frac{(-1)^{k+1}}{k^5 \binom{2k}{k}} &= 2\zeta(5) - \frac{4}{3}L^5 + \frac{8}{3}L^3\zeta(2) + 4L^2\zeta(3) \\ &+ 80 \sum_{n>0} \left( \frac{1}{(2n)^5} - \frac{L}{(2n)^4} \right) \rho^{2n}\end{aligned}\tag{3.2.18}$$

where  $L = \log(\rho)$  and  $\rho = (\sqrt{5} - 1)/2$ ; with similar formulae for  $A_4$  and  $A_6$  (series with alternating signs),  $S_5, S_6$  and  $S_7$  (series with nonalternating signs) [54].

A less well-known formula for  $\zeta(5)$  due to Koecher suggested generalizations for  $\zeta(7), \zeta(9), \zeta(11) \dots$ . Again the coefficients were found by integer relation algorithms. The technique of *bootstrapping* the earlier pattern kept size of the search space manageable. Note that the requisite sums converge relatively quickly, and so are easy to compute even to high precision, which are needed for large relation searches.

For example:

$$\zeta(7) = \frac{5}{2} \sum_{k=1}^{\infty} \frac{(-1)^{k+1}}{k^7 \binom{2k}{k}} + \frac{25}{2} \sum_{k=1}^{\infty} \frac{(-1)^{k+1}}{k^3 \binom{2k}{k}} \sum_{j=1}^{k-1} \frac{1}{j^4}. \quad (3.2.19)$$

The authors of [36] were able, by finding integer relations for  $n = 1, 2, \dots, 10$ , to encapsulate the formulae for  $\zeta(4n + 3)$  in a single conjectured generating function, (entirely *ex machina*):

**Theorem 3.2.1** *For any complex  $|z| < 1$ , we have, formally,*

$$\begin{aligned} \sum_{n=0}^{\infty} \zeta(4n + 3) z^{4n} &= \sum_{k=1}^{\infty} \frac{1}{k^3 (1 - z^4/k^4)} \\ &= \frac{5}{2} \sum_{k=1}^{\infty} \frac{(-1)^{k-1}}{k^3 \binom{2k}{k} (1 - z^4/k^4)} \prod_{m=1}^{k-1} \frac{1 + 4z^4/m^4}{1 - z^4/m^4}. \end{aligned} \quad (3.2.20)$$

The first “=” is easy. The second is quite unexpected in its form! Thus,  $z = 0$  yields Apéry’s formula for  $\zeta(3)$  and the coefficient of  $z^4$  yields (3.2.19).

**How Theorem 3.2.1 was discovered.** The first ten cases show (3.2.20) has the form

$$\frac{5}{2} \sum_{k \geq 1} \frac{(-1)^{k-1}}{k^3 \binom{2k}{k}} \frac{P_k(z)}{(1 - z^4/k^4)}$$

for undetermined  $P_k$ ; with abundant data to compute

$$P_k(z) = \prod_{m=1}^{k-1} \frac{1 + 4z^4/m^4}{1 - z^4/m^4}.$$

Many reformulations of (3.2.20) were found, including a marvellous finite sum:

$$\boxed{\sum_{k=1}^n \frac{2n^2}{k^2} \frac{\prod_{i=1}^{n-1} (4k^4 + i^4)}{\prod_{i=1, i \neq k}^n (k^4 - i^4)} = \binom{2n}{n}}. \quad (3.2.21)$$

This was obtained via Gosper's *telescoping algorithm* of Wilf-Zeilberger type after a mistake in an electronic Petri dish—when a  $\text{\TeX}$  “infty” was typed instead of “infinity” and *Maple* returned an answer that suggested it “knew” an algorithm for such finite sums.

This identity was subsequently proved by Almkvist and Granville [6] thus finishing the proof of (3.2.20) and giving a rapidly converging series for any  $\zeta(4N + 3)$  where  $N$  is positive integer. And perhaps shedding light on the irrationality of  $\zeta(7)$ ?

Paul Erdős, when shown (3.2.21) shortly before his death, rushed off. Twenty minutes later he returned saying he did not know how to prove it but if proven it would have implications for Apéry's result (“ $\zeta(3)$  is irrational”).

The failure to discover a similar function for  $\zeta(4n + 1)$  rests largely on the fact that *too many* relations were found by computer and no candidate to behave like (3.2.1) was isolated to generalize the initial cases such as

$$\zeta(5) = 2 \sum_{k=1}^{\infty} \frac{(-1)^{k+1}}{k^5 \binom{2k}{k}} - \frac{5}{2} \sum_{k=1}^{\infty} \frac{(-1)^{k+1}}{k^3 \binom{2k}{k}} \sum_{j=1}^{k-1} \frac{1}{j^2}. \quad (3.2.22)$$

### 3.2.3 A Taste of Ramanujan

Ramanujan obtained almost analogous evaluations of  $\zeta(2n + 1)$ . For  $M \equiv 3 \pmod{4}$ ,

$$\begin{aligned} \zeta(4N + 3) = & -2 \sum_{k \geq 1} \frac{1}{k^{4N+3} (e^{2\pi k} - 1)} \\ & + \frac{2}{\pi} \left\{ \frac{4N + 7}{4} \zeta(4N + 4) - \sum_{k=1}^N \zeta(4k) \zeta(4N + 4 - 4k) \right\}, \end{aligned}$$

where the interesting term is the rapidly convergent hyperbolic trigonometric series, while the term in braces is a rational multiple of  $\pi^{4N+4}$ . Correspondingly, for  $M \equiv 1 \pmod{4}$ ,

$$\begin{aligned} \zeta(4N+1) &= -\frac{2}{N} \sum_{k \geq 1} \frac{(\pi k + N)e^{2\pi k} - N}{k^{4N+1}(e^{2\pi k} - 1)^2} \\ &\quad + \frac{1}{2N\pi} \left\{ (2N+1)\zeta(4N+2) + \sum_{k=1}^{2N} (-1)^k 2k \zeta(2k) \zeta(4N+2-2k) \right\}. \end{aligned}$$

In each case, only a finite set of  $\zeta(2N)$  values is required, and the full precision value  $e^\pi$  is reused throughout. The number  $e^\pi$  is the easiest transcendental number to rapidly compute (see Problem 7 of Chapter 3 in the first volume).

For  $\zeta(4N+1)$ , a “nicer” series has recently been decoded and then proved from a few PSLQ experiments of Plouffe. It is equivalent to:

$$\left\{ 2 - (-4)^{-N} \right\} \sum_{k=1}^{\infty} \frac{\coth(k\pi)}{k^{4N+1}} - (4)^{-2N} \sum_{k=1}^{\infty} \frac{\tanh(k\pi)}{k^{4N+1}} = Q_N \times \pi^{4N+1}. \quad (3.2.23)$$

The quantity  $Q_N$  in (3.2.23) is an explicit rational:

$$Q_N = - \sum_{k=0}^{2N+1} \frac{B_{4N+2-2k} B_{2k}}{(4N+2-2k)!(2k)!} \times \left\{ (-1)^{\binom{k}{2}} (-4)^N 2^k + (-4)^k \right\}. \quad (3.2.24)$$

This was also discovered using integer relation methods. For instance,

$$\frac{9}{4} \sum_{k=1}^{\infty} \frac{\coth(\pi k)}{k^5} - \frac{1}{16} \sum_{k=1}^{\infty} \frac{\tanh(\pi k)}{k^5} = \frac{5}{672} \pi^5.$$

On substituting

$$\tanh(x) = 1 - \frac{2}{\exp(2x) + 1}, \quad \coth(x) = 1 + \frac{2}{\exp(2x) - 1}$$

in (3.2.23), one may solve for  $\zeta(4N+1)$ . For example:

$$\zeta(5) = \frac{1}{294} \pi^5 + \frac{2}{35} \sum_{k=1}^{\infty} \frac{1}{(1 + e^{2k\pi})k^5} + \frac{72}{35} \sum_{k=1}^{\infty} \frac{1}{(1 - e^{2k\pi})k^5},$$

and  $\zeta(5) - \pi^5/294 = -0.0039555\dots$

### 3.3 Other L-series

The function

$$\beta(s) = L_{-4}(s) = \sum_{n=0}^{\infty} \frac{(-1)^n}{(2n+1)^s}, \quad (3.3.25)$$

is sometimes known as the *Catalan zeta function*, since  $\beta(2) = G$  is Catalan's constant, perhaps the simplest number whose irrationality is unproven. It is independent of  $\zeta$  being based on the multiplicative character modulo 4 which takes values 0, 1, 0,  $-1$ . It is the simplest example of a primitive Dirichlet L-series as arrives when one studies primes in arithmetic progression.

In this case the generating function

$$\sec(z) = \sum_{m=0}^{\infty} \frac{|E_{2m}|z^{2m}}{(2m)!}, \quad |z| \leq \frac{\pi}{2}, \quad (3.3.26)$$

with  $E_{2n+1} = 0$  for  $n \geq 0$ , defines the *Euler numbers*—some authors label our  $E_{2n}$  as  $E_n$ , which we met in Chapter 2 of the first volume. Correspondingly,

$$\sum_{k=0}^n \binom{2n}{2k} E_{2n-2k} = 0, \quad (3.3.27)$$

and

$$\pi \sec(\pi z) = \sum_{m=0}^{\infty} 4^{m+1} \beta(2m+1) z^{2m}. \quad (3.3.28)$$

Comparing coefficients in (3.3.28) and (3.3.26) yields

$$\beta(2m+1) = |E_{2m}| \frac{\left(\frac{\pi}{2}\right)^{2m+1}}{2(2m)!}, \quad (3.3.29)$$

so that in this case it is the odd values that are tractable:

$$\beta(1) = \frac{\pi}{4}, \quad \beta(3) = \frac{\pi^3}{32}, \quad \beta(5) = \frac{5\pi^5}{1536}, \quad \beta(7) = \frac{61\pi^7}{184320},$$

while the first six even Euler numbers are

$$1, -1, 5, -61, 1385, -50521, 2702765.$$

For integer  $k$ , an important class of Dirichlet series is given by

$$L_{\pm k}(s) = \sum_{n=1}^{\infty} \frac{\left(\frac{\pm k}{n}\right)}{n^s}, \quad (3.3.30)$$

where  $\left(\frac{\pm k}{n}\right)$  is the Legendre-Jacobi symbol. When  $n$  is prime, the Legendre symbol is defined as  $\left(\frac{m}{n}\right) = 1$  if  $m$  is a quadratic residue modulo  $n$  (i.e.,  $m$  is a perfect square modulo  $n$ ), and  $= -1$  otherwise. When  $n = p_1 p_2 \cdots p_r$  for odd primes  $p_i$  not necessarily distinct, then  $\left(\frac{m}{n}\right)$  is defined as  $\left(\frac{m}{p_1}\right) \left(\frac{m}{p_2}\right) \cdots \left(\frac{m}{p_r}\right)$ . Using this notation,

$$L_{-8}(s) = 1 + \frac{1}{3^s} - \frac{1}{5^s} - \frac{1}{7^s} + \cdots$$

and

$$L_{+8}(s) = 1 - \frac{1}{3^s} - \frac{1}{5^s} + \frac{1}{7^s} + \cdots.$$

More generally, for any *multiplicative character*  $\chi$  one can define

$$L_{\chi}(s) = \sum_{n=1}^{\infty} \frac{\chi(n)}{n^s}. \quad (3.3.31)$$

The previous case corresponds to characters modulo  $k$ . Then  $L_{\chi}(s)$  has a corresponding Euler product

$$L_{\chi}(s) = \prod_p \left(1 - \frac{\chi(p)}{p^s}\right)^{-1}, \quad (3.3.32)$$

where  $p$  runs over the primes.

For *primitive* characters modulo  $d > 0$  (see [143, 44]) there is a functional equation analogous to that for zeta given in (3.1.9):

$$L_{-d}(s) = C(s) \cos\left(\frac{s\pi}{2}\right) L_{-d}(1-s) \quad (3.3.33)$$

$$L_{+d}(s) = C(s) \sin\left(\frac{s\pi}{2}\right) L_{+d}(1-s) \quad (3.3.34)$$

where  $C(s) = 2^s \pi^{s-1} k^{-s+1/2} \Gamma(1-s)$ . Moreover, the *Dirichlet class number formulae* are

$$L_{-d}(1) = \frac{2\pi h(-d)}{\sqrt{d} w(d)} \quad (3.3.35)$$

$$L_d(1) = \frac{2h(d)}{\sqrt{d}} \log(\varepsilon(d)), \quad (3.3.36)$$

where  $h(-d)$  is the class number of  $\mathbf{Q}(\sqrt{D})$ , where  $d = D$  when  $d$  is congruent to 1 modulo 4 and  $d = D$  otherwise,  $\varepsilon(d)$  is a fundamental unit in  $\mathbf{Q}(\sqrt{D})$ , and  $w(d) = 2$  except that  $w(1) = 4$  and  $w(3) = 6$ .

For our present purposes it suffices that we know that  $L_d(1)/\pi$  satisfies a quadratic equation and  $L_d(1)/\pi \sqrt{d}$  is a rational multiple of the logarithm of a quadratic surd, as we can explore what the values are with integer relation methods—and hunt for evaluations at other odd integers  $s$ .

We illustrate with a few cases. Working to less than 20 digits, we find  $L_{-8}(1) = \sqrt{3}\pi/9$ ,  $L_{-8}(3) = 4\sqrt{3}\pi^3/343$ ,  $L_{-8}(1) = \sqrt{2}\pi/4$ ,  $L_{-8}(3) = 3\sqrt{2}\pi^3/128$ ,  $L_{-12}(1) = \sqrt{3}\pi^3/6$ ,  $L_{-12}(3) = \sqrt{3}\pi^3/45$  and in confirmation  $L_{-4}(3) = \pi^3/32$ . Also,  $L_{-67}(1) = \sqrt{67}\pi/67$  and  $L_{-163}(1) = \sqrt{163}\pi/163$  are particularly simple, since these are the largest cases where the corresponding imaginary quadratic field has unique factorization.

Similarly,  $L_{+5}(1) = 2\sqrt{5}/5 \cdot \log((1+\sqrt{5})/2)$ ,  $L_{+8}(1) = \sqrt{2}/2 \cdot \log(1+\sqrt{2})$ , while  $L_{+13}(1) = 2\sqrt{13}/13 \cdot \log((3+\sqrt{13})/2)$  and  $L_{+29}(1) = 2\sqrt{29}/29 \cdot \log((35+\sqrt{29})/2)$ .

### 3.4 Multi-Zeta Values

*Euler sums* or *MZVs* (“multiple zeta values” or “multi zeta values”) are wonderful generalizations of the classical  $\zeta$  function. For natural numbers  $i_1, i_2, \dots, i_k$ ,

$$\zeta(i_1, i_2, \dots, i_k) = \sum_{n_1 > n_2 > \dots > n_k > 0} \frac{1}{n_1^{i_1} n_2^{i_2} \dots n_k^{i_k}}. \quad (3.4.37)$$

Thus  $\zeta(a) = \sum_{n \geq 1} n^{-a}$  is as before and

$$\zeta(a, b) = \sum_{n=1}^{\infty} \frac{1 + \frac{1}{2^b} + \dots + \frac{1}{(n-1)^b}}{n^a}.$$

In general, the integer  $k$  is the sum's *depth* and  $i_1 + i_2 + \cdots + i_k$  is its *weight*.

This definition (3.4.37) clearly extends to alternating and character sums as we shall see below. MZVs have recently found interesting interpretations in high energy physics, knot theory, combinatorics etc. Such MZVs satisfy many striking identities, of which

$$\zeta(2, 1) = \zeta(3) \quad \text{and} \quad 4\zeta(3, 1) = \zeta(4) \quad (3.4.38)$$

are the simplest. Thus these double zeta sums can be reduced to values of the classical zeta function! Does this happen just in a few special cases, or is this the tip of an iceberg, the beginning of a theory? We would like to answer questions such as: Which multiple zeta values can be reduced to simpler ones, i.e., to rational combinations of MZVs of lower depth? How many irreducible MZVs remain for given depth and weight? Can the relations between different MZVs be sorted, labeled and classified?

The needed computations quickly become very large scale: mixing fields, tools and interfaces such as Reduce, C++, Fortran, Pari, Snap, etc. A high precision *fast  $\zeta$ -convolution* allows use of integer relation algorithms leading to important dimensional (reducibility) conjectures and amazing identities. An algorithm for computing these values is given in Section 7.5.1. See also the Euler-zeta computation tool at the URL <http://www.cecm.sfu.ca/projects/ezface+>. Euler himself found and partially proved theorems on reducibility of depth 2 to depth 1  $\zeta$ 's ( $\zeta(6, 2)$  is the lowest weight "irreducible").

### 3.4.1 Various Methods of Attack

One of the pleasures of work in the area is that so many methods are useful: combinatorial, analytic (complex and real), algebraic, number theoretic, and numerical. This leads to amazing identities and important dimensional (reducibility) conjectures. Almost certainly, the simplest of our dimensional conjectures are not provable by currently known mathematical techniques. As we emphasized above, we can't yet determine whether  $\zeta(5)$ ,  $\zeta(7)$  or  $G$  are rational or irrational, much less transcendental.

We shall finish this section by establishing a conjecture of Zagier first published in [36, 52] to which we refer along with [51] for general information not given in the section. The proof we give is a refinement due to Zagier of one found by Broadhurst during the development of the joint corpus in [63, 52]. The



identity is

$$\zeta(\{3, 1\}_n) = \frac{1}{2n+1} \zeta(\{2\}_{2n}) \quad \left( = \frac{2\pi^{4n}}{(4n+2)!} \right). \quad (3.4.39)$$

Here  $\{s\}_n$  is the string  $s$  repeated  $n$  times. This is the *unique* non-commutative analogue of Euler's evaluation of  $\zeta(2n)$ .

We illustrate the diversity of the area with a deep conjecture that sits as a very special case of various dimensional conjectures we make below and provide evidence for:

**Conjecture.** (Drinfeld(1991)-Deligne) The graded Lie algebra of *Grothendieck & Teichmuller* has no more than one generator in odd degrees, and no generators in even degrees.

In the known “non-reducible” identities for Euler sums, all  $\zeta$ -terms have the same weight. This is of great importance for guided integer relation searches, as it dramatically reduces the size of the search space. It is, moreover, useful to consider more general sums:

$$\zeta(i_1, i_2, \dots, i_k; \sigma_1, \sigma_2, \dots, \sigma_k) = \sum_{n_1 > n_2 > \dots > n_k > 0} \frac{\sigma_1^{n_1} \sigma_2^{n_2} \dots \sigma_k^{n_k}}{n_1^{i_1} n_2^{i_2} \dots n_k^{i_k}}. \quad (3.4.40)$$

For general complex  $\sigma_i$ , (3.4.40) defines *Eulerian polylogarithms*, while  $\sigma_i \in \{1, -1\}$  produce *Euler sums*. We restrict the term *multi zeta value* (MZV) to:

$$\zeta(i_1, i_2, \dots, i_k) = \sum_{n_1 > n_2 > \dots > n_k > 0} \frac{1}{n_1^{i_1} n_2^{i_2} \dots n_k^{i_k}},$$

that is, when  $\sigma_i \equiv 1$ .

### 3.4.2 Reducibility and Dimensional Conjectures

As a first taste we prove the following lemma due to Euler:

#### Lemma 3.4.1

$$\zeta(a, b) + \zeta(b, a) = \zeta(a)\zeta(b) - \zeta(a+b),$$

for integer  $a > 1, b \geq 1$ . In particular

$$\zeta(a, a) = \frac{1}{2}\zeta(a)^2 - \frac{1}{2}\zeta(2a)$$

reduces  $\zeta(a, a)$ .

**Proof.** Observe that

$$\begin{aligned} \zeta(a, b) + \zeta(b, a) + \zeta(a + b) &= \sum_{n=1}^{\infty} \sum_{m>n}^{\infty} \frac{1}{n^a m^b} + \sum_{n=1}^{\infty} \sum_{m<n}^{\infty} \frac{1}{n^a m^b} + \sum_{n=1}^{\infty} \sum_{n=m}^{\infty} \frac{1}{n^a m^b} \\ &= \sum_{n=1}^{\infty} \sum_{n=1}^{\infty} \frac{1}{n^a m^b} = \zeta(a)\zeta(b). \end{aligned} \quad (3.4.41)$$

□

More complex versions of this sort of argument often involving partial fraction identities lead to sets of equations (and to the matrices in Section 4.5), and to algebraic proofs of many cognate MZV identities. For example:

$$\zeta(a, b, c) + \zeta(a, c, b) + \zeta(c, a, b) = \zeta(c)\zeta(a, b) - \zeta(a, b + c) - \zeta(a + c, b).$$

As a second taste, we show how a computer algebra system can “prove” Euler’s first significant result in the area: “generatingfunctionology” produces:

$$\frac{(-1)^m}{(m-1)!} \int_0^1 \frac{\log^{m-1}(t) \sum_{n>0} a_n t^n}{1-t} dt = \sum_{n>0} \frac{\sum_{k<n} a_k}{n^m}$$

and so represents  $\zeta(m, 1)$  as the following integral

$$\begin{aligned} \zeta(m, 1) &= \frac{(-1)^m}{(m-1)!} \int_0^1 \frac{\log^{m-1}(t) \log(1-t)}{1-t} dt \\ &= \frac{(-1)^m}{2(m-1)!} \int_0^1 \frac{(m-1) \log^{m-2}(t) \log^2(1-t)}{t} dt. \end{aligned} \quad (3.4.42)$$

Inspection of the definition of the beta function

$$\beta(x, y) = \int_0^1 t^{x-1} (1-t)^{y-1} dt \quad (3.4.43)$$

shows that the right side of (3.4.42) can be written as a  $\beta$ -function derivative

$$\zeta(m, 1) = \frac{(-1)^m}{2(m-2)!} B_1^{(m-2)}(0), \quad (3.4.44)$$

where  $B_1(y) = \frac{\partial^2}{\partial x^2} \beta(x, y) \Big|_{x=1}$ . Since

$$\frac{\partial^2}{\partial x^2} \beta(x, y) = \beta(x, y) [(\Psi(x) - \Psi(x+y))^2 + (\Psi'(x) - \Psi'(x+y))],$$

we have a *digamma* representation via

$$B_1(y) = \frac{1}{y} ((-\gamma - \Psi(y+1))^2 + (\zeta(2) - \Psi'(y+1))).$$

Indeed, without going beyond (3.4.44), we may implement (3.4.44) in *Maple* or *Mathematica* very painlessly and *discover* its Riemann  $\zeta$ -function reduction

$$\begin{aligned} \zeta(n, 1) &= \sum_{k=1}^{\infty} \left(1 + \frac{1}{2} + \cdots + \frac{1}{k}\right) (k+1)^{-n} \\ &= \frac{n\zeta(n+1)}{2} - \frac{1}{2} \sum_{k=1}^{n-2} \zeta(n-k)\zeta(k+1), \end{aligned}$$

from the first five or ten symbolic values. Moreover each case computed is a proof. Note that one wants to ensure that *Maple* or *Mathematica* does not evaluate  $\zeta(2n)$ , otherwise when  $n+1$  is even a convolution will not be exposed:

$$6\zeta^2(4) + 6\zeta(2)\zeta(6) = \frac{17}{9450} \pi^8 = 17\zeta(8).$$

To make the notion of reduction more explicit, a key problem is to find the *dimension* of a minimal generating set for a  $(\mathbb{Q}, +, \cdot)$ -algebra that contains

1. all Euler sums of weight  $n$  and depth  $k$ , generated by Euler sums,  $E_{n,k}$ ;
2. all MZVs of weight  $n$  and depth  $k$ , generated by Euler sums,  $E_{n,k}$ ; or
3. all MZVs of weight  $n$  and depth  $k$ , generated by MZVs,  $D_{n,k}$ .

Conjectured generating functions (due to Broadhurst-Kreimer, Zagier, and others) are:

$$\begin{aligned} \prod_{n \geq 3} \prod_{k \geq 1} (1 - x^n y^k)^{E_{n,k}} &\stackrel{?}{=} 1 - \frac{x^3 y}{(1-x^2)(1-xy)} \\ \prod_{n \geq 3} \prod_{k \geq 1} (1 - x^n y^k)^{M_{n,k}} &\stackrel{?}{=} 1 - \frac{x^3 y}{1-x^2} \\ \prod_{n \geq 3} \prod_{k \geq 1} (1 - x^n y^k)^{D_{n,k}} &\stackrel{?}{=} 1 - \frac{x^3 y}{1-x^2} + \frac{x^{12} y^2 (1-y^2)}{(1-x^4)(1-x^6)}. \end{aligned}$$

For  $k = 2$ ,  $n$  odd and  $k = 3$ ,  $n$  even, the result implicit for  $D_{n,k}$  can be proven by “elementary methods.” Note that  $D_{n,k}$  has a disconcertingly complicated conjectured rational generating function.

In the next example and elsewhere we sometimes write  $-s_i$  or  $\bar{s}_i$  to denote alternation in the  $i$ -th position.

**Example 3.4.2** *Two MZV reductions.*

**MZV over MZVs:** an example of a sum that reduces is

$$\zeta(4, 1, 3) = -\zeta(5, 3) + \frac{71}{36}\zeta(8) - \frac{5}{2}\zeta(5)\zeta(3) + \frac{1}{2}\zeta(3)^2\zeta(2).$$

**MZV over Euler sums:**  $\zeta(4, 2, 4, 2)$  is irreducible as an MZV but as an Euler sum we have:

$$\begin{aligned} \zeta(4, 2, 4, 2) &= -\frac{1024}{27}\zeta(-9, -3) - \frac{267991}{5528}\zeta(12) - \frac{1040}{27}\zeta(9, 3) - \frac{76}{3}\zeta(9)\zeta(3) \\ &\quad - \frac{160}{9}\zeta(7)\zeta(5) + 2\zeta(6)\zeta(3)^2 + 14\zeta(5, 3)\zeta(4) + 70\zeta(5)\zeta(4)\zeta(3) - \frac{1}{6}\zeta(3)^4. \end{aligned}$$

However,  $\zeta(5, 3)$ ,  $\zeta(-9, -3)$  are irreducible over the Euler sums. □

Tables 3.1, 3.2, and 3.3 give values of these three dimensions. The tools used included partial fractions, functional equations and the “shuffles algebra” [36, 52]. The generating functions have been confirmed *numerically*, and a sizable subset is proven symbolically, in the following ranges:

$E_{n,k}$	$k$	1	2	3	4	5	6
3		1					
4			1				
5		1		1			
6			1		1		
7		1		2		1	
8			2		2		1
9		1		3		3	
10			2		5		3
11		1		5		7	
12			3		8		9
13		1		7		14	
14			3		14		20
15		1		9		25	
16			4		20		42
17		1		12		42	
18			4		30		75
19		1		15		66	
20			5		40		132

Table 3.1: Values of  $E_{n,k}$  for various  $n$  and  $k$ .

1.  $E_{n,k}$  (with REDUCE and PSLQ):  $k = 2$  and  $n \leq 44$  and  $k = 7$  and  $n \leq 8$ .
2.  $M_{n,k}$  (with REDUCE and PSLQ):  $k = 2$  and  $n \leq 17$  and  $k = 7$  and  $n \leq 20$ .
3.  $D_{n,k}$ , modulo a big prime (with REDUCE and Fortran):  $k = 3$  and  $n \leq 141$  and  $k = 7$  and  $n \leq 21$ ; (with Fortran):  $k = 3$  and  $n \leq 161$  and  $k = 7$  and  $n \leq 23$ .

Hybrid code based on symbolic evaluation of identities and on PSLQ allowing exact reduction has been run for: (1) All alternating (Euler) sums to weight 9; (2) all MZV's to weight 14. Thus all these evaluations are fully established, as are many more scattered in the literature.

$M_{n,k}$	$k$	1	2	3	4	5	6
3	$n$	1					
4							
5		1					
6							
7		1					
8			1				
9		1					
10			1				
11		1		1			
12			2				
13		1		2			
14			2		1		
15		1		3			
16			3		2		
17		1		5		1	
18			3		5		
19		1		7		3	
20			4		8		1

Table 3.2: Values of  $M_{n,k}$  for various  $n$  and  $k$ .

$D_{n,k}$	$k$	1	2	3	4	5	6
3		1					
4							
5		1					
6							
7		1					
8			1				
9		1					
10			1				
11		1		1			
12			1		1		
13		1		2			
14			2		1		
15		1		2		1	
16			2		3		
17		1		4		2	
18			2		5		1
19		1		5		5	
20			3		7		3

Table 3.3: Values of  $D_{n,k}$  for various  $n$  and  $k$ .

### 3.5 Double Euler Sums

A natural first generalization of the  $\zeta$  function, initially studied by Euler, is to let  $\zeta(t, s) = \sum_{n=1}^{\infty} \sum_{k=1}^{n-1} k^{-s} n^{-t}$ . Such sums, which we first encountered in Lemma 3.4.1, are called *double Euler sums*.

As a more detailed foray into Euler sums, it is our intention to prove the following evaluation due to Euler, in a self contained fashion.

$$\begin{aligned} \zeta(t, s) &= \frac{1}{2} \left[ \binom{s+t}{s} - 1 \right] \zeta(s+t) + \zeta(s)\zeta(t) \\ &\quad - \sum_{j=1}^n \left[ \binom{2j-2}{s-1} + \binom{2j-2}{t-1} \right] \zeta(2j-1)\zeta(s+t-2j+1) \\ &\hspace{25em} \text{if } s \text{ is odd,} \\ \zeta(t, s) &= -\frac{1}{2} \left[ \binom{s+t}{s} + 1 \right] \zeta(s+t) \\ &\quad + \sum_{j=1}^n \left[ \binom{2j-2}{s-1} + \binom{2j-2}{t-1} \right] \zeta(2j-1)\zeta(s+t-2j+1) \\ &\hspace{25em} \text{if } s \text{ is even.} \end{aligned}$$

for  $s+t$  odd, as in ([30]). The terms involving  $\zeta(1)$  which he used here can be cancelled formally if  $t > 1$ . Euler obtained his evaluation by computing many examples ( $s+t \leq 13$ ) and then extrapolating the general formula, without actual proof. The proof we give is not the simplest, but it allows us to play interestingly with combinatorial matrices, a subject we revisit in the next chapter.

We need the following lemma which was known already to Euler and can be proved by induction on  $s+t$ .

**Lemma 3.5.1** *Define*

$$A_j^{(s,t)} = \binom{s+t-j-1}{s-j} \text{ and } B_j^{(s,t)} = \binom{s+t-j-1}{t-j}.$$

*Then we have the partial fraction decomposition*

$$\frac{1}{x^s(1-x)^t} = \sum_{j=1}^s \frac{A_j^{(s,t)}}{x^j} + \sum_{j=1}^t \frac{B_j^{(s,t)}}{(1-x)^j},$$

for  $s, t \geq 0, s+t \geq 1$ .



We will now derive systems of linear equations for the values  $\zeta(s, t)$  where  $s + t = N$ , a constant. First, as we have already seen (Lemma 3.4.1), there is a simple relation between  $\zeta(s, t)$  and  $\zeta(t, s)$ :

$$\zeta(s, t) + \zeta(t, s) = \zeta(s)\zeta(t) - \zeta(s + t)$$

for  $s, t \geq 2$ . We will refer to this equation and its equivalents as “reflection formulas.” It follows that  $2\zeta(s, s) = \zeta^2(s) - \zeta(2s)$ .

Second, we have

$$\begin{aligned} \zeta(s)\zeta(t) &= \left( \sum_{k=1}^{\infty} \frac{1}{k^s} \right) \cdot \left( \sum_{n=1}^{\infty} \frac{1}{n^t} \right) = \sum_{n=1}^{\infty} \sum_{k=1}^{n-1} \frac{1}{k^s(n-k)^t} \\ &= \sum_{n=1}^{\infty} \sum_{k=1}^{n-1} \left( \sum_{j=1}^s \frac{A_j^{(s,t)}}{n^{s+t-j}k^j} + \sum_{j=1}^t \frac{B_j^{(s,t)}}{n^{s+t-j}(n-k)^j} \right) \\ &= \sum_{j=1}^s A_j^{(s,t)} \zeta(s+t-j, j) + \sum_{j=1}^t B_j^{(s,t)} \zeta(s+t-j, j), \end{aligned}$$

for  $s, t \geq 2$ , with  $A_j^{(s,t)}$  and  $B_j^{(s,t)}$  defined as in Lemma 3.5.1. We will refer to these equations as “decomposition formulas.”

The following version of Euler’s equations can be used, together with the reflection formulas, to prove Euler’s formula (3.5.45) algebraically (which we do not do explicitly here):

$$\begin{aligned} \zeta(t, s) &= (-1)^s \sum_{j=2}^s (-1)^j A_j^{(s,t)} \zeta(j) \zeta(s+t-j) + (-1)^s \sum_{j=2}^t B_j^{(s,t)} \zeta(s+t-j, j) \\ &\quad + (-1)^s \binom{s+t-2}{s-1} (\zeta(s+t-1, 1) + \zeta(s+t)). \end{aligned}$$

We will now distinguish the two cases  $s+t$  odd and  $s+t$  even. First, we treat the case where  $s+t = 2n+1$ . We have  $2n-2$  equations in the  $2n-2$  unknowns  $\zeta(2, 2n-1), \zeta(3, 2n-2), \dots, \zeta(2n-3, 2)$ . We can reduce the  $\zeta(k, 2n+1-k)$  with  $k > n$  to  $\zeta(2n+1-k, k)$  by the reflection formulas. This leaves us with the  $n-1$  unknowns  $\zeta(2, 2n-1), \dots, \zeta(n, n+1)$ . The matrix which corresponds to these equations has the entries

$$(A_j^{(k, 2n+1-k)} + B_j^{(k, 2n+1-k)} - B_{2n+1-k}^{(k, 2n+1-k)})_{j,k=2, \dots, n}.$$

However, it will simplify matters considerably if we augment this matrix by allowing  $j, k$  to run from 1 to  $n$ , and then multiply alternate rows by  $-1$ .

Define, therefore, the  $n \times n$  matrices  $A, B, C, M$  by

$$A_{kj} = (-1)^{k+1} \binom{2n-j}{2n-k}, \quad B_{kj} = (-1)^{k+1} \binom{2n-j}{k-1}, \quad C_{kj} = (-1)^{k+1} \binom{j-1}{k-1}$$

( $k, j = 1, \dots, n$ ) and  $M = A + B - C$ . Define the  $n$  vector  $r$  by  $r_1 = 0$  and

$$r_k = (-1)^{k+1} \left[ \zeta(k)\zeta(2n+1-k) + \sum_{i=k}^n \binom{i-1}{k-1} (\zeta(2n+1) - \zeta(2n+1-i)\zeta(i)) \right]$$

for  $k = 2, \dots, n$ . We then have to solve the system  $Mx = r$ , where the vector  $x$  has  $x_1 = 0$ , and  $x_k = \zeta(k, 2n+1-k)$  for  $k > 1$ .

We shall need the following lemma which is known and can be proved by induction on  $m$ .

**Lemma 3.5.2** (i) For  $0 \leq \mu \leq m$ ,  $\sum_{i=0}^{\mu} \binom{m-\mu+i}{i} = \binom{m+1}{\mu}$ .

(ii) For  $0 \leq \nu, \mu \leq m-1$ ,  $\sum_{i=1}^m (-1)^{i+1} \binom{m-i}{\nu} \binom{m-\mu-1}{i-1} = \binom{\mu}{m-\nu-1}$ .

Setting  $\mu = 0$  in (ii) and changing the order of summation yields

$$\sum_{i=1}^m (-1)^{i+1} \binom{i-1}{\nu-1} \binom{m-1}{i-1} = (-1)^{m+1} \delta_{m\nu}.$$

Lemma 3.5.2(ii) can be used to prove the following matrix identities.

$$A^2 = C^2 = I, \tag{3.5.45}$$

$$A = BC, \quad B = AC, \quad C = AB, \tag{3.5.46}$$

$$B^2 = CA, \quad CB = BA. \tag{3.5.47}$$

It follows from these identities that  $B^3 = BCA = AA = I$ , and that the matrix group generated by  $A, B$  and  $C$  is the permutation group on 3 symbols. We shall not use this, but will discuss this from a different perspective in Section 4.5, where we also prove (3.5.45), (3.5.46), and (3.5.47).

These matrix identities now allow us to show that  $M$  is invertible; in fact, we have

$$\begin{aligned} M^2 &= AA + AB - AC + BA + BB - BC - CA - CB + CC \\ &= I + C - B - A + I = 2I - M, \end{aligned}$$

so that

$$M^{-1} = \frac{1}{2}(M + I).$$

Thus, to prove Euler's formula, it remains to determine  $M^{-1}r$ . For this purpose, define  $p_1 = p_{2n} = 0$ ,  $p_k = \zeta(k)\zeta(2n + 1 - k)$  for  $k = 2, \dots, 2n - 1$ ,  $p = (p_k)_{k=1, \dots, n}$ ,  $\bar{p} = ((-1)^{k+1}p_k)_{k=1, \dots, n}$  and  $e = (1, \dots, 1)$ . Then  $r_k = \bar{p}_k - (Cp)_k + \zeta(2n + 1)(Ce)_k$ . Now let  $k \geq 2$ . Then it follows that

$$\begin{aligned} (M^{-1}r)_k &= \left( \frac{1}{2}(M + I)r \right)_k \\ &= \frac{1}{2} [ ((A + B - C + I)\bar{p})_k - ((A + B - C + I)Cp)_k \\ &\quad + \zeta(2n + 1)((A + B - C + I)Ce)_k ] \\ &= \frac{1}{2} [ ((A + B - C + I)\bar{p})_k - ((A + B + C - I)p)_k \\ &\quad + \zeta(2n + 1)((A + B + C - I)e)_k ]. \end{aligned} \quad (3.5.48)$$

### 3.6 Duality Evaluations and Computations

For non-negative integers  $s_1, \dots, s_k$ , we consider

$$\zeta_a(s_1, \dots, s_k) = \sum_{n_j > n_{j+1} > 0} a^{-n_1} \prod_{j=1}^k n_j^{-s_j}, \quad (3.6.49)$$

a special case of the *multidimensional polylogarithm*. Note that

$$\zeta_a(s) = \sum_{n>0} \frac{1}{a^n n^s} = \text{Li}_s(a^{-1})$$

is the usual polylogarithm for  $s \in \mathbf{N}$  and  $|a| > 1$ . We write  $\zeta = \zeta_1$  and  $\kappa = \zeta_2$ .

We also define a *unit Euler sum* by

$$\rho(\sigma_1, \dots, \sigma_k) = \sum_{n_j > n_{j+1} > 0} \prod_{j=1}^k \frac{\sigma_j^{n_j}}{n_j}.$$

**Iterated integral representations.** Put

$$\omega_a = \frac{dx}{x-a}.$$

Integration over  $0 \leq x_1 \leq x_2 \leq \dots \leq x_s \leq 1$  allows us to write

$$\zeta_a(s_1, \dots, s_k) = (-1)^k \int_0^1 \prod_{j=1}^k \omega_0^{s_j-1} \omega_a,$$

and dually

$$\zeta_a(s_1, \dots, s_k) = (-1)^{s+k} \int_0^1 \prod_{j=k}^1 \omega_{1-a} \omega_1^{s_j-1},$$

follows on changing  $x \mapsto 1-x$  at each level. So:

$$(-1)^k \zeta_a(s_1 + 2, \{1\}_{r_1}, \dots, s_k + 2, \{1\}_{r_k}) = (-1)^r \int_0^1 \prod_{j=1}^k \omega_0^{s_j+1} w_a^{r_j+1}, \quad (3.6.50)$$

and dually

$$(-1)^k \zeta_a(s_1 + 2, \{1\}_{r_1}, \dots, s_k + 2, \{1\}_{r_k}) = (-1)^s \int_0^1 \prod_{j=k}^1 \omega_{1-a}^{r_j+1} w_1^{s_j+1}. \quad (3.6.51)$$

**Theorem 3.6.1** 1. *Setting  $a = 1$  gives the “duality for MZVs”:*

$$\zeta(s_1 + 2, \{1\}_{r_1}, \dots, s_k + 2, \{1\}_{r_k}) = \zeta(r_k + 2, \{1\}_{s_k}, \dots, r_1 + 2, \{1\}_{s_1}). \quad (3.6.52)$$

2. Setting  $a = 2$  gives a corresponding “kappa-to-unit-Euler” duality:

$$\begin{aligned} & \kappa(s_1 + 2, \{1\}_{r_1}, \dots, s_k + 2, \{1\}_{r_k}) \\ &= (-1)^{r+k} \zeta(\bar{1}, \{1\}_{r_k}, \bar{1}, \{1\}_{s_k}, \dots, \bar{1}, \{1\}_{r_1}, \bar{1}, \{1\}_{s_1}). \end{aligned} \tag{3.6.53}$$

3. A more general, less convenient, “kappa-to-unit-Euler” duality similarly derivable is

$$\kappa(s_1, \dots, s_k) = (-1)^k \rho(\tau_1, \tau_2/\tau_1, \tau_3/\tau_2, \dots, \tau_s/\tau_{s-1}),$$

where  $[\tau_1, \dots, \tau_s] = [-1, \{1\}_{s_k-1}, \dots, -1, \{1\}_{s_1-1}]$ .

For example, we immediately see from part 1. of Theorem 3.6.1 that  $\zeta(2, 1, 1, 1) = \zeta(5)$ , and that  $\zeta(\{2, 1\}_n) = \zeta(\{3\}_n)$  for all  $n$ , while  $\zeta(3, 1)$  is self dual. There is a profusion of nice specializations, some of which we now list.

**Some  $\kappa \leftrightarrow \rho$  duality examples.**

$$\kappa(1) = \sum_{n \geq 1} \frac{1}{n2^n} = -\log(1/2) = \sum_{n \geq 1} \frac{(-1)^{n+1}}{n} = -\zeta(\bar{1}),$$

$$\kappa(2) = \sum_{n \geq 1} \frac{1}{n^2 2^n} = \text{Li}_2(1/2) = \sum_{n \geq 1} \frac{(-1)^{n+1}}{n} \sum_{k=1}^{n-1} \frac{(-1)^k}{k} = -\zeta(\bar{1}, \bar{1}),$$

$$\kappa(r + 2) = \sum_{n \geq 1} \frac{1}{n^{r+2} 2^n} = \text{Li}_{r+2}(1/2) = -\zeta(\bar{1}, \bar{1}, \{1\}_r), \quad (r \geq 0)$$

$$\kappa(\{1\}_n) = (-1)^n \zeta(-1, \{1\}_{n-1}) = \frac{(\log 2)^n}{n!},$$

$$\kappa(2, \{1\}_n) = (-1)^{n+1} \zeta(-1, \{1\}_n, -1),$$

$$\kappa(\{1\}_{m+1}, 2, \{1\}_n) = (-1)^{m+n} \zeta(-1, \{1\}_n, \{-1\}_2, \{1\}_m),$$

$$\kappa(1, n+2) = \rho(-1, -1, \{1\}_n, -1),$$

$$\kappa(1, n) = \int_0^{\frac{1}{2}} \frac{\text{Li}_n(z)}{1-z} dz.$$

In particular,

$$\kappa(1, 2) = \frac{5}{7} \text{Li}_2\left(\frac{1}{2}\right) \text{Li}_1\left(\frac{1}{2}\right) - \frac{2}{7} \text{Li}_3\left(\frac{1}{2}\right) + \frac{5}{21} \text{Li}_1\left(\frac{1}{2}\right)^3$$

$$\kappa(1, 3) = \text{Li}_3\left(\frac{1}{2}\right) \text{Li}_1\left(\frac{1}{2}\right) - \frac{1}{2} \text{Li}_2\left(\frac{1}{2}\right)^2.$$

We note that differentiation proves  $\kappa(0, \{1\}_n) = \kappa(\{1\}_n)$ . Applied to  $\zeta(n+2)$  this provides a lovely closed form for  $\kappa(2, \{1\}_n)$ .

**Two  $\kappa$ -reductions.** Every MZV of depth  $N$  is a sum of  $2^N$   $\kappa$ 's of depth  $N$ , hence *easily* computed, using integral ideas similar to below:

A better method is to set  $\omega_0 = dx/x$ ,  $\omega_1 = -dx/(1-x)$ . Then

$$\zeta(s_1, \dots, s_k) = \sum_{n_j > n_{j+1} > 0} \prod_{j=1}^k n_j^{-s_j}$$

again has representation

$$\zeta(s_1, \dots, s_k) = (-1)^k \int_0^1 \omega_0^{s_1-1} \omega_1 \cdots \omega_0^{s_k-1} \omega_1.$$

The domain,  $1 > x_j > x_{j+1} > 0$ , in  $n = \sum_j s_j$  variables, splits into  $n+1$  parts: each being a product of regions  $1 > x_j > x_{j+1} > \lambda$ , for first  $r$  variables, and  $\lambda > x_j > x_{j+1} > 0$ , for rest. The substitution  $x_j \mapsto 1 - x_j$  replaces an integral of the former by the latter type, with  $\lambda$  replaced by  $\bar{\lambda} = 1 - \lambda$ .

Thence let  $S(\omega_0, \omega_1)$  be the  $n$ -string  $\omega_0^{s_1-1} \omega_1 \cdots \omega_0^{s_k-1} \omega_1$  specifying a MZV. Let  $T_r$  denote the substring of the first  $r$  letters and  $U_{n-r}$  the complementary substring, on the last  $n-r$  letters, so that  $S = T_r U_{n-r}$  for  $n \geq r \geq 0$ . Then

$$\zeta(s_1, \dots, s_k) = \int_0^1 S = \sum_{r=0}^n \pm \int_0^{\tilde{\lambda}} \tilde{T}_r \int_0^{\lambda} U_{n-r} \quad (3.6.54)$$

where  $\tilde{\phantom{x}}$  indicates reversal of letter order.

The alternate polylogarithmic integral

$$\zeta_z(s_1, \dots, s_k) = \sum_{n_j > n_{j+1} > 0} z^{-n_1} \prod_{j=1}^k n_j^{-s_j} = \int_0^z \omega_0^{s_1-1} \omega_1 \cdots \omega_0^{s_k-1} \omega_1 \quad (3.6.55)$$

applied to right side of (3.6.54) produces the MZV as the scalar product of two vectors, composed of  $\zeta_z$ -values with  $z = p$  and  $z = q$ , for any desired  $p, q > 1$  with  $1/p + 1/q = 1$ . We usually set  $p = q = 2$ , i.e.,  $\zeta_2 = \kappa$ .

**Example 3.6.2** *Hölder convolution example.*

For any  $1/p + 1/q = 1$

$$\begin{aligned} \zeta(2, 1, 2, 1, 1, 1) &= \zeta_p(2, 1, 2, 1, 1, 1) \\ &+ \zeta_p(1, 1, 2, 1, 1, 1) \zeta_q(1) + \zeta_p(1, 2, 1, 1, 1) \zeta_q(2) \\ &+ \zeta_p(2, 1, 1, 1, 1) \zeta_q(3) + \zeta_p(1, 1, 1, 1, 1) \zeta_q(1, 3) \\ &+ \zeta_p(1, 1, 1, 1) \zeta_q(2, 3) + \zeta_p(1, 1, 1) \zeta_q(3, 3) \\ &+ \zeta_p(1, 1, 1) \zeta_q(4, 3) + \zeta_p(1, 1) \zeta_q(5, 3) = \zeta(5, 3) \end{aligned}$$

This uses a homogenous combination of  $2s$  polylogs of no higher depth, where  $s$  is the weight of the MZV. It also provides another duality result on letting  $q \rightarrow \infty$ .  $\square$

The method, which is called Hölder convolution, because it relies on complementary  $p$  and  $q$  as in Hölder's inequality of Chapter 5 of the first volume, is easily programmed (see Section 7.5.1), as the following code partially illustrates:

```
Seq := proc(s, t) local k, n;
  if 1 < s[1] then
```

```

[s[1] - 1, seq(s[k], k = 2 .. nops(s))], [1, op(t)]
  else [seq(s[k], k = 2 .. nops(s))],
      [t[1] + 1, seq(t[k], k = 2 .. nops(t))]
  fi end

SEQ := proc(a) local w, k, s;
  w := convert(a, '+'); s := a, [];
  for k to w do s := Seq(s); print(s, k) od;
  s[2] end

```

```
>SEQ([5,3]);
```

```

      [4, 3], [1], 1
      [3, 3], [1, 1], 2
      [2, 3], [1, 1, 1], 3
      [1, 3], [1, 1, 1, 1], 4
      [3], [2, 1, 1, 1], 5
      [2], [1, 2, 1, 1, 1], 6
      [1], [1, 1, 2, 1, 1, 1], 7
      [], [2, 1, 2, 1, 1, 1], 8

      [2, 1, 2, 1, 1, 1]

```

The time to compute  $D$  digits for MZV  $\zeta(s_1, \dots, s_k)$ , of weight  $n$ , is roughly  $c(n)D$  precision  $D$  multiplications (with  $c(n) \propto n$ , for large  $n$ , whatever the depth,  $k$ ). This idea extends reasonably to all Euler sums. Thus, 100 digits of  $\zeta(5, 3)$  takes only a fraction of a second on a 2003-era system, and 1,000 digits only a few seconds. The MPFUN multiprecision software (a previous version of the ARPREC arbitrary precision computation package described in Chapter 6 of the first volume) required 47 minutes on a 2000-era system to compute 20,000 digits.



### 3.7 Proof of the Zagier Conjecture

For  $r \geq 1$  and  $n_1, \dots, n_r \geq 1$ , again specialize the *polylogarithm* to one variable (we write  $L(n_1, \dots, n_r; x) = \zeta_{x^{-1}}(n_1, \dots, n_r)$  to highlight dependence on  $x$ )

$$L(n_1, \dots, n_r; x) = \sum_{0 < m_r < \dots < m_1} \frac{x^{m_1}}{m_1^{n_1} \dots m_r^{n_r}}.$$

Thus

$$L(n; x) = \frac{x}{1^n} + \frac{x^2}{2^n} + \frac{x^3}{3^n} + \dots$$

is the classical *polylogarithm* (see Section 1.7), while

$$L(n, m; x) = \frac{1}{1^m} \frac{x^2}{2^n} + \left( \frac{1}{1^m} + \frac{1}{2^m} \right) \frac{x^3}{3^n} + \left( \frac{1}{1^m} + \frac{1}{2^m} + \frac{1}{3^m} \right) \frac{x^4}{4^n} + \dots,$$

and

$$L(n, m, l; x) = \frac{1}{1^l} \frac{1}{2^m} \frac{x^3}{3^n} + \left( \frac{1}{1^l} \frac{1}{2^m} + \frac{1}{1^l} \frac{1}{3^m} + \frac{1}{2^l} \frac{1}{3^m} \right) \frac{x^4}{4^n} + \dots.$$

These series converge absolutely for  $|x| < 1$  (conditionally on  $|x| = 1$  unless  $n_1 = 1$  and  $x = 1$ ). These polylogarithms are *determined uniquely* by the differential equations

$$\frac{d}{dx} L(n_1, \dots, n_r; x) = \frac{1}{x} L(n_1, n_2, \dots, n_r - 1; x)$$

if  $n_r \geq 2$ ; while for  $n_r = 1$ ,

$$\frac{d}{dx} L(n_1, \dots, n_r; x) = \frac{1}{1-x} L(n_1, \dots, n_{r-1}; x)$$

with the *initial conditions*  $L(n_1, \dots, n_r; 0) = 0$  for  $r \geq 1$  and  $L(\emptyset; x) \equiv 1$ .

It transpires that if  $\bar{s} = (s_1, s_2, \dots, s_r)$  and  $w = \sum s_i$ , every *periodic* polylogarithm leads to a function

$$L_{\bar{s}}(x, t) = \sum_n L(\{\bar{s}\}_n; x) t^{wn}$$

which solves an algebraic ordinary differential equation in  $x$ , and leads to nice *recurrence relations*.

In the simplest case, with  $r = 1$ , the ODE is  $D_s F = t^s F$  where

$$D_s = \left( (1-x) \frac{d}{dx} \right)^1 \left( x \frac{d}{dx} \right)^{s-1}$$

and the solution (by series) is a generalized hypergeometric function:

$$L_{\bar{s}}(x, t) = 1 + \sum_{n \geq 1} x^n \frac{t^s}{n^s} \prod_{k=1}^{n-1} \left( 1 + \frac{t^s}{k^s} \right),$$

as follows from considering  $D_s(x^n)$ . Similarly, for  $r = 1$  and negative integers

$$L_{-\bar{s}}(x, t) = 1 + \sum_{n \geq 1} (-x)^n \frac{t^s}{n^s} \prod_{k=1}^{n-1} \left( 1 + (-1)^k \frac{t^s}{k^s} \right),$$

and  $L_{-\bar{1}}(2x-1, t)$  solves a hypergeometric ODE.

Indeed

$$L_{-\bar{1}}(1, t) = \frac{1}{\beta(1+t/2, 1/2-t/2)}.$$

We correspondingly obtain ODEs for eventually periodic Euler sums. Thus  $L_{-2,1}(x, t)$  is a solution of

$$\begin{aligned} t^6 F &= x^2(x-1)^2(x+1)^2 D^6 F + x(x-1)(x+1)(15x^2 - 6x - 7) D^5 F \\ &+ (x-1)(65x^3 + 14x^2 - 41x - 8) D^4 F + (x-1)(90x^2 - 11x - 27) D^3 F \\ &+ (x-1)(31x - 10) D^2 F + (x-1) DF. \end{aligned}$$

This leads to four-term recursion for  $F = \sum c_n(t)x^n$  with initial values  $c_0 = 1, c_1 = 0, c_2 = t^3/4, c_3 = -t^3/6$ , and the ODE can be simplified.

We are now ready to prove Zagier's conjecture. Again, let  $F(a, b; c; x)$  denote the *hypergeometric function*. Then:

**Theorem 3.7.1** For  $|x|, |t| < 1$  and all positive integers  $n$

$$\begin{aligned} \sum_{n=0}^{\infty} L(\underbrace{3, 1, 3, 1, \dots, 3, 1}_{n\text{-fold}}; x) t^{4n} &= \tag{3.7.56} \\ F\left(\frac{t(1+i)}{2}, \frac{-t(1+i)}{2}; 1; x\right) &\times F\left(\frac{t(1-i)}{2}, \frac{-t(1-i)}{2}; 1; x\right). \end{aligned}$$

**Proof.** Both sides of the putative identity start

$$1 + \frac{t^4}{8} x^2 + \frac{t^4}{18} x^3 + \frac{t^8 + 44t^4}{1536} x^4 + \dots$$

and are *annihilated* by the differential operator

$$D_{31} = \left( (1-x) \frac{d}{dx} \right)^2 \left( x \frac{d}{dx} \right)^2 - t^4$$

Once discovered—and it was discovered after much computational evidence—this can be checked variously in *Mathematica* or *Maple* (e.g., in the package *gfun*)!  $\square$

**Maple code.**

```
deq:=proc(F) D(D(F))+A*D(F)+B*F; end;
eqns:= {(deq(H))(x),D((deq(H)))(x),D(D((deq(H))))(x)};
for p from 0 to 4 do eqns:=subs(('@@'(D,p))(H)(x)=y[p],eqns); od:
yi_sol:=solve(eqns,{seq(y[i], i=0..4)});
id:=x->x; T:=x->t;
# The annihilator to be checked for any product
A31:=proc(F) (1-id)*D((1-id)*D(id*D(id*D(F)))) - T^4*F; end;
Z:=expand(A31(F1*F2)(x)); for p from 0 to 4 do
  Z:=subs(('@@'(D,p))(F1)(x)=y[1,p],
    ('@@'(D,p))(F2)(x)=y[2,p],Z); od:
```

The annihilator to be checked for the given product is

```
a[1]:= x -> 1/x; b[1]:= x -> I*t^2/2*1/(x*(1-x)); a[2]:= x ->1/x;
b[2]:= x -> -I*t^2/2*1/(x*(1-x));
for i from 1 to 2 do for o from 2 to 4 do
  y[i,o]:=subs(subs(A=a[i],B=b[i],
  y[0]=y[i,0],y[1]=y[i,1],yi_sol),y[o]): od: od:

normal(Z);
```

The code returns zero showing the hypergeometric product solves the differential equation.

**Corollary 3.7.2 (Zagier Conjecture)**

$$\zeta(\underbrace{3, 1, 3, 1, \dots, 3, 1}_{n\text{-fold}}) = \frac{2\pi^{4n}}{(4n+2)!} \quad (3.7.57)$$

**Proof.** We have

$$F(a, -a; 1; 1) = \frac{1}{\Gamma(1-a)\Gamma(1+a)} = \frac{\sin \pi a}{\pi a}$$

where the first equality comes from Gauss's evaluation of  $F(a, b; c; 1)$  (see (6.7.51)) and the second was proved in Section 5.4 of the first volume. Hence, setting  $x = 1$ , in (3.7.56) produces

$$\begin{aligned} & F\left(\frac{t(1+i)}{2}, \frac{-t(1+i)}{2}; 1; 1\right) F\left(\frac{t(1-i)}{2}, \frac{-t(1-i)}{2}; 1; 1\right) \\ &= \frac{2}{\pi^2 t^2} \sin\left(\frac{1+i}{2}\pi t\right) \sin\left(\frac{1-i}{2}\pi t\right) \\ &= \frac{\cosh \pi t - \cos \pi t}{\pi^2 t^2} = \sum_{n=0}^{\infty} \frac{2\pi^{4n} t^{4n}}{(4n+2)!} \end{aligned}$$

on using the Taylor series of  $\cos$  and  $\cosh$ . Comparing coefficients in (3.7.56) completes the proof.  $\square$

If one suspects that Corollary 3.7.2 holds, once one can compute these sums well, it is very easy to verify many cases numerically and be entirely convinced.

### 3.8 Extensions and Discoveries

**Generalizations of the Zagier identity.** It is possible to arrive at the result without differential equations, just combinatorial manipulations of the iterated integral representations. This can be generalized to

$$\sum_{s \in \mathcal{I}} \zeta(s) = \frac{\pi^{4n+2}}{(4n+3)!} \quad (3.8.58)$$

where  $s$  runs over the set  $\mathcal{I}$  of all  $2n + 1$  possible insertions of the number 2 in the string  $\{3, 1\}_n$ . Actually (3.8.58) is just the beginning of a large family of conjectured identities that were determined intensively using PSLQ, not all of which are proved. Other broad generalizations are discussed in Exercise 15.

Compare the much easier result which follows from the product formula for  $\sin$

$$\sum_{n=0}^{\infty} L(\{2\}_n; x) t^{2n} = F(it, -it; 1; x)$$

and, more generally,

$$\sum_{n=0}^{\infty} L(\{p\}_n; x) t^{pn} = {}_pF_{p-1}(-\omega t, -\omega^3 t, \dots, -\omega^{2p-1} t; 1, \dots, 1; x)$$

where  $\omega^p = -1$ . In each case expanding the right hand-side as a power series is easy.

The amazing factorizations in the result for  $\zeta(\{3, 1\}_n)$  and

$$\zeta(\{3, 1\}_n) = 4^{-n} \zeta(\{4\}_n) = \frac{1}{2n+1} \zeta(\{2, 2\}_n)$$

beg the question “What other deep Clausen-like hypergeometric factorizations lurk within?”

Broadhurst and Lisoněk used one of the present authors’ implementation of PSLQ to search for Zagier generalizations. They found that “cycles” such as  $Z(m_1, m_2, \dots, m_{2n+1}) = \zeta(\{2\}_{m_1}, 3, \{2\}_{m_2}, 1, \{2\}_{m_3}, 3, \dots, 1, \{2\}_{m_{2n+1}})$  participate in many such identities. Checking PSLQ input vectors from all  $Z$  values of fixed weight ( $2K$ , say) along with the value  $\zeta(\{2\}_K)$  detected many identities, from which general patterns were “obvious.” This led to a *conjecture* (among many):

$$\sum_{i=0}^{2n} Z(C^i S) \stackrel{?}{=} \zeta(\{2\}_{M+2n}) \quad (3.8.59)$$

for  $S$  a string of  $2n + 1$  numbers summing to  $M$ , and  $C^i S$  the cyclic shift of  $S$  by  $i$  places. Zagier’s identity is the case of (3.8.59) with entries of  $S$  zero.

The symmetry in (3.8.59) highlighted that Zagier-type identities have serious combinatorial content. For  $M = 0, 1$  we could reduce (3.8.59) to evaluation of

combinatorial sums; and thence to truly combinatorial proofs. For  $M \geq 2$  we have no proofs, but very strong evidence.

Perhaps the most striking conjecture (open indeed for  $n > 2$ ) is tantalizingly easy to state, and to numerically verify, and has eluded proof for five years since its numerical discovery:

$$8^n \zeta(\{-2, 1\}_n) \stackrel{?}{=} \zeta(\{2, 1\}_n), \quad (3.8.60)$$

or equivalently that the functions

$$L_{-2,1}(1, 2t) = L_{2,1}(1, t) \quad (= L_3(1, t))$$

agree for small  $t$ . It appears to be the unique identification of an Euler sum with a distinct MZV of its type. Can just the case  $n = 2$  be proven symbolically as is the case for  $n = 1$  (see Exercise 16)?

To sum up, our simplest conjectures (on the number of irreducibles) are still beyond present proof techniques. Is  $\zeta(5)$  or  $G$  rational? Such questions may or may not be close to proof! Thus, the field appears wide open for numerical exploration.

Dimensional conjectures sometimes involve finding integer relations between hundreds of quantities and so demanding precision of thousands of digits—often of hard to compute objects. In that vein, one of the present authors and Broadhurst recently found a *polylogarithmic ladder* of length 17 (a record) with such “ultra-PSLQing” [19].

### 3.9 Multi-Clausen Values

We finish this chapter by returning briefly to binomial sums first detailed in Section 1.7. The study of so called *Deligne words* for multiple integrals generating *Multiple Clausen (or Multi-Clausen) Values* at  $\pi/3$  such as

$$\mu(a, b) = \sum_{n>m>0} \frac{\sin(n\frac{\pi}{3})}{n^a m^b},$$

seem quite fundamental. It leads to results like

$$S_3 = \sum_{k=1}^{\infty} \frac{1}{k^3 \binom{2k}{k}} = \frac{2\pi}{3} \mu(2) - \frac{4}{3} \zeta(3),$$

$$S_5 = \sum_{k=1}^{\infty} \frac{1}{k^5 \binom{2k}{k}} = 2\pi\mu(4) - \frac{19}{3}\zeta(5) + \frac{2}{3}\zeta(2)\zeta(3),$$

$$S_6 = \sum_{k=1}^{\infty} \frac{1}{k^6 \binom{2k}{k}} = -\frac{4\pi}{3}\mu(4, 1) + \frac{3341}{1296}\zeta(6) - \frac{4}{3}\zeta(3)^2.$$

Much more is detailed in [54]. This includes a generalization of MZV duality, and finishes with an accounting of alternating sums:

$$A_N = \sum_{k=1}^{\infty} \frac{(-1)^{k+1}}{k^N \binom{2k}{k}}.$$

In this setting, it is more convenient to work with the set

$$\mathcal{K}_k = \{L_k(\rho^p) \mid p \in \mathcal{C}\}$$

where  $\rho = (\sqrt{5} - 1)/2$ , of *Kummer-type polylogarithms*, of the form

$$L_k(x) = \frac{1}{(k-1)!} \int_0^x \frac{(-\log|y|)^{k-1} dy}{1-y} = \sum_{r=0}^{k-1} \frac{(-\log|x|)^r}{r!} \text{Li}_{k-r}(x)$$

where as before  $\text{Li}_k(x) = \sum_{n>0} x^n/n^k$ .

$$\tilde{L}_k(x) = L_k(x) - L_k(-x) = 2L_k(x) - 2^{1-k}L_k(x^2).$$

Then  $A_3$  is Apéry's sum,  $A_5$  was expressed in (3.2.18), and

$$A_4 = 4\tilde{L}_4(\rho) - \frac{1}{2}L^4 - 7\zeta(4).$$

In fact, there are five integer relations between  $\mathcal{K}_4$ ,  $L^4$ ,  $\zeta(4)$  and  $A_4$ . Another simple example is

$$A_4 = \frac{16}{9}\tilde{L}_4(\rho^3) - 2L^4 - \frac{23}{9}\zeta(4).$$

For  $9 \geq N \geq 6$  one gets corresponding integer relations but not enough to obtain closed forms for  $A_N$ .

### 3.10 Commentary and Additional Examples

1. **A binomial generalization of zeta.** Show that

$$\sum_{k=0}^{\infty} \frac{\binom{2k}{k}}{\binom{2k+4n}{k+2n} (2k+4n+1)(2k+2n+1)} = \binom{2n}{n}^2 \frac{\pi^2}{2^{8n+3}}, \quad (3.10.61)$$

for each integer  $n \geq 0$ . Hint: This is a sum that *Maple* can evaluate in closed form. Reason: The summand on the left of (3.10.61) equals

$$\begin{aligned} & \frac{(2k)!(k+2n)!^2}{k!^2(2k+4n)!(2k+4n+1)(2k+2n+1)} \\ &= \frac{[(k+1)(k+2)\cdots(k+2n)]^2}{(2k+2n+1)[(2k+1)(2k+2)\cdots[2k+4n+1]]} \\ &= \frac{(k+1)(k+2)\cdots(k+2n)}{2^{2n}(2k+2n+1)[(2k+1)(2k+3)\cdots(2k+4n+1)]} = f_n(k), \end{aligned}$$

where  $f_n$  is a rational function with numerator and denominator of degrees  $2n$  and  $2n+2$  respectively. The partial fraction expansion of  $f_n$  is thus

$$f_n(x) = \frac{a}{(2x+2n+1)^2} + \sum_{i=0}^{2n} \frac{b_i}{(2x+2i+1)}.$$

But  $f_n(x) = -f_n(-2n-1-x)$  and it follows that  $b_i = -b_{2n-i}$  for each  $i$ . In particular  $b_n = 0$ . Also

$$\begin{aligned} a &= \lim_{x \rightarrow -n-1/2} (2x+2n+1)^2 f_n(x) \\ &= \frac{(-n+1/2)(-n+3/2)\cdots(n-1/2)}{2^{2n}(-2n)(-2n+2)\cdots(-2)\cdot 2\cdot 4\cdots(2n)} \\ &= \frac{[1\cdot 3\cdots(2n-1)]^2}{2^{6n}n!^2} = \frac{(2n)!^2}{2^{8n}n!^4} = \frac{1}{2^{8n}} \binom{2n}{n}^2. \end{aligned}$$

For  $-n \leq k < 0$ ,  $f_n(k) = 0$  so that the sum in question equals

$$\begin{aligned} \sum_{k=-n}^{\infty} f_n(k) &= \frac{1}{2^{8n}} \binom{2n}{n}^2 \sum_{k=-n}^{\infty} \frac{1}{(2n+2k+1)^2} \\ &\quad + \sum_{i=0}^{n-1} b_i \sum_{k=-n}^{\infty} \left( \frac{1}{2k+2i+1} - \frac{1}{2k+4n-2i+1} \right). \end{aligned}$$



The sum

$$\sum_{k=-n}^{\infty} \left( \frac{1}{2k+2i+1} - \frac{1}{2k+4n-2i+1} \right)$$

telescopes and equals

$$\sum_{k=-n}^{n-2i-1} \frac{1}{2k+2i+1} = 0$$

since in this sum the  $k$ -th term cancels with the  $-2i-k-1$ -st term. As

$$\sum_{k=-n}^{\infty} \frac{1}{(2n+2k+1)^2} = \sum_{j=0}^{\infty} \frac{1}{(2j+1)^2} = \frac{\pi^2}{8},$$

the identity (3.10.61) follows.

2. **The Riemann-Siegel formula.** In this exercise we derive the principle term of an asymptotic formula which Riemann used to compute the roots of the zeta function,  $\zeta(s) = \sum_{n=1}^{\infty} n^{-s}$  ( $s \in \mathbb{C}$ ), whenever this sum converges. We then use the asymptotic formula to find roots of  $\zeta(s)$ .

- (i) We first derive an integral representation of  $\zeta(s)$  which remains valid for all  $s \in \mathbb{C}$ , as Riemann states without proof in his famous 1859 paper [182]. Use the Mellin transform

$$\int_0^{\infty} e^{-nx} x^{s-1} dx = \frac{\Gamma(s-1)}{n^s}$$

together with the Gamma function identities from Chapter 5 of the first volume

$$\Gamma(s) = s\Gamma(s-1), \quad \frac{\pi s}{\Gamma(s)\Gamma(-s)} = \sin \pi s, \quad (3.10.62)$$

and the standard identity  $e^{-Nx}(e^x - 1)^{-1} = \sum_{n=N+1}^{\infty} e^{-nx}$  to show that

$$\zeta(s) = \sum_{n=1}^N n^{-s} + \Gamma(-s)\pi^{-1} \sin(\pi s) \int_0^{\infty} \frac{x^{(s-1)} e^{-Nx}}{e^x - 1} dx. \quad (3.10.63)$$

Now, using the residue theorem and the identities (3.10.62), show that

$$\begin{aligned} \zeta(s) = \sum_{n=1}^N n^{-s} + \Gamma(-s)(2\pi)^{s-1} 2 \sin \frac{\pi s}{2} \sum_{n=1}^N n^{-(1-s)} \\ + \frac{\Gamma(-s)}{2\pi i} \int_{C_N} \frac{(-x)^s e^{-Nx}}{x(e^x - 1)} dx, \end{aligned} \quad (3.10.64)$$

where  $C_N$  is the contour whose path descends the real axis (or, technically, just above the real axis) from  $+\infty$ , traces the circle of radius  $2\pi(N+1)$  counter clockwise and returns to  $+\infty$ , and where  $(-x)^s = \exp[s \log(-x)]$  is defined to be the branch which is real for positive real  $x$ .

(ii) Next consider the auxiliary function  $Z(t)$  defined by

$$Z(t) = e^{i\theta(t)} \zeta(1/2 + it), \quad \theta(t) = (i \log(\Gamma(it/2 - 3/4)) - t/2 \log \pi).$$

Define  $\xi(s) = \Gamma(s/2)(s-1)\pi^{-s/2}\zeta(s)$ . Again using the identity (3.10.62), show that,  $\xi(1/2 + it) = r(t)Z(t)$ , where

$$r(t) = -e^{\operatorname{Re} \log \Gamma(it/2 - 3/4)} \frac{t^2 + 1/4}{2\pi^{1/4}}.$$

Substitute the integral expression (3.10.63) for  $\zeta(s)$  into the definition of  $\xi(s)$  above and use the identities  $\theta(-t) = -\theta(t)$ ,  $r(-t) = r(t)$ , and  $2i \sin(\pi s/2) = -e^{i\pi/4} e^{t\pi/2} (1 - ie^{-t\pi})$  to show that  $Z(t) = Z_0(t) + R(t)$  where

$$Z_0(t) = \sum_{n=1}^N n^{-1/2} 2 \cos(\theta(t) - t \log n), \quad (3.10.65)$$

and

$$R(t) = \frac{e^{-i\theta(t)} e^{-t\pi/2}}{(2\pi)^{1/2} (2\pi)^{it} e^{-i\pi/4} (1 - ie^{-t\pi})} \int_{C_N} \frac{(-x)^{it-1/2} e^{-Nx}}{e^x - 1} dx. \quad (3.10.66)$$

To evaluate the remainder term  $R$ , Riemann used the asymptotic expansion

$$R(t) \approx \tilde{R}(t) = (-1)^{N-1} \left(\frac{t}{2\pi}\right)^{-1/4} \sum_j \left(\frac{t}{2\pi}\right)^{-j/2} C_j(p), \quad (3.10.67)$$

where the integer  $N$  is chosen to be the integer part of  $(t/2\pi)^{1/2}$ , and  $p$  is the fractional part. The first three terms of the expansion are

$$C_0(p) = \Psi(p) = \frac{\cos(2\pi(p^2 - p - \frac{1}{16}))}{\cos(2\pi p)}, \quad C_1(p) = -\frac{1}{2^5 3 \pi^2} \Psi^{(3)}(p),$$

$$C_2(p) = \frac{1}{2^{11} 3^4 \pi^4} \Psi^{(6)}(p) + \frac{1}{2^6 \pi^2} \Psi^{(2)}(p).$$

The notation  $\Psi^{(n)}$  indicates the  $n$ th derivative. The expansion (3.10.67) is the Riemann-Siegel formula, so named for its originator and the mathematician Carl Siegel [190] who discovered the formula in Riemann's working papers some 70 years after the publication of Riemann's original paper. If the  $C_1$  term is the first term omitted in the Riemann-Siegel formula, then Titchmarsh [201, pg. 331] showed that for  $t > 250\pi$  the error  $|R(t) - \tilde{R}(t)|$  is bounded by  $(3/2)(t/2\pi)^{(-3/4)}$ . The formula refined is still in use today. See also [86] for an exercise (1.59) that lead the reader through zeta calculations, including some explicit parts on Riemann-Siegel.

- (iii) Using the Riemann-Siegel formula and Titchmarsh's estimate for the error, give a numerical proof for the existence of zeros of  $\zeta(1/2 + it)$  in the interval  $t \in [999.784, 999.799]$ . Note that only 12 terms in the main sum (3.10.65) are needed to calculate the estimate for  $Z(t)$ . To achieve comparable accuracy using the alternative Euler-Maclaurin formula (see [109, Ch. 6] or Chapter 7 in this volume) would require hundreds of terms.

3. **Some quadratic zeta functions.** Problem: Explore evaluations of sums of the form

$$\zeta(a, b, c) = \sum_{n, m \geq 0} \frac{n^{2a} m^{2b}}{(n^2 + m^2)^c}$$

for  $a, b, c$  nonnegative integers. (Here as before the summation avoids poles of the summand.)

Solution:

- (a) The following identity expresses  $\zeta(a, 0, c)$  by using a Bessel function expansion of the normalized Mellin transform

$$M_c(f) = \frac{1}{\Gamma(c)} \int_0^\infty f(t)t^{c-1} dt$$

of the function

$$t \rightarrow \sum_{n,m} n^{2a} q^{n^2+m^2} = \sum_n n^{2a} q^{n^2} \theta_3(q)$$

with  $q = \exp(-t)$  after using the theta transform (2.3.15)

$$\theta_3(\exp(\pi t)) = \sqrt{\frac{\pi}{t}} \theta_3\left(\exp\left(\frac{\pi}{t}\right)\right).$$

This leads to the identity

$$\zeta(a, 0, c) = 2\delta_{0a}\zeta(2c) + 2\beta\left(c - \frac{1}{2}, \frac{1}{2}\right)\zeta(2c - 2a - 1) + 4\sum_{p \geq 1} \sigma_{[2a+1-2c]}(p)\mathcal{E}_c(p). \quad (3.10.68)$$

This is valid for real  $ac$  with  $d = c - a > 1$  and presumably provides an analytic continuation of  $\zeta(a, 0, c)$ , for the sum over positive integers. Here  $\sigma$  is a divisor function

$$\sigma_{[d]}(p) = \sum_{n|p} n^d,$$

and

$$\mathcal{E}_c(p) = \frac{\sqrt{\pi}}{\Gamma(c)} 2(\pi p)^{c-1/2} K_{(c-1/2)}(2\pi p)$$

is derived from the modified Bessel function of half integer order,  $K_{(c-1/2)}$ , of the second kind.

When  $c = N$  is integer,  $\mathcal{E}_N(p)$  is of the form

$$\pi \exp(-2\pi p) P_N(\pi p)$$

where  $P_N$  is a rational polynomial of degree  $N - 1$  with positive coefficients:  $P_1(x) = 1, P_2(x) = x + 1/2, P_3(x) = x^2/2 + 3/4x + 3/8, P_4(x) = 1/6x^3 + 1/2x^2 + 5/8x + 5/16$ . In general

$$P_N(x) = \sum_{k=0}^{N-1} \frac{\binom{N+k-1}{N-1} x^{N-1-k}}{(N-k-1)!4^k}.$$

This allows one to very efficiently compute  $\zeta(a, 0, c)$  via (3.10.68), using roughly  $D/4$  terms for  $D$  digits. Note also that for fixed  $c$  and variable  $a$  only the powers in  $\sigma_{[2a+1-2c]}$  vary so most of the computation can be saved.

(b) Then the general integer case follows from

$$\zeta(a, b, c) = \sum_{k=0}^e (-1)^{e-k} \binom{e}{k} \zeta(a + b - k, 0, c - k)$$

where  $e = \min(a, b)$ .

(c) Similar developments are possible for the more general form

$$\zeta_N(a, b, c) = \sum'_{n,m} \frac{n^{2a} m^{2b}}{(Nn^2 + m^2)^c},$$

with  $N > 0$ . We write  $\zeta = \zeta_1$ . For example,

$$\zeta_N(a, 0, c) = 2 \delta_{0a} \zeta(2c) + N^{(1/2-c)} \left[ 2\beta \left( c - \frac{1}{2}, \frac{1}{2} \right) \zeta(2c - 2a - 1) + 4 \sum_{p \geq 1} \sigma_{[2a+1-2c]}(p) \mathcal{E}_c(\sqrt{N}p) \right]. \quad (3.10.69)$$

Note that we now lose a symmetry (and apparently have many fewer closed forms) but have:  $\zeta_{1/N}(a, b, c) = N^c \zeta_N(b, a, c)$ .

Also, for  $N = 1, 2, 3$ , and especially for those with *disjoint discriminants*, many special values may be computed via elliptic integrals in the corresponding singular values. This leads to closed forms such as

$$\zeta(0, 0, c) = 4\zeta(c)L_{-4}(c),$$

and

$$\zeta_2(0, 0, c) = 2\zeta(c)L_{-8}(c)$$

where  $L_\sigma(c) = \sum_{n \geq 1} \left(\frac{\sigma}{n}\right) n^{-c}$  is the corresponding primitive L-series and  $\left(\frac{\sigma}{n}\right)$  is the Legendre-Jacobi symbol. We obtain

$$\zeta(2, 0, 4) = \frac{1}{4} \pi^2 G + \frac{1}{480} \frac{\pi^6}{\Gamma(3/4)^8},$$

and

$$\zeta(2, 2, 6) = \frac{1}{64} \pi^2 G - \frac{1}{1920} \frac{\pi^6}{\Gamma(3/4)^8} + \frac{1}{92160} \frac{\pi^{10}}{\Gamma(3/4)^{16}}$$

where  $G = L_{-4}(2)$  is Catalan's constant. There is a similar expression for  $\zeta(a, b, a + b + 2)$  for all integers  $a, b \geq 0$ .

Then also

$$\zeta_2(1, 1, 4) = \frac{1}{8} \zeta(2) L_{-8}(2) - \frac{1}{18} (3 - 2\sqrt{2}) \left(\frac{1}{8} \beta\left(\frac{1}{8}, \frac{1}{8}\right)\right)^4,$$

and there is a similar evaluation of  $\zeta_P(1, 1, 4)$  and  $\zeta_{2P}(1, 1, 4)$  when  $P$  is respectively of "type 1"  $((1), 5, 13, 21, 33 \dots)$  or "type 2"  $(1, 3, 5, 11, 15, \dots)$ , as described in section 9.2 of [44]. Hence,

$$\zeta_6(0, 0, c) = \zeta(c) L_{-24}(c) + L_{-3}(c) L_8(c)$$

and

$$\begin{aligned} \zeta_6(1, 1, 4) &= \frac{1}{12} \zeta_6(0, 0, 2) \\ &\quad - \frac{1}{15} 2^{1/3} (35 + 16\sqrt{3} - 20\sqrt{2} - 14\sqrt{6}) \left(\frac{1}{24} \beta\left(\frac{1}{24}, \frac{1}{24}\right)\right)^4, \end{aligned}$$

while

$$\begin{aligned} \zeta_{10}(1, 1, 4) &= \frac{1}{80} (\zeta(2) L_{-40}(2) + L_5(2) L_{-8}(2)) - \\ &\quad \left(7725 + 3452\sqrt{5} - 5460\sqrt{2} - 2442\sqrt{10}\right) \frac{\left(\frac{1}{40}\beta\right)^4 \left(\frac{1}{40}, \frac{1}{40}\right) \beta^4\left(\frac{9}{40}, \frac{9}{40}\right)}{120 \beta^4\left(\frac{3}{8}, \frac{3}{8}\right)}. \end{aligned}$$

For comparison, we note that we may also write

$$\zeta_1(1, 1, 4) = \frac{1}{2} \zeta(2) L_{-4}(2) - \frac{1}{30} \left(\frac{1}{4} \beta\left(\frac{1}{4}, \frac{1}{4}\right)\right)^4,$$

and may use elliptic transformation formulae to derive

$$\zeta_4(1, 1, 4) = \frac{1}{32} \zeta(2) L_{-4}(2) - \frac{11}{64} \zeta_1(1, 1, 4).$$

Hence,

$$\begin{aligned} \sum'_{n,m} (-1)^n \frac{n^2 m^2}{(n^2 + m^2)^4} &= 8 \zeta_4(1, 1, 4) - \zeta_1(1, 1, 4) \\ &= \frac{1}{16} \zeta(2) L_{-4}(2) + \frac{1}{80} \left( \frac{1}{4} \beta \left( \frac{1}{4}, \frac{1}{4} \right) \right)^4, \end{aligned}$$

and

$$\sum'_{n,m} (-1)^{n+m} \frac{n^2 m^2}{(n^2 + m^2)^4} = 4 \sum'_{n,m} (-1)^n \frac{n^2 m^2}{(n^2 + m^2)^4}.$$

Note also that  $L_5(2) = 4\sqrt{5}\pi^2/125$  and  $L_8(2) = \sqrt{2}\pi^2/16$ .

4. **A multi-zeta evaluation.** Consider

$$\sigma_{n,m} = \sum_{s \in \mathcal{S}(n,m)} \zeta(s_1, s_2, \dots, s_m)$$

summed over all strings of length  $m$  consisting of nonnegative integers adding up to  $n$ , with  $s_1 > 1$  to insure convergence. Determine  $\sigma_{n,m}$  as a multiple of  $\zeta(n)$ . Hint: Note that  $\zeta(2, 1) = \zeta(3)$ ,  $\zeta(3, 1) + \zeta(2, 2) = \zeta(4)$ , etc. After working out a few more examples, the pattern can easily be observed. See also Item 11.

5. **Harmonic numbers.** Prove that

$$H_n = \sum_{k=1}^n \frac{1}{k}$$

is never integer for  $n > 1$ . Hint: A slick proof uses Bertrand's postulate: the proven fact that *there is always a prime  $p$  in the interval  $(n/2, n]$* . Now write

$$H_n = \frac{n!/1 + n!/2 + \dots + n!/n}{n!}.$$

Then  $p$  divides the numerator and all but one term of the denominator  $n!/p$  (as  $p > n/2$ ). One can alternately establish the result using more elementary methods.

6. **Calabi's unifying integral.** It is possible to unify the L-series evaluations in terms of Bernoulli numbers (3.2.13) and Euler numbers (3.3.26), as follows. Consider, for integer  $N \geq 1$

$$\sigma(N) = \sum_{k=-\infty}^{\infty} \frac{1}{(4k+1)^N}.$$

- (a) Show that the ordinary generating function of  $\sigma(N)$  is

$$\Sigma(z) = \frac{\pi z}{4} \left( \sec\left(\frac{\pi z}{2}\right) + \tan\left(\frac{\pi z}{2}\right) \right),$$

and obtain the explicit relations between the Euler numbers, the Bernoulli numbers, and  $\sigma(N)$ .

- (b) Show that  $\sigma(1) = \int_0^1 1/(1+x^2) dx = \arctan(1) = \pi/4$ , and that

$$\sigma(2) = \sum_{k=-\infty}^{\infty} \frac{1}{(4k+1)^2} = \sum_{k=0}^{\infty} \frac{1}{(2k+1)^2} = \int_0^1 \int_0^1 \frac{dx dy}{1-(xy)^2} = \frac{\pi^2}{8}.$$

(See also Exercise 11, parts (a) and (b), of Chapter 2.)

- (c) More generally for  $N > 1$ , show that

$$\sigma(N) = \sum_{k=0}^{\infty} \frac{(-1)^{kN}}{(2k+1)^N} = \int_0^1 \cdots \int_0^1 \frac{dx_1 \cdots dx_N}{1 \pm (x_1 \cdots x_N)^2} = \left(\frac{\pi}{2}\right)^N \nu(N),$$

where  $\nu(N)$  is the volume, necessarily rational, of the polytope

$$P_N = \{(v_1, v_2, \dots, v_N) : v_k \geq 0, v_k + v_{k+1} \leq 1, (1 \leq k \leq N)\}.$$

One way to obtain the volume is to use, and justify, the (one-to-one) change of variables

$$x_k = \frac{\sin u_k}{\cos u_{k+1}} \quad (1 \leq k \leq N),$$

where we set  $u_{N+1} = u_1, v_{N+1} = v_1$ .



This and more may be explored in Elkies' recent article [110]. It is a nice computer algebra challenge to symbolically invert the substitution for  $N = 2, 3, 4$ .

7. **A multi-dimensional polylogarithm extension.** A useful specialization of the general multidimensional polylogarithm, which is at the same time an extension of the polylogarithm, is the case in which each  $b_j = b$ . Under these circumstances, we write

$$\lambda_b(s_1, \dots, s_k) = \sum_{\nu_1, \nu_2, \dots, \nu_k \geq 1} \prod_{j=1}^k b^{-\nu_j} \left( \sum_{i=j}^k \nu_i \right)^{-s_j}. \quad (3.10.70)$$

When  $b = \pm 1$  this is an Euler sum.

Let  $|p| \geq 1$ . The double generating function equality

$$1 - \sum_{m=0}^{\infty} \sum_{n=0}^{\infty} x^{m+1} y^{n+1} \lambda_p(m+2, \{1\}_n) = F\left(y, -x; 1-x; \frac{1}{p}\right) \quad (3.10.71)$$

holds. Note that when  $p = 1$  the symmetry of the hypergeometric function produces a case of MZV duality:  $\zeta(m+2, \{1\}_n) = \zeta(n+2, \{1\}_m)$ , for all  $m$  and  $n$ , because

$$\begin{aligned} F(y, -x; 1-x; 1) &= \frac{\Gamma(1-x)\Gamma(1-y)}{\Gamma(1-x-y)} \quad (3.10.72) \\ &= \exp\left\{ \sum_{k=2}^{\infty} (x^k + y^k - (x+y)^k) \frac{\zeta(k)}{k} \right\}. \end{aligned}$$

Expanding the rightmost function gives a closed form for  $\zeta(m+2, \{1\}_n)$ .

**Proof.** (of (3.10.71)) By definition of  $\lambda_p$

$$\begin{aligned}
 \sum_{m=0}^{\infty} \sum_{n=0}^{\infty} x^{m+1} y^{n+1} \lambda_p(m+2, \{1\}_n) &= y \sum_{m=0}^{\infty} x^{m+1} \sum_{k=1}^{\infty} \frac{1}{k^{m+2} p^k} \prod_{j=1}^{k-1} \left(1 + \frac{y}{j}\right) \\
 &= \sum_{m=0}^{\infty} x^{m+1} \sum_{k=1}^{\infty} \frac{(y)_k}{k^{m+1} k! p^k} \\
 &= \sum_{k=1}^{\infty} \frac{(y)_k}{k! p^k} \left(\frac{x}{k-x}\right) \\
 &= - \sum_{k=1}^{\infty} \frac{(y)_k (-x)_k}{k! p^k (1-x)_k} \\
 &= 1 - F\left(y, -x; 1-x; \frac{1}{p}\right)
 \end{aligned}$$

as claimed. □

8. **A symbolic multi-dimensional zeta evaluation.** Use

$$1 - \exp \sum_{k=2}^{\infty} \zeta(k) \frac{(x^k + y^k - (x+y)^k)}{k}$$

in (3.10.72) to compute  $\zeta(n, \{1\}_m)$  symbolically for  $n+m < 9$ .

9. **Three proofs of an identity.**

$$\sum_{n>0} \frac{H_n^2}{n^2} = \frac{17}{4} \zeta(4).$$

Here, as before  $H_n = \sum_{k=1}^n 1/k$ .

(a) *Fourier analysis proof.* As in Chapter 2 of this volume, obtain the Fourier series of the function whose square integral is given by

$$\frac{1}{2\pi} \int_0^\pi (\pi - t)^2 \log^2(2 \sin \frac{t}{2}) dt = \sum_{n=1}^{\infty} \frac{(\sum_{k=1}^n \frac{1}{k})^2}{(n+1)^2}. \quad (3.10.73)$$

(b) *Algebraic proof.* Write

$$\begin{aligned}
 \sum_{n>0} \frac{H_n^2}{n^2} &= \sum_{n>0} \frac{\left(H_{n-1} + \frac{1}{n}\right)^2}{n^2} && (3.10.74) \\
 &= \sum_{n>0} \frac{H_{n-1}^2}{n^2} + 2\zeta(3, 1) + \zeta(4) \\
 &= 2\zeta(2, 1, 1) + \zeta(4) + 2\zeta(3, 1) + \zeta(4) \\
 &= 2\zeta(2, 1, 1) + \zeta(4) + 2\zeta(3, 1) + \zeta(4) = \left(4 + \frac{1}{4}\right) \zeta(4),
 \end{aligned}$$

since  $\zeta(2, 1, 1) = \zeta(4)$  (by MZV duality or the  $\zeta(m+2, \{1\}_n)$  special case), and  $4\zeta(3, 1) = \zeta(4)$  (by the first case of Zagier's evaluation or by the double Euler sum evaluation).

(c) *Residue theory proof.* Apply residue theory to

$$\phi_{p,q}(s) = \frac{\pi \cot(\pi s) \Psi^{(p-1)}(-s)}{2 s^q (p-1)!}$$

obtain  $\zeta(p, q)$  for  $p+q < 8$ , by integrating over circles of radius  $R \rightarrow \infty$  centered at the origin [112].

10. **Reduction to zeta values.** Reduce

$$\sum_{n>0} \frac{H_n^3}{n^3}$$

and

$$\sum_{n>0} \frac{H_n^2}{n^4}$$

to Riemann zeta values.

11. **The Ohno duality theorem.** Ohno [169] provides the following beautiful generalization of MZV duality (see Theorem 3.6.1). Define, for integers  $k_1, k_2, \dots, k_{n-1} \geq 1, k_n \geq 2, l \geq 0$ ,

$$Z(k_1, k_2, \dots, k_n; l) = \sum_{\substack{c_1+c_2+\dots+c_n=l \\ c_i \geq 0}} \zeta(k_1 + c_1, k_2 + c_2, \dots, k_n + c_n).$$

Also define, for integer  $a_1, b_1, a_2, b_2, \dots, a_s, b_s \geq 1$ , the dual sequences

$$\mathbf{k} = (\{1\}_{a_1-1}, b_1 + 1, \{1\}_{a_2-1}, b_2 + 1, \dots, \{1\}_{a_s-1}, b_s + 1)$$

and

$$\mathbf{k}' = (\{1\}_{b_s-1}, a_s + 1, \{1\}_{b_{s-1}-1}, a_{s-1} + 1, \dots, \{1\}_{b_1-1}, a_1 + 1).$$

Then for all such  $\mathbf{k}$ ,  $\mathbf{k}'$  and  $l$ ,

$$Z(\mathbf{k}, l) = Z(\mathbf{k}'; l). \quad (3.10.75)$$

- (a) Recover the MZV duality result from (3.10.75) with  $l = 0$ .
- (b) Apply (3.10.75) with  $k = n + 1$  and  $\mathbf{k}' = (\{1\}_{n-1}, 2)$  to obtain an evaluation of all legal  $\zeta$ -values of length  $n$  with weights summing to  $l$ .
- (c) Deduce

$$\frac{1}{\Gamma(s)} \int_0^\infty \frac{t^s - 1}{e^t - 1} \text{Li}_k(1 - e^{-t}) dt = \zeta(k + 1, \{1\}_{s-1}).$$

12. **The Ohno-Zagier generating function.** Ohno and Zagier provide the following impressive generating function. For multi-indices  $\mathbf{k} = (k_1, k_2, \dots, k_n)$  with  $k_i > 0$ , let  $I_0(k, n, s)$  denote those *admissible multi-indices* of weight  $k$ , depth  $n$  and height  $s = \#\{i : k_i > 1\}$ . Let

$$G_0(k, n, s) = \sum_{\mathbf{k} \in I_0(k, n, s)} \zeta(\mathbf{k}).$$

Note that  $I_0(k, n, s)$  is nonempty exactly if  $s > 0, n \geq s$ , and  $k \geq n + s$ .

Denote the generating function

$$\Phi_0(x, y, z) = \sum_{k, n, s} G_0(k, n, s) x^{k-n-s} y^{n-s} z^{s-1}.$$

Then

$$(xy - z) \Phi_0(x, y, z) = \left( 1 - \exp \left( \sum_{n>1} (x^n + y^n - \alpha^n - \beta^n) \frac{\zeta(n)}{n} \right) \right), \quad (3.10.76)$$

where  $\alpha, \beta$  are the roots of  $t^2 - t(x + y) = z$ .

- (a) Deduce that all the coefficients  $G_0(k, n, s)$  of  $\Phi_0$  are polynomials in  $\zeta(2), \zeta(3), \dots$  with rational coefficients.
- (b) Show (3.10.76) is equivalent to

$$1 - (xy - z) \Phi_0(x, y, z) = \prod_{m \geq 1} \left( 1 - \frac{xy - z}{(m-x)(m-y)} \right).$$

13. **Bertrand's postulate.** Show that  $\binom{2n}{n}$  is even for  $n > 0$  and use this to prove Bertrand's postulate that there is always a prime between  $n$  and  $2n$ .
14. **MZV stuffles.** Define a binary operation mapping pairs of ordered lists  $u = (u_1, \dots, u_m)$  and  $v = (v_1, \dots, v_n)$  (for non-negative integers  $m$  and  $n$ ) into multisets of ordered lists by the recursion

$$\begin{cases} () * u &= u * () = \{u\}, \\ au * bv &= a(u * bv) \cup b(au * v) \cup (a+b)(u * v), \end{cases}$$

where, for example,  $au = (a, u_1, \dots, u_m)$  and more generally, if  $M$  is a multiset of ordered lists, then  $aM$  denotes the multiset obtained by placing  $a$  at the front of each list in  $M$ .

- (a) Show that

$$\zeta(u)\zeta(v) = \sum_{w \in u*v} \zeta(w).$$

- (b) Let  $f(|u|, |v|)$  denote the number of lists (counting multiplicity) in  $u * v$ . Show that the formal power series identity

$$\sum_{m=0}^{\infty} \sum_{n=0}^{\infty} f(m, n) x^m y^n = \frac{1}{1 - x - y - xy}$$

holds.

(c) Hence, show [67] that

$$\begin{aligned} f(m, n) &= \sum_{k=0}^m \binom{m}{k} \binom{n+k}{m} = \sum_{k=0}^{\min(m, n)} \binom{n}{k} \binom{m}{k} 2^k \\ &= \left| \left\{ (b_1, \dots, b_m) \in \mathbb{Z}^m : \sum_{j=1}^m |b_j| \leq n \right\} \right| \\ &= \left| \left\{ (b_1, \dots, b_n) \in \mathbb{Z}^n : \sum_{j=1}^n |b_j| \leq m \right\} \right|. \end{aligned}$$

[The proof that the two sets of lattice points are the same size is an exercise in Pólya-Szegő's *Problems and Theorems in Analysis*.] There is an analogous, but simpler result to (a) for integral "shuffles."

15. **Extensions of Zagier's identity.** We sketch some quite broad extensions of the method of Theorem 3.7.56. Let  $f$  and  $g$  be differentiable univariate functions, and define differential operators  $D_f = f(x)d/dx$ ,  $D_g = g(x)d/dx$ . Fix a constant  $t$  and let  $U$  and  $V$  be sets of solutions to the respective differential equations

$$(D_f D_g - t)u = 0, \quad (D_f D_g + t)v = 0.$$

- (a) Prove [69, 67] that  $UV = \{uv : u \in U, v \in V\}$  is a set of solutions to the differential equation

$$(D_f^2 D_g^2 + 4t^2)w = 0,$$

and moreover [68],  $UV$  is a basis iff  $U$  and  $V$  are bases for the solution spaces of their respective equations.

Hint: With obvious notation, prove the modified Wronskian determinant identity [68]

$$\begin{vmatrix} u_1 v_1 & u_2 v_1 & u_1 v_2 & u_2 v_2 \\ D_g u_1 v_1 & D_g u_2 v_1 & D_g u_1 v_2 & D_g u_2 v_2 \\ D_g^2 u_1 v_1 & D_g^2 u_2 v_1 & D_g^2 u_1 v_2 & D_g^2 u_2 v_2 \\ D_f D_g^2 u_1 v_1 & D_f D_g^2 u_2 v_1 & D_f D_g^2 u_1 v_2 & D_f D_g^2 u_2 v_2 \end{vmatrix}$$

$$= 8t \left| \begin{matrix} u_1 & u_2 \\ D_g u_1 & D_g u_2 \end{matrix} \right|^2 \left| \begin{matrix} v_1 & v_2 \\ D_g v_1 & D_g v_2 \end{matrix} \right|^2.$$

(b) Generalize this result.

(c) **Applications:** For real  $x$  with  $0 \leq x \leq 1$ , positive integers  $s_j$ , and signs  $\sigma_j \in \{1, -1\}$ , let

$$\zeta_x(\sigma_1 s_1, \dots, \sigma_k s_k) = \sum_{n_1 > \dots > n_k > 0} x^{n_1} \prod_{j=1}^k n_j^{-s_j} \sigma_j^{n_j},$$

and set  $\zeta = \zeta_1$ . For  $0 \leq x \leq 1$  and complex  $z$ , let

$$\begin{aligned} Y_1(x, z) &= F(z, -z; 1; x), \\ Y_2(x, z) &= (1-x)F(1+z, 1-z; 2; 1-x), \\ G(z) &= \frac{1}{4} \{ \psi(1+iz) + \psi(1-iz) - \psi(1+z) - \psi(1-z) \}, \end{aligned}$$

where  $F$  is the Gaussian hypergeometric function, and  $\psi$  is the logarithmic derivative of the Euler gamma function:  $\psi(z) = \Gamma'(z)/\Gamma(z)$ .

Then [36]

$$\sum_{n=0}^{\infty} (-1)^n z^{4n} 4^n \zeta_x(\{3, 1\}_n) = Y_1(x, z) Y_1(x, iz),$$

and [69]

$$\begin{aligned} \sum_{n=0}^{\infty} (-1)^n z^{4n+2} 4^n \zeta_x(3, \{1, 3\}_n) &= G(z) Y_1(x, z) Y_1(x, iz) \\ &\quad - \frac{Y_1(x, iz) Y_2(x, z)}{4Y_1(1, z)} + \frac{Y_1(x, z) Y_2(x, iz)}{4Y_1(1, iz)} \end{aligned}$$

define entire functions of  $z$ .

(d) Rederive that for positive integers  $n$ ,

$$\zeta(\{3, 1\}_n) = 4^{-n} \zeta(\{4\}^n) = \frac{2\pi^{4n}}{(4n+2)!}.$$

Additionally show [69] that

$$\begin{aligned}\zeta(3, \{1, 3\}_n) &= 4^{-n} \sum_{k=0}^n \zeta(4k+3) \zeta(\{4\}^{n-k}) \\ &= \sum_{k=0}^n \frac{2\pi^{4k}}{(4k+2)!} \left(-\frac{1}{4}\right)^{n-k} \zeta(4n-4k+3),\end{aligned}$$

and

$$\begin{aligned}\zeta(2, \{1, 3\}_n) &= 4^{-n} \sum_{k=0}^n (-1)^k \zeta(\{4\}^{n-k}) \left\{ (4k+1)\zeta(4k+2) \right. \\ &\quad \left. - 4 \sum_{j=1}^k \zeta(4j-1)\zeta(4k-4j+3) \right\}.\end{aligned}$$

(e) For complex  $z$ , set

$$A(z) = \sum_{n=0}^{\infty} z^n \zeta(\{-1\}^n) = \prod_{j=1}^{\infty} \left(1 + \frac{(-1)^j z}{j}\right) = \frac{\Gamma(1/2)}{\Gamma(1+z/2)\Gamma(1/2-z/2)}.$$

Let  $t$  and  $z$  be related by  $z = (1+i)t/2$ , and set  $s = (1+x)/2$ , where  $0 \leq x \leq 1$ . Define  $U(s, z) = Y_1(s, z) - zY_2(s, z)$ , where  $Y_1$  and  $Y_2$  are the Gaussian hypergeometric functions previously defined. Then [69],

$$\sum_{n=0}^{\infty} \left[ t^{2n} \zeta_x(\{-1, 1\}^n) + t^{2n+1} \zeta_x(-1, \{1, -1\}^n) \right] = \frac{U(s, -z)U(s, iz)}{A(-z)A(iz)}$$

defines an entire function of  $z$ .

(f) Conclude that for all complex  $t$ ,

$$\sum_{n=0}^{\infty} \left[ t^{2n} \zeta(\{-1, 1\}^n) + t^{2n+1} \zeta(-1, \{1, -1\}_n) \right] = A\left(\frac{t}{1-i}\right) A\left(\frac{t}{1+i}\right) \quad (3.10.77)$$

and if  $z = (1+i)t/2$ , then

$$\begin{aligned}1 + \sum_{n=0}^{\infty} \left[ t^{2n+1} \zeta(-1, \{-1, 1\}^n) + t^{2n+2} \zeta(-1, -1, \{1, -1\}_n) \right] \\ = \frac{1}{2}(1+i)zA(z)A(-iz) \left\{ \pi \csc(\pi z) - i\pi \operatorname{csch}(\pi z) + 4G(z) \right\}.\end{aligned} \quad (3.10.78)$$



- (g) Deduce explicit formulas [69] for the alternating unit Euler sums appearing as coefficients in (3.10.77) and (3.10.78).
- (h) Find additional applications of these ideas to multiple polylogarithms or other special functions.

16. **A multi-zeta identity.** Consider the power series

$$J(x) = \sum_{n_1 > n_2 > 0} \frac{x^{n_1}}{n_1^2 n_2}.$$

- (a) Show for  $0 \leq x \leq 1$  that

$$\begin{aligned} J(x) &= \int_0^x \frac{\log^2(1-t)}{2t} dt = \zeta(3) + \frac{1}{2} \log^2(1-x) \log(x) \\ &+ \log(1-x) \text{Li}_2(x) - \text{Li}_3(x), \end{aligned}$$

- (b) and that

$$J(-x) = -J(x) + \frac{1}{4} J(x^2) + J\left(\frac{2x}{x+1}\right) - \frac{1}{8} J\left(\frac{4x}{(x+1)^2}\right). \quad (3.10.79)$$

- (c) Deduce that  $J(1) = 8J(-1)$ .
- (d) Evaluate  $J(1/2)$ .

This functional equation was found, once the ingredients were determined by inspection, by evaluating (3.10.79) (actually, a version of it with undetermined coefficients) at a random point and then using LLL. Another successful strategy is to evaluate each  $J$  function at enough specific values of  $x$  to enable one to solve linear equations for the unknown coefficients.

If  $L(x)$  and  $R(x)$  denote the left-hand and the right-hand sides of (3.10.79), respectively, then computer manipulations (for  $0 < x < 1$ ) show that  $dL/dx = dR/dx$ : mechanically differentiating both sides and using `simplify` reduces the difference between the two to zero. Now this completes a proof of  $\zeta(2, 1) = 8\zeta(\bar{2}, 1) = \zeta(3)$ , see (3.8.60). Even the next case of (3.8.60) has only been established indirectly.

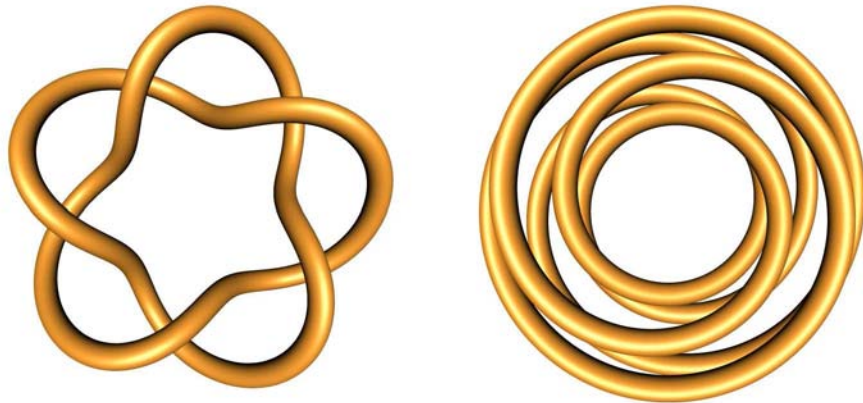


Figure 3.1: How does one identify two knots?

17. **Torus knots and zeta values.** For integers  $p, q > 1$  the  $p - q$  torus knot is the knot that transpires when string is wound  $p$  times around one way on the torus, while being wound  $q$  times in the other direction. Figure 3.1 shows the  $2 - 5$  and  $5 - 2$  torus knots in three dimensions. Despite looking very different this pair are clearly mathematically the same knot (the torus is the product of two circles and we just exchange generators.) There is, along the lines of the discussion in Section 2.6 of the first volume, a connection between quantum field theory and multi-zeta values on one hand and between quantum field theory and knot theory on the other.

This has an especially interesting consequence for torus knots. Indeed the  $2 - (2n + 1)$  and  $(2n + 1) - 2$  knots are indirectly but tightly coupled with  $\zeta(2n + 1)$  for each  $n \geq 1$ . This is intriguing since the standard knot invariants of Alexander or Jones attach a polynomial (*algebraic*) quantity to a given knot. It would be very interesting to see a direct and natural identification. By contrast the unknot is identified with  $\pi$  and the  $2 - (2n)$  knots are identified trivially with links via  $\pi^{2n}$  (Euler yet again!).

18. **More knotty problems.** Figure 3.2 shows the  $8_8$  knot (from a standard catalogue in KnotPlot) and the famous Reidemeister moves, which are used to rearrange knots.

The knots in Figure 3.3 were listed as separate knots in knot tables. These are  $10_{161}$  and  $10_{162}$ , in [185] which notes that in 1974 they were shown by

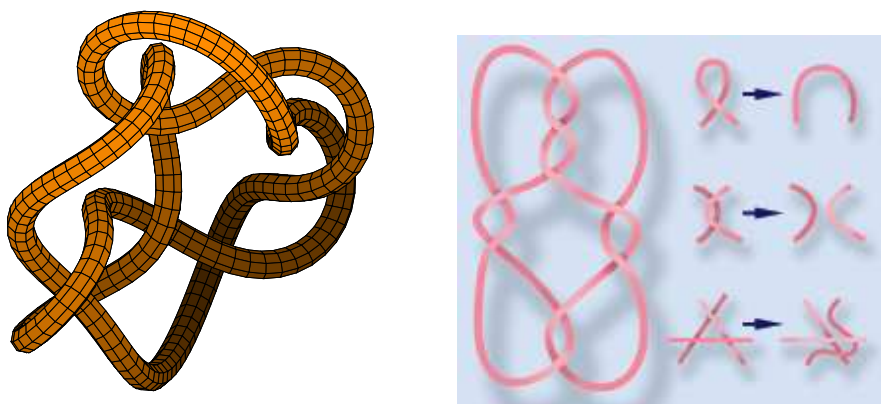


Figure 3.2: The knot  $8_8$  and Reidemeister moves

Perko to be equivalent [173]. The knots are still listed as separate in some knot tables, including the recent book by Kawauchi [145].

A lengthy sequence of images showing the equivalence is at the URL <http://www.cecm.sfu.ca/~scharein/projects/perko>, with the nice experimental mathematics connection that these deformations were performed entirely automatically using the KnotPlot tool, which is available at the URL <http://www.colab.sfu.ca/KnotPlot>. Indeed, both may be deformed to the knot in Figure 3.4.

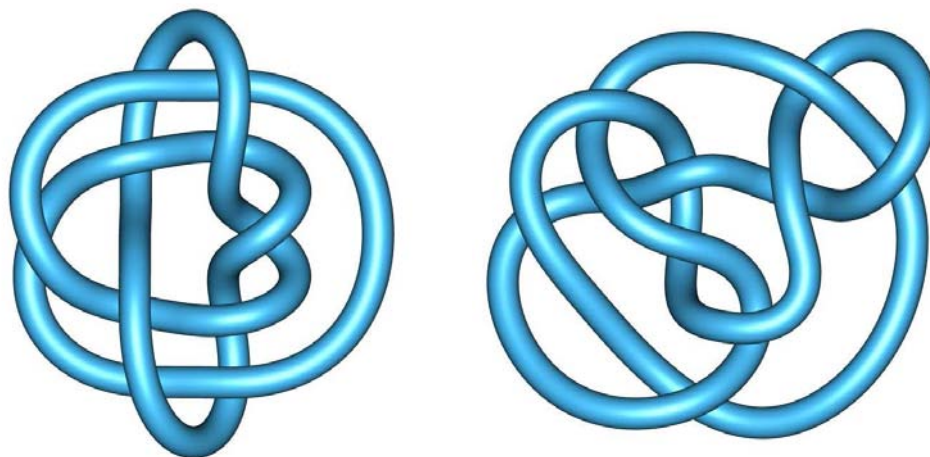


Figure 3.3: The knots  $10_{161}$  and  $10_{162}$

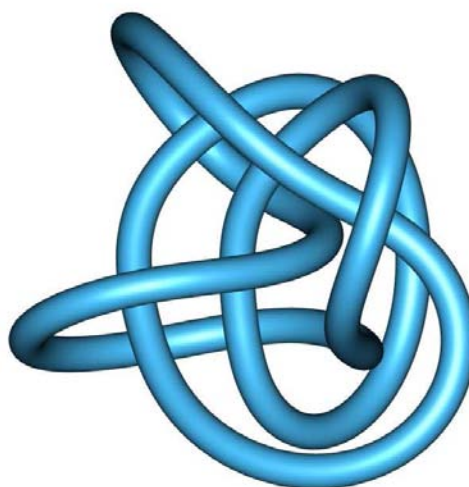


Figure 3.4: The knot equivalent to both  $10_{161}$  and  $10_{162}$

# Chapter 4

## Partitions and Powers

I'll be glad if I have succeeded in impressing the idea that it is not only pleasant to read at times the works of the old mathematical authors, but this may occasionally be of use for the actual advancement of science.

Constantin Carathéodory, speaking to an MAA meeting in 1936

In this chapter we address the theory of additive partitions and that of representations as sums of squares, from an experimental perspective. Both have a distinguished history. We will show that computational techniques can accelerate both solution and understanding. What's more, these techniques have a number of interesting applications, including, for example, Madelung's constant in physical chemistry.

### 4.1 Partition Functions

The number of *additive partitions* of  $n$ ,  $p(n)$ , is formally generated by

$$P(q) = 1 + \sum_{n \geq 1} p(n)q^n = \prod_{n \geq 1} (1 - q^n)^{-1}. \quad (4.1.1)$$

One ignores "0" and permutations. Thus  $p(5) = 7$  since

$$\begin{aligned} 5 &= 4 + 1 = 3 + 2 = 3 + 1 + 1 = 2 + 2 + 1 \\ &= 2 + 1 + 1 + 1 = 1 + 1 + 1 + 1 + 1. \end{aligned} \quad (4.1.2)$$

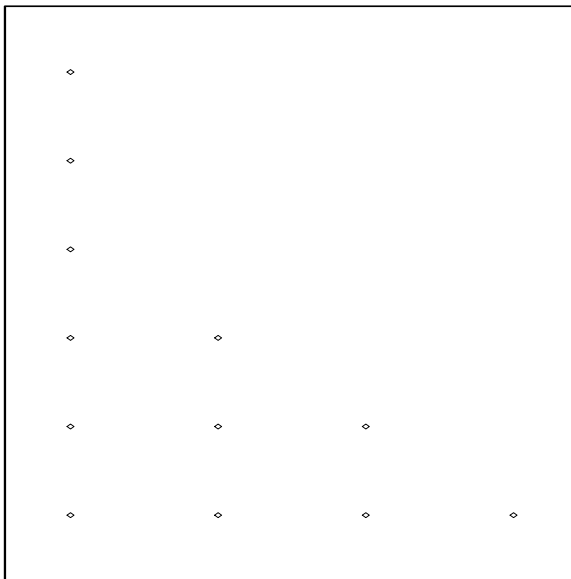


Figure 4.1: A Ferrer diagram

Additive partitions are less tractable than multiplicative ones as there is no analogue of unique prime factorization nor the corresponding structure.

Formula (4.1.1) is easily seen by expanding  $(1 - q^n)^{-1}$  and comparing coefficients. It is relatively easy to deduce that  $2^{\sqrt{n}} < p(n) < e^{\pi\sqrt{2n/3}}$  for  $n > 3$  (see [167]), and that the series is absolutely convergent for  $|q| < 1$ . We return to the analytic behavior of this series below.

Partitions provide a wonderful example of why Keith Devlin calls mathematics “the science of patterns” [102]. Many geometric representations exist. For example, the partition  $5 = 4 + 1$  can be represented as a point at  $(0, 0)$  and four points at  $(0, 1), (1, 1), (2, 1), (3, 1)$ . Read with axis reversed, this identifies  $1 + 4$  with  $2 + 1 + 1 + 1$  and so on. See Figure 4.1, which identifies  $1 + 1 + 1 + 2 + 3 + 4$  and  $6 + 3 + 2 + 1$ . Such techniques provide alternate ways to prove results such as *the number of partitions of  $n$  with all parts odd is the number of partitions of  $n$  into distinct parts*, (see Exercise 1).

A modern computational temperament leads to:

**Question:** How hard is  $p(n)$  to compute—in 1900 (for MacMahon the “father of combinatorial analysis”) or in 2000 (for *Maple* or *Mathematica*)?

**Answer:** The computation of  $p(200) = 3972999029388$  took MacMahon months and intelligence. Now, however, we can use the most naive approach: computing 200 terms of the series for the inverse product in (4.1.1) instantly produces the result using either *Mathematica* or *Maple*. Obtaining the result  $p(500) = 2300165032574323995027$  is not much more difficult, using the *Maple* code

```
> N:=500; coeff(series(1/product(1-q^n,n=1..N+1),q,N+1),q,N);
2300165032574323995027
```

□

### 4.1.1 Euler's Pentagonal Number Theorem

In early versions of *Maple*, computing  $P(q)$  was quite slow, while taking the series for the reciprocal of the series for  $Q(q) = \prod_{n \geq 1} (1 - q^n)$  was quite manageable? Why? Clearly the series for  $Q$  must have special properties. Indeed

$$Q(q) = 1 - q - q^2 + q^5 + q^7 - q^{12} - q^{15} + q^{22} + q^{26} - q^{35} - q^{40} + q^{51} + q^{57} - q^{70} - q^{77} + q^{92} + O(q^{100}). \quad (4.1.3)$$

If we do not immediately recognize these *pentagonal numbers*  $((3n \pm 1)n/2)$ , Sloane's online Encyclopedia of Integer Sequences, available on the Internet at <http://www.research.att.com/~njas/sequences>, again comes to the rescue, with abundant references to boot.

So, algorithmic analysis predicts *Euler's pentagonal number theorem*:

$$\prod_{n \geq 1} (1 - q^n) = \sum_{n=-\infty}^{\infty} (-1)^n q^{(3n+1)n/2}. \quad (4.1.4)$$

One would be less prone to look at  $Q$  on the way to  $P$  today when the computation is very snappy.

With this success under our belt, we might well ask what about powers of  $Q$ ? We obtain

$$\begin{aligned} Q^2(q) = & 1 - 2q - q^2 + 2q^3 + q^4 + 2q^5 - 2q^6 - 2q^8 - 2q^9 + q^{10} + 2q^{13} + 3q^{14} \\ & - 2q^{15} + 2q^{16} - 2q^{19} - 2q^{20} - 2q^{23}q^{24} + 2q^{26} + 2q^{27} - 2q^{28} \\ & + 2q^{29} + q^{30} + 2q^{31} + 2q^{33} - 2q^{34} - 2q^{35} + 2q^{36} - 2q^{38} \\ & - 4q^{40} + q^{44} - 2q^{45} + 2q^{48} + O(q^{50}) \end{aligned}$$

which is not nearly as lacunary; but

$$Q^3(q) = 1 - 3q + 5q^3 - 7q^6 + 9q^{10} - 11q^{15} + 13q^{21} - 15q^{28} + 17q^{36} - 19q^{45} + O(q^{51})$$

which has exposed another famous result—a special form of *Jacobi's triple product*. The general form is

$$\prod_{n \geq 1} (1 + xq^{2n-1})(1 + x^{-1}q^{2n-1})(1 - q^{2n}) = \sum_{n=-\infty}^{\infty} x^n q^{n^2}. \quad (4.1.5)$$

Then the formula implicit in (4.1.1) is

$$\prod_{n \geq 1} (1 - q^n)^3 = \sum_{m=0}^{\infty} (-1)^m (2m+1) q^{m(m+1)/2} \quad (4.1.6)$$

which may be obtained on replacing  $q$  by  $q^{1/2}$  and  $x$  by  $-wq^{1/2}$  in (4.1.5), differentiating with respect to  $w$ , and then letting  $w \rightarrow 1$  from below.

If we write  $P(q)Q(q) = 1$  in terms of the Cauchy convolution we have

$$\sum_{k \leq n} q_k p(n-k) = \delta_n \quad (4.1.7)$$

where  $q_k$  is the coefficient of  $q^k$  in (4.1.3), and  $\delta_n$  is the Kronecker function which is 1 when  $n = 0$  and 0 otherwise. It is a nice exercise to make this into Euler's explicit recursion for  $p(n)$  which only needs to compute  $O(\sqrt{n})$  smaller values of  $p(k)$ . One can similarly develop somewhat more efficient formulae relying on information such as (4.1.1).

## 4.1.2 Modular Properties of Partitions

Ramanujan used MacMahon's table of  $p(n)$ ,  $1 \leq n \leq 200$  to intuit remarkable and deep congruences such as

$$p(5n+4) \equiv 0 \pmod{5}$$

$$p(7n+5) \equiv 0 \pmod{7}$$



and

$$p(11n + 6) \equiv 0 \pmod{11},$$

from data like

$$\begin{aligned} P(q) = & 1 + q + 2q^2 + 3q^3 + 5q^4 + 7q^5 + 11q^6 + 15q^7 + 22q^8 + 30q^9 \\ & + 42q^{10} + 56q^{11} + 77q^{12} + 101q^{13} + 135q^{14} + 176q^{15} + 231q^{16} \\ & + 297q^{17} + 385q^{18} + 490q^{19} + 627q^{20} + 792q^{21} + 1002q^{22} \\ & + 1255q^{23} + \dots \end{aligned}$$

If one generates more terms of  $p(n)$  and displays them in an appropriately sized matrix this becomes much clearer:

$$\begin{bmatrix} 1 & 1 & 2 & 3 & 5 \\ 7 & 11 & 15 & 22 & 30 \\ 42 & 56 & 77 & 101 & 135 \\ 176 & 231 & 297 & 385 & 490 \\ 627 & 792 & 1002 & 1255 & 1575 \end{bmatrix} \quad (4.1.8)$$

shows clearly the congruence  $p(5n + 4) \equiv 0 \pmod{5}$  in the last column.

Correspondingly

$$\begin{bmatrix} 1 & 1 & 2 & 3 & 5 & 7 & 11 \\ 15 & 22 & 30 & 42 & 56 & 77 & 101 \\ 135 & 176 & 231 & 297 & 385 & 490 & 627 \\ 792 & 1002 & 1255 & 1575 & 1958 & 2436 & 3010 \\ 3718 & 4565 & 5604 & 6842 & 8349 & 10143 & 12310 \end{bmatrix} \quad (4.1.9)$$

shows clearly the congruence  $p(7n + 5) \equiv 0 \pmod{7}$  in the second last column.

Driven entirely by limited experimental data, Ramanujan conjectured an audacious set of correct modular identities, and not surprisingly over generalized! He conjectured that if  $d = 5^a 7^b 11^c$  and  $24n \equiv 1 \pmod{d}$  then  $p(n) \equiv 0 \pmod{d}$ . This is equivalent to the same conjectures for  $d$  a power of 5, 7, 11. This holds for  $a, b, c < 3$  and for all powers of 5, 11 but fails  $7^3$  as  $p(243) = 133978259344888$  since  $133978259344888 \equiv 245 \pmod{343}$  quickly shows in the 21st century while

$243 \cdot 24 = 17 \cdot 7^3 + 1$ . Such modular identities (see [132, 44]) and their extensions remain an active source of research today. The simplest case is described in Exercise 2.

### 4.1.3 The “Exact” Formula for $p(n)$

One of the signal achievements of early twentieth century analysis was Hardy and Ramanujan’s precise asymptotic for  $p(n)$ , [76]. It is based in part on an analysis of the *Dedekind  $\eta$ -function*  $\eta(q) = e^{\pi iz/12} \prod_{n \geq 1} (1 - e^{2\pi inz})$ . The function  $\eta$  is closely related to  $Q(q)$ , and  $\theta_3(q)$  discussed in the next section, and satisfies a modular equation like (4.2.21). Their asymptotic is

$$p(n) = \frac{e^{K\lambda_n}}{4\sqrt{3}\lambda_n^2} \left( 1 + O\left(\frac{1}{\sqrt{n}}\right) \right) \quad (4.1.10)$$

where  $K = \pi\sqrt{2/3}$  and  $\lambda_n = \sqrt{n - 1/24}$ .

This was subsequently refined by Rademacher to

$$p(n) = \frac{1}{\pi\sqrt{2}} \sum_{k=1}^{\infty} \alpha_k(n) \sqrt{k} \frac{d}{dx} \left[ \frac{\sinh\left(\frac{\pi}{k} \sqrt{\frac{2}{3}} \left(x - \frac{1}{24}\right)\right)}{\sqrt{\left(x - \frac{1}{24}\right)}} \right]_{x=n} \quad (4.1.11)$$

where

$$\alpha_k(n) = \sum_{(h,k)=1}^k \omega_{h,k} e^{-2\pi inh/k},$$

and  $\omega_{h,k} = \exp(\pi i \tau_{h,k})$  with

$$\tau_{h,k} = \sum_{m=1}^{k-1} \left( \frac{m}{k} - \left\lfloor \frac{m}{k} \right\rfloor - \frac{1}{2} \right) \left( \frac{hm}{k} - \left\lfloor \frac{hm}{k} \right\rfloor - \frac{1}{2} \right).$$

If order  $\sqrt{n}$  terms are appropriately used, the nearest integer is  $p(n)$ .

A mere five terms of this expansion provides  $p(200) \approx 3972999029387.86108$  and six terms yields  $p(500) \approx 2300165032574323995027.196661$ . As we have seen the underlying asymptotic is

$$p(n) \sim \frac{1}{4n\sqrt{3}} e^{\pi\sqrt{2n/3}}.$$

Later Erdős made an “elementary” derivation of the Hardy-Ramanujan formula (4.1.10). A recent discussion of this formula is given by Almkvist and Wilf in [8]. It is interesting to speculate how much corresponding beautiful mathematics is not done when computation becomes too easy—both *Maple* and *Mathematica* have good built in partition functions.

## 4.2 Singular Values

The Jacobian theta functions are a very rich source mine for experimentation—both as a tool to learning classical theory and to discover new phenomena. Further details of what follows are given fully in [44]. For our purposes we consider only the three classical  $\theta$ -functions:

$$\begin{aligned}\theta_3(q) &= \sum_{n=-\infty}^{\infty} q^{n^2}, & (4.2.12) \\ \theta_4(q) &= \sum_{n=-\infty}^{\infty} (-1)^n q^{n^2}, \\ \theta_2(q) &= \sum_{n=-\infty}^{\infty} q^{(n+1/2)^2},\end{aligned}$$

for  $|q| \leq 1$ . Note that  $\theta_3^2$  is the generating function for the number of ways of writing a number as a sum of two squares, counting order and sign. Similarly,  $\theta_2^2$  counts sums of two odd squares.

A beautiful result of Jacobi’s is

$$\theta_3^4(q) = \theta_2^4(q) + \theta_4^4(q). \quad (4.2.13)$$

If we write  $k = \theta_2^2/\theta_3^2$  and  $k' = \theta_4^2/\theta_3^2$ , we note that  $k^2 + (k')^2 = 1$ . It transpires that

$$(i) \quad \theta_3^2(q^2) = \frac{\theta_4^2(q) + \theta_3^2(q)}{2} \quad (ii) \quad \theta_4^2(q^2) = \theta_4(q) \theta_3(q). \quad (4.2.14)$$

Now (4.2.13) and (4.2.14) can be proved in many ways and can be “verified” symbolically in many more. For example, Jacobi’s triple product (4.1.5) with  $x =$

$\pm 1$  becomes product representations for  $\theta_3(q)$  and  $\theta_4(q)$  respectively. Multiplying these together yields (ii).

Even without access to the triple product, there is a simple algorithm (see [9, 10]) for converting a sum  $A(q) = 1 + \sum_{n \geq 1} a_n q^n$  to a product  $\prod_{n \geq 1} (1 - q^n)^{-b_n}$ , even preserving rationality of coefficients. To get an idea why, differentiating logarithmically and expanding the denominator of the right side leads to

$$\begin{aligned} \frac{q A'(q)}{A(q)} &= \sum_{n \geq 1} \frac{n b_n q^n}{1 - q^n} \\ &= \sum_{n, k \geq 1} n b_n q^{nk} = \sum_{n \geq 1} B_n^* q^n, \end{aligned} \quad (4.2.15)$$

where we set  $a_0 = 1$  and

$$B_n^* = \sum_{d|n} d b_d. \quad (4.2.16)$$

Thus,  $q A'(q) = A(q) \sum_{n > 0} B_n^* q^n$ , which is equivalent to the convolution

$$n a_n = \sum_{k=1}^n B_k^* a_{n-k} = \sum_{k=0}^{n-1} B_{n-k}^* a_k. \quad (4.2.17)$$

This clearly determines  $\{a_k\}$  given  $\{b_m\}$  and the converse obtains from the Möbius inversion formula of (4.3.34).

**Example 4.2.1** *Counting rooted trees.*

The number  $T_n$  of rooted trees with  $n$  branches is given by a formula due to Arthur Cayley (1821–1895)

$$T(x) = 1 + \sum_{k=1}^{\infty} T_{k+1} x^k = \prod_{n=1}^{\infty} (1 - x^n)^{-T_n}. \quad (4.2.18)$$

As we remarked in Section 1.6 of the first volume, the product and the sum share their coefficients. The recursion (4.2.17) for  $T_n$  becomes

$$T_{n+1} = \frac{1}{n} \sum_{k=1}^n T_k^* T_{n+1-k}, \quad T_n^* = \sum_{d|n} d T_d. \quad (4.2.19)$$

which starts:  $T_1 = 1, T_2 = 1, T_3 = 2, T_4 = 4, T_5 = 9, T_6 = 20, T_7 = 48 \dots$ .  $\square$

Applied to  $1 + 2q + 2q^4 + 2q^9 + 2q^{16}$  the algorithm produces

$$\frac{(1 - q^2)^3 (1 - q^4) (1 - q^6)^3 (1 - q^8) (1 - q^{10})^3 (1 - q^{12}) (1 - q^{14})^3 (1 - q^{16})}{(1 - q)^2 (1 - q^3)^2 (1 - q^5)^2 (1 - q^7)^2 (1 - q^9)^2 (1 - q^{11})^2 (1 - q^{13})^2 (1 - q^{15})^2},$$

from which a form of the product for  $\theta_3(q)$  can be read off. By contrast  $1 + 3q + 3q^4 + 3q^9 + 3q^{16}$  produces

$$\frac{(1 - q^2)^6 (1 - q^4)^{15} (1 - q^6)^{97} (1 - q^8)^{573} (1 - q^{10})^{3867} (1 - q^{12})^{26446}}{(1 - q)^3 (1 - q^3)^8 (1 - q^5)^{39} (1 - q^7)^{231} (1 - q^9)^{1485} (1 - q^{11})^{10056}} \cdot \frac{(1 - q^{14})^{187761} (1 - q^{16})^{1356198}}{(1 - q^{13})^{70305} (1 - q^{15})^{503384}}$$

which is pretty good evidence that no natural product exists. We will establish (4.2.14) (i) in the next section.

Also, one notes that it follows, from (4.2.14), that  $\theta_3^2$  and  $\theta_4^2$  parametrize the AGM and that

$$\begin{aligned} AG(\theta_3^2(q), \theta_4^2(q)) &= AG(\theta_3^2(q^2), \theta_4^2(q^2)) = AG(\theta_3^2(q^4), \theta_4^2(q^4)) = & (4.2.20) \\ \dots AG(\theta_3^2(q^{2^n}), \theta_4^2(q^{2^n})) \dots &= AG(\theta_3^2(0), \theta_4^2(0)) = AG(1, 1) = 1 \end{aligned}$$

since the iteration's limit is unchanged if one starts at the first or the second stage of the iteration, and since AG is continuous. Another marvellous fact that follows from *Poisson summation* is that

$$k(e^{-\pi s}) = k'(e^{-\pi/s}), \tag{4.2.21}$$

for  $s > 0$ . In particular,  $k(e^{-\pi}) = \sqrt{1/2}$ . Then (4.2.21) and (4.2.20), in conjunction with the already explored relationship between Elliptic integrals and the AGM (Section 5.6 of the first volume) shows that with the above definition of  $k$ ,

$$K(k(q)) = \frac{\pi}{2} \theta_3^2(q). \tag{4.2.22}$$

Now the classical theory of *modular equations* asserts that there is an algebraic relationship between  $k = k(q)$  and  $l = k(q^N)$  for each positive integer  $N$ . For

example, the quadratic equation, implicit in (4.2.14), is  $l' = 2\sqrt{k'}/(1+k')$ , while the *cubic modular equation* may be written as

$$\theta_3(q)\theta_3(q^3) = \theta_4(q)\theta_4(q^3) + \theta_2(q)\theta_2(q^3) \quad (4.2.23)$$

or equivalently

$$\sqrt{kl} + \sqrt{k'l'} = 1. \quad (4.2.24)$$

Similarly, for  $N = 7$  the equation is

$$\sqrt{\theta_3(q)\theta_3(q^7)} = \sqrt{\theta_4(q)\theta_4(q^7)} + \sqrt{\theta_2(q)\theta_2(q^7)} \quad (4.2.25)$$

or equivalently

$$\sqrt[4]{kl} + \sqrt[4]{k'l'} = 1. \quad (4.2.26)$$

The existence of such modular equations means that there is an algebraic relationship between  $K(k)$  and  $K(l)$  and in particular that  $k_N = k(e^{-\pi\sqrt{N}})$  is a (solvable) algebraic number, called the  $N$ -th *singular value*. It is also the case that two *invariants* used by Ramanujan reduce the degrees of these quantities. He used

$$G_N = (2kk')^{-1/12}, \quad g_N = (2k/k'^2)^{-1/12}, \quad (4.2.27)$$

and it transpires that  $G_N$  is better for odd  $N$  and  $g_N$  for even  $N$ .

From the equations above, since  $k' = l, l' = k$  in this case, we may read off the values  $G_1 = g_2 = 1, G_3 = 2^{1/12}$ , and  $G_7 = 2^{1/4}$ ; with a little more work we may obtain  $g_4 = 2^{1/8}, g_6^6 = \sqrt{2} + 1, g_8^8 = (\sqrt{2} + 1)/2$ , and  $G_9^6 = (2 + \sqrt{3})$ . From these evaluations in turn we may easily determine  $k_N$  for  $N = 1, 2, 3, 4, 6, 7, 8, 9$ . Had we supplied the quintic modular equation we could determine  $G_5^{12} = \sqrt{5} + 2, g_{10}^2 = (\sqrt{5} + 1)/2, G_{15}^3 = 2^{3/4}(\sqrt{5} + 1)/2$ , and  $G_{25} = (\sqrt{5} + 1)/2$ .

Each of these has a reworking as an infinite series evaluation. Thus,

$$\theta_3(e^{-\pi}) = \sqrt[4]{2}\theta_4(e^{-\pi}) = \sqrt[4]{2}\theta_2(e^{-\pi}). \quad (4.2.28)$$

But this is not the main point of this section. We have sketched that (modular) functions such as

$$N \mapsto \frac{\theta_2(q)}{\theta_3(q)}, \quad q = e^{-\pi\sqrt{N}}, \quad (4.2.29)$$

are guaranteed to have algebraic values; and by their nature they are very rapidly computable to high precision. Thus, they provide excellent test beds for (i) recovering *minimal polynomials* from numerical data, and (ii) for simplifying the radicals so obtained.

For example, working to 15 places, the “MinimalPolynomial” feature of *Maple*, which uses lattice basis reduction, returns  $x^2 - 2x - 1$  for  $g_{22}^2$ , returns  $x^2 - 12x - 1$  for  $G_{37}^4$ , and returns  $1 + 10x + 23x^2 - 10x^3 + x^4$  for  $G_{58}^2$ —leading to three of the cleanest singular values. Correspondingly,  $G_{11}^4$  solves the cubic  $x^3 - 4x^2 + 4x - 1 = 0$ ,  $g_{12}^4 = 2^{1/6}(\sqrt{3} + 1)$ ,  $G_{13}^4 = (\sqrt{13} + 1)/2$ , and  $g_{14}^2$  yields the polynomial  $x^4 - 2x^3 + 4x^2 - x + 1$ , which gives  $g_{14}^2 + g_{14}^{-2} = \sqrt{2} + 1$ . Also  $G_{17}^{12} + G_{17}^{-12} = 40 + 10\sqrt{17}$ . In each case, the root or radical obtained using a low precision “hunt” can be checked almost instantly to many hundreds or thousands of digits precision.

For instance, we may discover that  $x = G_{47}^4/2$  is a root of the solvable irreducible quintic  $x^5 - 10x^4 + 9x^3 - 4x^2 - 1 = 0$ . In cases of degree less than ten, *Maple* can provide the Galois group (in this case the dihedral group  $D_5$ ),

and also the enormous radical:

$$\begin{aligned}
5x = 10 + & \left[ \frac{273625}{4} + \frac{66025}{4} \sqrt{5} - \frac{53885}{772} \sqrt{45355 - 16826\sqrt{5}} \right. \\
& \left. - \frac{1847377065}{772} \frac{\sqrt{5}}{\sqrt{45355 - 16826\sqrt{5}}} \right]^{1/5} \\
& + \frac{6285 + \frac{1265}{2} \sqrt{5} - \frac{2255}{386} \sqrt{45355 - 16826\sqrt{5}} - \frac{20331965}{193} \frac{\sqrt{5}}{\sqrt{45355 - 16826\sqrt{5}}}}{\left( \frac{273625}{4} + \frac{66025}{4} \sqrt{5} - \frac{53885}{772} \sqrt{45355 - 16826\sqrt{5}} - \frac{1847377065}{772} \frac{\sqrt{5}}{\sqrt{45355 - 16826\sqrt{5}}} \right)^{3/5}} \\
& + \frac{\frac{1375}{2} + 35\sqrt{5} + \frac{11}{386} \sqrt{45355 - 16826\sqrt{5}} - \frac{1767012}{193} \frac{\sqrt{5}}{\sqrt{45355 - 16826\sqrt{5}}}}{\left( \frac{273625}{4} + \frac{66025}{4} \sqrt{5} - \frac{53885}{772} \sqrt{45355 - 16826\sqrt{5}} - \frac{1847377065}{772} \frac{\sqrt{5}}{\sqrt{45355 - 16826\sqrt{5}}} \right)^{2/5}} \\
& + \frac{\frac{155}{2} + \frac{13}{2} \sqrt{5}}{\sqrt[5]{\frac{273625}{4} + \frac{66025}{4} \sqrt{5} - \frac{53885}{772} \sqrt{45355 - 16826\sqrt{5}} - \frac{1847377065}{772} \frac{\sqrt{5}}{\sqrt{45355 - 16826\sqrt{5}}}}},
\end{aligned}$$

which repeated massaging reduces to

$$\begin{aligned}
x = 2 + & \sqrt[5]{\frac{2189}{100} + \frac{2641}{500} \sqrt{5} - \frac{1}{2500} \sqrt{1436961550 + 641957866 \sqrt{5}}} \\
& + \sqrt[5]{\frac{2189}{100} - \frac{2641}{500} \sqrt{5} - \frac{1}{2500} \sqrt{1436961550 - 641957866 \sqrt{5}}} \\
& + \sqrt[5]{\frac{2189}{100} - \frac{2641}{500} \sqrt{5} + \frac{1}{2500} \sqrt{1436961550 - 641957866 \sqrt{5}}} \\
& + \sqrt[5]{\frac{2189}{100} + \frac{2641}{500} \sqrt{5} + \frac{1}{2500} \sqrt{1436961550 + 641957866 \sqrt{5}}}.
\end{aligned}$$

Likewise, Ramanujan's celebrated singular value, sent in his letter to Hardy, is

$$\begin{aligned}
k_{210} = & (\sqrt{2} - 1)^2 (2 - \sqrt{3}) (\sqrt{7} - \sqrt{6})^2 (8 - \sqrt{63}) \\
& \times (\sqrt{10} - 3)^2 (4 - \sqrt{15})^2 (\sqrt{15} - \sqrt{14}) (6 - \sqrt{35}).
\end{aligned} \tag{4.2.30}$$



Indeed,  $k_{330}$  and  $k_{462}$  have a similar form involving fundamental solutions to Pell's equation (units of real quadratic fields).

Finally we note that for small  $N$  the elliptic integral  $K(k_N)$  is correspondingly susceptible to evaluation in terms of Gamma functions. Thus, to go along with our previous evaluation of  $K(k_1)$  we have

$$K(k_3) = \frac{3^{1/4} \Gamma(\frac{1}{3})^3}{2^{7/3} \pi} \quad \text{and} \quad K(k_7) = \frac{\Gamma(\frac{1}{7}) \Gamma(\frac{2}{7}) \Gamma(\frac{4}{7})}{4 \pi \sqrt[4]{7}}. \quad (4.2.31)$$

In each case, there is a neater expression in terms of the  $\beta$ -function waiting to be disinterred.

### 4.3 Crystal Sums and Madelung's Constant

We have seen the power of converting series to products and making other changes of representation. We now introduce Lambert series which are representations of the form

$$\sum_{n=1}^{\infty} f(n) \frac{x^n}{1-x^n} = \sum_{n=1}^{\infty} F(n) x^n \quad (4.3.32)$$

where

$$F(n) = \sum_{d|n} f(d) \quad (4.3.33)$$

summed over all positive divisors of  $n$ , due originally to Laguerre. The identity (4.3.32) is established by using the binomial theorem and gathering up terms, much as with the partition function above.

Thus, for  $f(n) \equiv 1$  we have  $F(n) = \tau(n) = \sigma_0(n)$ , the number of divisors of  $n$ , while  $f(n) = n^k$  ( $k \neq 0$ ) yields  $\sigma_k(n)$  the  $k$ -th power sum of the divisors. Recall that the Möbius function is defined by  $\mu(1) = 1$ ,  $\mu(n) = (-1)^m$  if  $n$  is the product of  $m$  distinct prime factors in  $n$ , and zero otherwise. Then the *Möbius inversion theorem* says that

$$\sum_{d|n} F(d) \mu(n/d) = f(n) \quad (4.3.34)$$

for *any* arithmetic function  $f$ .

This is an analogue of Cauchy convolution.

### 4.3.1 Sums of Squares

Let us observe that

$$\theta_3^m(q) = 1 + \sum_{n \geq 1} r_m(n)q^n \quad (4.3.35)$$

where  $\theta_3$  is defined by (4.2.12) and  $r_m(n)$  counts the number of ways of writing  $n = \sum_{k=1}^m n_k^2$ , again distinguishing order and sign of the integers used.

It is easy to compute a significant number of terms by merely expanding truncations of the series on the right-hand side of (4.3.35). This is quite effective for small even numbers of squares.

**Example 4.3.1** *Two squares.*

The first 60 terms of  $r_2(n)/4$  are

$$\begin{aligned} &1, 1, 0, 1, 2, 0, 0, 1, 1, 2, 0, 0, 2, 0, 0, 1, 2, 1, 0, 2, 0, 0, 0, 0, 3, 2, 0, 0, 2, 0, \\ &0, 1, 0, 2, 0, 1, 2, 0, 0, 2, 2, 0, 0, 0, 2, 0, 0, 0, 1, 3, 0, 2, 2, 0, 0, 0, 2, 0, 0 \end{aligned}$$

which does not immediately show a clear pattern. However, applying (4.3.34) to the first 30 terms yields

$$1, 0, -1, 0, 1, 0, -1, 0, 1, 0, -1, 0, 1, 0, -1, 0, 1, 0, -1, 0, 1, 0, -1, 0, 1, 0, -1, 0, 1, 0, -1, 0, 1, 0,$$

and the formula is immediately evident. It is

$$r_2(n) = 4(d_1(n) - d_3(n)), \quad (4.3.36)$$

where  $d_k$  is the number of divisors of  $n$  congruent to  $k$  modulo four. Equivalently,

$$\theta_3^2(q) - 1 = 4 \sum_{n \geq 0} (-1)^n \frac{q^{2n+1}}{1 - q^{2n+1}}. \quad (4.3.37)$$

□

**Example 4.3.2** *Four squares.*

The series grows much faster ( $r_2(n)$  is  $O(n^\delta)$  for any  $\delta > 0$ ) and the first twenty terms of  $r_4(n)/8$  are

$$1, 3, 4, 3, 6, 12, 8, 3, 13, 18, 12, 12, 14, 24, 24, 3, 18, 39, 20, 18,$$

while Möbius inversion produces

$$1, 2, 3, 0, 5, 6, 7, 0, 9, 10, 11, 0, 13, 14, 15, 0, 17, 18, 19, 0,$$

from which it is obvious that

$$r_4(n) = 8 \sum_{d|n, 4 \nmid d} d \quad (4.3.38)$$

and a nice corollary is that since  $1|n$ ,  $r_4(n)$  is always positive (Lagrange's famous theorem).  $\square$

**Example 4.3.3** *Six squares.*

Möbius inversion produces

$$12, 48, 148, 192, 300, 336, 948, 768, 716, 1200, 2388, 1344, 2028, 2256, 3700, \\ 3072, 3468, 3120, 7188, 4800.$$

There is clearly structure but what? We leave this as a challenge and turn to eight squares.  $\square$

**Example 4.3.4** *Eight squares.*

Möbius inversion now produces

$$11, 6, 27, 64, 125, 162, 343, 512, 729, 750, 1331, 1728, 2197, 2058, 3375, 4096.$$

There is again clearly structure but what? If we apply inversion to  $(-1)^d r_8(d)$  (this is using  $\theta_4$  instead of  $\theta_3$ ) we are rewarded with

$$-1, 8, -27, 64, -125, 216, -343, 512, -729, 1000, -1331, 1728, -2197, 2744, \\ -3375, 4096.$$

Thus,

$$\theta_3^8(-q) = \theta_4^8(q) = 1 + 16 \sum_{n \geq 1} \frac{(-1)^n n^3 q^n}{1 - q^n}. \quad (4.3.39)$$

□

We end this subsection by deriving (4.2.14) (i) as promised. Indeed we note that this is equivalent to

$$\sum_{n > 0} r_2(n) q^{2n} = \sum_{n > 0} \frac{(1 + (-1)^n)}{2} r_2(n) q^n.$$

This follows immediately from  $r_2(2n) = r_2(n)$ , given that (4.3.36) shows  $r_2(n)$  only depends on the odd part of  $n$ . Of course we have not proven any of these representations, but only uncovered them.

### 4.3.2 Multidimensional Sums

Consider the sums

$$\mathcal{M}_2(s) = \sum_{\substack{m, n \in \mathbb{Z} \\ (m, n) \neq 0}} \frac{(-1)^{m+n}}{(m^2 + n^2)^{s/2}}, \quad (4.3.40)$$

$$\mathcal{M}_3(s) = \sum_{\substack{(m, n, p) \in \mathbb{Z} \\ (m, n, p) \neq 0}} \frac{(-1)^{m+n+p}}{(m^2 + n^2 + p^2)^{s/2}}, \quad (4.3.41)$$

and higher dimensional versions  $\mathcal{M}_N(s)$  defined analogously. In future we write  $\sum'$  to denote that poles of the summatory are left out. We are primarily interested in the value  $\mathcal{M}_3(1)$  which is called *Madelung's constant* for sodium chloride, as it is an attempt to count the potential at the origin if alternating charges are placed at all other points of an integer cubic lattice. The physical-chemical literature generally treats (4.3.40) and (4.3.41) as well-defined objects. Mathematically this is far from so. Since these numbers are often computed, something more must be said (see [32]).

**Example 4.3.5** *Convergence over increasing circles, squares and diamonds.*

1. **Squares.** The sum  $b_2^N(s) = \sum_{n=1}^N \sum_{m=1}^N (n^2 + m^2)^{-s/2}$  converges to an analytic function  $b_2(s)$ , for  $\operatorname{Re}(s) > 0$ .
2. **Circles.** The sum  $\sum_{n^2+m^2 \leq N} (n^2 + m^2)^{-s/2} = \sum_{k \leq N} r_2(k)/k^{s/2}$  converges to an analytic function  $r_2(s)$ , for  $\operatorname{Re}(s) > 1/3$ , but fails to converge somewhere above  $1/4$ . This relies on the fact that the average order of  $r_2(n)$  is quite well understood, [133]. Where the limit exists it must agree with that for squares by uniqueness of analytic continuation.
3. **Diamonds.** Consider adding up over increasing diamonds:  $|n| + |m| = N$ . Then the contribution of each shell is  $\sum_{m=0}^N 1/\sqrt{N^2 - 2Nm + 2m^2}$  and the limit is the Riemann integral  $\int_0^1 1/\sqrt{1-2t+2t^2} dt = \sqrt{2} \log(1 + \sqrt{2})$ . So the terms of the series do not even go to zero.

Thus, the order of summation matters even for these “natural sums.” In fact, if we always add these sums over increasing hypercubes, they converge to an analytic limit for  $\operatorname{Re}(s) > 0$ . So we shall take this as the default meaning of the sum [44, 31, 32]. We also note that these sums converge very slowly, so that direct summation methods are to be avoided. We shall return to the evaluation of  $\mathcal{M}_3(1)$  in the next section.

The normalized *Mellin transform* (see also Item 3a) is a special form of the Laplace transform, which makes the link between theta functions and zeta functions, as we saw in Chapter 3. In this setting, we recall that

$$\mu_s(f) = \frac{1}{\Gamma(s)} \int_0^\infty f(t)t^{s-1} dt. \quad (4.3.42)$$

Then it is easy to check that  $\mu_s(e^{-tn}) = n^{-s}$  and also that

$$\mathcal{M}_2(2s) = \mu_s(\theta_4^2(e^{-t}) - 1), \quad (4.3.43)$$

and using (4.3.37) the fact that  $\theta_4^2(q) - 1 = \theta_3^2(-q) - 1 = 4 \sum_{n>0} (-1)^n (-q)^{2n+1} / (1 - (-q)^{2n+1}) = 4 \sum_{n,m \geq 1} (-1)^{n+m-1} q^{m(2n-1)}$  implies that

$$\begin{aligned} \mathcal{M}_2(2s) &= 4 \sum_{n,m \geq 1} \frac{(-1)^{n+m-1}}{m(2n-1)^s} \\ &= 4 \sum_{n \geq 0} \frac{(-1)^n}{(2n+1)^s} \sum_{m \geq 1} \frac{(-1)^m}{m^s} = -4 \alpha(s) \beta(s), \end{aligned} \quad (4.3.44)$$

where as before  $\alpha(s) = \sum_{m \geq 1} (-1)^{m+1}/m^s$ , the alternating zeta function, and  $\beta(s) = \sum_{n \geq 0} (-1)^n/(2n+1)^s$  is the Catalan zeta function ( $\beta(2) = G$  is Catalan's constant).

Similar arguments based on (4.3.38) lead to

$$\mathcal{M}_4(2s) = -8\alpha(s)\alpha(s-1), \quad (4.3.45)$$

and to a corresponding formula for  $M_8(s)$ . In each of these cases the values are easily computed from the analytic continuations of the underlying zeta functions as given in Chapter 2. In particular  $\mathcal{M}_2(1) = -1.61554262671282472386\dots$  and  $\mathcal{M}_4(1) = -1.83939908404504706623$ . Moreover, various closed forms exist such as  $\mathcal{M}_4(2) = -4\log(2)$ .

We complete this subsection by listing some other formulae. Define  $\mathcal{L}_N(2s) = \sum_{n=1}^{\infty} r_N(n)/n^s$ . The corresponding  $\mathcal{L}_N(2s)$  are known for  $N = 2, 4, 6$  and  $8$ ,

$$\begin{aligned} \mathcal{L}_2(2s) &= 4\zeta(s)\beta(s), \\ \mathcal{L}_4(2s) &= 8(1 - 4^{1-s})\zeta(s)\zeta(s-1), \\ \mathcal{L}_6(2s) &= 16\zeta(s-2)L_{-4}(s) - 4\zeta(s)\beta(s-2), \\ \mathcal{L}_8(2s) &= 16(1 - 2^{1-s} + 4^{2-s})\zeta(s)\zeta(s-3). \end{aligned}$$

From  $\mathcal{L}_6(2s)$  one can reverse the steps we employed and discover the formula for  $r_6(n)$  left open in (4.3.39).

Many more formulae are discussed in [44, 38] and references therein. For example, as discovered by Zucker, Glasser and Robertson, we have similar closed forms for L-series based on the quadratic form  $x^2 + 2Py^2$ . We let  $r_{2,2p}(n)$  be the number of representations of  $n = m^2 + 2Pk^2$  and let  $\mathcal{L}_{2,2P}(2s) = \sum_{(n,m) \neq 0} (m^2 + 2Pn^2)^{-s} = \sum_{n>0} r_{2,2P}(n)n^{-s}$ . Then

$$\mathcal{L}_{2,2P}(2s) = 2^{1-t} \sum_{\mu|P} \mathcal{L}_{\epsilon_\mu\mu}(s)\mathcal{L}_{-8P\epsilon_\mu/\mu}(s)$$

for the *type two* integers

$$P = 1, 3, 5, 11, 15, 21, 29, 35, 39, 51, 65, 95, 105, 165, 231,$$

which we will meet again in Section 5.2. Here  $\epsilon_\mu = \left(\frac{-1}{\mu}\right)$  and  $L_\mu(s) = \sum_{n \geq 1} \left(\frac{\mu}{n}\right) n^{-s}$ , where  $\left(\frac{\mu}{n}\right)$  is the Legendre-Jacobi symbol. Thus  $L_1(s) = \zeta(s)$  and  $L_{-4}(s) = \beta(s)$ . See also Section 3.3.

### 4.3.3 Madelung's Constant

Odd squares are notoriously less amenable to closed forms. Following Hardy, Bateman in [21] gives the following formula for  $r_3(n)$ . Let

$$\chi_2(n) = \begin{cases} 0 & \text{if } 4^{-a}n \equiv 7 \pmod{8} \\ 2^{-a} & \text{if } 4^{-a}n \equiv 3 \pmod{8} \\ 3 \cdot 2^{-1-a} & \text{if } 4^{-a}n \equiv 1, 2, 5, 6 \pmod{8}, \end{cases}$$

where  $a$  is the highest power of 4 dividing  $n$ . Then

$$r_3(n) = \frac{16\sqrt{n}}{\pi} L_{-4n}(1) \chi_2(n) \times \prod_{p^2|n} \left( \frac{p^{-\tau} - 1}{p^{-1} - 1} + p^{-\tau} \left( 1 - \frac{1}{p} \left( \frac{-p^{-2\tau}n}{p} \right) \right)^{-1} \right) \quad (4.3.46)$$

where  $\tau = \tau_p$  is the highest power of  $p^2$  dividing  $n$ .

The corresponding formula for  $\mathcal{M}_3(s)$  or for  $\mathcal{L}_3(s)$  is thus not tractable. We turn to Bessel functions and let  $K_s$  be the *modified Bessel function of the second kind*. Then

$$\mathcal{L}_3(2s) = \frac{6\pi}{s} \zeta(2s-2) + \frac{12\pi^{s+1}}{\Gamma(s+1)} \sum_{m=1}^{\infty} r_2(m) m^{s/2} \sum_{n=1}^{\infty} \frac{1}{n^{s-2}} K_s(2\pi n \sqrt{m}). \quad (4.3.47)$$

The second term of (4.3.47) can be rewritten as

$$\frac{12\pi^{s+1}}{\Gamma(s+1)} \sum_{k>0} k^{\frac{s}{2}} K_s(2\pi\sqrt{k}) \sum_{n^2|k} \frac{r_2(k/n^2)}{n^{2s-2}}.$$

Moreover, these Bessel functions are elementary when  $s$  is a half-integer. Most nicely, for “jellium,” which is the Wigner sum analogue of Madelung's constant which arises when one considers bathing a positively charged cubic crystal in a continuous background charge, we have

$$\mathcal{L}_3(1) = -\pi + 3\pi \sum_{m>0} r_2(m) \operatorname{cosech}^2(\pi\sqrt{m}),$$

and the exponential rate of convergence is apparent. Exactly analogous is the most accessible expansion for Madelung's constant due to Benson and proved in [44]

$$\mathcal{M}_3(1) = -12\pi \sum_{m,n \geq 0} \operatorname{sech}^2 \left( \frac{\pi}{2} ((2m+1)^2 + (2n+1)^2)^{1/2} \right), \quad (4.3.48)$$

in which again the convergence is exponential. Summing for  $m, n \leq 3$  produces 1.747564594... correct to 8 digits. If we write  $o_3(n)$  for the number of ways of writing  $n$  as a sum of two odd squares, this becomes

$$\mathcal{M}_3(1) = -3\pi \sum_{m > 0} o_3(m) \operatorname{sech}^2 \left( \frac{\pi}{2} \sqrt{m} \right). \quad (4.3.49)$$

There is a corresponding formula for  $\mathcal{M}_3(s)$ .

There is also beautiful formula for  $\theta_2^3$  due to Andrews (given with a typographical error in [44]):

$$\theta_2^3(q) = 8 \sum_{n=0}^{\infty} \sum_{j=0}^{2n} \left( \frac{1 + q^{4n+2}}{1 - q^{4n+2}} \right) q^{(2n+1)^2 - (j+1/2)^2}. \quad (4.3.50)$$

From (4.3.50) the reader will be able to derive almost immediately Gauss's result that *every number is the sum of three triangular numbers*, and is challenged to apply (4.3.50) to the study of  $\mathcal{M}_3(1)$ .

Another related class of physically meaningful integrals are the *logarithmic Watson integrals*,  $L_d$  which arise in the study of polymers and are studied in [141]:

$$L_d = \frac{1}{\pi^d} \int_0^\pi \cdots \int_0^\pi \log \left( d - \sum_1^d \cos(s_k) \right) ds_1 \cdots ds_d. \quad (4.3.51)$$

For  $d = 1, 2$ , these reduce to  $L_1 = (1/\pi) \int_0^\pi \log(1 - \cos(t)) dt = -\log 2$  and  $L_2 = (1/\pi^2) \int_0^\pi \int_0^\pi \log(2 - \cos(t) - \cos(s)) dt ds = 4\beta(2)/\pi - \log(2)$ . The evaluation of  $L_2$  is equivalent to

$$\sum_{n=0}^{\infty} \frac{\binom{2n}{n}^2}{4^{2n+1} (2n+1)} = \frac{\beta(2)}{\pi}.$$



No closed form is known for  $d > 2$ . The prior evaluations and numerical exploration are facilitated by the lovely one-dimensional representation

$$L_d = \int_0^\infty \frac{e^{-t} - e^{-dt} I_0(t)^d}{t} dt, \quad (4.3.52)$$

where  $I_0$  is a Bessel function of the first kind. A number of additional results related to Madelung's constant can be found in a series of papers by Richard Crandall [90, 93, 91, 92].

## 4.4 Some Fibonacci Sums

Theta functions turn up in quite unexpected places as we now show.

The *Fibonacci sequence*,

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

takes its name from its first appearance in print, which seems to have been in the famous book *Liber Abaci* published by Leonardo Fibonacci (also known as Leonardo of Pisa) in 1202. He asked:

How many pairs of rabbits can be produced from a single pair in a year if every month each pair begets a new pair which from the second month on becomes productive?

Lest one thinks the problem is imprecise, Fibonacci describes the solution in the text and in the margin. There one finds written vertically

parium 1 primus 2 Secundus 3 tercius 5 Quartus 8 Quintus 13 Sestus  
21 Septimus 34 Octauus 55 Nonus 89 Decimus 144 Undecimus 233  
Duodecimus 377.

We leave it to the reader to decide that this indeed leads to the Fibonacci sequence, but we do note that “the proof is left as an exercise” seems to have occurred first in *De Triangulis Omnimodis* by Regiomontanus, written in 1464 (but published in 1533). He is quoted as saying “This is seen to be the converse of the preceding. Moreover, it has a straightforward proof, as did the preceding. Whereupon I leave it to you for homework.”

Among its many other contributions such as popularizing Hindu-Arabic notation in the west, the Liber Abaci contains methods for extracting cube roots and for solving quadratics and the lovely identity  $(a^2 + b^2)(c^2 + d^2) = (ac \pm bd)^2 + (ad \mp bc)^2$ , which show the product of sums of two squares is such a sum.

The Fibonacci sequence occurs in many contexts both serious and quirky. For example 144 is the only Fibonacci square. A moment's inspection shows that it is generated by

$$F_0 = 1, \quad F_1 = 1, \quad F_{n+1} = F_n + F_{n-1}. \quad (4.4.53)$$

It clearly grows quickly (like rabbits), indeed it is monotonic and so  $F_{n+2} > 2F_n$ . More precisely what is happening? If we look numerically at  $F_{n+1}/F_n$  for  $n = 10, 20, 30, 40$  we see 1.61818181818, 1.61803399852, 1.61803398875, 1.61803398875, and either the human eye or constant detection reveals this to be the *Golden Mean*:

$$G = \frac{\sqrt{5} + 1}{2},$$

to the precision used.

Indeed the standard theory of two term linear recurrence relations leads to

$$F_n = \frac{\left(\frac{\sqrt{5}+1}{2}\right)^n - \left(\frac{1-\sqrt{5}}{2}\right)^n}{\sqrt{5}} \quad (4.4.54)$$

where  $-g = (1 - \sqrt{5})/2$  is the other root of  $x^2 = x + 1$ .

It is easy to check that the sequence in (4.4.54) satisfies the recursion in (4.4.53), and has the correct initial conditions. Since  $|g| < 1$  it is also easy to see that  $F_{n+1}/F_n \rightarrow G$ , as claimed, and to deduce many other identities such as  $F_{n+1}F_{n-1} = F_n^2 + (-1)^n$  for  $n \geq 2$ .

There is a slightly less well known companion *Lucas sequence* named after the French number theorist Edouard Lucas (1842–1891):

$$L_0 = 2, \quad L_1 = 1, \quad L_{n+1} = L_n + L_{n-1}, \quad (4.4.55)$$

which is correspondingly solved by

$$L_n = \left(\frac{\sqrt{5} + 1}{2}\right)^n + \left(\frac{1 - \sqrt{5}}{2}\right)^n. \quad (4.4.56)$$

As both Fibonacci and Lucas sequences are built of geometric sequences, it is clear that we can easily evaluate sums like  $\sum_{n=1}^N F_n^k$  for positive integer  $k$ . What happens for negative integers is more interesting.

A preparatory lemma is useful:

**Lemma 4.4.1** ([44]) For  $0 < \beta < \alpha$  with  $\alpha\beta = 1$ ,

$$\sum_{n=1}^{\infty} \frac{1}{\alpha^n + \beta^n} = \sum_{n=1}^{\infty} \frac{\beta^n}{1 + \beta^{2n}} = \theta_3^2(\beta), \quad (4.4.57)$$

$$\sum_{n=0}^{\infty} \frac{1}{\alpha^{2n+1} + \beta^{2n+1}} = \sum_{n=0}^{\infty} \frac{\beta^{2n+1}}{1 + \beta^{2n+1}} = \frac{1}{4} \theta_2^3(\beta^2). \quad (4.4.58)$$

**Proof.** The proof of the first formula is a consequence of (4.3.37), discovered in our discussion of sums of squares. This relies on confirming that

$$\sum_{n=1}^{\infty} \frac{\beta^n}{1 + \beta^{2n}} = \sum_{n=0}^{\infty} (-1)^n \frac{\beta^{2n+1}}{1 - \beta^{2n+1}}. \quad (4.4.59)$$

(Try expanding both sides as double sums.)

The second formula then follows by applying the first to  $\alpha^2$  and  $\beta^2$ , and then subtracting that result from the first to obtain  $(\theta_3^2(\beta) - \theta_3^2(\beta^2))/4$ , which equals  $\theta_2^3(\beta^2)/4$ .  $\square$

Two immediate consequences are

$$\sum_{n=0}^{\infty} \frac{1}{F_{2n+1}} = \frac{\sqrt{5}}{4} \theta_2^2 \left( \frac{3 - \sqrt{5}}{2} \right) \quad (4.4.60)$$

$$\sum_{n=1}^{\infty} \frac{1}{L_{2n}} = \frac{1}{4} \theta_3^2 \left( \frac{3 - \sqrt{5}}{2} \right) + \frac{1}{4}. \quad (4.4.61)$$

Two somewhat more elaborate derivations, (see [44] section 3.7), lead to

$$\sum_{n=1}^{\infty} \frac{1}{F_n^2} = \frac{5}{24} \left( \theta_2^4 \left( \frac{3 - \sqrt{5}}{2} \right) - \theta_4^4 \left( \frac{3 - \sqrt{5}}{2} \right) + 1 \right) \quad (4.4.62)$$

$$\sum_{n=1}^{\infty} \frac{1}{L_n^2} = \frac{1}{8} \left( \theta_3^4 \left( \frac{3 - \sqrt{5}}{2} \right) - 1 \right). \quad (4.4.63)$$

Since it is known that the classical theta functions are transcendental for algebraic values  $q, 0 < |q| < 1$ , we discover the far from obvious result that the left-hand side of each of (4.4.60), (4.4.61), (4.4.63) is a transcendental number, as probably is (4.4.62).

Moreover, since both the initial sums and especially the theta functions are easy to compute numerically, we can hunt for other such identities using integer relation methods. In this way, we find:

$$\sum_{n=1}^{\infty} \frac{(-1)^n}{F_n^2} = \frac{5}{48} \left( 2 - \theta_2^4 \left( \frac{3 - \sqrt{5}}{2} \right) - 2 \theta_4^4 \left( \frac{3 - \sqrt{5}}{2} \right) \right), \quad (4.4.64)$$

and a host of more recondite identities.

By contrast, a remarkable elementary identity is

$$\sum_{n=0}^{\infty} \frac{1}{F_{2n+1} + F_{2k-1}} = \frac{(2k-1)\sqrt{5}}{2F_{2k-1}}, \quad (4.4.65)$$

for  $k = 1, 2, 3, \dots$ . So while  $\sum_{n=0}^{\infty} F_{2n+1}^{-1}$  is transcendental,  $\sum_{n=0}^{\infty} (F_{2n+1} + 1)^{-1} = \sqrt{5}/2$ . If we compute the corresponding continued fractions of the two sums, we obtain the quite different results  $[1, 1, 4, 1, 2, 3, 6, 2, 1, 3, 1, 189, 1, 3, 12]$  and  $[1, 8, 2, 8, 2, 8, 2, 8, 2, 8]$  in partial confirmation.

## 4.5 A Characteristic Polynomial Triumph

We illustrate the possibilities of computing with symbolic characteristic polynomials, with an example arising in partial factorizations relating to double Euler sums (see Section 3.5). The rationale for looking at these matrices was discussed in the previous chapter. Consider  $n \times n$  matrices  $A, B, C, M$ :

$$A_{kj} = (-1)^{k+1} \binom{2n-j}{2n-k}, \quad B_{kj} = (-1)^{k+1} \binom{2n-j}{k-1},$$

$$C_{kj} = (-1)^{k+1} \binom{j-1}{k-1}$$

( $k, j = 1, \dots, n$ ) and a composite matrix

$$\boxed{M = A + B - C.}$$

We aim to prove  $M$  is invertible, indeed that

$$\boxed{M^{-1} = \frac{M + I}{2}.}$$

The key is discovering

$$A^2 = C^2 = I, \quad BC = A, \quad B^2 = CA. \quad (4.5.66)$$

It follows that  $B^3 = BCA = AA = I$ , and that the *group generated by  $A, B$  and  $C$  is the symmetric group  $S_3$* . Once (4.5.66) is discovered, combinatorial proofs are quite routine—either for a human or a computer—as we now show. It will help to look at Lemma 3.5.2.

*Proof of  $A^2 = I$ :*

$$\begin{aligned} (A^2)_{kj} &= (-1)^{k+1} \sum_{i=1}^n (-1)^{i+1} \binom{2n-i}{2n-k} \binom{2n-j}{2n-i} \\ &= (-1)^{k+1} \sum_{i=n+1}^{2n} (-1)^i \binom{i-1}{2n-k} \binom{2n-j}{i-1} = (-1)^{k+1} (-1)^{j+1} \delta_{kj}. \end{aligned}$$

*Proof of  $C^2 = I$ :*

$$(C^2)_{kj} = (-1)^{k+1} \sum_{i=1}^n (-1)^{i+1} \binom{i-1}{k-1} \binom{j-1}{i-1} = (-1)^{k+1} (-1)^{j+1} \delta_{kj}.$$

*Proof of  $BC = A$ :*

$$(BC)_{kj} = (-1)^{k+1} \sum_{i=1}^n (-1)^{i+1} \binom{2n-i}{k-1} \binom{j-1}{i-1} = (-1)^{k+1} \binom{2n-j}{2n-k}.$$

It follows also that  $AC = BC^2 = B$ , and similarly  $AB = AAC = C$ .

*Proof of  $B^2 = CA$ :*

$$\begin{aligned}
 (B^2)_{kj} - (CA)_{kj} &= (-1)^{k+1} \sum_{i=1}^n (-1)^{i+1} \binom{2n-i}{k-1} \binom{2n-j}{i-1} \\
 &\quad - (-1)^{k+1} \sum_{i=1}^n (-1)^{i+1} \binom{i-1}{k-1} \binom{2n-j}{2n-i} \\
 &= (-1)^{k+1} \sum_{i=1}^n (-1)^{i+1} \binom{2n-i}{k-1} \binom{2n-j}{i-1} \\
 &\quad + (-1)^{k+1} \sum_{i=n+1}^{2n} (-1)^{i+1} \binom{2n-i}{k-1} \binom{2n-j}{i-1} \\
 &= (-1)^{k+1} \binom{j-1}{2n-k} = 0.
 \end{aligned}$$

Then  $B^3 = BCA = A^2 = I$  follows from the other identities. Finally,  $CB = AB^2 = ACA = BA$ , and the group is indeed  $S_3$ .

Additionally, one now easily shows

$$\boxed{M^2 + M = 2I}$$

as formal algebra, using (4.5.66) and its consequences, since  $M = A + B - C$ .

The truth is that after unsuccessfully peering at various instances of  $M$  the authors of ([50]) decided to look at instances of “*minpoly*( $M, x$ )” and then emboldened tried “*minpoly*( $B, x$ )” in *Maple*, when the minimal polynomial for all  $n < 8$  of  $M$  was the same quadratic  $t^2 + t - 2$ . By contrast random  $n \times n$  matrices have full degree minimal polynomials, as is guaranteed by the Cayley-Hamilton theorem.

By chance a much weaker related fact appeared as *American Mathematical Monthly* Problem 01735 in 1999.

**Problem.** If  $L_n$  is the  $n$ -by- $n$  matrix with  $i, j$ -entry equal to  $\binom{i-1}{j-1}$  then  $L_n^2 \equiv I_n \pmod{2}$ , where  $I_n$  is the  $n$ -by- $n$  identity matrix. Show that if  $R_n$  is the  $n$ -by- $n$  matrix with  $i, j$ -entry equal to  $\binom{i-1}{n-j}$  then  $R_n^3 \equiv I_n \pmod{2}$ .

**Solution:** Let  $A, B, C$  be the  $n \times n$  matrices with  $i, j$ -entries given by

$$A_{ij} = (-1)^j \binom{n-i}{n-j}, \quad B_{ij} = (-1)^j \binom{i-1}{n-j}, \quad C_{ij} = (-1)^j \binom{i-1}{j-1}.$$

Since  $L_n \equiv C \pmod{2}$  and  $R_n \equiv B \pmod{2}$ , it suffices to prove that  $C^2 = I_n$  and  $B^3 = -I_n$ , which is entirely analogous to the proof of (4.5.66) given above, and also  $A^2 = I_n$ . By contrast the minimum polynomial of  $L_n$  is  $t \mapsto (t-1)^n$  and that for  $R_n$  is less elegant.

In a related analysis, for  $n > 3$ , however, the corresponding  $(n-1) \times (n-1)$  matrix  $\tilde{M} = \tilde{A} + \tilde{B} - \tilde{C}$ , with

$$\begin{aligned}\tilde{A}_{kj} &= (-1)^{k+1} \binom{2n-j-1}{2n-k-1}, & \tilde{B}_{kj} &= (-1)^{k+1} \binom{2n-j-1}{k-1}, \\ \tilde{C}_{kj} &= (-1)^{k+1} \binom{j-1}{k-1}, & \text{for } j, k &= 1, \dots, n-1,\end{aligned}$$

arose, and has minimal polynomial  $\tilde{M}^3 + 2\tilde{M}^2 - 3\tilde{M} = 0$ .

This may be proved much in the same way as in the previous case. It follows from analysis of the trace of  $\tilde{M}$  and of  $\tilde{M}^2$  that the number of null eigenvalues is  $\lfloor (n-1)/3 \rfloor$  and, since the minimal polynomial has no repeated roots, that the dimension of the null space is  $\lfloor (n-1)/3 \rfloor$ .  $\square$

The characteristic or the minimal polynomial, like partial fractions, is an object brought fully to life by computation. In much the same way *Jordan Forms* and other normal forms can be productively used to study singular values—in the matricial sense!

## 4.6 Commentary and Additional Examples

1. **Partitions with all parts odd.** Prove—analytically and combinatorially—that the number of partitions of  $n$  with all parts odd equals the number of partitions of  $n$  into distinct parts.
2. **Partition count of  $5n+4$  is divisible by 5.**

**Proof Sketch.** With  $Q$  as in the text above, we obtain

$$\begin{aligned}qQ^4(q) &= qQ(q)q^3(q) && (4.6.67) \\ &= \sum_{m \geq 0} \sum_{n = -\infty}^{\infty} (-1)^{n+m} (2m+1) q^{1+(3n+1)n/2+m(m+1)/2},\end{aligned}$$

from the triple product and pentagonal number theorems.

Now consider when  $k$  is a multiple of 5, and discover this can only happen if  $2m + 1$  is divisible by 5 as is the coefficient of  $q^{5m+5}$  in  $qQ^4(q)$ . Then by the binomial theorem

$$(1 - q)^{-5} \equiv (1 - q^5)^{-1} \pmod{5}$$

and so the coefficient of the same term in  $qQ(q^5)/Q(q)$  is divisible by 5. Finally

$$q + \sum_{n>1} p(n-1)q^n = qQ^{-1}(q) = \frac{qQ(q^5)}{Q(q)} \prod_{m=1}^{\infty} \sum_{n=0}^{\infty} q^{5mn},$$

as claimed.

3. **A combinatorial determinant problem.** Find the determinant of

$$\begin{bmatrix} \binom{n}{p} & \binom{n}{p+1} & \binom{n}{p+2} \\ \binom{n+1}{p} & \binom{n+1}{p+1} & \binom{n+1}{p+2} \\ \binom{n+2}{p} & \binom{n+2}{p+1} & \binom{n+2}{p+2} \end{bmatrix}$$

$$\begin{bmatrix} \binom{n}{p} & \binom{n}{p+1} & \binom{n}{p+2} & \binom{n}{p+3} \\ \binom{n+1}{p} & \binom{n+1}{p+1} & \binom{n+1}{p+2} & \binom{n+1}{p+3} \\ \binom{n+2}{p} & \binom{n+2}{p+1} & \binom{n+2}{p+2} & \binom{n+2}{p+3} \\ \binom{n+3}{p} & \binom{n+3}{p+1} & \binom{n+3}{p+2} & \binom{n+3}{p+3} \end{bmatrix}$$

and its  $q$ -dimensional extension as a function of  $n, p, q$ . (Taken from [127].)

Solution: The pattern is clear from the first few cases on simplifying in *Maple* or *Mathematica*.

4. **A sum-of-powers determinant.** Problem: Find the determinant of

$$\begin{bmatrix} \sum_{k=0}^1 k^4 & \sum_{k=0}^1 k^4 & \sum_{k=0}^1 k^4 & \sum_{k=0}^1 k^4 \\ \sum_{k=0}^1 k^4 & \sum_{k=0}^2 k^4 & \sum_{k=0}^2 k^4 & \sum_{k=0}^2 k^4 \\ \sum_{k=0}^1 k^4 & \sum_{k=0}^2 k^4 & \sum_{k=0}^3 k^4 & \sum_{k=0}^3 k^4 \\ \sum_{k=0}^1 k^4 & \sum_{k=0}^2 k^4 & \sum_{k=0}^3 k^4 & \sum_{k=0}^4 k^4 \end{bmatrix}$$



and its  $q$ -dimensional extension. (Taken from [127].)

Solution: The first few instances of this sequence are

$$1, 4, 216, 331776, 24883200000, 139314069504000000,$$

which can be quickly identified as  $(q!)^q$ , using the Sloane online sequence recognition tool.

This fact can be proved by taking cofactors on the last row, and observing that only the final two entries have nonzero cofactors with value  $(q-1)!^{q-1}$ .

5. **Putnam problem 1995–B3.** For each positive integer with  $n^2$  digits write the digits as a square matrix in order row by row. Thus 2354 becomes  $\begin{bmatrix} 2 & 3 \\ 5 & 4 \end{bmatrix}$ . Find, as a function of  $n$ , the sum of all  $9 \cdot 10^{n^2-1}$  such determinants, which arise on assuming that leading digits are non-zero.

Hint: With the help of a symbolic math program, observe that almost all sample matrices of this form have zero determinant. Then use multilinearity of the determinant to reduce the problem to computing the determinant of just one  $n \times n$  matrix.

Answer: For  $n = 1$  the answer is 45. For  $n = 2$  the matrix may be taken to be  $\begin{bmatrix} 450 & 405 \\ 450 & 450 \end{bmatrix}$ , with determinant 20250. For  $n > 2$  the value is zero.

6. **Crandall's integral representation for Madelung's constant.** The following identity is both beautiful and effective—though less effective for computational purposes than Benson's formula. For example, sixty digits of  $\mathcal{M}_3(1)$  can be obtained in seconds in *Maple* or *Mathematica* using Benson's identity, while using the numerical quadrature tools of Section 7.4 to compute the integral to the same 60 digit precision takes roughly one hour run time. Crandall's formula is derived in [90] from the Andrews formula for  $\theta_4^3$ . It is

$$\begin{aligned} \mathcal{M}_3(1) &= -\frac{2}{\pi} \int_0^1 r \, dr \int_{-\pi}^{\pi} \frac{1 + 2/(1 + r^{2(1-\sin \theta)})}{(1 + r^{1+\cos \theta})(1 + r^{1-\cos \theta})} \, d\theta \\ &= -1.7475645946332 \dots \end{aligned} \tag{4.6.68}$$

7. **Repeated exponential integrals.** Show that

(a)

$$\sqrt{\frac{2}{\pi}} \int_0^\infty \frac{e^{-x} \sin(ax)}{\sqrt{x}} dx = \frac{\sqrt{\sqrt{1+a^2}-1}}{\sqrt{1+a^2}},$$

(b)

$$\sqrt{\frac{2}{\pi}} \int_0^\infty \int_0^\infty \frac{e^{-x-y} \sin(ax+by)}{\sqrt{x+y}} dx dy = \frac{\frac{\sqrt{1+\sqrt{1+b^2}}}{\sqrt{1+b^2}} - \frac{\sqrt{1+\sqrt{1+a^2}}}{\sqrt{1+a^2}}}{a-b},$$

and evaluate

(c)

$$\sqrt{\frac{2}{\pi}} \int_0^\infty \int_0^\infty \int_0^\infty \frac{e^{-x-y-z} \sin(ax+by+cz)}{\sqrt{x+y+z}} dx dy dz$$

for real coefficients  $a, b$  and  $c$ .

(d) Generalize the results above, by dimension and to cosines.

(e) Evaluate

$$\int_0^\infty \cdots \int_0^\infty \frac{e^{-(\sum_{k=1}^n x_k)}}{\sqrt{\sum_{k=1}^n x_k}} dx_1 \cdots dx_n,$$

for  $n = 1, 2, \dots$ . This shows the prior integrals are absolutely convergent.

8. **Andrew's convolution.** Prove or disprove that (4.2.17) preserves integrality of coefficients in both directions.

Hint: To determine whether  $\{b_n\}$  is integer if and only if  $\{a_k\}$  is, in one direction expand  $\prod_{k>0} (1-q^k)^{-b_k}$  by the binomial theorem and note that the coefficients are integers when the values of  $b_k$  are. In the other direction, observe that  $b_1 = a_1$  and inductively consider

$$\prod_{k=1}^n (1-q^k)^{b_k} \left( 1 + \sum_{m=1}^{\infty} a_m q^m \right) = \prod_{n+1}^{\infty} (1-q^k)^{-b_k}.$$

This is the basis for an efficient algorithm but, in a modern computational package, (4.2.17) is very easy to program and likely to be more efficient,

especially as one will rarely want more than a few hundred terms of the product.

9. **Berkeley problem 6.13.15.** Determine the final digit of  $23^{23^{23^{23}}}$ . Answer: The last digit is a “7.” Hint: *Maple* or *Mathematica* can verify that  $23^{23} \equiv 7 \pmod{4}$  and  $23^{23^{23}} \equiv 7 \pmod{4}$ . To prove this observed trend, work modulo four and observe that as  $\phi(10) = 4$ ,  $3^r \equiv 3^s \pmod{10}$  when  $r \equiv s \pmod{4}$ . Then use  $3^{23^{23}} \equiv -1 \pmod{4}$ .
10. **A series with binomial coefficients.** Prove that for all  $n \geq 0$

$$\sum_{k=0}^n \frac{1}{\binom{n}{k}} = (n+1) \sum_{k=0}^n \frac{1}{(n-k+1)2^k}$$

Hint: Consider computing—in two different ways—the electrical resistance between two points distance  $n+1$  apart in the  $n$ -dimensional unit cube if every edge has unit resistance.

11. **A binomial coefficient inequality.** (From [136, pg. 137]). Show inductively for  $n > 1$  that

$$\frac{4^n}{n+1} < \binom{2n}{n} < 4^n,$$

and for  $n > 6$  that

$$\left(\frac{n}{3}\right)^n < n! < \left(\frac{n}{2}\right)^n.$$

12. **An  $n$ -th root inequality.** (From [136, pg. 162]). Show that for all nonnegative numbers

$$\sqrt[n]{(1+a_1)(1+a_2)\cdots(1+a_n)} \geq 1 + \sqrt[n]{a_1 a_2 \cdots a_n}.$$

13. **A polygon problem.** Count (i) the number of ways a polygon with  $n+2$  sides can be cut into  $n$  triangles, (ii) the number of ways in which parentheses can be placed in a sequence of numbers to be multiplied, two at a time; and (iii) the number of paths of length  $2n$  through an  $n$ -by- $n$  grid that do not rise above the main diagonal (*Dijk paths*).

Hint: In each case the sequence starts

$$1, 2, 5, 14, 42, 132, 429, 1430, 4862.$$

The “gfun” package returns the ordinary generating function  $4(1 + \sqrt{1 - 4x})^{-2}$  and the recursion  $(4n + 6)u(n) = (n + 3)u(n + 1)$ , which gives rise to the Catalan numbers  $\frac{1}{n+1} \binom{2n}{n}$  named after Eugène Charles Catalan (1814–1894).

14. **A cubic theta function identity.** If we define

$$\begin{aligned} a(q) &= \sum_{m,n \in \mathbb{Z}} q^{m^2 + mn + n^2} & b(q) &= \sum_{m,n \in \mathbb{Z}} \omega^{n-m} q^{m^2 + mn + n^2} \\ c(q) &= \sum_{m,n \in \mathbb{Z}} q^{(n+1/3)^2 + (n+1/3)(m+1/3) + (m+1/3)^2} \end{aligned}$$

where  $\omega = \exp(2\pi i/3)$ , then we have a remarkable cubic identity parallel to Jacobi’s quartic identity:

$$a^3 = b^3 + c^3 \tag{4.6.69}$$

and a lovely parameterization of the  ${}_2F_1$  hypergeometric function [49]:

$$F\left(\frac{1}{3}, \frac{2}{3}, 1; \frac{c^3}{a^3}\right) = a, \tag{4.6.70}$$

which we met in other guise in (6.1.16).

- (a) Choosing  $q = \exp(-2\pi\sqrt{N/3})$  for rational  $N$ , it can be shown that  $s_N = c/a$  is an algebraic number expressible by radicals; see [49]. If  $N$  is a positive integer, then  $s_N$  is the  $N$ -th *cubic singular value*. As above what can we discover computationally about  $s_N$ ? For example, can we determine radical formulae for the higher order cubic singular values? The following helps the computations. It is known that in terms of the classical theta functions

$$\begin{aligned} a(q) &= \theta_3(q)\theta_3(q^3) + \theta_2(q)\theta_2(q^3) \\ b(q) &= (3a(q^3) - a(q))/2 & c(q) &= (a(q^{1/3}) - a(q))/2. \end{aligned}$$

The lacunarity of these series allows for very rapid computation.

- (b) Compute the product formula for  $a$ —it is not very pretty.

$$a(q) = \frac{(1 - q^2)^{21} (1 - q^4)^{345} (1 - q^6)^{8906} (1 - q^8)^{250257} (1 - q^{10})^{7538421}}{(1 - q)^6 (1 - q^3)^{76} (1 - q^5)^{1734} (1 - q^7)^{46662} (1 - q^9)^{1365388} \dots}$$

(c) While  $a$  does not have a nice product, one should persevere

$$\begin{aligned} b(q) &= (1-q)^3 (1-q^2)^3 (1-q^3)^2 (1-q^4)^3 (1-q^5)^3 (1-q^6)^2 \\ &\times (1-q^7)^3 (1-q^8)^3 (1-q^9)^2 (1-q^{10})^3 (1-q^{11})^3 (1-q^{12})^2 \\ &\times (1-q^{13})^3 (1-q^{14})^3 (1-q^{15})^2 \dots \end{aligned}$$

This turns out to be the key in providing the computer-guided but very intuitive proof given in [49].

15. **Triangles inscribed in a sphere.** Show that  $3\sqrt{3}/4$  is the maximum area for triangles inscribed in a unit sphere, and is attained only by equilateral triangles inscribed in a great circle of the sphere.

16. **Nests of radicals.** Identify the limits of the following infinite nested radicals and establish a rigorous sense in which the evaluations are justified.

(a)

$$\sqrt{1 + 2\sqrt{1 + 3\sqrt{1 + 4\sqrt{1 + 5\cdots}}}}$$

(b)

$$\sqrt{6 + 2\sqrt{7 + 6\sqrt{2 + \sqrt{9 + 5\cdots}}}}$$

(c)

$$\sqrt[3]{4 + \sqrt[3]{10 + 9\sqrt[3]{16 + 25\sqrt[3]{22 + \cdots}}}}$$

(d)

$$\sqrt[3]{a + \sqrt[3]{a + \sqrt[3]{a + \cdots}}}$$

for  $a = \sqrt{5/3}$ .

(e)

$$\sqrt[p]{a + \sqrt[p]{a + \sqrt[p]{a + \cdots}}}$$

for  $p > 1, a > 0$ .

(f)

$$\sqrt{2 - \sqrt{2 - \sqrt{2 + \sqrt{2 - \sqrt{2 - \sqrt{2 + \sqrt{2 - \sqrt{2 - \dots}}}}}}}}}$$

Hint: Find a functional equation for a *large* class of functions so that iteration in that class solves the functional equation uniquely. Many such equations were evaluated informally by Ramanujan. More details are given in [35]. Answers: (a) 3; (b) 4; (c) 2; (d)  $\frac{2}{\sqrt{3}} \left( \sqrt[3]{\frac{1}{2}\sqrt{5} + \frac{1}{2}} + \sqrt[3]{\frac{1}{2}\sqrt{5} - \frac{1}{2}} \right)$ ; (e) the positive root of  $x^p = x + a$ ; (f) the positive root of  $x^3 + x^2 - 2x = 1$ .

17. **Some unconditional sums.** Problem: Evaluate

(a)

$$\sum_{k=1}^{\infty} \sum_{l=1}^{\infty} \frac{1}{(4l-1)^{2k}} \quad \left( = \frac{1}{4} \log(2) \right)$$

(b)

$$\sum_{k=1}^{\infty} \sum_{l=1}^{\infty} \frac{1}{(4l-1)^{2k+1}} \quad \left( = \frac{1}{8} \pi - \frac{1}{2} \log(2) \right)$$

(c)

$$\sum_{k=1}^{\infty} \sum_{l=1}^{\infty} \frac{1}{(4l-2)^{2k}} \quad \left( = \frac{1}{8} \pi \right)$$

18. **Some conditional sums.** Problem: Evaluate

(a)

$$\sum_{n=1}^{\infty} \sum_{m=1}^{\infty} \frac{(-1)^{m+n}}{(m+n)^s} \quad (= \alpha(s) - \alpha(s-1))$$

(where  $\alpha(s) = (1 - 2^{1-s})\zeta(s)$ ) and thence justify the conditional evaluation  $\sum_{n=1}^{\infty} \sum_{m=1}^{\infty} (-1)^{m+n} / (m+n) = \log(2) - \frac{1}{2}$ .

$$(b) \quad \sum_{n=1}^{\infty} \sum_{m=1}^{\infty} (-1)^{m+n} \frac{mn}{(m+n)^2} \quad \left( = \frac{1}{6} \log(2) - \frac{1}{24} \right)$$

$$(c) \quad \sum_{n=1}^{\infty} \sum_{m=1}^{\infty} \frac{m^2 - n^2}{(m^2 + n^2)^2} \quad \left( = \frac{\pi}{4} \right)$$

(Note that exchanging the order changes the sign of the answer.)

19. **A matrix problem.** Let

$$M = \begin{bmatrix} p & q & 1-p-q \\ 1-p-q & p & q \\ q & 1-p-q & p \end{bmatrix}$$

Determine the behavior of  $M^n$  when  $p+q < 1, p > 0, q > 0$ .

Answer:  $3M^n \rightarrow E$ , the matrix with all entries 1.

20. **Theta and self-similarity.** For  $k = 2$  or  $4$  plot the set

$$\mathcal{K}_k(\alpha) = \left\{ |q| < 1 : \left| \frac{\theta_k}{\theta_3}(q) \right| > \alpha, q \in \mathbf{C} \right\}$$

for various values of  $\alpha > 0$ . For appropriate  $k$ , this is shown in Figure 20 for  $\alpha = 1, 3/4$ . Recall that  $\frac{\theta_2}{\theta_3}(e^{-\pi s}) = \frac{\theta_4}{\theta_3}(e^{-\pi/s})$  for  $\text{Re}(s) > 0$ .

Use the complex AGM iteration [44] in  $\theta$ -form to explore the structures and relations suggested in Figure 20. More details on these and like images are given in [95].

David Mumford and his colleagues' book *Indra's Pearls* [165] offers a wealth of important and visually enticing material. In the gloss to their book they write:

It is the story of our computer aided explorations of a family of unusually symmetrical shapes, which arise when two spiral motions of a very special kind are allowed to interact. These shapes

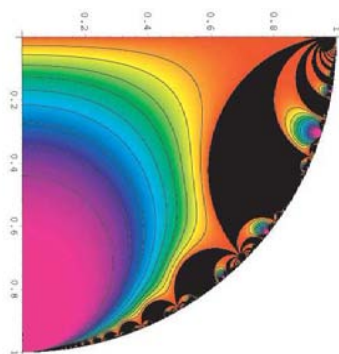
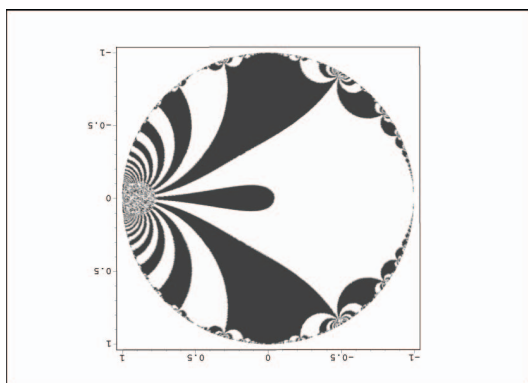


Figure 4.2: Where  $|\theta_2/\theta_3| > 1$  and where  $|\theta_4/\theta_3| > 1$  (first quadrant)



display intricate “fractal” complexity on every scale from very large to very small. Their visualisation forms part of a century-old dream conceived by the great German geometer Felix Klein.

21. **Putnam problem 1994–B4.** Let  $d_n$  be the greatest common divisor of the entries of  $A^n - I$  where  $A = \begin{bmatrix} 3 & 2 \\ 4 & 3 \end{bmatrix}$ . Show that  $d_n \rightarrow \infty$  with  $n$ . Hint: Observe numerically, then prove by induction, that  $A^n$  has determinant 1 and is of the form  $\begin{bmatrix} a_n & b_n \\ 2b_n & a_n \end{bmatrix}$ . Hence,  $a_n - 1 \mid 2b_n^2$ . Then write  $A^n$  explicitly via the Cayley-Hamilton theorem, which tells us that  $A^{n+1} = 6A^n - A^{n-1}$ .
22. **Putnam problem 1999–B5.** Evaluate the determinant of  $I + A$  where  $A_n$  is the  $n \times n$  matrix with entries  $\cos((j+k)\pi/n)$ .  
 Answer:  $A_n$  has determinant  $1 - n^2/4$ . Hint: The determinant equals  $\prod_{k=1}^n (1 + \lambda_k)$  where  $\lambda_k$  ranges over the eigenvalues of  $A_n$ . One may discover numerically that  $A_n$  has eigenvalue zero with multiplicity  $n - 2$  and remaining eigenvalues  $\pm n^2/4$ . Let  $v^{(m)}$  denote the vector with  $v_k^{(m)} = \exp(ikm2\pi/n)$ . One may also be lead to discover that the eigenvectors are  $v^{(0)}, v^{(2)}, v^{(3)}, \dots, v^{(n-2)}, v^{(1)} \pm v^{(n-1)}$ . This is then easy to formally confirm.
23. **Fibonacci and Lucas numbers in terms of hyperbolic functions.** Show that

$$F_n = \frac{2}{\sqrt{5}} i^{-n} \sinh(n\theta) \quad \text{and} \quad L_n = 2 i^{-n} \cosh(n\theta)$$

where

$$\theta = \log \left( \frac{\sqrt{5} + 1}{2} \right) + i \frac{\pi}{2}.$$

Many Fibonacci formulas are then easy to obtain from the addition formulas for  $\sinh$  and  $\cosh$ —for example consider  $5F_n^2 - L_n^2$ . (See [103], which should be consulted whenever one “discovers” a result in classical number theory.)

24. **Berkeley problem 7.5.25.** Let  $M_n^2$  be the  $n \times n$  tri-diagonal matrix with  $a_{ij} = 1$  if  $|i - j| = 1$  and all other entries zero. (a) Find the determinant of  $M_n^2$  and (b) show that the eigenvalues are symmetric around the origin. Answer: (a)  $\det(M_n^2) = \cos(n\pi/2)$ . (b) Compute the characteristic polynomial inductively and observe that it contains only odd (resp. even) powers for  $n$  odd (resp. even).

25. **Gersgorin circles.** Let  $A = \{a_{ij}\}$  be a real  $n \times n$  matrix.

(a) Show that if

$$|a_{ii}| > \sum_{\substack{t=1 \\ t \neq i}}^n |a_{it}|,$$

then  $A$  is nonsingular.

(b) Deduce that each eigenvalue of  $A$  lies in one of the discs

$$|z - a_{ii}| \leq \sum_{\substack{t=1 \\ t \neq i}}^n |a_{it}|.$$

Hint: (a) Consider the largest coordinate in absolute value of an element,  $x$ , in the null space of  $A$ .

26. **Putnam problem 1996–B4.** For a square matrix  $A$  define  $\sin(A)$  via the power series

$$\sin(A) = \sum_{n=0}^{\infty} \frac{(-1)^n}{(2n+1)!} A^{2n+1} \quad (4.6.71)$$

Prove or disprove that  $\begin{bmatrix} 1 & 1996 \\ 0 & 1 \end{bmatrix}$  is in the range of (4.6.71). Answer: It is not. Which matrices are?

Hint: Program the above in a symbolic math program, and observe what the range looks like for various input  $2 \times 2$  matrices. The result can be obtained by considering the normal form of any  $A$  with  $\sin(A) = \begin{bmatrix} 1 & 1996 \\ 0 & 1 \end{bmatrix}$ ,

which shows  $A$  can not be diagonalizable. Thus,  $A$  must have equal eigenvalues and so is conjugate to a matrix of the form  $B = \begin{bmatrix} a & b \\ 0 & a \end{bmatrix}$ .

Now determine that  $\sin(B) = \begin{bmatrix} \sin(x) & y \cos(x) \\ 0 & \sin(x) \end{bmatrix}$ . Since  $\det(\sin(A)) = \det(\sin(B))$  deduce that  $|\sin(x)| = 1$  and so  $\cos(x) = 0$ .

27. **Formula for zeta(2,1).** Obtain the formula for  $\zeta(2, 1)$  using the Cauchy-Lindelöf theorem applied to  $\pi \cot(\pi z)\Psi(-z)$  and  $1/z^2$ .

28. **A log-trig integral.** Show the following due to Victor Adamchik

$$\begin{aligned} & -\frac{1}{4\pi^2} \int_0^{2\pi} \int_0^{2\pi} \ln(3 - [\cos(x) + \cos(y) + \cos(x+y)]) \, dx \, dy \\ & = \frac{\pi}{\sqrt{3}} + \ln(2) - \frac{\Psi'(\frac{1}{6})}{2\sqrt{3}\pi}. \end{aligned}$$

29. **Repeated exponentiation.** Recursions like  $x_1 = t > 0$  and  $x_n = t^{x_{n-1}}$  for  $n > 0$  have been subject to considerable scrutiny.

(a) Study the existence and behavior of

$$\bar{x}_\infty = \lim_{n \rightarrow \infty} x_{2n}, \quad \underline{x}_\infty = \lim_{n \rightarrow \infty} x_{2n+1},$$

for  $0 \leq x \leq 1$ . Note that, in Figure 4.3, the even approximations on  $[0, 1]$  decrease while the odd ones increase. Thus, the limits are taken uniformly by Dini's theorem, which asserts that the monotone limit of continuous functions on a compact set is uniform if and only if the limit is continuous. (Compare Exercises 3 and 10 of Chapter 1.)

Hint: show that (i) the solution to  $t^x = x$  is  $t \mapsto -W(-\log t)/\log t$  and that (ii)  $\bar{x}_\infty(t)$  and  $\underline{x}_\infty(t)$  are the two solutions to  $t^{t^x} = x$ , which bifurcate at  $\hat{b} = \exp(-\exp(1)) \approx 0.06598803584$ . In terms of the real branches of the Lambert  $W$  function, they are portions of  $\exp((W_k(t \log t))/t)$  for  $k = 0, -1$  on  $[0, \exp(-1)]$  and on  $[\exp(-1), 1]$  respectively. The shared component is  $-W(-\log t)/\log t$  on  $[\hat{b}, 1]$  (see Figure 4.4). What happens for  $t > 1$ ? (The righthand asymptote in Figure 4.4 is at approximately 1.44466786100976613366.) This is discussed in [81].

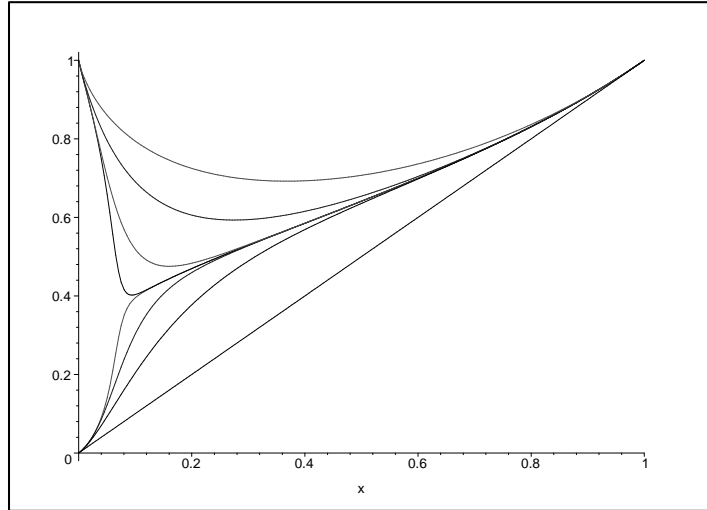


Figure 4.3: Approximations to  $x^{x^{x^{x^{\dots}}}}$

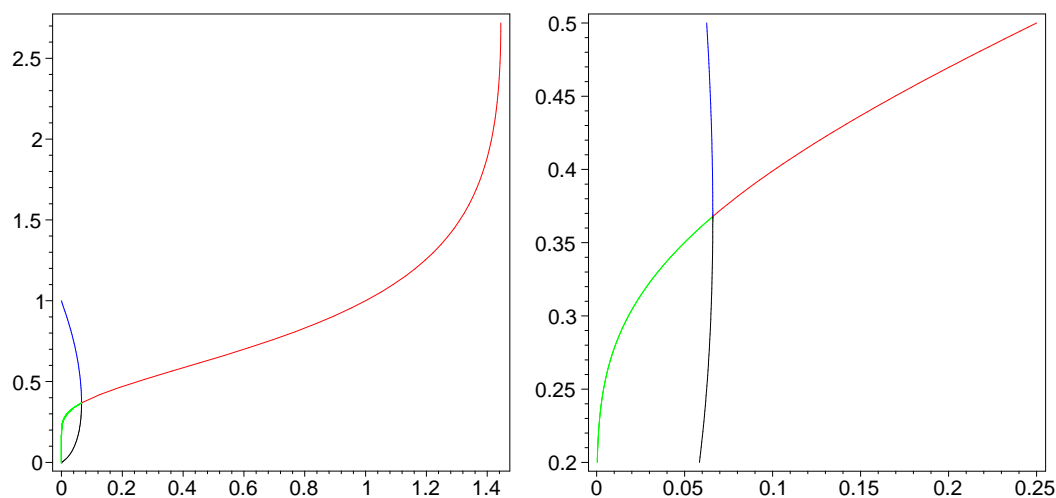
- (b) Estimate  $\int_0^1 \bar{x}_\infty(t) dt$  and  $\int_0^1 \underline{x}_\infty(t) dt$ .
- (c) How many distinct meanings may be assigned to the  $n$ -fold exponentiation  $x^{*n} = x^{x^{x^{\dots x}}}$ ?
- (d) Show that

$$\int_0^1 (x^x)^x dx = 1 + \frac{1}{2} \sum_{n=1}^{\infty} \frac{1}{2^n} \sum_{k=1}^n \frac{(-1)^{k+1}}{(k+1)^{n+1}} \binom{n}{k}$$

and that

$$\int_0^1 x^{(x^x)} dx = \frac{1}{2} + \sum_{n=1}^{\infty} (-1)^n \sum_{k=1}^n \frac{(n-k)^k}{(k+1)^{n+1}} \binom{n}{k} \quad (4.6.72)$$

where the sums are absolutely convergent. Try to further refine both these sums. A useful source for classical combinatorial identities is [183]

Figure 4.4: Solutions to  $x^{x^{x^{\dots}}}$ .

(e) Investigate  $\int_0^1 x^{*n} dx$  for more general  $n$ .

**Proof of (4.6.72).** Note that  $0 < x < x^{x^x} < 1$  when  $0 < x < 1$ , and hence, by dominated convergence, that

$$\lim_{\varepsilon \rightarrow 0^+} \int_0^1 x^\varepsilon x^{x^x} dx = \int_0^1 x^{x^x} dx.$$

In what follows suppose that  $\varepsilon > 0$ . We will prove that

$$\int_0^1 x^\varepsilon x^{x^x} dx = \sum_{n=0}^{\infty} \sum_{k=0}^{\infty} (-1)^{k+n} \frac{n^k}{(k+1+\varepsilon)^{n+k+1}} \binom{n+k}{k}, \quad (4.6.73)$$

where the double sum is absolutely convergent.

We have

$$\begin{aligned}
 \int_0^1 x^\varepsilon x^{x^x} dx &= \int_0^1 x^\varepsilon e^{x^x \log x} dx = \sum_{n=0}^{\infty} \frac{1}{n!} \int_0^1 x^\varepsilon (x^x \log x)^n dx \\
 &= \sum_{n=0}^{\infty} \frac{1}{n!} \sum_{k=0}^{\infty} \int_0^1 x^\varepsilon \log^n x \frac{(xn \log x)^k}{k!} dx \\
 &= \sum_{n=0}^{\infty} \sum_{k=0}^{\infty} \frac{n^k}{n!k!} \int_0^1 x^{k+\varepsilon} \log^{n+k} x dx.
 \end{aligned}$$

Observe next that

$$\begin{aligned}
 \sum_{n=0}^{\infty} \sum_{k=0}^{\infty} \frac{n^k}{n!k!} \int_0^1 x^{k+\varepsilon} |\log^{n+k} x| dx &= \sum_{n=0}^{\infty} \sum_{k=0}^{\infty} \frac{n^k}{n!k!} \int_0^1 x^{k+\varepsilon} \log^{n+k} \left( \frac{1}{x} \right) dx \\
 &= \sum_{n=0}^{\infty} \sum_{k=0}^{\infty} \frac{n^k}{n!k!} \int_1^{\infty} t^{-k-\varepsilon-2} \log^{n+k} t dt \\
 &= \int_1^{\infty} \frac{t^{t^{\frac{1}{t}}}}{t^{2+\varepsilon}} dt < \infty, \tag{4.6.74}
 \end{aligned}$$

since  $t^{t^{\frac{1}{t}}} \sim t$  as  $t \rightarrow \infty$ . It follows from (4.6.74) that the changes in order of sums and integrals in (4.6.74) are valid; and hence, on making the change of variable  $(k+1+\varepsilon) \log x = -u$  in the final integral in (4.6.74), that

$$\int_0^1 x^\varepsilon x^{x^x} dx = \sum_{n=0}^{\infty} \sum_{k=0}^{\infty} \frac{(-1)^{n+k} n^k}{n!k!(k+1+\varepsilon)^{n+k+1}} \int_0^{\infty} u^{n+k} e^{-u} du,$$

which establishes (4.6.73).

We have shown that

$$\int_0^1 x^x dx = \lim_{\varepsilon \rightarrow 0^+} \sum_{n=0}^{\infty} \sum_{k=0}^{\infty} (-1)^{k+n} \frac{n^k}{(k+1+\varepsilon)^{n+k+1}} \binom{n+k}{k}.$$

We can not automatically replace the limit by

$$\sum_{n=0}^{\infty} \sum_{k=0}^{\infty} (-1)^{k+n} \frac{n^k}{(k+1)^{n+k+1}} \binom{n+k}{k},$$

for this double sum is not absolutely convergent (as is shown by (4.6.74) with  $\varepsilon = 0$ ) but on treating the terms with  $n = 0$  and  $n > 0$  separately (as above) one does obtain (4.6.72).

30. **Fibonacci squares.** Decide whether or not 144 is the only square Fibonacci number. Relatedly,  $1729 = 12^3 + 1 = 10^3 + 9^3$  is the smallest integer with two distinct representations as a sum of two cubes. What is the next instance?
31. **Fibonacci.** A more precise description of Fibonacci's question mentioned in Section 3.4 is taken from *Scritti di Leonardo Pisano*, pubblicati da Baldassare Boncompagni, Roma, 1857 (Vol. 1), Liber Abaci, chapter XII. This corresponds to folio pages 123 verso and 124 recto of the manuscript *Conversi Soppressi C.1. nr 2616, Codice Magliabechiano* (Biblioteca di Firenze):

Quot paria coniculorum in uno anno ex uno pario germinentur. Quidam posuit unum par coniculorum in quodam loco, qui erat undique pariete circondatus, ut sciret, quot ex eo paria germinerentur in uno anno: cum natura eorum sit per singulum mensem aliud par germinare; et in secundo mense ab eorum natiuitate germinant. Quia suprascriptum par in primo mense germinat, duplicabis ipsum, erunt paria duo in uno mense. Ex quibus unum, scilicet primum, in secundo mense geminat; et sic sunt in secundo mense paria 3; ex quibus, in uno mense duo pregnantur; et ...

L.E. Sigler translates this as [191]:

How Many Pairs of Rabbits Are Created by One Pair in One Year.

A certain man had one pair of rabbits together in a certain enclosed place, and one wishes to know how many are created

from the pair in one year when it is the nature of them in a single month to bear another pair, and in the second month those born to bear also. Because the above written pair in the first month bore, you will double it; there will be two pairs in one month. One of these, namely the first, bears in the second month, and thus there are in the second month 3 pairs; of these in one month two are pregnant, and ...

Note that it is assumed rabbits are immortal! Alternatively, assume pairs breed after one month and die after two reproductions.



# Chapter 5

## Primes and Polynomials

I hope that . . . I have communicated a certain impression of the immense beauty of the prime numbers and the endless surprises which they have in store for us.

Don Zagier, 1977 [211]

Many computational number theory problems involve careful combinatorial search techniques. To be successful the search space must be reasonably sized. Often a mathematical insight is essential in trimming the search space, and thus much more helpful than the ability to do fast computations.

### 5.1 Giuga's Prime Number Conjecture

Giuga's prime number conjecture was formulated in 1950 by G. Giuga [119].

**Conjecture 5.1.1 Giuga's conjecture.** *A number  $n \in \mathbb{N}$ ,  $n > 1$ , is a prime if and only if*

$$s_n = \sum_{k=1}^{n-1} k^{n-1} \equiv n - 1 \pmod{n}.$$

Whenever  $n$  is a prime, then  $s_n$  must be congruent to  $n - 1$ . This is a consequence of Fermat's little theorem: *If  $p$  is a prime, then  $k^{p-1} \equiv 1 \pmod{p}$  for all  $k = 1, \dots, p - 1$ .* Thus the question is if there are composite numbers  $n$  with  $s_n \equiv n - 1 \pmod{n}$ . It is known that such a number  $n$  (a

counterexample) must have at least 14,000 decimal digits, if it exists at all. This is a huge number! It is clear that this bound has not been computed by checking every composite number  $n$  up to  $n = 10^{14,000}$ . Instead, Giuga's problem has been converted to a combinatorial search problem, which we shall now describe in detail.

### 5.1.1 Computation of Exclusion Bounds

The following reformulation of the problem, given already by Giuga, is essential.

**Theorem 5.1.2** *Let  $n \in \mathbb{N}$ ,  $n \geq 1$ , be given. Then  $s_n \equiv n-1 \pmod{n}$  if and only if for every prime divisor  $p$  of  $n$  the relations  $(p-1) \mid (n/p-1)$  and  $p \mid (n/p-1)$  hold.*

**Proof.** As a preliminary consideration, write  $n = p \cdot q$  and note that  $s_n \equiv q \cdot \sum_{k=1}^n k^{n-1} \equiv -q \pmod{p}$  if  $(p-1) \mid (n-1)$  by Fermat's little theorem, and  $s_n \equiv 0 \pmod{p}$  if  $(p-1) \nmid (n-1)$  since the multiplicative group mod  $p$  is cyclic.

Now assume that  $s_n \equiv -1 \pmod{n}$ . Then the preliminary consideration implies that  $(p-1) \mid (n-1)$  and  $s_n \equiv -1 \equiv -q \pmod{p}$  since  $s_n \equiv 0 \pmod{p}$  is impossible. Thus also  $(p-1) \mid (q-1)$  and  $p \mid (q-1)$ .

On the other hand, assume that  $(p-1) \mid (q-1)$  and  $p \mid (q-1)$ . Then also  $(p-1) \mid (n-1)$ , and thus  $s_n \equiv -q \pmod{p}$  by the preliminary consideration. The second assumption  $q \equiv 1 \pmod{p}$  implies that  $s_n \equiv -1 \pmod{p}$ , thus  $p \mid (s_n + 1)$ . This holds for every prime divisor  $p$  of  $n$ , and since  $n$  must be squarefree (as  $p \mid (q-1)$  implies  $p^2 \nmid n$ ), we have that  $n \mid (s_n + 1)$ , thus  $s_n \equiv -1 \pmod{n}$ .  $\square$

Composite squarefree numbers  $n$  which satisfy  $(p-1) \mid (n-1)$  (or equivalently  $(p-1) \mid (n/p-1)$ ), for each prime divisor  $p$  of  $n$ , are called *Carmichael numbers*. Interestingly, they are pseudo-primes in the sense that Fermat's little theorem is satisfied by them in the following form: The Carmichael numbers are precisely the composite numbers  $n$  for which  $n \mid (k^n - k)$ , for every  $k \in \mathbb{N}$ . This is Korselt's criterion (1899). The smallest Carmichael numbers are 561, 1105, and 1729. It has only recently been proved by Alfors, Granville and Pomerance [5] that there are infinitely many Carmichael numbers. Note that if  $p$  is a prime divisor of a Carmichael number  $n$ , then for no  $k \in \mathbb{N}$  can  $kp + 1$  be a prime divisor of  $n$ ;

otherwise, we would have  $kp \mid (n - 1)$ , a contradiction to  $p \mid n$ . In particular, every Carmichael number is odd.

Correspondingly, we call an integer  $n$  with  $p \mid (n/p - 1)$  for every prime divisor  $p$  of  $n$  a *Giuga number*. As seen in the proof of Theorem 5.1.2, every Giuga number is squarefree. The smallest examples are 30, 858, and 1722; at the moment only 13 Giuga numbers are known. The largest Giuga number found so far has 97 digits and 10 prime factors (one has 35 digits). It is not known if there are infinitely many Giuga numbers. Nor is it known if there is an odd Giuga number or if any  $n$  can be Giuga and Carmichael at the same time.

The following characterization of Giuga numbers is very helpful for finding such numbers. It is moreover needed in the computation of exclusion bounds for counterexamples.

**Theorem 5.1.3** *An  $n \in \mathbb{N}$  with  $n > 1$  is a Giuga number if and only if it is squarefree and satisfies*

$$\sum_{p|n} \frac{1}{p} - \prod_{p|n} \frac{1}{p} \in \mathbb{N} \quad (5.1.1)$$

where the sum and the product go over all prime divisors  $p$  of  $n$ .

**Proof.** Write  $n = p_1 \cdots p_m$  and  $q_i = n/p_i$ . Then  $n$  is a Giuga number iff  $p_i \mid (q_i - 1)$  for all  $i$  iff  $p_i \mid (q_1 + \cdots + q_m - 1)$  for all  $i$  iff  $n \mid (q_1 + \cdots + q_m - 1)$  iff  $(q_1 + \cdots + q_m - 1)/n = \sum \frac{1}{p} - \prod \frac{1}{p} \in \mathbb{N}$ .  $\square$

By Theorem 5.1.3, the sum over the reciprocals of the prime divisors of a Giuga number must be greater than one. Thus an odd Giuga number must have at least 9 distinct prime factors. Since the product of the 9 smallest odd primes has 10 decimal digits, this already proves, without much computation, and certainly without checking billions of numbers, that a counterexample to Giuga's conjecture must be greater than  $10^9$ . This proof even does not take into account that any counterexample must also be a Carmichael number. Since this condition further restricts the possible prime factors in any counterexample, its systematic use will increase the lower bound significantly. This idea is precisely what lead to the computation of the lower bound of  $10^{14,000}$ , and therefore we shall now explain it in more detail. From now on, denote by  $q_k$  the  $k$ -th odd prime; thus  $q_1 = 3, q_2 = 5, q_3 = 7, \dots$ . Denote further  $Q = \{3, 5, 7, 11, \dots\}$  and  $Q_k = \{q_1, q_2, \dots, q_{k-1}\}$ .

The precise criterion that is used for these computations follows, as we have seen above, from the Carmichael- and Giuga-ness of any counterexample: *If a number  $n = p_1 \cdots p_m$  with  $m > 1$  prime factors  $p_i$  is a counterexample to Giuga's conjecture (i.e., satisfies  $s_n \equiv n - 1 \pmod{n}$ ), then  $p_i \neq p_j$  and  $p_i \not\equiv 1 \pmod{p_j}$  for  $i \neq j$ , and  $\sum_{i=1}^m \frac{1}{p_i} > 1$ .*

**Definition 5.1.4** *A set of primes is normal if it contains no primes  $p, q$  such that  $p$  divides  $q - 1$ .*

Thus if 3 is in a normal set of primes, then 7, 13, 19,  $\dots$  cannot be prime, and if 5 belongs then 11, 31, 41,  $\dots$  are excluded. The prime factors of any counterexample form a normal set.

Denote by  $\mathcal{N}_k$  the system of all normal subsets of  $Q_k$ . For example,  $\mathcal{N}_1 = \{\{\}\}$  and  $\mathcal{N}_2 = \{\{\}, \{3\}\}$ . For fixed  $k$  and  $N \in \mathcal{N}_k$ , denote by  $T_k(N)$  the subset of  $Q$  determined by the following algorithm: (1) start with  $T = N$  and  $j = k$ ; (2) while  $\sum_{p \in T} \frac{1}{p} \leq 1$  do: if  $N \cup \{q_j\}$  is normal then  $T = T \cup \{q_j\}$ ; fi;  $j = j + 1$ ; od; (3) return  $T_k(N) = T$ . By Exercise 4 below, this algorithm always terminates and produces a set  $T_k(N)$ .

It is clear that the sets  $T_k(N)$  have the following properties:

- (i)  $N \subseteq T_k(N)$ ,
- (ii) if  $p \in T_k(N) \setminus N$ , then  $p \geq q_k$  and  $N \cup \{p\}$  is normal,
- (iii)  $\sum_{p \in T_k(N)} \frac{1}{p} > 1$ , but  $\sum_{\substack{p \in T_k(N) \\ p \neq q}} \frac{1}{p} \leq 1$  for every prime  $q \in T_k(N) \setminus N$ .

We will be mainly interested in the number of elements of the sets  $T_k(N)$  as well as their product; thus set  $j_k(N) = |T_k(N)|$ ,  $P_k(N) = \prod_{p \in T_k(N)} p$  and  $j_k = \min\{r_k(N) : N \in \mathcal{N}_k\}$ ,  $P_k = \min\{P_k(N) : N \in \mathcal{N}_k\}$ .

For example,

$$\begin{aligned} T_1(\{\}) &= \{3, 5, 7, 11, 13, 17, 19, 23, 29\}, & j_1(\{\}) &= 9 & \text{and } P_1(\{\}) &> 10^9, \\ T_2(\{\}) &= \{5, 7, 11, 13, 17, \dots, 107, 109\}, & j_2(\{\}) &= 27 & \text{and } P_2(\{\}) &> 10^{42}, \\ T_2(\{3\}) &= \{3, 5, 11, 17, \dots, 317, 347\}, & j_2(\{3\}) &= 36 & \text{and } P_2(\{3\}) &> 10^{71}. \end{aligned}$$

Thus,  $j_1 = 9$ ,  $P_1 > 10^9$  and  $j_2 = 27$ ,  $P_2 > 10^{42}$ . We have already noted that, since the sum over the reciprocals of the prime factors of a counterexample  $n$  must exceed one,  $n$  must contain at least 9 prime factors and have at least 10 decimal digits. This is recaptured by the computation of  $j_1$  and  $P_1$  above. However, the computation of  $j_2$  and  $P_2$  now gives a better exclusion bound: Any

counterexample  $n$  either does or doesn't contain the prime factor 3. If it doesn't, then for every  $j$ , the  $j$ -th largest element of  $T_2(\{\})$  is a lower bound for the  $j$ -th largest prime factor of  $n$ . Therefore  $n$  must have at least 27 different prime factors and must be larger than  $10^{42}$ . If  $n$  does contain the prime factor 3, then the elements of  $T_2(\{3\})$  are lower bounds for the prime factors of  $n$ . Therefore, in this case  $n$  has at least 36 distinct prime factors and must be larger than  $10^{71}$ .

In general, we can conclude that every counterexample  $n$  has at least  $j_k$  distinct prime factors and exceeds  $P_k$ , for every  $k \in \mathbb{N}$ . Both  $j_k$  and  $P_k$  are increasing in  $k$ , as we will see shortly. To find exclusion bounds for counterexamples, one therefore has to compute  $j_k$  (then  $n \geq \prod_{j=1}^{j_k} q_j$ ) or  $P_k$  (then  $n \geq P_k$ ), for values of  $k$  as high as possible. On the surface, this is a daunting task, since the number of normal sets,  $|\mathcal{N}_k|$ , probably grows exponentially in  $k$ . Matters can be accelerated slightly by noting that, since  $j_k$  and  $P_k$  are minimal values, we do not have to continue the computation of a certain  $j_k(N)$  or  $P_k(N)$  if their intermediate values already exceed a previously computed  $j_k(\tilde{N})$  or  $P_k(\tilde{N})$ .

In this way, with 1950's technology Giuga [119] computed  $j_8 = 323$  and estimated  $j_9 > 361$ ; this gives an exclusion bound of  $\prod_{j=1}^{361} q_j > 10^{1039}$ . Later, in 1985, E. Bedocchi [22] computed  $j_9 = 554$ ; this gives a bound of  $\prod_{j=1}^{554} q_j > 10^{1716}$ . In 1994, one of the present authors, with three coauthors [29], computed by the same method  $j_{19} = 825$ , and thus a lower bound of  $10^{2722}$ . But at this point the method seems exhausted; because of the exponential growth, not much more information can be expected from additional computational.

However, appearances are deceiving—the story does not end here. With one additional insight, the exclusion bounds can be forced up significantly. This insight is to see that the systems  $\mathcal{N}_k$  have a tree structure. Consider a normal set  $N \in \mathcal{N}_k$ . Then this set has one or two successors in  $\mathcal{N}_{k+1}$ . The first successor is  $N$  itself; it is normal and thus contained in  $\mathcal{N}_{k+1}$ . The second successor may be the set  $N' = N \cup \{q_k\}$ ; this set is contained in  $\mathcal{N}_{k+1}$  iff it is normal. The insight now is that the  $j$  and  $P$  values always increase from  $N$  to its successors.

**Theorem 5.1.5** *For each  $k$ ,  $j_{k+1}(N) \geq j_k(N)$ ,  $j_{k+1}(N') \geq j_k(N)$  and  $P_{k+1}(N) \geq P_k(N)$ ,  $P_{k+1}(N') \geq P_k(N)$ .*

**Proof.** If  $\sum_{p \in N} > 1$ , then the assertion is trivial. There is something to prove only when  $N \subsetneq T_k(N)$ . We have to distinguish two cases.

(1) If  $N \cap \{q_k\}$  is normal, then  $N$  has the two successors  $N$  and  $N'$  in  $\mathcal{N}_{k+1}$ .

Regarding  $N$ , the prime  $q_k$  is included in  $T_k(N)$ , but not in  $T_{k+1}(N)$ . Other than that, we have  $T_k(N) \setminus \{q_k\} \subseteq T_{k+1}(N)$ , since the normality condition for including primes is the same for both sets. Because of property (iii) above, the prime  $q_k$ , missing in  $T_{k+1}(N)$ , must be compensated by at least one prime higher than any element of  $T_k(N)$ . Therefore  $j_{k+1}(N) = |T_{k+1}(N)| \geq |T_k(N)| = j_k(N)$ , and similarly  $P_{k+1}(N) \geq P_k(N)$ .

Regarding  $N'$ ,  $T_k(N)$  can contain primes which are congruent to 1 modulo  $q_k$ , but  $T_{k+1}(N')$  cannot.  $T_k(N)$  minus these primes is a subset of  $T_{k+1}(N')$ . Therefore each of these primes has to be compensated by at least one higher prime. Again, we get  $j_{k+1}(N') \geq j_k(N)$  and  $P_{k+1}(N') \geq P_k(N)$ .

(2) If  $N \cap \{q_k\}$  is not normal, then the only successor of  $N$  in  $\mathcal{N}_{k+1}$  is  $N$  itself. Since  $q_k$  neither appears in  $T_k(N)$  nor in  $T_{k+1}(N)$ , these sets are equal and we have  $j_{k+1}(N) = j_k(N)$  and  $P_{k+1}(N) = P_k(N)$ .  $\square$

The consequence from this theorem is that the computation of the numbers  $j_k$  and  $P_k$  can be done recursively: Assume that an upper bound  $u$  for, say,  $j_k$  is already known. Then it is not necessary to compute the  $j$ -value of any successor of a normal set with a  $j$ -value exceeding  $u$ . In other words, the tree of the  $\mathcal{N}_k$ 's can be partially trimmed at an early level. This indeed significantly reduces the number of cases to be checked. Our run-times seem to suggest that this reduces an exponential algorithm to a polynomial algorithm. We cannot prove this reduction, but of course we can happily run the computations.

How do we get upper bounds for  $j_k$  and  $P_k$ , needed for the algorithm? Since these values are minima, it is enough to compute  $j_k(N)$  and  $P_k(N)$  for one normal set  $N \in \mathcal{N}_k$ ; these will then be upper bounds. Of course, the smaller these bounds are, the faster the algorithm will be, because phony branches of the tree will be trimmed at an earlier stage. Which normal sets give small  $j_k$  and  $P_k$  values? From his computations, Giuga noticed that sets  $L_k$ , defined below, always in fact seem to have the minimal value. Although we cannot prove that these sets always give the minimum, we can at least use their values as good upper bounds. In our computations we never have found a set with smaller values.

These sets are given by

$$L_5 = \{5, 7\} \quad \text{and} \quad L_{k+1} = \begin{cases} L_k \cup \{q_k\} & \text{if this set is normal,} \\ L_k & \text{otherwise.} \end{cases}$$

With this improved, recursive algorithm, we could increase the exclusion bounds significantly. In 1994, using *Maple* on a workstation, we could compute the values  $j_k$  up to  $k = 100$ . For values of  $k$  around 100, this took a few CPU hours for each  $k$ . We got  $j_{100} = 3050$ , leading to an exclusion bound of  $10^{12,054}$ . We then continued the computations in C++ (the re-coding took two months) and could extend the range up to  $k = 135$  with the result  $j_{135} = 3459$  and an exclusion bound of  $10^{13,886}$ . This algorithm then crashed (irrevocably for linguistic reasons) in the Tokyo Computer Centre before doing any new work. We see here forcibly the dilemma of when to use high-level languages or to opt for computational speed.

We did not compute the  $P_k$ -values in 1994, since the slightly higher exclusion bounds they could give us were more than offset by the additional cost to compute many products. This disadvantage only disappears for high  $k$  levels, and recently Holger Rauhut has computed  $P_{106}$  on a Sun workstation, leading to an exclusion bound of  $10^{14,164}$  (achieved by  $L_{106}$ , of course). This computation took about five days.

Note that the set  $L_{27,692}$  is normal, has 8,135 elements and satisfies

$$\sum_{q \in L_{27,692}} 1/q > 1.$$

Therefore,  $j_k \leq 8135$  for all  $k \geq 27,692$ , and the method would be used up at this level. It can never lead to higher and higher exclusion bounds. But with current technology we are still far away from exhausting the algorithm.

### 5.1.2 Giuga Sequences

As we have seen, if there is a counterexample to Giuga's conjecture, then it is to be found among the Giuga numbers: Composite integers  $n$  with  $p \mid (n/p - 1)$  for every prime divisor  $p$  of  $n$ ; or, equivalently, squarefree composite integers  $n$  with  $\sum_{p|n} 1/p - \prod_{p|n} 1/p \in \mathbb{N}$ .

The problem of finding Giuga numbers can be relaxed to a *combinatorial* problem by dropping the requirement that all factors must be prime. More precisely, we define a *Giuga sequence* to be a finite sequence of integers,  $[n_1, \dots, n_m]$ , satisfying  $n_j \mid (\prod_{i \neq j} n_i - 1)$ . As in the prime case, it follows from this definition that the  $n_j$  in a Giuga sequence must be relatively prime, and an equivalent definition is: *A sequence of integers  $[n_1, \dots, n_m]$  is a Giuga sequence if and only*

if the  $n_j$  are relatively prime and satisfy  $\sum 1/n_j - \prod 1/n_j \in \mathbb{N}$ . An example of a non-prime Giuga sequence is

$$\frac{1}{2} + \frac{1}{3} + \frac{1}{7} + \frac{1}{83} + \frac{1}{5 \times 17} - \frac{1}{296310} = 1.$$

For all known examples, the “sum minus product” value is one; to reach any higher value, as we saw, the sequence would have to have at least 59 factors. To find all Giuga sequences of a given length, one could check in principle all sequences of this length whose elements are not too large (the sum over their reciprocals must be greater than one to be ruled out). However, the number of these grows exponentially; even for length seven there are too many to check them all.

Luckily, we have the following reasonably effective theorem which tells us how to find all Giuga sequences of length  $m$  with a given initial segment of length  $m - 2$ .

**Theorem 5.1.6** (a) Take an initial sequence of length  $m - 2$ ,  $[n_1, \dots, n_{m-2}]$ . Let

$$P = n_1 \cdots n_{m-2}, \quad S = 1/n_1 + \cdots + 1/n_{m-2}.$$

Fix an integer  $v > S$  (this will be the sum minus product value). Take any integers  $a, b$  with  $a \cdot b = P(P + S - v)$  and  $b > a$ . Let

$$n_{m-1} = (P + a)/P(v - S), \quad n_m = (P + b)/P(v - S).$$

Then

$$S + 1/n_{m-1} + 1/n_m - 1/Pn_{m-1}n_m = v.$$

The sequence  $[n_1, \dots, n_{m-1}, n_m]$  is a Giuga sequence if and only if  $n_{m-1}$  is an integer.

(b) Conversely, if  $[n_1, \dots, n_{m-1}, n_m]$  is a Giuga sequence with sum minus product value  $v$ , and if we define

$$a = n_{m-1}P(v - S) - P, \quad b = n_mP(v - S) - P$$

(with  $P$  and  $S$  the product and the sum of the first  $m - 2$  terms) then  $a$  and  $b$  are integers and  $a \cdot b = P(P + S - v)$ .

To conclude the section on Giuga’s conjecture, we note Agoh’s conjecture (1995), which is equivalent:

$$nB_{n-1} \equiv -1 \pmod{n} \quad \text{if and only if } n \text{ is prime;}$$

here  $B_n$  is a *Bernoulli number*.



### 5.1.3 Lehmer's Problem

**Conjecture 5.1.7** Lehmer's conjecture (1932).

$$\phi(n) \mid n - 1$$

if and only if  $n$  is prime.

Lehmer called this

A problem as hard as existence of odd perfect numbers.

For Lehmer's conjecture, the set of prime factors of any counterexample  $n$  is a normal family. Lehmer's conjecture has now been verified for up to 14 prime factors of  $n$ . The related condition

$$\phi(n) \mid n + 1,$$

is known to have eight solutions with up to six prime factors:  $2, F_0, \dots, F_4$  (the *Fermat primes*) and a rogue pair: 4919055 and 6992962672132095. Fermat primes are primes of the form  $F_n = 2^{2^n} + 1$ . As an early example of experimental error, the sequence starts 3, 5, 17, 257, 65537, all of which are prime. On this inductive basis Fermat conjectured all Fermat numbers were prime, despite the fact that  $F_5 = 4294967297$  is divisible by 641, as Euler discovered.

Recently this result was extended by one to seven prime factors—by dint of a heap of factorizations! But the next cases of Lehmer's two problems (15 and 8 respectively) are much too large for current methods and machines. The *curse of exponentiality* strikes again!

## 5.2 Disjoint Genera

The role of experimentation pattern recognition is reiterated in the description of how the following theorem [55] was found, in response to a question posed by Richard Crandall in his analysis of the Madelung constant [93].

**Theorem 5.2.1** *There are at most 19 positive integers not of the form of  $xy + yz + xz$  with  $x, y, z \geq 1$ . The only non-square-free cases are 4 and 18. The first 16 square-free cases are*

$$1, 2, 6, 10, 22, 30, 42, 58, 70, 78, 102, 130, 190, \mathbf{210}, 330, 462, \quad (5.2.2)$$

which correspond to “discriminants with one quadratic form per genus.”

If the 19th exists, it is greater than  $10^{11}$ , which the Generalized Riemann Hypothesis excludes.

These exceptions were found with a crude Matlab program, which also showed there were no others less than 50,000. One may then note that the largest three numbers correspond to three very special singular values ( $k_{210}, k_{330}, k_{462}$  of Section 4.2). Further inspection showed that the square free solutions in (5.2.2) corresponded precisely to quadratic forms  $Q_{2P}(n, m) = n^2 + 2Pm^2$ , with precisely one quadratic form per genus, see Section 4.3. In this manner was the theorem discovered.

After the research was nearly finished, the authors remembered to consult Sloane’s online Encyclopedia and learned that the theorem was true! This was based only on email communications, though indeed there are now several published proofs. Moreover, if you now consult the Encyclopedia, it will tell you the numbers are those with one form per genus, and give details! Had the authors consulted the database earlier, they would have considered Crandall’s question answered. This would have saved time, but left the database and literature poorer.

### 5.3 Gröbner Bases and Metric Invariants

In this section we introduce *Gröbner bases* and the Pedersen-Roy-Szpirglas real solution counting method. We apply these modern tools together with ideas from classical distance geometry (the Cayley-Menger determinant) to show how Petr Lisoněk settled two open problems in Euclidean geometry posed in the *American Mathematical Monthly* in 1999 [160].

By analyzing systems of algebraic equations satisfied by metric invariants of a tetrahedron, we shall conclude (i) that, in general, the four face areas, circumradius and volume together do not uniquely determine a tetrahedron, and that (ii) there exist non-regular tetrahedra that are uniquely determined just by the four face areas and circumradius.

The development outlines the main steps of [154], which can be obtained (along with a *Maple* worksheet containing all computations) on the Internet at <http://www.cecm.sfu.ca/~lisonek/tetrahedron.html>. Executing all computations in *Maple* takes about 40 seconds CPU time on a 2003-era computer.

We should add here that the *Magma* tool is also very useful for performing this type of algebraic computation. Information on *Magma* is available at <http://magma.maths.usyd.edu.au/magma>.

### 5.3.1 Formulation of the Polynomial System

Let  $\mathbb{R}^n$  denote  $n$ -dimensional Euclidean space with the Euclidean distance between points  $v$  and  $w$  be denoted by  $d(v, w)$ . Consider a general tetrahedron  $T$  in  $\mathbb{R}^3$  and let  $v_i$  ( $i = 1, 2, 3, 4$ ) be the vertices of  $T$ . Clearly,  $T$  is determined uniquely (up to a rigid motion) if the lengths  $d(v_i, v_j)$  of its six edges are given. We shall analyze whether  $T$  can be determined uniquely by sets of *metric invariants* other than that of all edge lengths.

Denote the *squared edge lengths* by

$$s_{i,j} = d(v_i, v_j)^2, \quad (5.3.3)$$

the volume by  $V$  and the circumradius (i.e., radius of the circumscribed sphere) by  $R$ . We let the area of the face  $v_i v_j v_k$  be denoted  $A_l$ , where  $\{i, j, k\} \cup \{l\} = \{1, 2, 3, 4\}$ .

Let  $P_i$  ( $1 \leq i \leq n$ ) be  $n$  points in  $\mathbb{R}^{n-1}$  and denote  $d_{i,j} = d(P_i, P_j)$ . The *Cayley-Menger determinant* ([27], §40) associated with the points  $P_i$  ( $1 \leq i \leq n$ ) is the determinant of the  $(n+1) \times (n+1)$  matrix

$$D(P_1, \dots, P_n) = \begin{vmatrix} 0 & 1 & 1 & \cdots & 1 \\ 1 & 0 & d_{1,2}^2 & \cdots & d_{1,n}^2 \\ 1 & d_{2,1}^2 & 0 & \cdots & d_{2,n}^2 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & d_{n,1}^2 & d_{n,2}^2 & \cdots & 0 \end{vmatrix}. \quad (5.3.4)$$

The  $(n-1)$ -dimensional volume  $\text{Vol}(P_1, \dots, P_n)$  of the convex hull of  $P_1, \dots, P_n$  satisfies the equality ([27], p. 98)

$$\text{Vol}(P_1, \dots, P_n)^2 = \frac{(-1)^n}{2^{n-1}((n-1)!)^2} D(P_1, \dots, P_n). \quad (5.3.5)$$

Hence, the facial areas and the volume of the tetrahedron can be expressed as low degree polynomials in the squared edge lengths  $s_{i,j}$ . In the case of a triangle,  $T$ , (5.3.5) is a disguised version of the *Heron's formula*:

$$A(T) = \sqrt{s(s-a)(s-b)(s-c)} \quad (5.3.6)$$

where  $s = (a + b + c)/2$  is the *semi-perimeter* of the circle and  $a, b, c$  are the lengths of the sides (Exercise 13).

Let  $C$  denote the center of the circumscribed sphere of our tetrahedron. The five points  $v_1, v_2, v_3, v_4, C$  lie in a three-dimensional space (a hyperplane of  $\mathbb{R}^4$ ) and therefore  $\text{Vol}(v_1, v_2, v_3, v_4, C) = 0$ . Hence, the algebraic equation for the circumradius  $R$  is

$$D(v_1, v_2, v_3, v_4, C) = 0 \quad (5.3.7)$$

where  $d(v_i, C) = R$  for  $1 \leq i \leq 4$ . Consider seven positive real numbers  $a_{1,2}, a_{1,3}, \dots, a_{3,4}, W$ . Clearly, the statement

*The volume of the tetrahedron with edge lengths  $\sqrt{a_{1,2}}, \sqrt{a_{1,3}}, \dots, \sqrt{a_{3,4}}$  is  $W$ .*

implies the statement

*The point  $(a_{1,2}, a_{1,3}, \dots, a_{3,4}, W)$  is a zero of the polynomial  $f_1(s_{1,2}, s_{1,3}, \dots, s_{3,4}, V)$ .*

Here the explicit form of  $f_1$  can be easily extracted from (5.3.3–5.3.5). (Exercise 14.)

The analogous statements for the other metric invariants of the tetrahedron (e.g., circumradius, four face areas) imply corresponding statements about zeros of five other computable polynomials  $f_2, \dots, f_6$  with rational coefficients.

Since we transform a geometric problem into an algebraic one, we should make sure that we are working with concrete algebraic objects that have a meaning (interpretation) back in the original geometric domain. The essential question here is, for which positive sextuples  $(s_{1,2}, s_{1,3}, \dots, s_{3,4})$  in  $\mathbb{R}^6$  does there exist a tetrahedron whose squared edge lengths are the values  $s_{i,j}$ ? It is shown in [27], §40 that this is the case exactly when all squared volumes (5.3.5) evaluate to non-negative, which therefore is not only a necessary but also a sufficient condition for the existence of the tetrahedron.

The important consequence is that whenever there exists a positive solution  $(a_{1,2}, a_{1,3}, \dots, a_{3,4})$  in  $\mathbb{R}^6$  to a system  $\{f_1^* = 0, \dots, f_6^* = 0\}$ , where  $f_i^*$  is  $f_i$  with  $V, R, A_1, \dots, A_4$  replaced by positive real constants, then this solution does have a geometric interpretation and corresponds to a tetrahedron in  $\mathbb{R}^3$ , whose edge lengths are  $\sqrt{a_{i,j}}$ .

Since scaling does not affect the answer to our problem in any way, we typically allow ourselves to normalize the value of one of the variables, say  $R$ .

### 5.3.2 Gröbner Bases

Let  $\mathbf{Q}[\mathbf{x}_1, \dots, \mathbf{x}_n]$  denote the ring of  $n$ -variable polynomials with rational coefficients in indeterminates  $x_1, \dots, x_n$ . Let  $f_1, \dots, f_s \in \mathbf{Q}[\mathbf{x}_1, \dots, \mathbf{x}_n]$ . The *ideal generated by*  $f_1, \dots, f_s$  is defined by

$$\langle f_1, \dots, f_s \rangle = \left\{ \sum_{i=1}^s h_i f_i : h_i \in \mathbf{Q}[\mathbf{x}_1, \dots, \mathbf{x}_n] \right\} .$$

The (affine) *variety* determined by  $f_1, \dots, f_s$  is  $\mathbf{V}(f_1, \dots, f_s) \subset C^n$  defined as

$$\mathbf{V}(f_1, \dots, f_s) = \{x \in C^n : f_1(x) = \dots = f_s(x) = 0\} .$$

If  $I = \langle f_1, \dots, f_s \rangle$ , then we say that  $\{f_1, \dots, f_s\}$  is a *basis* of  $I$ . Clearly, if  $\{f_1, \dots, f_s\}$  and  $\{g_1, \dots, g_t\}$  are two bases of the same ideal, then  $\mathbf{V}(f_1, \dots, f_s) = \mathbf{V}(g_1, \dots, g_t)$ . *That said*, one of the bases may be much more suitable than the other for assessing properties of the variety (such as (non-)emptiness, cardinality, finiteness, dimension, characterizing all points in the variety, etc.).

Roughly *Gröbner bases* are those special bases of polynomial ideals which are especially suited for studying properties of affine varieties such as those listed above, and for answering questions about the ideal itself—such as determining whether or not a given polynomial belongs to the ideal. There are many different Gröbner bases for each polynomial ideal all of which can be computed using Bruno *Buchberger's algorithm*. Which Gröbner basis is most suitable in a given situation depends on the question to be resolved for the given variety or ideal. The book [85] is a very accessible introduction to polynomial ideals, their varieties and the algorithms operating on them.

## 5.4 A Sextuple of Metric Invariants

As noted the question as to whether a tetrahedron is uniquely determined by its volume, circumradius and face areas was posed as an open problem in [160]. In this section we constructively answer this question negatively, by producing two (or more) tetrahedra that share the same volume, circumradius and face areas.

Consider a tetrahedron  $T$  given by squared edge lengths. Equations for the volume, circumradius and face areas of  $T$  may be obtained on substituting the values of  $s_{i,j}$  into the polynomials  $f_1, \dots, f_6$  introduced earlier. Conversely, by

substituting the numerical values of  $V, R, A_1, \dots, A_4$  into the  $f_i$ 's we set up a polynomial system in the squared edge lengths as unknowns—let us call this system  $F^* = \{f_1^*, \dots, f_6^*\}$ .

Thus, we hope for distinct positive solutions to  $F^*$ . To determine these solutions we first find  $G$ , a *grevlex Gröbner basis* for  $F^*$ . In all numerical examples that Lisoněk and Israel [154] tried, they found that  $\langle F^* \rangle$  was a zero-dimensional ideal. They then used the method of Pedersen, Roy and Szpirglas to count all real solutions of  $F^*$ . (See Chapter 2 of [84]; in particular we adopt the terminology and notation introduced therein.)

Let  $p_{i,j}$  be the generators for the univariate elimination ideals; that is,  $\langle p_{i,j} \rangle = \mathbb{C}[s_{i,j}] \cap \langle F^* \rangle$ . Using the theory in Chapter 2, Section 2 of [84], first one finds the  $p_{i,j}$ 's from the grevlex Gröbner basis  $G$  by working in the algebra  $A = \mathbb{C}[s_{1,2}, s_{1,3}, \dots, s_{3,4}] / \langle G \rangle$ . Second, one uses interval arithmetic (together with knowledge of the total number of real solutions) to isolate these real solutions of  $F^*$ . Third, one selects the positive solutions.

It turns out to be easy to obtain examples in which several different tetrahedra sharing the same volume, circumradius and face areas, as is illustrated in the following numerical example. Some intermediate expressions are omitted because of their large size.

**Example 5.4.1** *Tetrahedral volume.*

Consider the tetrahedron defined by

$$(s_{1,2}, s_{1,3}, s_{1,4}, s_{2,3}, s_{2,4}, s_{3,4}) = (1, 1, 2, 2, 1, 2).$$

We find that  $V = \sqrt{1/48}$ ,  $R = \sqrt{7/12}$ ,  $A_1 = A_2 = \sqrt{7/16}$  and  $A_3 = A_4 = 1/2$ .

1. By substituting these six values in the polynomials  $f_1, \dots, f_6$  we obtain the set  $F^* = \{f_1^*, \dots, f_6^*\}$ . Next compute  $G$ , the Gröbner basis for  $F^*$  with respect to the grevlex ordering induced by  $s_{1,2} > s_{1,3} > \dots > s_{3,4}$ , and the basis for the algebra  $A = \mathbb{C}[s_{1,2}, s_{1,3}, \dots, s_{3,4}] / \langle F^* \rangle$ .
2. It turns out that the dimension of  $A$  is 8, and that a monomial basis of  $A$  is  $B = (1, s_{1,2}, s_{2,3}, s_{2,4}, s_{2,4}^2, s_{1,2}s_{2,4}, s_{3,4}, s_{2,4}s_{3,4})$ . Let  $h$  be the constant function 1 and construct the symmetric bilinear form  $S_1$  as described in [84, page 65].

3. Compute the characteristic polynomial of  $M_1$ , the matrix of  $S_1$  with respect to  $B$ . It turns out that there are six sign variations in the coefficient list of this characteristic polynomial, whence by Descartes rule of signs the signature of  $S_1$  is four, and there are four distinct real solutions to the system  $F^*$  by [84, Theorem 5.2].
4. Using the *Maple* procedure `Groebner[univpoly]` (which implements the Faugère-Gianni-Lazard-Mora basis conversion method), we use  $G$  to find the generators  $p_{i,j}$  for the univariate elimination ideals  $\langle p_{i,j} \rangle = \mathbb{C}[s_{i,j}] \cap \langle F^* \rangle$ . They turn out to be the following polynomials:

$$\begin{aligned} p_{1,2}(x) &= (x-1)a(x) \\ p_{1,3}(x) &= p_{1,4}(x) = p_{2,3}(x) = p_{2,4}(x) \\ &= (x-1)(x-2)b(x) \\ p_{3,4}(x) &= (x-2)c(x) \end{aligned}$$

where

$$\begin{aligned} a(x) &= 9x^3 - 123x^2 + 491x - 249 \\ b(x) &= 81x^6 - 1485x^5 + 10215x^4 - 25803x^3 + 22865x^2 \\ &\quad - 3303x + 162 \\ c(x) &= 9x^3 - 150x^2 + 1028x - 960. \end{aligned}$$

5. Using Sturm's theorem (the *Maple* procedure `realroot`), determine the isolating intervals for all real roots of  $a$ ,  $b$  and  $c$ . It turns out that  $a$  and  $c$  have one real root each, namely  $\alpha_1 \in [75/128, 19/32]$  and  $\gamma_1 \in [35/32, 141/128]$ , respectively, while  $b$  has two real roots  $\beta_1 \in [219/128, 55/32]$  and  $\beta_2 \in [267/128, 67/32]$ .
6. These isolating intervals are narrow enough to prove (using interval arithmetic) that only four boxes in  $\mathbb{R}^6$  (Cartesian products of the separating intervals) can possibly contain a solution of  $F^*$ .
7. Since we know that there are exactly four real solutions, we have isolated all solutions of  $F^*$ . Their values  $(s_{1,2}, s_{1,3}, s_{1,4}, s_{2,3}, s_{2,4}, s_{3,4})$  are  $(1, 2, 1, 1, 2, 2)$ ,  $(1, 1, 2, 2, 1, 2)$ ,  $(\alpha_1, \beta_1, \beta_2, \beta_2, \beta_1, \gamma_1)$  and  $(\alpha_1, \beta_2, \beta_1, \beta_1, \beta_2, \gamma_1)$ .

8. As each  $s_{i,j}$  value is positive for all four solutions, it follows (cf. the discussion at the end of Section 5.3.1) that we have constructed four different tetrahedra that share the same volume, circumradius and face areas. On inspecting the four solutions we see that the four tetrahedra pair up as two pairs, with the elements of each pair related by interchanging the labels of the vertices  $v_1$  and  $v_2$ . Thus, we have obtained two essentially different tetrahedra.

□

## 5.5 A Quintuple of Related Invariants

It was noted in [160] part (b) that the quintuple of values  $(A_1, A_2, A_3, A_4, R)$  uniquely determines any *regular* tetrahedron. In this case, of course  $A_1 = A_2 = A_3 = A_4 = \sqrt{4/3}R^2$ . As to whether there exist *non-regular* tetrahedra uniquely determined by  $(A_1, A_2, A_3, A_4, R)$ , was posed as an open problem in [160]. This question is answered affirmatively on showing:

**Example 5.5.1** *Every non-degenerate tetrahedron, having a face that is an equilateral triangle inscribed in a great circle of the circumscribed sphere, is determined uniquely by its four face areas and circumradius.*

1. Assume without loss of generality that  $R = 1$ . It is easy to see that  $3\sqrt{3}/4$  is the maximum area among all triangles inscribed in a unit sphere, and it is attained only by equilateral triangles inscribed in a great circle of the sphere (Exercise 15).
2. Therefore the shape of one of the faces of the tetrahedron is determined by requiring its area to be equal to  $3\sqrt{3}/4$ . Assume this face is  $v_1v_2v_3$ . Then  $R = 1$ ,  $s_{1,2} = s_{1,3} = s_{2,3} = 3$  and  $A_4 = 3\sqrt{3}/4$ .
3. Let  $F'$  be the set of polynomials  $f_2, \dots, f_6$  from Section 5.3.1 with the values from the previous sentence substituted.

It takes only a few seconds to compute the Gröbner basis for the ideal generated by  $F'$  for the lexicographic ordering induced by  $s_{3,4} > s_{2,4} >$



$A_1 > A_3 > A_2 > s_{1,4}$ . This basis contains (among others) the polynomials  $f^2$  and  $g^2$ , where

$$\begin{aligned} f &= 81s_{1,4}^4 - 432s_{1,4}^3 + 288(A_2^2 + A_3^2)s_{1,4}^2 \\ &\quad + 256A_2^4 + 256A_3^4 - 512A_2^2A_3^2 \\ g &= 9s_{1,4}^2 - 54s_{1,4} - 16A_1^2 + 32A_2^2 + 32A_3^2 + 27. \end{aligned}$$

4. Further

$$f = Q \cdot g + S \tag{5.5.8}$$

where

$$\begin{aligned} Q &= 9s_{1,4}^2 + 6s_{1,4} + 16A_1^2 + 9 \\ S &= (960A_1^2 - 192A_2^2 - 192A_3^2 + 324)s_{1,4} \\ &\quad + 256(A_1^4 + A_2^4 + A_3^4) - 512(A_1^2A_2^2 + A_1^2A_3^2 + A_2^2A_3^2) \\ &\quad - 288(A_1^2 + A_2^2 + A_3^2) - 243. \end{aligned}$$

If both  $f$  and  $g$  vanish, then  $S$  must vanish by (5.5.8). The equation  $S = 0$  determines  $s_{1,4}$  uniquely if  $L$ , the coefficient at  $s_{1,4}$  in  $S$ , is non-zero.

5. The discriminant of  $g$  as a quadratic polynomial in  $s_{1,4}$  is

$$Z = 576A_1^2 - 1152A_2^2 - 1152A_3^2 + 1944,$$

and observe that  $6 \cdot L = 5184A_1^2 + Z$ .

6. If there exists a tetrahedron with face areas  $A_1, A_2, A_3, 3\sqrt{3}/4$  and circumradius 1, then  $g$  must have a real (positive) root, hence  $Z \geq 0$  and consequently  $L > 0$  by non-degeneracy, in particular  $L \neq 0$ . Therefore  $s_{1,4}$  is determined uniquely by  $S = 0$ .

7. By applying analogous arguments to  $s_{2,4}$  and  $s_{3,4}$  we prove that all  $s_{j,4}$  are determined uniquely ( $1 \leq j \leq 3$ ). Since also  $s_{1,2} = s_{1,3} = s_{2,3} = 3$  are determined uniquely by  $A_4 = 3\sqrt{3}/4$ , we have established that the tetrahedron is determined uniquely.

□

### 5.5.1 Some Open Questions on Invariants

An intriguing question is whether, for *every* set of positive real constants  $V, R, A_1, \dots, A_4$ , there are *only finitely many* tetrahedra, all having these values as their respective metric invariants.

If the answer is affirmative, what is then the *maximal number* of such tetrahedra? For example, the values  $(s_{1,2}, s_{1,3}, s_{1,4}, s_{2,3}, s_{2,4}, s_{3,4}) = (16, 25, 9, 9, 33, 54)$  yield a family of six such tetrahedra, but it is not known whether this is maximal.

The bounds obtained by applying general theorems from algebraic geometry seem much larger than the empirical results obtained in [154] by running the algorithm outlined above on a number of different examples.

In cases when the number of solutions to the system  $F^*$  is finite one can apply *Bézout's theorem*, and so obtain  $3 \cdot 4 \cdot 2^4 = 192$  as an upper bound on the number of solutions. A theorem of Milnor [162] gives 9,375 as an upper bound on the number of real solutions of  $F^*$  that are isolated points. Here, the central issue is whether the variety is always zero-dimensional (is a finite set of isolated points), or whether it can have (components of) positive dimension. In the zero-dimensional case, Bézout's theorem applies, while Milnor's result unconditionally bounds the number of isolated real points.

## 5.6 Sloane's Harmonic Designs

In this final section we describe and illustrate an ambitious and successful experimental approach to finding spherical designs used by N. J. A. Sloane, R. H. Hardin and P. Cara [130, 131, 192, 193].

A set of  $N$  points  $\{P_1, \dots, P_N\}$  on the unit sphere  $\Omega_n = S^{n-1} = \{x = (x_1, \dots, x_n) \in \mathbf{R}^n : x \cdot x = 1\}$  forms a *spherical  $t$ -design* if the identity

$$\int_{\Omega_n} f(x) d\mu(x) = \frac{1}{N} \sum_{i=1}^N f(P_i) \quad (5.6.9)$$

holds for all polynomials  $f$  of degree  $\leq t$ , where  $\mu$  is uniform measure on the sphere normalized to have total measure one ([100, 120, 121, 122] and [79, §3.2]). A spherical  $t$ -design is also a  $t'$ -design for all  $t' \leq t$ . The largest  $t$  for which the points form a  $t$ -design is called the *strength* of the design.

It is known that if  $N$  is large enough then a spherical  $t$ -design in  $\Omega_n$  always exists [188]. The problem is to find the smallest value of  $N$  for a given strength and dimension, or equivalently to find the largest strength  $t$  that can be achieved with  $N$  points in  $\Omega_n$ .

Not surprisingly, the main application is to numerical integration, but spherical designs also have applications to the design of experiments in statistics. Their study has involved many interesting questions in algebra and group theory. For searching for a spherical  $t$ -design, and for verifying that a set of points does form a spherical  $t$ -design, the following equivalent condition is more useful than the definition:

$P_1, \dots, P_N$  forms a spherical  $t$ -design if and only if the polynomial identities

$$\frac{1}{N} \sum_{i=1}^N (P_i \cdot x)^{2s} = \left( \prod_{j=0}^{s-1} \frac{2j+1}{2j+n} \right) (x \cdot x)^s, \quad (5.6.10)$$

and

$$\frac{1}{N} \sum_{i=1}^N (P_i \cdot x)^{2\bar{s}+1} = 0, \quad (5.6.11)$$

hold, where  $s$  and  $\bar{s}$  are defined by  $\{2s, 2\bar{s}+1\} = \{t-1, t\}$  ([121]; [179, p. 114], [180]).

The approach taken by Sloane et al. follows the following three stages.

1. *Experimentation.* Given the dimension  $n$ , a specified number of points  $N$ , and a desired value of the strength  $t$ , search for a set of points satisfying (5.6.10) and (5.6.11). If no solution seems to exist, decrease  $t$  and try again. This is repeated a large number of times (with different starting configurations) until a collection of putative designs has been assembled.

At this stage, these are only numerical approximations to the desired designs; that is, numerical coordinates for points which appear to satisfy (5.6.10) and (5.6.11), with an error which is less than  $10^{-10}$ .

The search algorithm used was a modification of the "pattern search" of Hooke and Jeeves [138], a distant cousin of the conjugate gradient method.

2. *Beautification.* They now attempt to show that there *is* a spherical  $t$ -design in the neighborhood of the numerical points, that is, to find algebraic

expressions for the coordinates of points, such that (5.6.10) and (5.6.11) are satisfied exactly.

This step involves a considerable amount of guesswork, guided by computations of the geometrical structure of the computer-produced points, such as their apparent automorphism group.

The crucial step in the beautification process is to use knowledge of the automorphism group to reduce the number of unknowns in the design. If the points appear to fall into  $k$  orbits under the group, the number of unknowns is reduced from  $N(n-1)$  (the number of degrees of freedom in the original design) to  $k(n-1)$ . With luck, they are now able to solve equations (5.6.10) and (5.6.11) exactly.

3. **Generalization.** Try to find infinite families of designs that generalize those found at step 2.

One example will serve to illustrate the process, a four-dimensional spherical 6-design (in  $\Omega_4$ ) with 42 points. After beautification, this turned out to consist of six heptagons, each in a different plane, with a group of order 42 acting transitively on the 42 points. In other words, the computer-produced design was suggesting that they should choose six planes in  $\mathbf{R}^4$ , i.e., six points in the Grassmann manifold  $G(4,2)$ , and draw a heptagon in each plane. Sloane and his colleagues found this an appealing idea, in view of the recent work on finding packings and designs in Grassmann manifolds (see [15, 73, 78, 189])! It immediately suggested several general constructions, one of which is the following.

Let  $\Pi_1, \dots, \Pi_6$  be the planes in four dimensions spanned by the rows of the following six matrices:

$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix},$$

$$\begin{bmatrix} s & 0 & -h & h \\ 0 & s & -h & -h \end{bmatrix}, \begin{bmatrix} -h & h & s & 0 \\ -h & -h & 0 & s \end{bmatrix},$$

$$\begin{bmatrix} -s & 0 & -h & -h \\ 0 & -s & h & -h \end{bmatrix}, \begin{bmatrix} -h & -h & -s & 0 \\ h & -h & 0 & -s \end{bmatrix},$$

where  $s = 1/\sqrt{2}$ ,  $h = 1/2$ .

**Theorem 5.6.1** *Let  $p$  be an integer  $\geq 3$  and draw regular  $p$ -gons in each plane. The resulting  $N = 6p$  points*

$$\{\cos(j\theta)u_i + \sin(j\theta)v_i : 0 \leq j < p, 1 \leq i \leq M\}, \tag{5.6.12}$$

where  $u_i$  and  $v_i$  span  $\Pi_i$ , form a  $6p$ -point spherical  $t$ -design with  $t = \min\{p-1, 7\}$ .

This is an interesting result, since there are very few general constructions known for infinite families of spherical designs.

The theorem yields the following  $t$ -designs.

$p$	3	4	5	6	7	8	9	10	11	12	$\dots$
$N$	18	24	30	36	42	48	54	60	66	72	$\dots$
$t$	2	3	4	5	6	7	7	7	7	7	$\dots$

The papers [131, 192, 193] contain many other examples.

## 5.7 Commentary and Additional Examples

1. **A prime generating number.** Let

$$\alpha = \sum_{m=1}^{\infty} \frac{p_m}{10^{2^m}} = 0.020300050000000700000000000000110\dots,$$

where  $p_m$  denotes the  $m$ -th prime. Show that

$$p_n = \lfloor 10^{2^n} \alpha \rfloor - 10^{2^{n-1}} \lfloor 10^{2^{n-1}} \alpha \rfloor,$$

which would be useful if one could find a method of computing  $\alpha$  to the needed precision to obtain  $p_n$  without knowing  $p_n$ . This seems unlikely.

2. **Carmichael and Lucas-Carmichael numbers.** The Carmichael numbers are usually defined as those *pseudo-primes* such that  $a^{n-1} \equiv 1 \pmod n$  for each  $a$  relatively prime to  $n$ . Korelt (1899) showed this coincides with the definition in the text: Namely  $n$  is square-free and  $(p-1)|(n-1)$  whenever  $p|n$ . The smallest examples are 561, 1105, 1729, and a much larger one is  $17 \cdot 37 \cdot 41 \cdot 131 \cdot 251 \cdot 571 \cdot 4159 = 2013745337604001$ .

We suggest you consult Sloane's table (or the on-line version) to learn about the sequence starting

$$399, 935, 2015, 2915, 4991, 5719, 7055.$$

3. **Odd perfect numbers.** It is known that there are no odd perfect number with seven or fewer prime factors. Show that any odd perfect number with eight prime factors must be divisible by 3, 5 or 7. More generally Servais in 1888 proved that the smallest prime factor can not exceed the number of prime factors.
4. **Primes in arithmetic progression.** Let  $\pi_{d,r}(x)$  denote the number of primes of the form  $nd + r \leq x$ , with  $d$  and  $r$  relatively prime. Then Dirichlet proved the number of primes in each such progression is infinite. Indeed,

$$\pi_{d,r}(x) \sim \frac{1}{\phi(d)} \frac{x}{\log x}.$$

This is due to de la Vallée Poussin [181, pg.149].

- (a) Use this to show that every normal sequence of primes can be extended so that the sum of its reciprocals exceeds one.
- (b) It is conjectured that one can find arbitrarily long arithmetic progressions all of whose members are prime [181, pg. 153].

The longest known arithmetic sequence of primes is currently 22, starting with the prime 11410337850553 and continuing with common difference 4609098694200 [176]. The longest known sequence of *consecutive* primes in arithmetic progression is ten. It starts with the 93-digit prime

100996972469714247637786655587969840329509324689190041803603  
417758904341703348882159067229719,

and has difference 210.

5. **A fourth degree polynomial problem.** (From [136, pg. 87]). Let  $\alpha_1, \alpha_2, \alpha_3, \alpha_4$  be the roots of the polynomial

$$P(x) = x^4 + px^3 + qx^2 + rx + 1.$$

Show that

$$(1 + \alpha_1^4)(1 + \alpha_2^4)(1 + \alpha_3^4)(1 + \alpha_4^4) = (p^2 + r^2)^2 + q^4 - 4pq^2r.$$

Hint: consider  $\prod_{k=1}^4 P(e^{(2k-1)i\pi/4})$ .

6. **Putnam problem 1991–B4.** Show that for  $p$  an odd prime

$$\sum_{j=0}^p \binom{p}{j} \binom{p+j}{j} \equiv 2^p + 1 \pmod{p^2}.$$

7. **1-additive sequences.** Given  $(u, v)$ , the 1-additive sequence generated by  $(u, v)$  is defined as  $a_1 = u$ ,  $a_2 = v$ , and for  $n \geq 3$ ,  $a_n$  is the least integer exceeding  $a_{n-1}$  and possessing a unique representation of the form  $a_i + a_j$  for  $i < j$ . Many 1-additive sequences behave quite erratically. The sequence generated by  $(2, 3)$  for example defies any simple characterization. Finch conjectured, and later Schmerl and Spiegel proved, that the sequence generated by  $(2, v)$  for odd  $v \geq 5$  has precisely two even terms, so that the sequence of successive differences is eventually periodic. Finch has further conjectured, based on extensive computations, that for odd  $v \geq 5$

**Conjecture 5.7.1 (Finch)** *If  $v \neq 2^m - 1$  for any  $m \geq 3$ , the sequence generated by  $(4, v)$  has precisely three even terms:  $4, 2v + 4$  and  $4v + 4$ . When  $v = 2^m - 1$  for some  $m \geq 3$ , then the sequence generated by  $(4, v)$  has precisely four even terms:  $4, 2v + 4, 4v + 4$  and  $4v^2 + 2v - 4$ .*

Cassaigne and Finch have proven this conjecture for the case  $v \equiv 1 \pmod{4}$ , but the question is open for  $v \equiv 3 \pmod{4}$ . See [75] for further details.

8. **Amicable numbers.** Two numbers are *amicable* if, like 220 and 284, each is the sum of the others proper divisors, Thabit ibn Kurrah (ca. A.D. 850) noted that if  $n > 1$  and each of  $p = 3 \cdot 2n - 1 - 1$ ,  $q = 3 \cdot 2n - 1$ , and  $r = 9 \cdot 2n - 1 - 1$  are prime, then  $2npq$  and  $2nr$  are amicable numbers. It was many years until this formula led to a second and third pair of amicable numbers! Fermat provided the pair 17,296 and 18,416 ( $n=4$ ) in a letter to Mersenne in 1636. Computer searches have found all such numbers with 10 or fewer digits. It is unknown if there are infinitely many amicable pairs or any relatively prime pair. (Such a pair must be more than twenty-five digits long, and the product must be divisible by at least 22 distinct primes.)

9. **More on amicable numbers.** The smallest example of an amicable four-cycle is

$$\begin{aligned} n_1 &= 2^2 \cdot 5 \cdot 17 \cdot 3719, & n_3 &= 2^2 \cdot 521 \cdot 829, \\ n_2 &= 2^2 \cdot 5 \cdot 193 \cdot 401, & n_4 &= 2^5 \cdot 40787, \end{aligned}$$

discovered by H. Cohen in 1970 during an exhaustive search up to sixty thousand. Very recently Blankenagel, Borho and vom Stein have obtained fifty new amicable four-cycles by a seed-and-complete method akin to the way in which Giuga sequences were generated.

10. **Aliquot sequences.** Consider the iteration  $n \mapsto s(n) = \sigma(n) - n$ , where  $\sigma$  is the divisor function. It is unknown whether this iteration must eventually become periodic, as is clearly the case for an amicable pair. There are five numbers less than 1000 whose status is unsettled (“the Lehmer five”: 276, 552, 564, 660, and 966). Such Aliquot sequences can grow very rapidly before subsiding. One may consider iterating many other arithmetic functions with irregular growth. Reference: [129].
11. **Prime power problem.** Let  $\Lambda(n) = \log(p)$  for  $p = n^m$  a prime power and be zero otherwise. Show that

$$\sum_{n=1}^{\infty} \frac{\Lambda(n)}{n^s} = \frac{\zeta'(s)}{\zeta(s)}$$

and so that

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

12. **Putnam problem 1988–B1. (disjoint genera)** Show that every composite number can be expressed in the form  $xy + yz + zx + 1$  for positive integers  $x, y, z$ . Compare what is true in the case of primes.
13. Verify that (5.3.6) is a specialization of (5.3.5).
14. Confirm that the geometric statement of subsection 5.3.1 implies the algebraic one.



15. Show that a positive sextuple  $(s_{1,2}, s_{1,3}, \dots, s_{3,4})$  in  $\mathbb{R}^6$  comes from a tetrahedron whose squared edge lengths are the values  $s_{i,j}$  exactly when all the squared volumes in (5.3.5) are nonnegative.
16. **Putnam problem 1992–B5.** For each  $n$ , evaluate the determinant,  $D_n$  of the  $(n-1) \times (n-1)$  matrix of which the  $5 \times 5$  case is

$$\begin{bmatrix} 3 & 1 & 1 & 1 & 1 \\ 1 & 4 & 1 & 1 & 1 \\ 1 & 1 & 5 & 1 & 1 \\ 1 & 1 & 1 & 6 & 1 \\ 1 & 1 & 1 & 1 & 7 \end{bmatrix}.$$

Generalize this result to the case of arbitrary entries  $1 + a_k$  along the diagonal. Hint: Note that  $D_n = n D_{n-1} + (n-1)!$ .

17. **Wilson's theorem.** Wilson's theorem is the assertion that if  $p$  is prime,  $(p-1)! \equiv -1 \pmod{p}$ . (The Lagrange converse is also true; namely, the congruence is necessary for primality.) This can be proved using a group argument, as follows. Note first that only  $1, -1$  are self-inverses modulo  $p$ , so that the product of *all other residues* consists of element-inverse pairs  $aa^{-1}$ . Thus  $\prod_{a=2}^{p-2} a \equiv 1 \pmod{p}$  and Wilson's theorem follows when you include  $1, -1$  in the product. Such observations actually lead to certain computational advantages in the evaluation of very large factorials, as in [94].
18. **Prouhet-Tarry-Escott problem.** Given positive integers  $n$  and  $k$ , this Diophantine problem asks for non trivial solutions to

$$\begin{aligned} \alpha_1 + \alpha_2 + \dots + \alpha_n &= \beta_1 + \beta_2 + \dots + \beta_n \\ \alpha_1^2 + \alpha_2^2 + \dots + \alpha_n^2 &= \beta_1^2 + \beta_2^2 + \dots + \beta_n^2 \\ &\dots \\ \alpha_1^k + \alpha_2^k + \dots + \alpha_n^k &= \beta_1^k + \beta_2^k + \dots + \beta_n^k. \end{aligned}$$

A solution is abbreviated as  $[\alpha] =_k [\beta]$ . For example

$$[-2, -1, 3] =_2 [2, 1, -3], \quad [-5, -1, 2, 6] =_3 [-4, -2, 4, 5]$$

$$[-8, -7, 1, 5, 9] =_4 [8, 7, -1, -5, -9],$$

and such *ideal solutions* with  $k = n - 1$  are known for  $n < 12$ . This can be equivalently rewritten as a question about  $\pm$  polynomials dividing  $(z - 1)^n$  of minimal length (sum of the absolute values of the coefficients) [65]. A good deal is known about the problem computationally, and the main open questions include:

- (a) Find a second inequivalent solution for  $n = 12$  where only

$$[\pm 151, \pm 140, \pm 127, \pm 186, \pm 61, \pm 22] =_{11} [\pm 148, \pm 146, \pm 1271, \pm 94, \pm 47, \pm 35]$$

is known. (Any other symmetric solution has some entry exceeding 1000 in absolute value.)

- (b) Find ideal solutions with  $n > 12$ .  
 (c) Find ideal solutions for each  $n$  or find a  $n$  with no ideal solution (more likely).

19.  $\zeta$  and arithmetic functions. Show that

$$\zeta(s) \sum_{n=1}^{\infty} \frac{a_n}{n^s} = \sum_{n=1}^{\infty} \frac{b_n}{n^s}$$

if and only if

$$\sum_{n=1}^{\infty} a_n \frac{x^n}{1 - x^n} = \sum_{n=1}^{\infty} b_n x^n,$$

$$\alpha(s) \sum_{n=1}^{\infty} \frac{a_n}{n^s} = \sum_{n=1}^{\infty} \frac{b_n}{n^s}$$

if and only if

$$\sum_{n=1}^{\infty} a_n \frac{x^n}{1 + x^n} = \sum_{n=1}^{\infty} b_n x^n.$$

Let  $\phi$  be Euler's totient function and denote  $\sigma_k(n) = \sum_{d|n} d^k$ . Write  $\sigma = \sigma_1$  and  $\tau = \sigma_0$  the number of divisors. Let  $\lambda$  be the number of prime factors of

$n$  counting multiplicity. Let  $q(n) = |\mu(n)|$ , where  $\mu$  is the Möbius function, so that  $q(n)$  is 1 when  $n$  is *quadratifrei* (squarefree) and 0 otherwise.

Deduce, using the above and facts such as Euler's product for  $\zeta$  that for  $s$  large enough to assure convergence of the Dirichlet series, the following hold.

(a)

$$\sum_{n=1}^{\infty} \frac{\phi(n)}{n^s} = \frac{\zeta(s-1)}{\zeta(s)}$$

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

(b)

$$\sum_{n=1}^{\infty} \frac{\tau(n)}{n^s} = \zeta^2(s)$$

and more generally

(c)

$$\sum_{n=1}^{\infty} \frac{\sigma_k(n)}{n^s} = \zeta(s-k)\zeta(s)$$

(d)

$$\sum_{n=1}^{\infty} \frac{\lambda(n)}{n^s} = \frac{\zeta(2s)}{\zeta(s)}$$

(e)

$$\sum_{n=1}^{\infty} \frac{\tau^2(n)}{n^s} = \frac{\zeta^4(s)}{\zeta(2s)}$$

(f)

$$\sum_{n=1}^{\infty} \frac{q(n)}{n^s} = \frac{\zeta(s)}{\zeta(2s)},$$

and more generally discover what  $d_k$  and  $q_k$  must count if

(g)

$$\sum_{n=1}^{\infty} \frac{q_k(n)}{n^s} = \frac{\zeta(s)}{\zeta(ks)}$$

and

(h)

$$\sum_{n=1}^{\infty} \frac{d_k(n)}{n^s} = \zeta^k(s).$$

If  $\sigma_k^*(n) = \sum_{d|n} (-1)^k d^k$ , consider similar formulae in terms of  $\alpha$ . Note that in each case it is very easy to check the claimed identities numerically for reasonable large  $s$ .

20. **Hurwitz's results on three and five squares.** If  $n$  is a square then

$$r_3(n) = 6 \prod_p \left[ \frac{p^{\lambda_p/2+1} - 1}{p-1} - (-1)^{(p-1)/2} \frac{p^{\lambda_p/2} - 1}{p-1} \right],$$

and

$$r_5(n) = 10 \frac{2^{3\lambda/2+3} - 1}{2^3 - 1} \prod_p \left[ \frac{p^{3\lambda_p/2+1} - 1}{p^3 - 1} - p \frac{p^{\lambda_p/2} - 1}{p^3 - 1} \right],$$

where  $p$  ranges over the odd prime factors of  $n$  and  $\lambda_p$  is its multiplicity. This and much more is discussed in [80].

21. **Berkeley problem 6.11.33.** Let  $n$  be a positive integer and let  $P$  be a given polynomial of degree  $n$ . Explicitly compute a non-trivial polynomial  $Q$  of the form

$$Q(x) = \sum_{i=0}^n a_i x^{2^i}$$

containing  $P$  as a factor. Hint: Using a computer algebra system, perform the Euclidean algorithm to write

$$x^{2^i} = Q_i(x)P(x) + R_i(x)$$

for each  $0 \leq i \leq n$ , where  $\text{degree}(R_i) \leq n-1$ . Since  $\{R_0, R_1, \dots, R_n\}$  are dependent there are  $a_i$ , not all zero, with  $\sum_{i=0}^n a_i R_i = 0$ .

22. **A polynomial problem.** (From [136, pg. 86]). Find all real numbers  $|r| < 2$  such that  $x^{14} + rx^7 + 1$  divides  $x^{154} - rx^{77} + 1$ .
23. **Putnam problem 1989–A3.** Show all roots of

$$11z^{10} + 10iz^9 + 10iz - 11 = 0$$

lie on the unit circle. Hint: This can be solved explicitly using *Maple*, *Mathematica*, or a custom-written root-finding program that employs Newton iterations (see Chapter 7).

24. **Berkeley problem 6.11.23.** Prove that  $P(x) = 1 + x + \cdots + x^{p-2} + x^{p-1}$  is irreducible over  $\mathbf{Q}$  when  $p$  is prime. Is this true more generally? Hint: Using a computer algebra system, compute  $P(y+1)$  and use Eisenstein's criterion: *it suffices to find a prime  $q$  which divides all but the leading coefficient while  $q^2$  does not divide the constant coefficient.*
25. **The Hilbert matrix.** The  $n$ -dimensional Hilbert matrix  $H(n)$  is the banded Hankel (constant on off diagonals) matrix whose entry is  $H_{i,j} = 1/(i+j-1)$ . Thus

$$H(3) = \begin{bmatrix} 1 & 1/2 & 1/3 \\ 1/2 & 1/3 & 1/4 \\ 1/3 & 1/4 & 1/5 \end{bmatrix}.$$

- (a) Let  $d_n = \det^{-1}(H(n))$ . Then the sequence starts

$$1, 12, 2160, 6048000, 266716800000, 186313420339200000.$$

Determine the closed form for this sequence, say using Sloane's encyclopedia, and prove it. Similarly consider  $K(n)$  with  $K_{i,j} = 1/(i+j)$ .

- (b) Discover the structure of  $H^{-1}(n)$ . For example

$$H^{-1}(3) = \begin{bmatrix} 9 & -36 & 30 \\ -36 & 192 & -180 \\ 30 & -180 & 180 \end{bmatrix}.$$

- (c) Hilbert matrices are very poorly conditioned, and so make good tests of numerical routines. Let  $H_D(N)$  denote the floating point evaluation of  $H(N)$  using  $D$  digits. Compare  $\det(H(15))$  and  $\det(H_D(15))$  for  $D = 10^k, k = 1, \dots, 5$ .
- (d) The corresponding infinite Hilbert matrix

$$H = \left( \frac{1}{i+j-1} \right)_{i,j=1}^{\infty}$$

induces a bounded linear operator on the square summable sequences, whose operator norm is no greater than  $\pi$ .

- (e) Indeed  $\|H\| = \pi$ .

A lovely paper on many aspects of the Hilbert matrix is [77].

Answers: (a) The value is  $d_n = \prod_{k=1}^{n-1} (2k+1) \binom{2k}{k}^2$ .

(b) The general term of the  $n \times n$  inverse is

$$(-1)^{i+j} (i+j-1) \binom{i+j-2}{i-1}^2 \binom{n+i-1}{n-j} \binom{n+j-1}{n-i}.$$

To prove this it may help to use the determinant formula

$$\det \left( \left[ \frac{1}{x_i + y_j} \right] \right) = \frac{\prod_{i>j} (x_i - x_j)(y_i - y_j)}{\prod_{i,j} (x_i + y_j)},$$

and Cramer's rule.

(c) This dramatically highlights the difference between computing symbolically and numerically.

(d) Let  $\Lambda = (\lambda_{i-j})$  be the doubly-infinite matrix where  $\lambda_0 = 0$  and  $\lambda_k = 1/k$  otherwise for  $k \in \mathbb{Z}$ . Note that one may write

$$\Lambda = \begin{bmatrix} -H & L \\ L^* & H \end{bmatrix},$$

so that  $\Lambda$  is a dilation of  $H$ . It follows that  $\|H\| \leq \|\Lambda\| = \sqrt{\|\Lambda^2\|}$ , since  $\Lambda$  is symmetric. Now direct calculation shows  $\Lambda^2 = (a_{i+j})$  is a non-negative

doubly-infinite Toeplitz matrix with  $a_0 = 2\zeta(2)$  and  $a_k = 2/k^2$  otherwise. As each row sums to  $\sum_{i \in \mathbb{Z}} a_i = 6\zeta(2) = \pi^2$ , we are done when once we observe that  $\Lambda^2 = \sum_{k \in \mathbb{Z}} a_k J_k$  where the operator  $J(k)_{i,j} = \delta_{i-j}(k)$  is zero except when  $|i - j| = k$ , and 1 in that case. Thus,

$$\|\Lambda^2\| \leq \sum_k |a_k| \|J(k)\| = \sum_k |a_k| = \pi^2$$

—Jensen and Euler strike again!

(e) Let  $T(n)$  be the Hankel matrix  $(\tau_{i-j+1})_{i,j=1}^n$  with  $\tau_k = 1/k$  for  $k > 0$  and is zero otherwise. Observe that  $A(n) \geq T(n) \geq 0$  coordinate-wise and so  $\|H\| \geq \|H(n)\| \geq \|T(n)\|$ . It thus suffices to show  $\liminf_n \|T(n)\| \geq \pi$ . To see this, use  $\|T(n)\| \geq \langle T(n)v_n, v_n \rangle / \|v_n\|^2$  for  $v_n = (1, 1/\sqrt{2}, \dots, 1/\sqrt{n})$ . Then with  $\alpha_k = \sum_{j=1}^{k-1} (j(k-j))^{-1/2}$  and  $H_n = \sum_{k=1}^n 1/k$ ,

$$\begin{aligned} \|H\| &\geq \liminf_n \frac{1}{H_n} \sum_{k=1}^n \alpha_k/k = \liminf_k \alpha_k \\ &= \liminf_k \frac{1}{k} \sum_{j=1}^{k-1} \left(\frac{j}{k}\right)^{-1/2} \left(1 - \frac{j}{k}\right)^{-1/2}, \end{aligned}$$

which is a Riemann sum for  $\int_0^1 x^{-1/2}(1-x)^{-1/2} dx = \beta(1/2, 1/2) = \pi$ .

26. **The holes in the argument.** Figure 5.1 plots all roots of polynomials,  $\mathcal{B}_N$ , with coefficients in  $\{0, 1, -1\}$  up to degree  $N = 18$ . The graphically displayed information would be very hard to digest numerically. The zeros in Figure 5.1 are colored by their local density normalized to the range of densities; from red for low density to yellow for high density.

The fractal structure and the holes around the roots are of different shapes and precise locations. This and more is described in [66]. For example, when  $\alpha$  is a Pisot number

$$C_1 \frac{1}{\alpha^N} \leq \min_{\alpha \neq \beta \in \mathcal{B}_N} |\alpha - \beta| \leq C_2 \frac{1}{\alpha^N},$$

for constants  $C_1, C_2 > 0$ . Similarly, for  $\alpha$  a  $d$ -th root of unity,

$$C_3 \frac{1}{N^{(1+\phi(d)/2)(k+1)}} \leq \min_{\alpha \neq \beta \in \mathcal{B}_N} |\alpha - \beta| \leq C_4 \frac{1}{N^{(1+\phi(d)/2)(k+1)}},$$

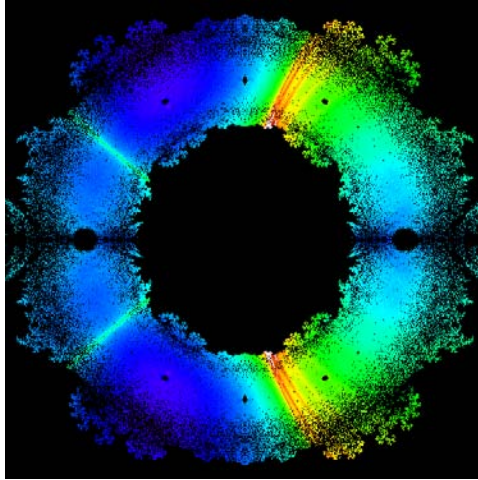


Figure 5.1: Zeros of zero-one polynomials

for constants  $C_1, C_2 > 0$ . Unexplained phenomena can be seen in images at the URL <http://www.cecm.sfu.ca/personal/loki/Projects/Roots/Book>.

27. **Rudin-Shapiro polynomials.** These are defined recursively by  $P_0(z) = Q_0(z) = 1$  and

$$\begin{aligned} P_{n+1}(z) &= P_n(z) + z^{2^n} Q_n(z), \\ Q_{n+1}(z) &= P_n(z) - z^{2^n} Q_n(z). \end{aligned}$$

- (a)  $P_n$  and  $Q_n$  are even of degree  $2^n - 1$  and have only  $\pm 1$  coefficients.  
 (b) For  $|z| = 1$

$$|P_{n+1}(z)|^2 + |P_{n+2}(z)|^2 = 2(|P_n(z)|^2 + |P_{n+1}(z)|^2)$$

$$\text{and so } |P_n(z)|, |Q_n(z)| \leq 2^{(n+1)/2} \leq \sqrt{2}(2^n - 1).$$

In consequence, these are examples of  $\pm 1$  polynomials  $p_n$  that grow no faster than  $\sqrt{2} \text{degree}(p_n)$  on the unit disk. A famous problem of Littlewood [65] is to find polynomials  $p_n$  of degree  $n$  with

$$c_1 \sqrt{n} \leq |p_n(z)| \leq c_2 \sqrt{n},$$

for all  $n$ .



28. **Hyperbolic polynomials and self-concordant barriers for hyperbolic means.** Following the development in [153], we shall construct two classes of self-concordant barrier functions on natural convex sets.

Let  $p : \mathbf{R}^n \mapsto \mathbf{R}$  be a polynomial on  $n$  variables, *homogeneous* of degree  $m$ , that is,  $p(tx) = t^m p(x)$  for every  $t \in \mathbf{R}$  and  $x \in \mathbf{R}^n$ . We say that  $p(x)$  is *hyperbolic* in the direction  $d \in \mathbf{R}^n$  if the polynomial  $t \mapsto p(x + td)$  has  $m$  real roots for every  $x$ . Denote the negative of these roots, which depend on  $x$  and  $d$ , by  $t_i(x, d)$ ,  $i = 1, \dots, m$ .

The following are three important examples of hyperbolic polynomials.

- (a)  $p_1 : x \in \mathbf{R}^n \mapsto \prod_{i=1}^n x_i$  with respect to the direction  $d = (1, \dots, 1)$ .
- (b)  $p_2 : x \in \mathbf{R}^n \mapsto x_1^2 - \sum_{i=2}^n x_i^2$  with respect to the direction  $d = (1, 0, \dots, 0)$ .
- (c) Let  $\mathcal{S}^n$  be the space of  $n \times n$  symmetric matrices and recall that it is isomorphic to  $\mathbf{R}^{n(n+1)/2}$ . The polynomial in the entries of the symmetric matrix given by  $p_3 : X \in \mathcal{S}^n \mapsto \det(X)$  is hyperbolic with respect to the direction  $d = I$  (the  $n \times n$  identity matrix).

Define the sets

$$C(p, d) = \{x \in \mathbf{R}^n : p(x + td) \neq 0, \forall t \geq 0\},$$

$$C_a(p, d) = \{x \in \mathbf{C} : p(x) > a\}, \text{ for } a \geq 0.$$

It is known (see [114]) that  $C(p, d)$  is an open, convex cone; that  $p$  is hyperbolic in the direction of any vector in  $C(p, d)$ ; also for any  $c \in C(p, d)$ ,  $C(p, c) = C(p, d)$ ; and that  $p(x)^{1/m}$  is a concave function on  $C(p, d)$  and zero on the boundary. (See Figure 5.2.)

On the other hand, let  $Q$  be a convex set in  $\mathbf{R}$ , and  $F$  a real valued function defined on  $Q$ . We denote the boundary of  $Q$  by  $\partial Q$ , and by  $D^k F(x)[h, \dots, h] = \left. \frac{d^k}{dt^k} F(x + th) \right|_{t=0}$  the  $k$ -th directional derivative of  $F$  in the direction  $h$  evaluated at zero. We say that  $F$  is a  $\theta$ -self-concordant barrier on  $Q$  if the following three conditions are satisfied:

- (i)  $|DF(x)[h]| \leq \sqrt{\theta} (D^2 F(x)[h, h])^{1/2}$
- (ii)  $|D^3 F(x)[h, h, h]| \leq 2(D^2 F(x)[h, h])^{3/2}$ , and
- (iii)  $F(x^r) \rightarrow \infty$  for any sequence  $x^r \rightarrow x \in \partial Q$ .

- (a) Following the steps given below, prove that:
- i. The function  $F(x) = -\log(p(x))$  is an  $m$ -self-concordant barrier on the set  $C(p, d)$ , (see [128]).
  - ii. For any  $a \geq 0$ , the function  $F(x) = -m \log(p(x) - a)$  is an  $m^2$ -self-concordant barrier on the set  $C_a(p, d)$ , (see [153]).

Self concordance plays a central role in interior point methods of modern optimization.

- (b) Suggested steps on the path.
- i. The barrier condition (iii) should be clear from the mentioned properties of hyperbolic polynomials on the set  $C(p, d)$ .
  - ii. Show that the following representation holds

$$p(x + r d) = p(d) \prod_{i=1}^m (r + t_i(x, d)).$$

The part that requires thought is to determine the constant  $p(d)$  on the right-hand side. (Keep in mind that  $\{t_i(x, d)\}_{i=1}^m$  are the negative of the roots of the hyperbolic polynomial.)

- iii. Use the above factorization and the homogeneity of  $p$  to show that for any  $x \in C(p, d)$  and any  $h \in \mathbf{R}^n$  we have

$$p(x + r h) = p(x) \prod_{i=1}^m (1 + r t_i(h, x)).$$

- iv. Use this representation to obtain the derivatives of  $r \mapsto p(x + r h)$  for any  $x \in C(p, d)$  and  $h \in \mathbf{R}^n$ :

$$\frac{d}{dr} p(x + r h) = p(x + r h) \sum_{i=1}^m \frac{t_i(h, x)}{1 + r t_i(h, x)}.$$

- v. Calculate the directional derivatives of  $F(x) = -\log(p(x))$  and prove the self-concordant inequalities hold with  $\theta = m$ .
- vi. For the second part, consider first the case  $a = 0$ . Next, use a linear substitution to argue without loss of generality that  $a = 1$ .

vii. Set  $t_i = t_i(h, x)$  and define  $\alpha = p(x) - 1$ ,

$$C_1 = \sum_{i=1}^m t_i, C_2 = \sum_{i=1}^m t_i^2, C_3 = \sum_{i=1}^m t_i^3.$$

Show that for  $F(x) = -m \log(p(x) - 1)$  we have

$$\begin{aligned} DF(x)[h] &= -m \frac{\alpha + 1}{\alpha} C_1, \\ D^2F(x)[h, h] &= m \frac{\alpha + 1}{\alpha^2} C_1^2 + m \frac{\alpha + 1}{\alpha} C_2, \text{ and} \\ D^3F(x)[h, h, h] &= -m \frac{(\alpha + 1)(\alpha + 2)}{\alpha^3} C_1^3 - 3m \frac{(\alpha + 1)}{\alpha^2} C_1 C_2 \\ &\quad - 2m \frac{\alpha + 1}{\alpha} C_3. \end{aligned}$$

viii. Use the Cauchy-Schwarz inequality to show that the first inequality in the definition of self-concordancy holds, with  $\theta = m^2$ .

ix. To show the second inequality in the definition, square both sides and group terms with respect to powers of  $\alpha$ . Notice that the inequality is homogeneous of degree one with respect to the vector  $(t_1, \dots, t_m)$ , thus without loss  $C_1 = \pm 1$ . Notice also that  $\alpha > 0$  since  $x \in C_a(p, d)$ , and  $a = 1$ . Now, show that all coefficients in front of the different powers of  $\alpha$  are positive.

(c) Verify the three examples of hyperbolic polynomials given at the beginning of this item.

(d) Try to prove the inequalities in a computer algebra system.

29. **The Lax conjecture is true.** An elegant 1958 conjecture of Lax concerning hyperbolic polynomials has recently been settled by Lewis, Parrilo and Ramana [152]:

**Theorem 5.7.2** *A polynomial  $p$  in three real variables is hyperbolic of degree  $d$  with respect to  $e = (1, 0, 0)$ , with  $p(e) = 1$ , if and only if there are symmetric  $d \times d$  matrices  $B, C$  such that*

$$p(x, y, z) = \det(xI + yB + zC).$$

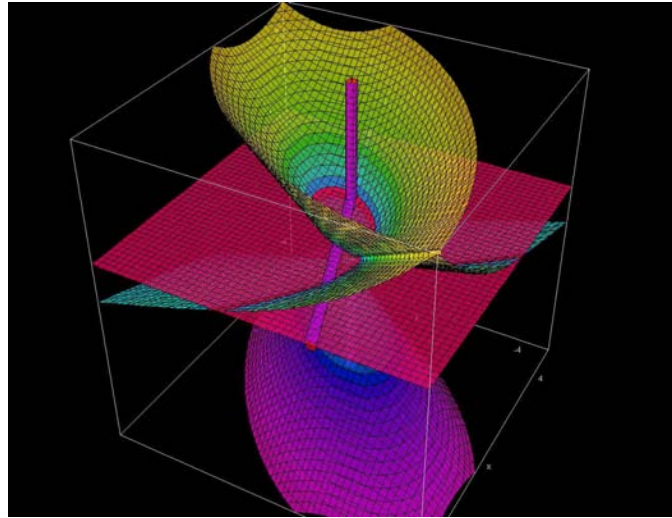


Figure 5.2: A hyperbolic cone

Deduce what characterizes the corresponding hyperbolic polynomials in two variables.

Figure 5.2 shows a third-order hyperbolic cone. It was drawn by Pablo Parrilo in a fine personal computer drawing package called *DPGraph* (for “dynamic photorealistic graphing”). The cone is actually the part on the top of the figure, all the rest is the “hidden.” There is a horizontal plane at  $z = 1$ , and on that plane a nice set with the “rigid convexity” property that contains the point  $(0, 0)$  near the center. The vertical line is in the direction  $e = (0, 0, 1)$  of hyperbolicity.

30. **Inverse problems.** Inverse problems provide a host of challenges for the computationally and experimentally minded. Examples, both explicit and implicit, have occurred throughout our volumes (e.g., maximum entropy problems, the self-concordant barriers above, the nuclear magnetic resonance example, and while exploring JPEG compression or the Watson integrals).
31. **Nonnegative polynomials.** The nonnegative polynomials of degree  $d$ ,  $P_+^d(I)$ , on an interval  $I$ , form an interesting convex cone in the vector space

of all polynomials  $P^d(I)$ . Recall that an *extreme ray* of a convex cone  $K$  is an element with the property that  $e = p + q, p, q \in K$  implies  $q$  and  $p$  are nonnegative multiples of  $e$ .

- (a) Characterize the extreme rays of  $P_+^d([0, \infty))$ , using the fundamental theorem of algebra. Deduce that every polynomial  $p$  that is nonnegative on  $\mathbb{R}^+$  can be expressed as

$$p(t) = t \sum_I p_i^2(t) + \sum_J p_j^2(t),$$

where  $p_i, p_j$  are finitely many real polynomials with nonnegative roots.

- (b) Deduce that every polynomial  $p$  that is nonnegative on  $\mathbb{R}$  can be expressed as

$$p(t) = \sum_I p_i^2(t),$$

where  $p_i$  are polynomials with real roots. Prove also that when  $p$  has even degree, it is the square of at most two polynomials.

- (c) Characterize the extreme rays of  $P_+^d(I)$ , for a finite interval.
- (d) **Hilbert's theorem.** A nonnegative polynomial in three variables, homogeneous of degree four, is expressible as a sum of squares of quadratic forms (in the three variables). The hypotheses are needed as the nonnegative polynomials below show.

- i.  $x^4y^2 + x^2y^4 + z^6 - 3x^2y^2z^2$  (Motzkin);
- ii.  $w^4 + x^2y^2 + y^2z^2 + 2z^2x^2 - 4xyzw$  (Choi and Lam).

Hint: A boundary polynomial must have a zero in the interval. An extreme polynomial must be of full degree with all roots in the interval. In each case, every polynomial in the cone is in the closed conical hull of the extreme directions. See [20] for details.

32. **Roots of a polynomial.** (From [136, pg. 87]). Let  $\alpha_1, \alpha_2, \alpha_3, \alpha_4$  be the roots of the polynomial

$$P(x) = x^4 + px^3 + qx^2 + rx + 1.$$

Show that

$$(1 + \alpha_1^4)(1 + \alpha_2^4)(1 + \alpha_3^4)(1 + \alpha_4^4) = (p^2 + r^2)^2 + q^4 - 4pq^2r.$$

Hint: consider  $\prod_{k=1}^4 P(e^{(2k-1)i\pi/4})$ .

33. **The Ehrhardt polynomial of a polytope.** An *integer polytope* is the convex hull of finitely many points with integer coefficients. The following result neatly links convexity, lattices and polynomials.

**Theorem 5.7.3 Ehrhardt polynomial.** *Let  $P$  be an integer polytope in  $\mathbb{R}^d$ . There exists a univariate polynomial of degree at most  $d$ ,  $p_P$ , such that*

$$p_P(k) = |(kP) \cap \mathbb{Z}^d|,$$

for all nonnegative integers  $k$ .

Let  $P$  be a given integer polytope.

- (a) Show that  $\dim P = \text{degree } p_P$ .
- (b) For integer polygons, use Pick's theorem to establish the existence of the Ehrhardt polynomial and moreover it has nonnegative coefficients.
- (c) Show that

$$p_P(-k) = (-1)^{\dim P} |\text{relint}(kP) \cap \mathbb{Z}^d|,$$

for all positive integers  $k$ . Here "relint" denotes the relative interior of  $P$  in its affine span (that is, the inside).

- (d) Show that the constant term of  $p_P$  is 1 and the coefficient of  $t^d$  is the volume of  $P$  (which could be zero). The coefficient of  $t^{d-1}$  is half of a "surface area."
- (e) Fix a positive integer  $m$ . Consider the Ehrhardt polynomial,  $p_m$ , of the tetrahedron with vertices  $(0, 0, 0)$ ,  $(1, 0, 0)$ ,  $(0, 1, 0)$  and  $(1, 1, m)$ . Show that

$$p_m(k) = \frac{m}{6}k^3 + k^2 + \frac{12-m}{6}k + 1.$$

This is described in Barvinok's fine recent book [20].

34. **The Schur functions.** We consider the symmetric functions,  $\Lambda$ , that is, the polynomials in indeterminates  $x_r$ , for  $r$  in  $\mathbb{N}$ . (See [156].)

- (a) Let  $\lambda = (l_1, l_2, \dots, l_n)$  be a partition of length  $n$ . Write  $l(\lambda)$  for the length of  $\lambda$ . The corresponding *Schur function* is defined by

$$s_\lambda = \det \left( x_i^{l_j + n - j} \right) / \det \left( x_i^{n - j} \right)_{1 \leq i, j \leq n}.$$

- (b) Show that the *elementary symmetric functions*,  $e_0 = 1$  and  $e_r = \sum_{l_1 < l_2 < \dots < l_r} x_{l_1} x_{l_2} \cdots x_{l_r}$ , for  $r \geq 1$ , have an ordinary generating function

$$E(t) = \sum_{r \geq 0} e_r t^r = \prod_{i=1}^n (1 + tx_i).$$

Correspondingly, the *complete symmetric functions*,  $h_r$ , have the generating function

$$H(t) = \sum_{r \geq 0} h_r t^r = \prod_{i=1}^n (1 - tx_i)^{-1}.$$

- (c) For  $r \geq 1$ , the formal  $r$ -th *power sum* is defined by  $p_r = \sum x_i^r$ . Show that the ordinary generating function

$$P(t) = \sum_{r \geq 1} p_r t^{r-1} = \sum_{i \geq 1} \frac{x_i}{1 - x_i t}$$

satisfies  $P(t) = E'(t)/E(t)$  and  $P(-t) = H'(t)/H(t)$ .

- (d) Show that the determinant expression for  $s_\lambda$  in terms of the  $h_r$  is

$$s_\lambda = \det (h_{l_i + i - j})_{1 \leq i, j \leq n},$$

for any  $n \geq l(\lambda)$ . Find a similar formula in terms of the  $e_r$ .

- (e) Conclude that every symmetric function is uniquely expressible as a polynomial in the elementary symmetric, the complete symmetric functions or in the Schur functions. That is, each class is a  $\mathbb{Z}$ -basis for the ring  $\Lambda$ .

- (f) Since the  $h_r$  and the  $e_r$  are each algebraically independent, it is possible to substitute as one wishes. For example

$$H(t) = \prod_{i \geq 0} \frac{1 - bq^i t}{1 - aq^i t}$$

leads to  $h_r = \prod_{i=1}^r (a - bq^{i-1}) / (1 - q^i)$ ,  $e_r = \prod_{i=1}^r (aq^{i-1} - 1) / (1 - q^i)$  and  $p_r = (a^r - b^r) / (1 - q^r)$ .

- (g) The *natural partial ordering* orders partitions by  $\lambda \geq \mu$  if  $\sum \lambda_i \lambda_i \geq \sum \mu_i$  for all  $i$ . Show that for two partitions of  $n$ ,  $\lambda \geq \mu$  if and only if there is a  $n \times n$  doubly stochastic matrix  $S$  with  $\mu = S\lambda$ . Represent this result graphically.

35. **Schur functions and Young tableaux.** Another often more useful representation is the *Young diagram* in which the points of Ferrer's diagram are replaced by squares. The *conjugate partition*,  $\lambda'$ , arises on exchanging rows and columns of the diagram. Shading squares allows one to usefully represent adding one partition to another, and much else. We now sketch a combinatorial approach to Schur functions following [196].

A tableau of shape  $\lambda$  is an array  $T = (T_{ij})$  of positive integers such that  $1 \leq i \leq \ell(\lambda)$ ,  $1 \leq j \leq \lambda_i$  that is nondecreasing in each row and strictly increasing in each column. We define the *type*  $\alpha(T) = (\alpha_1, \alpha_2, \dots)$  to be the number of occurrences of  $i$  in  $T$ . Thus, Figure 5.3 shows a (*semistandard*) *Young tableau*  $\hat{T}$  of shape  $(6, 5, 3, 3)$  and of type  $(3, 1, 1, 4, 4, 4, 1, 0, 2)$ . A *standard tableau* has each row and column strictly increasing.

A natural combinatorial definition of the Schur functions is to define

$$s_\lambda(x) = \sum_T x^T$$

summed over all monomials  $x^T = x^{\alpha_1(T)} x^{\alpha_2(T)} \dots$  of shape  $\lambda$ . For the tableau illustrated,  $x^{\hat{T}} = x_1^3 x_2 x_3 x_4^4 x_5^5 x_6 x_7 x_9^2$ .

- (a) Write out  $s_{(2,1)}$  explicitly.  
 (b) Let  $f^\lambda$  denote the number of standard tableaux of shape  $\lambda$ . Show that  $f^{(3,2)} = 5$  and then show that  $f^\lambda$  counts the number of lattice



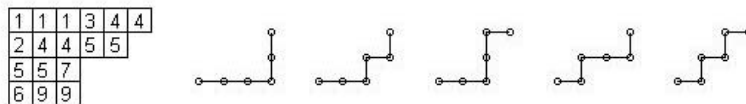


Figure 5.3: A Young tableau along with all lattice paths of shape  $(3,2)$

paths from  $0$  to  $\lambda$  in  $\mathbb{R}^{\ell(\lambda)}$  so that each step is a unit vector staying in the cone  $x_1 \geq x_2 \geq \dots \geq x_l \geq 0$ . These paths for  $\lambda = (3,2)$  are drawn in Figure 5.3.

- (c) Prove that this definition of Schur function coincides with the definition of Exercise 34.

John Stembridge's *Maple* package, <http://www.math.lsa.umich.edu/~jrs/> computes with Schur functions and much else.

36. **Small gaps between consecutive primes.** As this book was being completed, a breakthrough was announced in a long-standing question of the spacings of prime numbers [123]. The prime number theorem can be paraphrased as saying that the average size of  $p_{n+1} - p_n$  is  $\log p_n$ , where  $p_n$  denotes the  $n$ -th prime, so that

$$\Delta = \liminf_{n \rightarrow \infty} \frac{p_{n+1} - p_n}{\log p_n} \leq 1.$$

In 1926 Hardy and Littlewood showed that assuming the generalized Riemann hypothesis,  $\Delta \leq 2/3$ . Other researchers in the intervening lowered this figure, until 1986 when Maier established the result  $\Delta \leq 0.2486$ .

In a dramatic new development, Daniel Goldston and Cem Yildirim announced in April 2003 that not only is  $\Delta = 0$ , but also

$$p_{n+1} - p_n < (\log p_n)^{8/9}$$

holds for infinitely many  $n$ . What's more for any fixed integer  $r$  the inequality

$$p_{n+r} - p_n < (\log p_n)^{(8r)/(8r+1)}$$

holds for infinitely many  $n$ . They anticipated that even stronger results of this type can be established (such as decreasing  $8/9$  to  $4/5$ ). This report was widely hailed in the mathematical community and the press, after being circulated in preprint and described at conferences.

Unfortunately, Andrew Granville and K. Soundararajan subsequently found a problem in one of the arguments—some “small” error terms are actually of the same order of magnitude as the main term. As of this date the difficulty remains unresolved.

37. **More from Littlewood's miscellany.** [155] We conclude the chapter with a series of quotes and observations from Littlewood.

(a) (Page 56)

“(A. S. Besicovitch) A mathematician's reputation rests on the number of bad proofs he has given. (Pioneer work is clumsy.)”

(b) (Page 60)

“A precision professor had the habit of saying ‘... quartic polynomial  $ax^4 + bx^3 + cx^2 + dx + e$ , where  $e$  need not be the base of the natural logarithms.’ (It might be.)”

(c) (Page 61)

“I read in the proof-sheets of Hardy on Ramanujan: ‘As someone said, each of the positive integers was one of his personal friends.’ My reaction was, ‘I wonder who said that; I wish I had.’ In the next proof-sheets I read what now stands, ‘It was Littlewood who said ...’

(What had happened was that Hardy had received the remark in silence and with poker face, and I wrote it off as a dud. I later taxed Hardy with this habit; on which he replied: ‘Well, what is one to do, is one always to be saying “damned good’?’ To which the answer is ‘yes.’)”

- (d) (Pages 118–120) *Random Jottings on G. H. Hardy* (after a 35 year collaboration) “His spelling was not immaculate.” . . . “He preferred the Oxford atmosphere and said they took him seriously, unlike Cambridge.” . . . “He took a sensual pleasure in “calligraphy” and it would have been a deprivation if he didn’t make the final copy of a joint paper. (My standard role in a joint paper was to make the logical skeleton, in shorthand—no distinction between  $r$  and  $r^2$ ,  $2\pi$  and 1, etc., etc. But when I said ‘Lemma 17’ it stayed Lemma 17.)” . . . “He was indifferent to noise; very rare in creative workers at least when no longer young.”
- (e) (Page 149) “Creative workers need drink at night, ‘Roses and dung’. (Or: mathematicians read ‘rubbish’.) An experimentalist, having spent the day looking for a leak, has had a perfect mental rest by dinner time, and overflows with minor mental activity.”
- (f) (Page 164) “ ‘Always verify references.’ This is so absurd in mathematics that I used to say provocatively: ‘never . . .’.”



## Chapter 6

# The Power of Constructive Proofs II

Mathematical proofs, like diamonds, are hard as well as clear, and will be touched with nothing but strict reasoning.

John Locke, 1690 [106, page 115]

As we noted in the first volume, a computational or constructive approach can often make a proof significantly more understandable. What's more, the key concepts involved are made more apparent, and are more likely to be remembered later. Here we present some additional examples of computer-aided proofs, including some examples of “variational” methods of proof.

### 6.1 A More General AGM Iteration

We first study a more general form of the arithmetic-geometric mean iteration. This material has been adapted from [63]. Let  $a$  and  $b$  be real numbers,  $a > b > 0$ , and let  $N$  be an integer greater than 1. By the *iteration*  $AG_N$  we mean the following two-term recursion:  $a_0 = a$ ,  $b_0 = b$ , and, for any  $k \geq 0$ ,

$$a_{k+1} = \frac{a_k + (N-1)b_k}{N} \tag{6.1.1}$$

$$b_{k+1} = \frac{\sqrt[N]{(a_k + (N-1)b_k)^N - (a_k - b_k)^N}}{N}. \tag{6.1.2}$$

In the case  $N = 2$  we get the standard arithmetic-geometric mean (AGM) iteration discussed in the previous section. The case  $N = 3$  was studied in detail in [46], [49] and [45].

We have

$$a_{k+1}^N - b_{k+1}^N = \left( \frac{a_k - b_k}{N} \right)^N \quad \text{for any } k \geq 0. \quad (6.1.3)$$

This shows both the global convergence and local  $N$ th-order convergence of the iteration. So there is a common limit of  $(a_k)$  and  $(b_k)$ ,

$$M_N(a_0, b_0) = \lim_{k \rightarrow \infty} a_k = \lim_{k \rightarrow \infty} b_k. \quad (6.1.4)$$

Let  $*$  be the involution on  $[0, 1]$  defined by  $x^* = \sqrt[N]{1 - x^N}$ . Whenever we use the function symbol  $*$ , the respective value of  $N$  is clear from the context.

Because of the homogeneity

$$M_N(\lambda a, \lambda b) = \lambda M_N(a, b) \quad \text{for } \lambda > 0, \quad (6.1.5)$$

it is enough to investigate

$$A_N(x^N) = \frac{1}{M_N(1, x^*)}, \quad (6.1.6)$$

for  $0 < x < 1$ .

From (6.1.3) we can argue that  $a_k^N(z)$  and  $b_k^N(z)$  are analytic in the unit disk and converge uniformly to  $M_N(1, z)$  therein. Thus  $A_N(x^N)$  is analytic in a neighborhood of zero. Notice that  $A_N(0) = 1$ .

### 6.1.1 The Functional Equation for $A_N$

For any  $N > 1$  and any  $0 < x < 1$  we have, applying one step of the  $AG_N$  iteration,

$$M_N(1 + (N - 1)x, 1 - x) = M_N(1, x^*). \quad (6.1.7)$$

Further, by (6.1.5) we have

$$M_N(1 + (N - 1)x, 1 - x) = (1 + (N - 1)x) \cdot M_N \left( 1, \frac{1 - x}{1 + (N - 1)x} \right). \quad (6.1.8)$$

Combining the right-hand sides of (6.1.8) and (6.1.7), we arrive at the following functional equation for  $A_N$ :

$$(1 + (N - 1)x) \cdot A_N(x^N) = A_N \left( 1 - \left( \frac{1 - x}{1 + (N - 1)x} \right)^N \right). \quad (6.1.9)$$

Notice that (6.1.9) uniquely determines the Taylor series for  $A_N$  at 0 and hence has a *unique* analytic solution

$$A_N(x^N) = \frac{1}{M_N(1, x^*)}. \quad (6.1.10)$$

### 6.1.2 The Quadratic Case Recovered

For  $N = 2$ , the equation (6.1.9) specializes to

$$(1 + x) \cdot A_2(x^2) = A_2 \left( \frac{4x}{(1 + x)^2} \right), \quad (6.1.11)$$

whose solution was found by Gauss (see [44]) in the form

$$A_2(x^2) = \sum_{i=0}^{\infty} \binom{2i}{i}^2 \left( \frac{x}{4} \right)^{2i} = {}_2F_1 \left( \begin{matrix} 1/2, 1/2 \\ 1 \end{matrix} \middle| x^2 \right). \quad (6.1.12)$$

Gauss discovered this closed form after having been inspired by a great number of experimental numerical calculations (see [44, page 5, 7]).

Another way of expressing the  $AG_2$  limit is by means of the following definite integral:

$$I_2(a_0, b_0) = \int_0^{\infty} \frac{dt}{\sqrt{(t^2 + a_0^2)(t^2 + b_0^2)}}. \quad (6.1.13)$$

Then it is straightforward to prove (see [44]) that

$$M_2(a_0, b_0) = \frac{I_2(1, 1)}{I_2(a_0, b_0)} = \frac{\pi}{I_2(a_0, b_0)}. \quad (6.1.14)$$

The core step of this proof, namely showing the  $AG_2$ -invariance

$$I_2(a, b) = I_2 \left( \frac{a + b}{2}, \sqrt{ab} \right) \quad (6.1.15)$$

follows fairly easily by the substitution  $u = \frac{1}{2}(t - ab/t)$ . We leave this as a *Maple* or *Mathematica* exercise.

Yet another way of proving the  $AG_2$  limit formula is by identifying  $M_2(1, x)$  as a solution of a second-order linear differential equation, see [44].

### 6.1.3 The Cubic Case Solved

The closed form for  $M_3(1, x^*)$  was identified and proved in several ways in [46, 49, 45].

Nowadays, the discovery of the closed form for  $A_3(x^3)$  as a formal power series again is a routine task using any computer mathematics software. We obtain

$$\frac{1}{M_3(1, x^*)} = A_3(x^3) = {}_2F_1\left(\begin{matrix} 1/3, 2/3 \\ 1 \end{matrix} \mid x^3\right). \quad (6.1.16)$$

The cubic counterpart of (6.1.13) is

$$I_3(a_0, b_0) = \int_0^\infty \frac{t \, dt}{\sqrt[3]{(t^3 + a_0^3)(t^3 + b_0^3)^2}}. \quad (6.1.17)$$

We have (see [40, 42])

$$M_3(a_0, b_0) = \frac{I_3(1, 1)}{I_3(a_0, b_0)} = \frac{2\pi}{\sqrt{27}} \cdot \frac{1}{I_3(a_0, b_0)}. \quad (6.1.18)$$

Again, the crucial part of the proof of (6.1.18) is to show that the integral is invariant with respect to the  $AG_3$  iteration, that is,

$$I_3(a, b) = I_3\left(\frac{a + 2b}{3}, \sqrt[3]{b \frac{a^2 + ab + b^2}{3}}\right), \quad (6.1.19)$$

for all  $a > b > 0$ . The proof of (6.1.19) was proposed by one of the present authors [40] as Part (a) of the *American Mathematical Monthly* Problem 10281. The solutions published by the *Monthly* are concluded with the editorial comment which points out that "... There is still no self-contained proof that avoids exploiting the identification with a hypergeometric function." ([42, pg. 183])



Recently, John A. Macdonald found a proof of (6.1.19) that consists of a chain of variable substitutions together with the split of the integration range at the point 1. Here we present a variation of this proof, in which all integral substitutions have been simplified to the extent that they can be checked by a computer. The proof was announced in the “Revivals” section of the Monthly [41].

A large portion of the proof of (6.1.19) is encapsulated in the following lemma.

**Lemma 6.1.1** *For any  $\gamma \in (0, 1]$ ,*

$$I_3(1, \sqrt[3]{1-\gamma^3}) = \int_1^\infty \frac{dx}{\sqrt{(x-1)((x+3)x^2-4\gamma^3)}}. \quad (6.1.20)$$

**Proof.** We first note that

$$I_3(1, \sqrt[3]{1-\gamma^3}) = \int_0^\infty \frac{u \, du}{\sqrt{(u^3-1)^2+4(1-\gamma^3)u^3}}, \quad (6.1.21)$$

using the substitution  $\boxed{u^3 = t^3(1+t^3)/((1-\gamma^3)+t^3)}$ . Further, the equality

$$\int_0^\infty \frac{u \, du}{\sqrt{(u^3-1)^2+4(1-\gamma^3)u^3}} = \int_1^\infty \frac{(u+1) \, du}{\sqrt{(u^3-1)^2+4(1-\gamma^3)u^3}}$$

follows from

$$\int_0^\infty \frac{u \, du}{\sqrt{(u^3-1)^2+4(1-\gamma^3)u^3}} = \int_0^\infty \frac{du}{\sqrt{(u^3-1)^2+4(1-\gamma^3)u^3}}$$

and

$$\int_0^1 \frac{(u+1) \, du}{\sqrt{(u^3-1)^2+4(1-\gamma^3)u^3}} = \int_1^\infty \frac{(u+1) \, du}{\sqrt{(u^3-1)^2+4(1-\gamma^3)u^3}},$$

which both are easily verified by substituting  $\boxed{u = 1/u}$ .

Now

$$\int_1^\infty \frac{dx}{\sqrt{(x-1)((x+3)x^2-4\gamma^3)}} = \int_1^\infty \frac{(u+1) \, du}{\sqrt{(u^3-1)^2+4(1-\gamma^3)u^3}}$$

using the substitution  $x = u + 1/u - 1$ . This completes the proof of the lemma.  $\square$

For any  $y, z > 0$ , we have

$$I_3(y, z) = \frac{1}{y} \cdot I_3\left(1, \frac{z}{y}\right). \quad (6.1.22)$$

Let  $c = b/a$ . It can be verified easily that, in view of (6.1.22), we can rewrite (6.1.19) as

$$\frac{2c+1}{3} \cdot I_3(1, c) = I_3(1, c^\wedge), \quad (6.1.23)$$

where  $x \mapsto x^\wedge$  is defined by

$$x^\wedge = \sqrt[3]{\frac{9x(1+x+x^2)}{(1+2x)^3}}, \quad (6.1.24)$$

for any  $x \in [0, 1]$ .

We will denote function composition in the obvious way, e.g., by  $c^{*\wedge}$  we mean  $(c^*)^\wedge$ . One can check easily that, for any  $c \in [0, 1]$ ,

$$(2c+1)(2c^{\wedge*}+1) = 3, \quad (6.1.25)$$

or, in other words,

$$c^{\wedge*} = \frac{1-c}{1+2c}. \quad (6.1.26)$$

Recall that  $*$  is an involution on  $[0, 1]$ . From (6.1.26), it follows that also  $^{\wedge*}$  is an involution on  $[0, 1]$ . Thus,

$$^{\wedge*\wedge} = * \quad (6.1.27)$$

as functions on  $[0, 1]$ .

By an inspection of the function  $^{\wedge*}$ , we can see that the statement

$$\forall c \in (0, 1] \quad \frac{2c+1}{3} \cdot I_3(1, c) = I_3(1, c^\wedge) \quad (6.1.28)$$

is equivalent to

$$\forall C \in [0, 1) \quad \frac{2C^{\wedge*} + 1}{3} \cdot I_3(1, C^{\wedge*}) = I_3(1, C^{\wedge*\wedge}). \quad (6.1.29)$$

Taking into account (6.1.27), (6.1.25) and swapping the sides, we see that the last statement is equivalent to

$$\forall C \in [0, 1) \quad (2C + 1) \cdot I_3(1, C^*) = I_3(1, C^{\wedge*}). \quad (6.1.30)$$

Using (6.1.20), the equality (6.1.30) translates to

$$(2C + 1) \cdot \int_1^\infty \frac{dx}{\sqrt{(x-1)((x+3)x^2 - 4C^3)}} \quad (6.1.31)$$

$$= \int_1^\infty \frac{dz}{\sqrt{(z-1)\left((z+3)z^2 - 36\frac{C(1+C+C^2)}{(2C+1)^3}\right)}}, \quad (6.1.32)$$

which can be proven by the substitution

$$z = \frac{(x-C)^3(x-1)}{(x^3-C^3)(2C+1)} + 1.$$

This completes the proof of (6.1.23) and thus also the proof of (6.1.19).

## 6.2 Variational Methods and Proofs

Whenever one can formulate a mathematical problem in variational form (i.e., the solution is the minimum of some function with or without constraints), one has access to a variety of constructive tools.

A key and representative tool is the next result. It states that if a closed function (a function that is defined on a closed set, lower semicontinuous, somewhere finite and nowhere negative infinity) attains a value close to its infimum at some point then a nearby point minimizes a slightly perturbed function.

**Theorem 6.2.1 (Ekeland variational principle)** *Suppose the function  $f : \mathbb{R}^n \rightarrow (-\infty, +\infty]$  is closed and the point  $x \in \mathbb{R}^n$  satisfies  $f(x) \leq \inf f + \epsilon$  for some real  $\epsilon > 0$ . Then for any real  $\lambda > 0$  there is a point  $v \in \mathbb{R}^n$  satisfying the conditions*

- (a)  $\|x - v\| \leq \lambda$ ,
- (b)  $f(v) \leq f(x)$ , and
- (c)  $v$  is the unique minimizer of the function  $f(\cdot) + (\epsilon/\lambda)\|\cdot - v\|$ .

**Proof.** We can assume  $f$  is proper, and by assumption it is bounded below. Since the function

$$f(\cdot) + \frac{\epsilon}{\lambda}\|\cdot - x\|$$

therefore has compact level sets, its set of minimizers  $M \subset \mathbb{R}^n$  is nonempty and compact. Choose a minimizer  $v$  for  $f$  on  $M$ . Then for points  $z \neq v$  in  $M$  we know

$$f(v) \leq f(z) < f(z) + \frac{\epsilon}{\lambda}\|z - v\|,$$

while for  $z$  not in  $M$  we have

$$f(v) + \frac{\epsilon}{\lambda}\|v - x\| < f(z) + \frac{\epsilon}{\lambda}\|z - x\|.$$

Part (c) follows by the triangle inequality. Since  $v$  lies in  $M$  we have

$$f(z) + \frac{\epsilon}{\lambda}\|z - x\| \geq f(v) + \frac{\epsilon}{\lambda}\|v - x\| \quad \text{for all } z \text{ in } \mathbb{R}^n.$$

Setting  $z = x$  shows the inequalities

$$f(v) + \epsilon \geq \inf f + \epsilon \geq f(x) \geq f(v) + \frac{\epsilon}{\lambda}\|v - x\|.$$

Properties (a) and (b) follow. □

An immediate counterpart, far from easy before the advent of Ekeland's principle, is the following.

**Proposition 6.2.2** *Let  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  be differentiable and bounded below. Let  $\epsilon > 0$  be given. Then  $f$  has a  $\epsilon$ -critical point: a point  $v$  with*

$$\|\nabla f(v)\| \leq \epsilon. \tag{6.2.33}$$

**Proof.** Since  $f$  is bounded below  $f_\varepsilon = f + \varepsilon\|\cdot\|$  has bounded lower level sets and so achieves its infimum at some point  $v$ . Then we may check that

$$\|\nabla f(v)\| \leq \|\nabla f_\varepsilon(v)\| + \varepsilon = \varepsilon. \quad (6.2.34)$$

□

In many cases, this idea allows one ultimately to establish that the infimum does exist. Nonetheless, it is instructive to apply this result to a function, such as  $\exp$ , that does not achieve its infimum.

Given a set  $C \subset \mathbb{R}^n$  and a continuous *self map*  $f : C \rightarrow C$ , we ask whether  $f$  has a *fixed point*:  $f(x) = x$ . Ekeland's principle also has important, constructive applications to fixed point theory. We call  $f$  a *contraction map* if there is a real constant  $\gamma_f < 1$  such that

$$\|f(x) - f(y)\| \leq \gamma_f \|x - y\| \quad \text{for all } x, y \in C. \quad (6.2.35)$$

We are now able to painlessly establish a version of one of the most important theorems in applied analysis.

**Theorem 6.2.3 (Banach contraction)** *Any contraction on a closed subset of  $\mathbb{R}^n$  has a unique fixed point.*

**Proof.** Suppose the set  $C \subset \mathbb{R}^n$  is closed and the function  $f : C \rightarrow C$  satisfies the contraction condition (6.2.35). We apply the Ekeland variational principle (6.2.1) to the function

$$z \in \mathbb{R}^n \mapsto \begin{cases} \|z - f(z)\| & \text{if } z \in C \\ +\infty & \text{otherwise} \end{cases}$$

at an arbitrary point  $x$  in  $C$ , with the choice of constants

$$\epsilon = \|x - f(x)\| \quad \text{and} \quad \lambda = \frac{\epsilon}{1 - \gamma_f}.$$

This shows there is a point  $v$  in  $C$  satisfying

$$\|v - f(v)\| < \|z - f(z)\| + (1 - \gamma_f)\|z - v\|$$

for all points  $z \neq v$  in  $C$ . Hence  $v$  is a fixed point, since otherwise choosing  $z = f(v)$  gives a contradiction. The uniqueness is easy.  $\square$

With a little more work, we may estimate how far the fixed point is from our initial guess. Now, what if the map  $f$  is not a contraction? A very useful weakening of the notion is the idea of a *nonexpansive map*, which is to say a self map  $f$  satisfying

$$\|f(x) - f(y)\| \leq \|x - y\| \quad \text{for all } x, y.$$

A nonexpansive map on a nonempty compact set or a nonempty closed convex set may not have a fixed point, as simple examples like translations on  $\mathbb{R}$  or rotations of the unit circle show. On the other hand, a straightforward argument using the Banach contraction theorem shows this cannot happen if the set is nonempty, compact, *and* convex. Indeed, it suffices to pick some  $x_0 \in C$  and to consider the perturbed contraction mapping

$$f_\tau(x) = (1 - \tau)f(x) + \tau x_0. \quad (6.2.36)$$

This map must have a fixed point  $x_\tau$  with

$$\|f(x_\tau) - x_\tau\| \leq \tau \operatorname{diam}(C). \quad (6.2.37)$$

Now again taking limits completes the argument. In practice, this is often not a good way to find the fixed point, but it is a very good way to start.

Another example where the variational method shines is:

**Example 6.2.4** *The Rayleigh Quotient.*

Let  $A$  be a  $n \times n$  symmetric matrix. Consider the function

$$f_A(x) = \frac{\langle Ax, x \rangle}{\|x\|^2}, \quad (6.2.38)$$

defined on the open set  $x \neq 0$  in  $\mathbb{R}^n$ . Consider the minimum (or maximum) value of this function, which must exist since  $f_A$  is positively homogeneous and is continuous on the unit sphere.

Now the upshot is that, using the quotient rule, any such maximal (minimal) point,  $v$ , is an eigen-vector:  $Av = \lambda v$  for some real  $\lambda$  which is a maximal (minimal) eigenvalue. Alternatively, the reader familiar with Lagrange multipliers, can reach the same conclusion from studying  $\min\{f_A(x) : \|x\| = 1\}$ . From this the whole spectral theory of symmetric matrices can be obtained.  $\square$

### 6.3 Maximum Entropy Optimization

Maximum entropy methods are widely used in fields including crystallography, image reconstruction and the like. We consider the convex function  $p : \mathbb{R} \mapsto (-\infty, +\infty]$  given by

$$p(u) = \begin{cases} u \log u - u & \text{if } u > 0 \\ 0 & \text{if } u = 0 \\ +\infty & \text{if } u < 0 \end{cases}$$

and the associated convex function (the negative of the *Boltzmann-Shannon entropy*)  $f : \mathbb{R}^n \mapsto (-\infty, +\infty]$  by

$$f(x) = \sum_{i=1}^n p(x_i).$$

Then  $f$  is strictly convex on  $\mathbb{R}_+^n$ , with compact lower level sets. Moreover, it is differentiable on the interior of the orthant while the directional derivative  $f'(x; \hat{x} - x) = -\infty$  for any point  $x$  on the boundary of  $\mathbb{R}_+^n$ . This “barrier” property makes the entropy highly effective in many variational settings [61].

We consider a linear map  $G : \mathbb{R}^n \mapsto \mathbb{R}^m$  such that  $G\hat{x} = b$  for some  $\hat{x}$ . Then for any vector  $c$  in  $\mathbb{R}^n$ , the minimization problem

$$\inf\{f(x) + \langle c, x \rangle \mid Gx = b, x \in \mathbb{R}^n\}$$

has a unique optimal solution,  $\bar{x}$ , and all its coordinates are strictly positive. Moreover, some vector  $\lambda$  in  $\mathbb{R}^m$  satisfies

$$\nabla f(\bar{x}) = G^* \lambda - c \quad \text{that is} \quad \bar{x}_i = \exp(G^* \lambda - c)_i, \quad (6.3.39)$$

for all  $i$ . This can be achieved by solving  $G(\exp(G^*\lambda - c)) = b$  for  $\bar{\lambda}$ , using any nonlinear solver one wishes and setting  $\bar{x} = \exp(G^*\bar{\lambda}) - c$ . Here  $G^*$  is the transpose matrix.

As a striking example of the variational method, we consider the problem of determining when a square matrix  $A$  can be pre- and post-multiplied by diagonal matrices  $D_1$  and  $D_2$ , so that  $D_1AD_2$  is *doubly stochastic* (each row and column sums to one and all entries are non-negative). Such problems (called “DAD problems”) arise in actuarial science and afford a fine example of variational methods at work.

Suppose the  $k \times k$  matrix  $A$  has each entry  $a_{ij}$  nonnegative. We say  $A$  has *doubly stochastic pattern* if there is a doubly stochastic matrix with exactly the same zero entries as  $A$ . Define a set  $Z = \{(i, j) | a_{ij} > 0\}$ , and let  $\mathbb{R}^Z$  denote the set of vectors with components indexed by  $Z$ , and let  $\mathbb{R}_+^Z$  denote those vectors in  $\mathbb{R}^Z$  with all nonnegative components.

Consider the maximum entropy problem (P):

$$\begin{aligned} \inf & \sum_{(i,j) \in Z} (p(x_{ij}) - x_{ij} \log a_{ij}) \\ \text{subject to} & \sum_{i:(i,j) \in Z} x_{ij} = 1 \quad \text{for } j = 1, 2, \dots, k \\ & \sum_{j:(i,j) \in Z} x_{ij} = 1 \quad \text{for } i = 1, 2, \dots, k \\ & x \in \mathbb{R}^Z. \end{aligned}$$

The next result answers the question of when diagonalization is possible.

**Theorem 6.3.1 (DAD).** *Suppose  $A$  has doubly stochastic pattern. Then there is a point  $\hat{x}$  in the interior of  $\mathbb{R}_+^Z$  which is feasible for the problem above. Hence the problem has a unique optimal solution  $\bar{x}$ , and, for some vectors  $\lambda$  and  $\mu$  in  $\mathbb{R}^k$ ,  $\bar{x}$  satisfies*

$$\bar{x}_{ij} = a_{ij} \exp(\lambda_i + \mu_j) \quad \text{for } (i, j) \in Z.$$

*Moreover,  $A$  has doubly stochastic pattern if and only if there are diagonal matrices  $D_1$  and  $D_2$  with strictly positive diagonal entries and  $D_1AD_2$  doubly stochastic.*

Note that the best case is when all entries of  $A$  are strictly positive. It is good fun to formally use the classical method of Lagrange multipliers ( $\lambda$  and  $\mu$  above) to obtain this result, the prior discussion legitimates the process. A very satisfactory byproduct is that we have an algorithm for diagonalization, either by directly minimizing (P) or by using the dual system implicit in (6.3.39). (See Exercise 23.)



## 6.4 A Magnetic Resonance Entropy

The *Hoch and Stern information measure*, or *neg-entropy*, arises in nuclear magnetic resonance (NMR) analysis. It is defined in complex  $n$ -space by

$$H(z) = \sum_{j=1}^n h(z_j/b),$$

where  $h$  is convex and given (for scaling  $b$ ) by:

$$h(z) = |z| \log \left( |z| + \sqrt{1 + |z|^2} \right) - \sqrt{1 + |z|^2}$$

for quantum theoretic reasons.

It is easy to check by hand or computer (as was indeed the case) that

$$h^*(z) = \cosh(|z|).$$

By comparison the *Boltzmann-Shannon entropy* is

$$(z \log z - z)^* = \exp(z).$$

Efficient *dual algorithms* now may be constructed that are not at all apparent from the original formulation

Knowing “closed forms” helps:

$$(\exp \exp)^*(y) = y \log(y) - y \{W(y) + W(y)^{-1}\}$$

where *Maple* or *Mathematica* knows the complex *Lambert W* function, which is the solution of

$$W(x)e^{W(x)} = x.$$

Thus, the conjugate’s series is as well known to the computer algebra system as that for  $\exp$ :

$$-1 + (\log(y) - 1)y - \frac{1}{2}y^2 + \frac{1}{3}y^3 - \frac{3}{8}y^4 + \frac{8}{15}y^5 + O(y^6).$$

The  $W$  function arises usefully in many places. It was only recently that the function was named and then began to have a literature, largely because it had a name and existed in *Maple* and *Mathematica*. Unnamed objects are unlikely to be studied in an organized fashion.

## 6.5 Computational Complex Analysis

We would be remiss not to indicate how complex analysis comes to life in the presence of symbolic and numeric computation. We start by stating three fundamental theorems, each of which is most useful heuristically and formally.

**Theorem 6.5.1 (Argument principle)** *Suppose  $f$  is meromorphic inside a bounded region containing a simple closed curve  $C$  which contains no poles or zeros of  $f$ . Then*

$$\frac{1}{2\pi i} \int_C \frac{f'(z)}{f(z)} dz = N(C) - M(C), \quad (6.5.40)$$

where  $N(C)$  is the number of zeros and  $M(C)$  is the number of poles inside  $C$ , and the integral is taken counter-clockwise.

This is directly accessible to computation. We illustrate this at length in Exercise 2 at the end of this chapter. A shorter example is:

**Example 6.5.2** *Find the number of zeros of  $P(z) = z^3 + 3z^2 + 6z + 4$  with  $\operatorname{Re}(z) > 0$ .*

We note that  $P$  has no zeros on the positive real axis and has one or three on the negative real axis (by Descartes rule of signs). If we integrate over  $C(r) = \{z : |z| \leq r\}$ , to six places we get  $N(C(1.5)) = 1.00000$  and  $N(C(2.5)) = 3.00001$ , so we have located two complex zeros in the annulus between these two circles. Correspondingly,  $N(91/24 + C(109/24)) = 0$  (to eight digits) and thus with three circles we have answered the question, as in Figure 6.1.

In fact the zeros are at  $-1$  and  $-1 \pm \sqrt{3}i$ . □

A fairly direct consequence of the *maximum principle*, namely that the maximum of the modulus of a nonconstant analytic function occurs only on the boundary of whatever region we are considering, is:

**Theorem 6.5.3 (Rouché's theorem)** *Let  $f(z)$  and  $g(z)$  be analytic inside and on a simple closed curve  $C$ . Suppose that  $|g(z)| < |f(z)|$  on  $C$ . Then  $f(z)$  and  $f(z) - g(z)$  have the same number of zeros inside  $C$ .*

Computation can assure us that  $f$  and  $g$  do indeed satisfy both the hypotheses and the conclusions of the theorem. We leave some examples as exercises.

The third more specialized result is a very useful consequence of the Cauchy residue theorem due to Lindelöf (1905).

**Theorem 6.5.4 (Cauchy-Lindelöf)** *Let  $k(z)$  and  $r(z)$  be meromorphic in the complex plane. Suppose that  $r(z)$  is a rational function which is  $O(z^{-2})$  at infinity while  $k(z)$  is  $o(z)$  over an infinite set of circles  $|z| = R_n \rightarrow \infty$ . Then*

$$-\sum_{p \in \mathcal{P}} \operatorname{Res}(r k(s), s = p) = \sum_{q \in \mathcal{Q}} \operatorname{Res}(r k(s), s = q) \quad (6.5.41)$$

where  $\mathcal{P}$  denotes the poles of  $r$  and  $\mathcal{Q}$  denotes the poles of  $k$  that are not poles of  $r$ .

Recall that the *residue*,  $\operatorname{Res}(f(s), s = a)$ , is defined as the coefficient of  $(x - a)^{-1}$  in the Laurent series expansion of  $f$ . Again, they are most accessible to assisted computation.

Let  $k(z) = \pi / \sin(\pi z)$  and  $r(z) = 1/(z^2 + 1)$ , which clearly satisfy the hypotheses of the Cauchy-Lindelöf theorem. Then  $\mathcal{Q}$  consists of the zeros of  $\sin(\pi z)$ , thus  $\mathcal{Q} = \mathbb{Z}$ , and  $\mathcal{P}$  is  $\{\pm i\}$ . *Maple* computes the left side of (6.5.41) to be  $\pi / \sinh(\pi)$ . Correspondingly we can easily check in *Mathematica* that the terms of the right side from -6 to 6 are

$$1/37, -1/26, 1/17, -1/10, 1/5, -1/2, 1, -1/2, 1/5, -1/10, 1/17, -1/26, 1/37$$

so the general term is clearly  $\frac{(-1)^n}{n^2+1}$  and we have evaluated

$$1 + 2 \sum_{n=1}^{\infty} \frac{(-1)^n}{n^2 + 1} = \frac{\pi}{\sinh(\pi)}.$$

Similarly with  $r(z) = 1/z^2$  we reobtain

$$-\frac{1}{6}\pi^2 = 2 \sum_{n=1}^{\infty} \frac{(-1)^n}{n^2}.$$

More examples are scattered through the book.

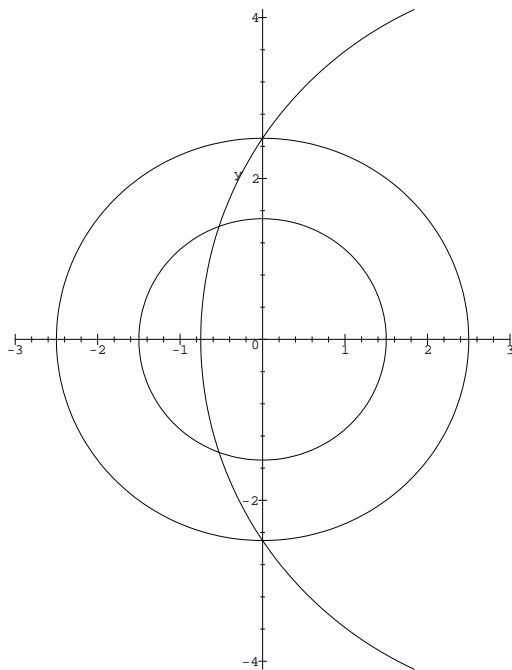


Figure 6.1: Trapping the zeros

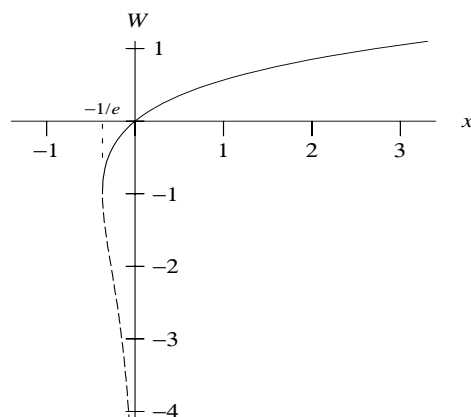


Figure 6.2: Real branches of Lambert's function that satisfy  $W \exp W = x$

It is a nice exercise now to obtain the sin product from the Cauchy-Lindelöf theorem applied to  $\cot(z) - 1/z$  and  $\frac{1}{(z-w)z}$  to deduce first that

$$w \cot(w) - 1 = 2w^2 \sum_{n=1}^{\infty} \frac{1}{n^2 \pi^2 - w^2},$$

and to find corresponding formulae for other trig functions.

## 6.6 The Lambert Function

As noted in Section 6.4, the Lambert  $W$  function satisfies

$$W(x)e^{W(x)} = x. \quad (6.6.42)$$

We give a short primer on  $W$  in Section 6.6.3, and a survey of history, properties and applications of  $W$  is to be found in [81, 83].

There is a branch point of  $W$  at  $x = -1/e$ , where  $W(x) = -1$ . See Figure 6.2, which can be produced in *Maple* by the command

```
> plot( [ t*exp(t), t, t=-5..1 ], -1..3, -4..1 );
```

The two real-valued branches of  $W$  are denoted  $W_0(x)$  and  $W_{-1}(x)$ ; we also refer to  $W_0(x)$  as the *principal branch*. To understand the function near the branch point at  $x = -1/e$ , after various experiments, we decide to compute the series of  $W_0(-e^{-1-z^2/2})$ . We obtain, very quickly, that

$$\begin{aligned} W_0(-\exp(-1 - z^2/2)) \\ &= -1 + z - \frac{1}{3}z^2 + \frac{1}{36}z^3 + \frac{1}{270}z^4 + \frac{1}{4320}z^5 - \frac{1}{17010}z^6 \\ &\quad - \frac{139}{5443200}z^7 - \frac{1}{204120}z^8 - \frac{571}{2351462400}z^9 + O(z^{10}). \end{aligned}$$

### 6.6.1 The Lambert Function and Stirling's Formula

We look up the sequence of denominators 1, 3, 36, 270, 4320,  $\dots$ , in [194]. [Sometimes denominators have nontrivial common factors with numerators. Cancellation of these common factors makes any “guessing” procedure more difficult.] Thus, reference [158] would not easily be found by a normal citation search. We find out in [158] that equation (6.6.43) gives coefficients needed in Stirling's formula for  $n!$ , which begins

$$n! \sim \sqrt{2\pi n} n^n e^{-n} \left( 1 + \frac{1}{12n} + \frac{1}{288n^2} - \frac{139}{51840n^3} - \frac{571}{2488320n^4} + O\left(\frac{1}{n^5}\right) \right).$$

The connection we discover (without doing any work ourselves) is that if

$$W_0(-e^{-1-z^2/2}) = \sum_{k \geq 0} (-1)^{k-1} a_k z^k,$$

then

$$n! \sim \sqrt{2\pi n} n^n e^{-n} \sum_{k \geq 0} \frac{1 \cdot 3 \cdot 5 \cdots (2k+1)}{n^k} a_{2k+1}.$$

Moreover, there is a lovely (and useful!) recurrence relation for the  $a_k$ 's, namely  $a_0 = 1$ ,  $a_1 = 1$ , and

$$a_n = \frac{1}{(n+1)a_1} \left( a_{n-1} - \sum_{k=2}^{n-1} k a_k a_{n+1-k} \right).$$

### 6.6.2 The Lambert Function and Riemann Surfaces

As we have seen, tools such as MATLAB and *Mathematica* permit easy accurate visualization of many objects, including Riemann surfaces [82, 204]. We now describe a simple technique for visualization of Riemann surfaces, namely to make 3-d plots of  $\operatorname{Re}f(z)$  or  $\operatorname{Im}f(z)$ .

It is necessary to *prove* something—namely, that we do get a good picture of the Riemann surface and not just a 3-d plot of an imaginary or real part. The key point is that given  $w = u + iv = f(z) = f(x + iy)$ , then we get an accurate Riemann surface by plotting, say,  $(x, y, v)$  *if and only if* the missing piece of information (here,  $u$ ) is completely determined once  $x$ ,  $y$ , and  $v$  are given.

This is simple, if not obvious: once we have a smooth three-dimensional surface, each point of which can be associated with a unique value (i.e., ordered pair) of the map  $z \mapsto w = f(z)$ , then we have a representation of the Riemann surface of  $f$ . The exact association is not automatic. For example, if  $w = \log(z)$  and we plot  $(x, y, u)$ , then we do *not* get a picture of the Riemann surface for logarithm, because the branch of  $v = \Im(w) = \arg(z)$  is not determined from  $u = \log(x^2 + y^2)/2$ ,  $x$ , and  $y$ . If we plot  $(x, y, v)$ , we *do* recover the classical picture of the Riemann surface for  $\log(z)$ , because given  $x$ ,  $y$ , and  $v$  we can easily find  $u$ . Figure 6.3 gives a static representation of the Riemann surface for the Lambert  $W$  function. Many others are graphed in [82] and [204].

**A one-to-one correspondence proof.** Given  $x$ ,  $y$ , and  $v$ , we have to solve for  $u$ . Of course, one takes the existence of  $(u, v)$  for a given  $(x, y)$ , [81]. We have

$$(u + iv)e^{u+iv} = x + iy,$$

which gives

$$ue^u + ive^u = (x + iy)e^{-iv} = (x + iy)(\cos v - i \sin v);$$

therefore,

$$ue^u + ive^u = (x \cos v + y \sin v) + i(y \cos v - x \sin v).$$

If  $v \neq 0$ , and moreover  $y \cos v - x \sin v \neq 0$ , then dividing the real part by the imaginary part gives  $u$  in terms of  $x$ ,  $y$ , and  $v$ :

$$u = \frac{v(x \cos v + y \sin v)}{y \cos v - x \sin v}.$$

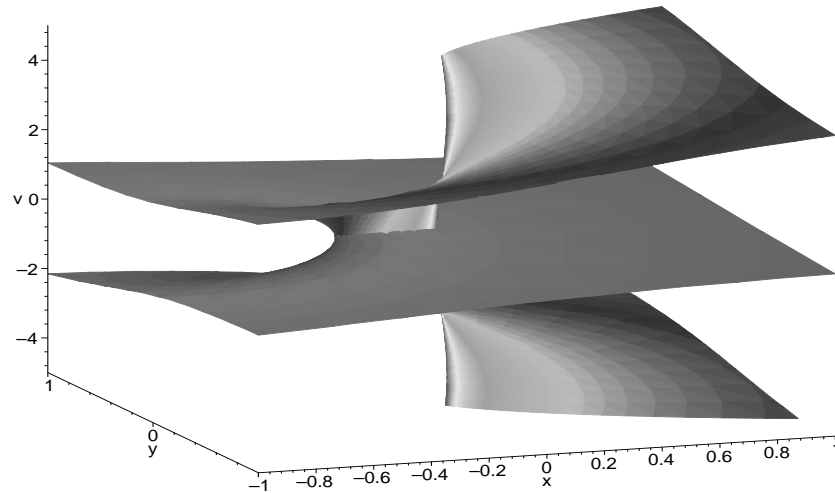


Figure 6.3: The Riemann surface for the Lambert function.

This solution is unique. Investigation of the exceptional conditions  $v = 0$  or  $y \cos v - x \sin v = 0$  leads to  $u \exp u = x$ , which has two solutions if and only if  $-1/e \leq x < 0$ , in the case  $v = 0$ , and to the singular condition  $u = -\infty$  and  $x = y = 0$ .

This is precisely what we observe in the graph: two sheets intersect only if  $-1/e \leq x < 0$  (note in color, the colors are different and hence, the corresponding sheets on the Riemann surface do not “really” intersect), and all sheets have a singularity at the origin, except the central one, which contains  $v = 0$ . This is as good a representation of the Riemann surface for the Lambert  $W$  function as can be produced in three dimensions. However, Figure 6.3 is nowhere near as intelligible as a live plot. On a personal computer, the use of `OpenGL` by *Maple* allows the plot to be rotated by direct mouse control. This helps give a good sense of what the surface is really like, in three dimensions.



### 6.6.3 The Lambert Function in Brief

If you have used *Maple* or *Mathematica* to solve transcendental equations, you may already have encountered the Lambert  $W$  function, defined by (6.6.42). The history and some of the properties of this remarkable function are described in [81]. The function provides a beautiful new look at much of undergraduate mathematics, and much of intrinsic interest. Here are some of the elementary properties of  $W$ .

1. On  $0 \leq x < \infty$  there is one real-valued branch  $W(x) \geq 0$  (see Figure 6.2). On  $-1/e < x < 0$  there are two real-valued branches. We call the branch that has  $W(0) = 0$  the principal branch. On this branch, it is easy to see that  $W(e) = W(1 \cdot e^1) = 1$ .
2. The derivative of  $W$  can be found by implicit differentiation to be

$$\begin{aligned} \frac{d}{dx}W(x) &= \frac{1}{(1+W(x))e^{W(x)}} \\ &= \frac{W(x)}{(1+W(x))x}, \end{aligned}$$

where the second formula follows on using  $\exp W(x) = x/W(x)$ , and holds if  $x \neq 0$ . We may use the first formula to find the value of the derivative at  $x = 0$ , and we see the singularity is just a removable one.

3. The function  $y = W(\exp z)$  satisfies

$$y + \log y = z.$$

This function appears, for example, in convex optimization. Consider the convex conjugate,  $f^*(s) = \sup_r rs - f(r)$ , of the function  $f(r) = r \log(r/(1-r)) - r$ . Show that  $f^*(s) = W(\exp s)$ .

4.  $W(x)$  has a Taylor series about  $x = 0$  with rational coefficients. Similarly,  $W(\exp z)$  has a Taylor series with rational coefficients about  $z = 1$ . The

first few terms are

$$\begin{aligned} W(e^z) = & 1 + \frac{1}{2}(z-1) + \frac{1}{16}(z-1)^2 - \frac{1}{192}(z-1)^3 - \frac{1}{3072}(z-1)^4 \\ & + \frac{13}{61440}(z-1)^5 - \frac{47}{1474560}(z-1)^6 - \frac{73}{41287680}(z-1)^7 \\ & + \frac{2447}{1321205760}(z-1)^8 - \frac{16811}{47563407360}(z-1)^9 \\ & - \frac{15551}{1902536294400}(z-1)^{10} + O((z-1)^{11}). \end{aligned}$$

5. There is an exact formula for the coefficients of the  $n$ th derivative of  $W(\exp z)$ , in terms of second-order Eulerian numbers  $\langle\langle n \rangle\rangle_k$  [126]. This formula comes from the following expression for the  $n$ th derivative of  $W(\exp z)$ , which is stated in [81]. Once the answer is known, the proof is an easy induction, which we leave for the reader.
6. The derivatives of  $W(\exp z)$  are

$$\frac{d^n}{dz^n} W(e^z) = \frac{q_n(W(e^z))}{(1+W(e^z))^{2n-1}}, \quad (6.6.43)$$

where  $q_n(w)$  is a polynomial of degree  $n$  satisfying the recurrence relation

$$q_{n+1}(w) = -(2n-1)wq_n(w) + (w+w^2)q'_n(w), \quad n > 1 \quad (6.6.44)$$

and having the explicit expression

$$q_n(w) = \sum_{k=0}^{n-1} \langle\langle n-1 \rangle\rangle_k (-1)^k w^{k+1}. \quad (6.6.45)$$

If  $n = 1$  we have  $q_1(w) = w$ , and it is convenient to put  $q_0(w) = w/(1+w)$ ; this isn't a polynomial, but it makes things work out right. This means the series for  $W(\exp z)$  about  $z = 1$  is just

$$W(e^z) = \sum_{n \geq 0} \frac{q_n(1)}{n! 2^{2n-1}} (z-1)^n. \quad (6.6.46)$$

## 6.7 Commentary and Additional Examples

1. **The Selberg integral.** The Selberg integral (see [11]) is an  $N$ -dimensional extension of Euler's beta integral

$$S_N(\lambda_1, \lambda_2, \lambda) = \left( \prod_{l=1}^N \int_0^1 dt_l t_l^{\lambda_1} (1-t_l)^{\lambda_2} \right) \prod_{1 \leq j < k \leq N} |t_k - t_j|^{2\lambda}$$

(the beta integral is the case  $N = 1$ ). Selberg evaluated this integral as a product of gamma functions:

$$S_N(\lambda_1, \lambda_2, \lambda) = \prod_{j=0}^{N-1} \frac{\Gamma(\lambda_1 + 1 + j\lambda)\Gamma(\lambda_2 + 1 + j\lambda)\Gamma(1 + (j+1)\lambda)}{\Gamma(\lambda_1 + \lambda_2 + 2 + (N+j-1)\lambda)\Gamma(1 + \lambda)}.$$

For example,

$$\begin{aligned} & \int_0^1 \int_0^1 x^a (1-x)^b y^a (1-y)^b (|x-y|)^{2c} dx dy \quad (6.7.47) \\ &= \frac{\Gamma(a+1)\Gamma(b+1)\Gamma(a+1+c)\Gamma(b+1+c)\Gamma(1+2c)}{\Gamma(a+b+2+c)\Gamma(a+b+2+2c)\Gamma(c+1)}. \end{aligned}$$

The Selberg integral gives rise to many variations and to interesting tests of multi-dimensional integration routines. A beautiful and illustrative application is to evaluate Gaussian ensembles arising in statistical mechanics such as

$$\int_0^\infty \cdots \int_0^\infty \exp\left(-\frac{1}{2} \sum_{i=1}^n x_i^2\right) \prod_{1 \leq i < j \leq n} |x_i - x_j|^{2\gamma} dx = \prod_{j=1}^n \frac{\Gamma(\gamma j + 1)}{\Gamma(\gamma + 1)}.$$

2. **Establishing inequalities numerically via the argument principle.** This refers to Theorem 6.5.1. We consider the *logarithmic* and  $\frac{2}{3}$ -*power* means,

$$\mathcal{L}(x, y) = \frac{x-y}{\log(x) - \log(y)}, \quad \mathcal{M}(x, y) = \sqrt[3]{\frac{x^{\frac{2}{3}} + y^{\frac{2}{3}}}{2}}.$$

A delicate estimate of an *elliptic integral* was reduced to establishing the elementary inequalities:

$$\mathcal{L}^{-1}(\mathcal{M}(x, 1), \sqrt{x}) < \mathcal{L}^{-1}(x, 1) < \mathcal{L}^{-1}(\mathcal{M}(x, 1), 1) \text{ for } 0 < x < 1.$$

- (a) To prove the right-hand inequality, we may draw on *graphic/symbolic* assistance to establish:

$$\mathcal{F}(x) = \mathcal{L}^{-1}(\mathcal{M}(x, 1), 1) - \mathcal{L}^{-1}(x, 1) > 0 \text{ for } 0 < x < 1.$$

Using any plotting routine, we will see that  $\mathcal{M}$  is a mean (or in other words,  $\min(a, b) \leq \mathcal{M}(a, b) \leq \max(a, b)$ ): as  $x < \mathcal{M}(x, 1) < 1$  illustrates. Also we observe that  $\mathcal{L}^{-1}$  is decreasing. These two graphical hints lead us directly to a proof of the right-hand inequality.

- (b) The left-hand inequality is equivalent to the following:

$$\mathcal{E}(x) = \mathcal{L}^{-1}(x, 1) - \mathcal{L}^{-1}(\mathcal{M}(x, 1), \sqrt{x}) > 0 \text{ for } 0 < x < 1.$$

This can be accomplished with a mix of *numeric/symbolic methods*:

- i. establishing that  $\lim_{x \rightarrow 0^+} \mathcal{E}(x) = \infty$ ;
- ii. using a Newton-like iteration to show that  $\mathcal{E}(x) > 0$  on  $[0.0, 0.9]$ ;
- iii. using a Taylor series expansion to show  $\mathcal{E}(x)$  has 4 zeroes at 1; and then
- iv. using the Argument Principle to establish there are no more zeros inside  $C = \{z : |z - 1| = \frac{1}{4}\}$ :

$$\frac{1}{2\pi i} \int_C \frac{\mathcal{E}'}{\mathcal{E}} = \#(\mathcal{E}^{-1}(0); C)$$

(the number of zeros inside  $C$ ).

These steps can each be made effective, and so constitute a proof, the only one so far known. It is the last step (iv) that requires care (to use a numerical quadrature rule in which one can ensure the integral is actually to one significant place as claimed. Just because we can compute it correctly to twenty places does not ensure that.)

### 3. Some limits of integrals. Evaluate:

- (a)

$$\lim_{x \rightarrow \infty} \frac{x^4}{e^{3x}} \int_0^x \int_0^{x-u} e^{u^3+v^3} dv du \quad \left( = \frac{2}{9} \right).$$

(b)

$$\lim_{n \rightarrow \infty} \frac{1}{n^4} \prod_{i=1}^{2n} (n^2 + i^2)^{1/n} \quad (= 25 \exp(2 \arctan(2)) - 4).$$

4. **Carleman's inequality.** Determine

$$\sup_S \frac{\sum_{n=1}^{\infty} (x_1 x_2 \cdots x_n)^{1/n}}{\sum_{n=1}^{\infty} x_n} \quad (= e),$$

where  $S$  denotes all non-negative non-zero sequences. Hint: consider sequences of the form  $n^n / (n+1)^{(n-1)}$ . The recent survey [105] contains a fine selection of proofs and of extensions of Carleman's inequality.

5. **The origin of the elliptic integrals.** The elliptic integrals take their name from the fact that  $E$  provides the arclength of an ellipse.(a) Show that the arclength,  $\mathcal{L}$ , of an ellipse with major semi-axis  $a$  and minor semi-axis  $b$  is

$$\mathcal{L} = 4a E' \left( \frac{b}{a} \right) = \int_{-\pi}^{\pi} \sqrt{a^2 \cos^2(\theta) + b^2 \sin^2(\theta)} d\theta.$$

(b) The period,  $p$ , of a pendulum with amplitude  $\alpha$  and length  $L$  is

$$p = 4 \sqrt{\frac{L}{g}} K \left( \sin \left( \frac{\alpha}{2} \right) \right),$$

where  $g$  is the gravitational constant. Deduce that for small amplitude  $p \approx 2\pi \sqrt{L/g}$ , as is the case of simple harmonic oscillation.

(c) Use the AGM iteration

$$\frac{\pi/2}{M(a, b)} = \int_0^{\pi/2} \frac{dt}{\sqrt{a^2 \cos^2(t) + b^2 \sin^2(t)}}. \quad (6.7.48)$$

to write a fast algorithm to compute  $p$  to arbitrary precision. Combine this with Legendre's identity

$$E(k) K(k') + K(k) E'(k) - K(k) K'(k) = \frac{\pi}{2}.$$

to compute the arclength of an ellipse.

Recall that a function is elementary if it can be realized in a finite number of steps from compositions of algebraic, exponential and trigonometric functions and their inverses. Liouville in 1835 was able to show that  $E$  and  $K$  are non-elementary transcendental functions. By contrast, the formula for the area of an ellipse was known to Archimedes.

6. **Euler's integral for hypergeometric functions.** Establish Euler's integral for the hypergeometric function

$$F(a, b, c; x) = \frac{\Gamma(c)\Gamma(b)}{\Gamma(c-b)} \int_0^1 t^{b-1} (1-t)^{c-b-1} (1-tx)^{-a} dt \quad (6.7.49)$$

for  $\operatorname{Re}(c) > \operatorname{Re}(b) > 0$ . Appropriately interpreted the right side of (6.7.49) is valid outside of  $[1, \infty)$  and so provides an analytic continuation of  $F$ . Hint: Expand  $(1-tx)^a$  by the binomial theorem and observe that term-by-term one has a  $\beta$  integral which as we saw can be written in terms of gamma functions.

7. **A hypergeometric function identity.** Show that

$$\frac{1}{x} F(1-y, 1, 1+x; -1) + \frac{1}{y} F(1-x, 1, 1+y; -1) = 2^{x+y-1} \beta(x, y).$$

Hint: Use (6.7.49).

8. **Multiple hypergeometric functions.** There are many extensions to hypergeometric functions in several complex variables. We list one as an example

$$F(\alpha, \beta, \beta', \gamma; x, y) = \sum_{m, n \geq 0} \frac{(\alpha)_{m+n} (\beta)_m (\beta')_n}{(\gamma)_{m+n} m! n!} x^m y^n.$$

This function (and three similar ones) satisfies a pair of second order partial differential equations. For various choices of parameters it reduces to the classical hypergeometric function:

$$F(\alpha, \beta, \beta', \gamma; x, x) = F(\alpha, \beta + \beta', \gamma; x).$$

It also has a double integral of Euler type

$$F(\alpha, \beta, \beta', \gamma; x, y) = \frac{\Gamma(\gamma)}{\Gamma(\beta)\Gamma(\beta')\Gamma(\gamma - \beta - \beta')} \quad (6.7.50)$$

$$\times \int_{\Omega} u^{\beta-1} v^{\beta'-1} (1-u-v)^{\gamma-\beta-\beta'-1} (1-ux-vy)^{-\alpha} dudv,$$

where  $\Omega$  is the positive quadrant of the unit circle,  $\{(x, y) : x \geq 0, y \geq 0, x^2 + y^2 \leq 1\}$ , and  $\operatorname{Re} \beta > 0, \operatorname{Re} \beta' > 0, \operatorname{Re} \gamma - \beta - \beta' > 0$ . Attempt to numerically validate (6.7.50) and to prove the identity.

9. **Another hypergeometric function evaluation.** Following Gauss, evaluate

$$F(a, b, c; 1) = \frac{\Gamma(c)\Gamma(c-a-b)}{\Gamma(c-a)\Gamma(c-b)} \quad (6.7.51)$$

and determine the region of validity.

10. **Berkeley problem 2.2.13.** Find the singular matrix  $S$  nearest in Euclidean norm to  $\begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix}$ . Hint: Directly solve the implied optimization

problem using a computer algebra program. Answer:  $\mathcal{S} = \begin{bmatrix} 0 & 0 \\ 0 & 2 \end{bmatrix}$ .

11. **Kirchhoff's law.** Consider a finite, undirected, connected graph with vertex set  $V$  and edge set  $E$ . Suppose that  $\alpha$  and  $\beta$  in  $V$  are distinct vertices and that each edge  $ij$  in  $E$  has an associated "resistance"  $r_{ij} > 0$  in  $\mathbb{R}$ . We consider the effect of applying a unit "potential difference" between the vertices  $\alpha$  and  $\beta$ . Let  $V_0 = V \setminus \{\alpha, \beta\}$ , and for "potentials"  $x$  in  $\mathbb{R}^{V_0}$  we define the "power"  $p : \mathbb{R}^{V_0} \rightarrow \mathbb{R}$  by

$$p(x) = \sum_{ij \in E} \frac{(x_i - x_j)^2}{2r_{ij}},$$

where we set  $x_\alpha = 0$  and  $x_\beta = 1$ .

- (a) Prove the power function  $p$  has compact level sets.

- (b) Deduce the existence of a solution to the following equations (describing “conservation of current”):

$$\begin{aligned} \sum_{j : ij \in E} \frac{x_i - x_j}{r_{ij}} &= 0 \text{ for } i \text{ in } V_0 \\ x_\alpha &= 0 \\ x_\beta &= 1. \end{aligned}$$

- (c) Prove the power function  $p$  is strictly convex.
- (d) Deduce that the conservation of current equations in part (b) have a unique solution.
12. **Berkeley problem 7.1.14.** Show that the functions  $t \mapsto \exp(\alpha_k t)$  are linearly independent over  $\mathbb{R}$  when the corresponding real numbers  $\alpha_1, \alpha_2, \dots, \alpha_n$  are distinct. Hint: Use *Maple* or *Mathematica* to observe the relative growth of the exponential values, given some sample values of  $\alpha_i$ .

13. **A central binomial series.** Show that

$$\arcsin^2(x) = \frac{1}{2} \sum_{m=1}^{\infty} \frac{(2x)^{2m}}{m \binom{2m}{m}}$$

by showing that both sides satisfy the same differential equation.

14. **Branches of arcsin.** Consider

$$u(r) = \arcsin \left( 4 \sqrt{\sin(r)} \frac{1 - \sin(r)}{1 + \sin^2(r)} \right).$$

Determine the behavior of  $u$  and  $u^{(2)}$  on  $[0, \pi/2]$ , and do the same for  $v(r) = u(2r)$  on  $[0, \pi/4]$ . Hint: Start by plotting  $u$  and  $u^{(2)}$  and the identity.

15. **Integrating  $\Gamma^{-1}$ .**

- (a) Evaluate

$$\int_0^\infty \frac{1}{x [\pi^2 + \log^2(x)]} dx.$$



(b) Show

$$\int_0^\infty \frac{1}{\Gamma(x+1)} dx = e - \frac{1}{\pi} \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} e^{-e^\pi \tan(\theta)} d\theta,$$

and

$$\int_0^\infty \frac{1}{\Gamma(x)} dx = e + \frac{1}{\pi} \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} e^{\pi \tan(\theta)} e^{-e^\pi \tan(\theta)} d\theta.$$

(c) For integer  $n \geq 0$ , show

$$\begin{aligned} \int_{-n}^{-n+1} \frac{1}{\Gamma(x)} dx &= (-1)^n \int_0^\infty e^{-x} \frac{(x+1)x^{n-1}}{\pi^2 + \log^2(x)} dx \\ &= \frac{(-1)^n}{\pi} \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} (e^{\pi \tan(\theta)} + 1) \frac{e^{n\pi \tan(\theta)}}{e^{e^\pi \tan(\theta)}} d\theta \end{aligned}$$

and

$$\int_{-n}^0 \frac{1}{\Gamma(x)} dx = \int_0^\infty e^{-x} \frac{(-x)^n - 1}{\pi^2 + \log^2(x)} dx.$$

Hence estimate the size of  $\int_{-n}^0 \Gamma(x)^{-1} dx$  asymptotically. In each case the rightmost expression is much easier to compute.

One way to do this is to use *Laplace's method* which is suited to asymptotics of integrals of the form  $\mathcal{I}(x) = \int_a^b e^{-xp(t)} q(t) dt$ , for appropriate positive  $p(t)$  and  $q(t)$  for which the bulk of the value of  $(x)$  is determined by the minimum value  $p(c)$ , and one estimates that  $\mathcal{I}(x) \approx \int_a^b e^{-x(p(c)+(t-c)p'(c))} q(c) dt$  which can be integrated explicitly [170]. Note that it is easy to see that the value is “just less” than  $n!$ . References: [128, 153, 114].

16. **Hankel loop integral for  $\Gamma$ .** Show that for any complex  $z$ ,

$$\Gamma(z) = \int_{-\infty}^{(0+)} e^t t^{-z} dt, \quad (6.7.52)$$

where the notation denotes the integral over a counterclockwise contour that starts at  $t = -\infty$ , circles around  $t = 0$  once and returns to  $t = -\infty$ . (Here  $t^{-z}$  is taken to be the principal value where the contour crosses the

positive real axis and to be continuous elsewhere.) Use (6.7.52) to show that

$$\int_0^\infty \frac{1}{\Gamma(x)} dx = e - \int_0^\infty \frac{e^{-t}}{\pi^2 + \log^2(t)} dt.$$

Compare part (b) of the previous exercise.

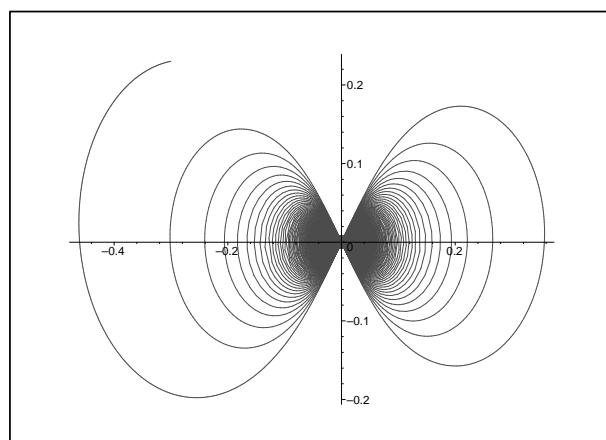
17. **A “gfun” proof for  $AG_2$ .** It is possible to entirely implement the evaluation of (6.1.12) and (6.1.16). We sketch Bruno Salvy’s derivation within Murray, Salvy and Zimmermann’s package “gfun” <http://algo.inria.fr/libraries/software.html>.

- (a) From (6.1.9), obtain the first dozen or so terms of the power series for  $A_2(x^2)$ .
- (b) The “gfun” program will guess a differential equation for  $A_2(x^2)$ .
- (c) It will then also provide a recursion for said differential equation.
- (d) This recursion is solvable in *Maple* to produce

$$u_n = \frac{\Gamma^2\left(n + \frac{1}{2}\right)^2}{\Gamma^2(n+1) \pi}$$

and summing  $\sum_{n \geq 0} u_n x^n$  produces the desired hypergeometric evaluation.

- (e) Provide the details of a similar albeit a little more elaborate computer proof of hypergeometric evaluation for  $AG_3$  of Section 6.1.3.
  - (f) Attempt to do likewise for  $N = 4$  and  $N = 7$ .
18. **Some complex plots from Guenard and Lemberg.** Plot the following functions  $\phi_k : (0, 1] \mapsto \mathbb{C}$  for  $k = 1, 2, \dots, 6$  in random order on an interval of the origin and determine from the plots which function came from which

Figure 6.4: Which function  $\phi_k$  is this?

plot [127].

$$\begin{aligned}\phi_1(t) &= t \sin(t^{-2}) e^{i \cos(t^4)/\sqrt{t}} \\ \phi_2(t) &= t \sin(t^{-1}) e^{i \cos(t^{-1})} \\ \phi_3(t) &= t \sin(t^{-2}) e^{i \cos(t^{-2})} \\ \phi_4(t) &= t \sin(t^{-1}) e^{i \cos^2(t^{-1})} \\ \phi_5(t) &= t \sin(t^{-1}) e^{i \sqrt{|\cos(t^{-1})|}} \\ \phi_6(t) &= \sqrt[3]{t} \sin^2(t^{-1/2}) e^{i \sin(\frac{1}{\sqrt{t}})}.\end{aligned}$$

One of the functions is plotted in Figure 6.4. Answer:  $\phi_3$ .

19. **Some ODE plots from Guenard and Lemberg.** Plot the families of solutions, in random order, to the following differential equations  $\mathcal{E}_k$  for  $k = 1, 2, \dots, 5$  and determine from the plots which differential equation generated which plot. One of the solutions is plotted in Figure 6.5 [127].

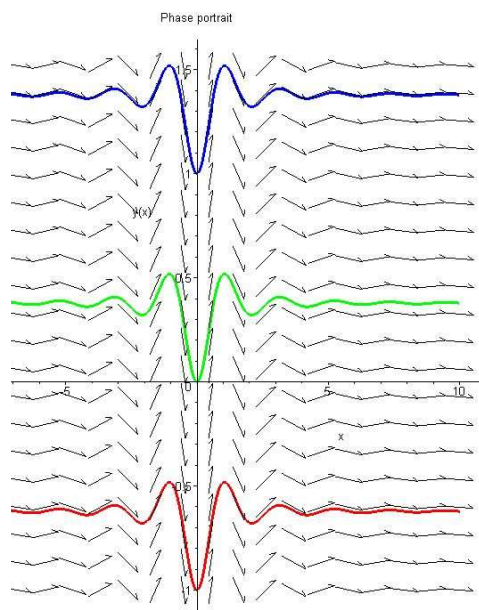


Figure 6.5: Which equation  $\mathcal{E}_k$  generated this?

$$\begin{aligned} \mathcal{E}_1 : \quad y' &= x \sin(xy) \\ \mathcal{E}_2 : \quad y' &= \sin(x) \sin(y) \\ \mathcal{E}_3 : \quad y' &= x \sin(y) \\ \mathcal{E}_4 : \quad y' &= y \sin(x) \\ \mathcal{E}_5 : \quad y' &= \frac{\sin(3x)}{1+x^2}. \end{aligned}$$

Answer:  $\mathcal{E}_3$ .

## 20. The Maclaurin series for W.

(a) Show that the Lambert W function satisfies

$$W(z) = \sum_{k=1}^{\infty} \frac{(-k)^{k-1}}{k!} z^k$$

and determine the radius of convergence.

(b) Show that

$$\sum_{k=1}^{\infty} \frac{k^k}{k!} x^k = \frac{1}{2},$$

is solved by  $x^* = 0.238843770191263\dots$

(c) Determine the analytic form of this number. This evaluation was discovered by Harold Boas [28], using the Inverse Symbolic Calculator, while studying multidimensional versions of Harald Bohr's 1914 result that if  $|\sum_{k=0}^{\infty} c_k z^k| < 1$  for  $|z| < 1$  then  $\sum_{k=0}^{\infty} |c_k z^k| < 1$  for  $|z| < 1/3$  (the best possible radius).

21. **Berkeley problem 4.3.7.** Does there exist a continuous solution on  $[0, 1]$  to  $\mathcal{T}(f) = f$ , where

$$\mathcal{T}(g)(x) = \sin(x) + \int_0^1 \frac{g(y)}{e^{x+y+1}} dy?$$

Answer: Yes. Hint:  $\mathcal{T}$  is a Banach contraction with constant at most  $\exp(-1) - \exp(-2)$ . Try plotting  $\mathcal{T}^n(0)$  for  $n \leq 3$ .

22. **Torczon's Multidirectional Search.** As computing paradigms change so do algorithms of choice—for good reasons such as increased speed, storage and parallelism and for less good reasons of vogue. Multidirectional search techniques were popular in the early days of modern computing because of their low overhead but their relative lack of accuracy and speed drove them out of fashion. The need to handle extremely large problems, in which “noisy” gradients are either unavailable or very costly, coupled with the shift to parallel computation has reawakened interest [101].

(a) Standard methods in numerical optimization, such as the general classes of line search and trust region methods [168], use gradient calculations to determine *descent directions*, or directions in which the objective function decreases. A general class of methods that has (re)gained popularity recently is *direct search* or *pattern search* methods, which use only function evaluations to find stationary points.

The basic direct search method is the simplex method of Nelder and Mead [166], first published in 1965; this method starts with a *simplex*

(a set of  $n + 1$  non-degenerate points in  $R^n$ ), and at each iteration replaces the simplex point with the largest objective function value with a new point with a smaller objective function value by reflecting across the centroid of the other simplex points, or contracting the entire simplex if such a point is not found.

Renewed interest in direct search methods began with the development in 1989 of the *multidirectional search* (MDS) algorithm of Virginia Torczon [202, 203] in which the update and contraction rules of Nelder/Mead are combined with additional rules to maintain the angles between the simplex points. The key achievement of the MDS algorithm over Nelder/Mead is a new set of convergence results; counter-examples exist where the Nelder/Mead algorithm converges to a non-stationary point [161]. Torczon generalized the MDS algorithm to a class of *pattern search* algorithms, where the set of grid-points are expanded from a simplex to any set containing a *positive basis*, provided the new iterate, chosen to be the “center” of the grid, and the amount that the grid is expanded/contracted, both satisfy very general properties that are very easily verified.

Direct search algorithms benefit from being easy to implement and applicable to a wide range of problems for which gradient-based methods are inappropriate. In addition, direct search methods are also easy to demonstrate in practice, and provide an example of where visualization tools greatly aid the research and learning process.

- (b) **Match contour plots to iterations.** Figure 6.6 shows contour plots of three sets of three iterates (circles) of the Nelder/Mead, multidirectional search, and pattern search methods, each applied to the *Rosenbrock function*:

$$f_K(x, y) = K(y - x^2)^2 + (1 - x)^2$$

where  $K = 100$  is most frequently used ( $K = 5$  is used in the illustrations). Using the general descriptions of each of these algorithms, match the algorithm to the set of iterations. When executed in parallel it is instructive to color code the computations performed by different processors. Rosenbrock’s function causes pain for many minimization methods because of the shape of its contours.

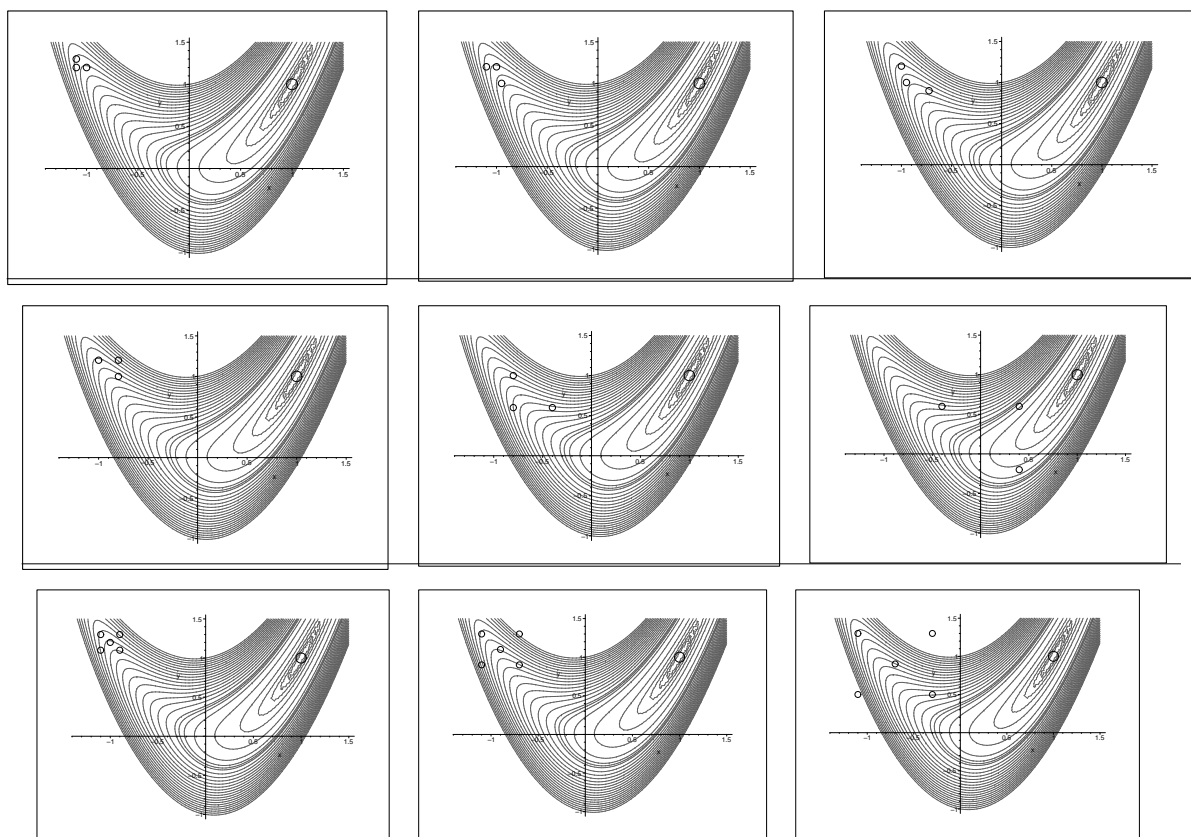


Figure 6.6: Comparison of three direct search methods

23. **Computing with DAD problems.** Find diagonal matrices  $D_1$  and  $D_2$  with strictly positive diagonal entries such that  $D_1AD_2$  is doubly stochastic, where

$$A = \begin{pmatrix} 2 & 10 & 6 \\ 1 & 0 & 4 \\ 7 & 15 & 1 \end{pmatrix}.$$

Hint: Observe that such matrices  $D_1$  and  $D_2$  must exist since  $A$  has doubly stochastic pattern; for example, consider

$$\begin{pmatrix} 1/3 & 1/3 & 1/3 \\ 1/2 & 0 & 1/2 \\ 1/6 & 2/3 & 1/6 \end{pmatrix}.$$

- (a) Construct the desired matrices by solving the minimization problem from Theorem 6.3.1.
- (b) Alternatively, solve this minimization problem using one of the direct search methods outlined in Exercise 22. For example, the following *Maple* code applies Torczon's pattern search algorithm to this specific matrix, by considering convex combinations of the  $3 \times 3$  permutation matrices with the same zero entry pattern as  $A$ .

```
p:=proc(u)
RETURN(piecewise(u>0,u*log(u)-u,u=0,0,u<0,infinity));end:

a:=proc(X,i,j)
RETURN(piecewise(A[i,j]>0,p(X[i,j])-X[i,j]*log(A[i,j]),0)); end:

obj:=proc(X,n) local i,j;
RETURN(evalf(sum(sum(a(X,i,j),i=1..n),j=1..n))); end:

define_matrix:=proc(a) local X,an,i;
X[1]:=matrix(3,3,[1,0,0,0,0,1,0,1,0]);
X[2]:=matrix(3,3,[0,1,0,1,0,0,0,0,1]);
X[3]:=matrix(3,3,[0,1,0,0,0,1,1,0,0]);
X[4]:=matrix(3,3,[0,0,1,1,0,0,0,1,0]); an:=sum(a[i],i=1..4);
RETURN(evalm(sum(a[i]/an*X[i],i=1..4))); end:

inds=[[1,0,0,0],[0,1,0,0],[0,0,1,0],[0,0,0,1],[0,0,0,0],
```



```

      [-1,0,0,0],[0,-1,0,0],[0,0,-1,0],[0,0,0,-1]];
numinds:=17; Digits:=50; numiters:=1000;
objvals:=[infinity\ $numiters]; stepsize:=1.1; decfactor:=0.5;
incfactor:=4.0; minval:=infinity; oldminval:=infinity;
curriter:=[1,1,1,1]; for iter from 1 t to numiters do
  currind:=0;
  oldminval:=minval;
  for i from 1 to numinds do
    X1[i]:=define_matrix(evalm(stepsize*inds[i]+curriter));
    objvals[i]:=obj(X1[i],3);
    if objvals[i] < minval then
      minval:=objvals[i];
      currind:=i;
    fi;
  od;
  if currind > 0 then
    curriter:=stepsize*inds[currind]+curriter;
  fi;
  if minval = oldminval then
    stepsize:=decfactor*stepsize;
  else
    stepsize:=incfactor*stepsize;
  fi;
od:

```

The solution is:

$$X = \begin{pmatrix} 1/5 & 1/2 & 3/10 \\ 1/3 & 0 & 2/3 \\ 7/15 & 1/2 & 1/30 \end{pmatrix}, \quad D_1 = \begin{pmatrix} 1/10 & 0 & 0 \\ 0 & 1/3 & 0 \\ 0 & 0 & 1/15 \end{pmatrix},$$

$$D_2 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1/2 & 0 \\ 0 & 0 & 1/2 \end{pmatrix}.$$

24. **Brouwer, eponymy and the fundamental theorem.** The Dutch mathematician and philosopher Luitzen Egbertus Jan Brouwer (1881–1966) is equally well known as one of the builders of modern topology and as the father of *intuitionism* (“Mathematics is nothing more, nothing

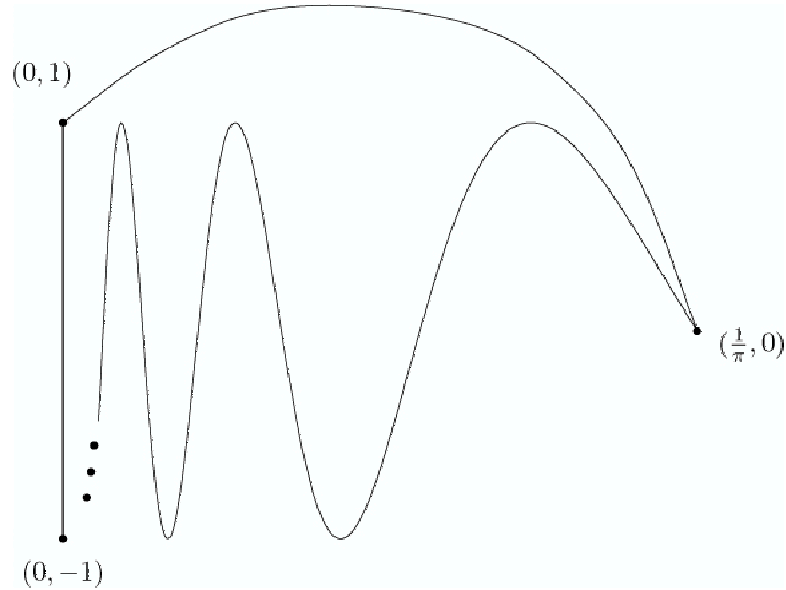


Figure 6.7: A nonconvex set with the fixed point property

less, than the exact part of our thinking.” . . . “The construction itself is an art, its application to the world an evil parasite.”). Intuitionism rejects the principle of the excluded middle (“A or not A”), unless one can effectively determine which case happens, and looks for fully analyzed proofs broken into “intuitively” compelling steps. It shares certain concerns with Bishop’s version of constructivism.

Several observations are worth making:

(i) Brouwer’s fixed point theorem (1912), that *every continuous self-map of a closed convex bounded set in  $\mathbb{R}^n$  has a fixed point*, was certainly prefigured by others such as Hadamard and Poincaré thereby illustrating Stigler’s principle of eponymy.

**The  $\sin(1/x)$  circle.** One connects the segments in Figure 6.7 to the  $\sin(1/x)$  curve. This produces a connected, but not simply, highly non-convex set,  $\gamma$ , such that every continuous self-map  $g$  on  $\gamma$  has a fixed point. Try to prove this by considering what happens if  $g(0) \neq 0$ .

(ii) The philosopher Brouwer soon disavowed much of the mathematical

tool-set used in his and other early proofs of his eponymous theorem.

(iii) In modern effective incarnations due to Scarf (1973) and others, the theorem is used computationally in optimization and mathematical economics. Many wonderful results such as the von Neumann minimax theorem, the Nash equilibrium theorem and Arrow-Debreu's proof of the existence of market equilibria used Brouwer's theorem, or its extensions, in their first proofs.

(iv) There have been several false proofs—some by distinguished mathematicians—of the Fundamental Theorem of Algebra based on Brouwer's theorem. In one case the result was applied to a discontinuous self-map! A correct proof, based on the existence of continuous  $n$ -th roots for continuous maps from the disc to the sphere, appears in [113] but citations of false ones continue.

(v) In this light, in her Simon Fraser Masters thesis, Tara Stuckless recently wrote:

In a 1949 paper [12], B. H. Arnold wrote "... it has been known for some time that the fundamental theorem of algebra could be derived from Brouwer's fixed point theorem.". He continued to give a simple one page proof that a polynomial with degree  $n \geq 1$  has at least one complex root. Unfortunately, he did this by applying Brouwer's theorem to a discontinuous function. This error was spotted and a correction was published less than two years later [13]. In 1951, M. K. Fort followed up with a brief proof, using the existence of continuous  $n$ th roots of a continuous non zero function on a disk in the complex plane to show that Brouwer can be used to prove the fundamental theorem [113].

In this author's opinion, except for an ill-fated choice of titles, this would have probably been the end of the incorrect proof by Arnold. Arnold's paper was boldly called "A Topological Proof of the Fundamental Theorem of Algebra", whereas Fort's paper was modestly titled "Some Properties of Continuous Functions". Since 1949, citations of Arnold's paper have popped up from time to time as proof of the existence of a topological proof of the fundamental theorem of algebra, including in a four hun-

dred page volume on fixed point theory published in 1981, a talk given in 2000 at a meeting of the Association for Symbolic Logic, as well as in newsgroup discussions that are at least as recent as 1998. On the other hand, this author had a difficult time tracking down Fort's paper, even knowing before hand that it did exist. No amount of searching electronic databases with relevant key words would produce it, though Arnold's paper was invariably returned. Perhaps it would have been useful for Fort to rename his work "A Correct Topological Proof of the Fundamental Theorem of Algebra". It seems certain that Arnold would have preferred this to having his blunder quoted so long after he made his apologies.

# Chapter 7

## Numerical Techniques II

Another thing I must point out is that you cannot prove a vague theory wrong . . . Also, if the process of computing the consequences is indefinite, then with a little skill any experimental result can be made to look like the expected consequences.

Richard Feynman, 1964 [199]

In this chapter, we will examine in more detail some additional underlying computational techniques that are useful in experimental mathematics. In particular, we shall briefly examine techniques for theorem proving, prime number computations, polynomial root finding, numerical quadrature, and infinite series summation. As in the first volume, we focus here on *practical* algorithms and techniques. In some cases, there are known techniques that have superior efficiencies or other characteristics, but for various reasons are not considered suitable for practical implementation. We acknowledge the existence of such algorithms but do not, in most cases, devote space to them.

### 7.1 The Wilf-Zeilberger Algorithm

One fascinating non-numerical algorithm is the Wilf-Zeilberger (WZ) algorithm, which employs “creative telescoping” to show that a sum (with either finitely or infinitely many terms) is zero. Below is example of WZ proof of  $(1 + 1)^n = 2^n$ . This proof is from Doron Zeilberger’s original *Maple* program, which in turn is inspired by the proof in [212].

Let  $F(n, k) = \binom{n}{k} 2^{-n}$ . We wish to show that  $L(n) = \sum_k F(n, k) = 1$  for every  $n$ . To this end we construct, using the WZ algorithm, the function

$$G(n, k) = \frac{-1}{2^{(n+1)}} \binom{n}{k-1} \left( = \frac{-k}{2(n-k+1)} F(n, k) \right), \quad (7.1.1)$$

and observe that

$$F(n+1, k) - F(n, k) = G(n, k+1) - G(n, k). \quad (7.1.2)$$

By applying the obvious telescoping property of these functions, we can write

$$\begin{aligned} \sum_k F(n+1, k) - \sum_k F(n, k) &= \sum_k (G(n, k+1) - G(n, k)) \\ &= 0, \end{aligned} \quad (7.1.3)$$

which establishes that  $L(n+1) - L(n) = 0$ . The fact that  $L(0) = 1$  follows from the fact that  $F(0, 0) = 1$  and 0 otherwise.

Obviously this proof does not provide much insight, since the difficult part of the result is buried in the construction of 7.1.1. In other words, this is an instance where computers provide “proofs,” but these “proofs” tend to be uninteresting. Nonetheless the extremely general nature of this scheme is of interest. It possibly presages a future in which a wide class of such “proofs” can automatically be obtained in a computer algebra system. Details and additional applications of this algorithm are given in [212].

## 7.2 Prime Number Computations

In Chapter 2 of the first volume we mentioned the connection between prime numbers and the Riemann zeta function. Prime numbers crop up in numerous other arenas of mathematical research, and often even in commercial applications, with the rise of RSA encryption methods on the Internet. Inasmuch as this research topic is certain to be of great interest for the foreseeable future, we mention here some of the techniques for counting, generating and testing prime numbers.

The prime counting function  $\pi(x)$  mentioned in Chapter 2 of the first volume is of central interest in this research. It is clear from even a cursory glance at the

data in Table 2.2 of the first volume, that the researchers who have produced these counts are not literally testing every integer up to  $10^{22}$  for primality—that would require much more computation than the combined power of all computers worldwide, even using the best known methods to test individual primes. Indeed, some very sophisticated techniques have been employed, which unfortunately are too technical to be presented in detail here. We refer interested readers to the discussion of this topic in the new book *Prime Numbers: A Computational Perspective* by Richard Crandall and Carl Pomerance [86, 140-150]. Readers who wish to informally explore the behavior of  $\pi(x)$  may use the following algorithm, which is a variant of a scheme originally presented by Eratosthenes of Cyrene about 200 BCE [86, pg. 114]:

**Algorithm 1** *Blocked Sieve of Eratosthenes.*

We are given an interval  $(L, R)$ , where  $L$  and  $R$  are even integers, where  $L > P = \lceil \sqrt{R} \rceil$  and a blocksize  $B$  is assumed to divide  $R - L$ . We assume that a table of primes  $p_k$  up to  $P$  of size  $Q = \pi(P)$  is available.

Initialize: For  $k := 2$  to  $Q$  set  $q_k := -(L + 1 + p_k)/2 \bmod p_k$ .

Process blocks: For  $T := L$  to  $R - 1$  step  $2 \cdot B$  do: for  $j := 0$  to  $B - 1$  set  $b_j := 1$ ; for  $k := 2$  to  $Q$  do: for  $j := q_k$  to  $B - 1$  step  $p_k$  do: set  $b_j := 0$ ; enddo; set  $q_k := (q_k - B) \bmod p_k$ ; enddo; for  $j := 0$  to  $B - 1$  do: if  $b_j = 1$  then output  $T + 2j + 1$ ; enddo; enddo.  $\square$

The Sieve of Eratosthenes is only efficient if one wants to find all primes up to a given point, or all primes in a (fairly large) interval. For testing a single integer, or a few integers, faster means are available. We summarize some of these individual primality tests here.

The simplest scheme, from a programming point of view, of testing whether an integer  $n$  is prime is simply to generate a table of primes  $(p_1, p_2, \dots, p_n)$  up to  $p_n > \sqrt{n}$ , using the Sieve of Eratosthenes, and then to iteratively divide  $n$  by each  $p_i$ . This actually works quite well for modest-sized integers, but becomes infeasible for  $n$  beyond about  $10^{16}$ .

If one does not require certainty, but only high probability that a number is prime, some very efficient probabilistic primality tests have been discovered in the past few decades. In fact, these schemes are now routinely used to generate primes for RSA encryption in Internet commerce. When you type in your Visa

or MasterCard number in a secure website to purchase a book or computer accessory, somewhere in the process it is quite likely that two large prime numbers have been generated, which were certified as prime using one of these schemes.

The most widely used probabilistic primality test is the following, which was originally suggested by Artjuhov in 1966, although it was not appreciated until it was rediscovered and popularized by Selfridge in the 1970s [86].

**Algorithm 2** *Strong probable prime test.*

Given an integer  $n = 1 + 2^s t$ , for integers  $s$  and  $t$  (and  $t$  odd), select an integer  $a$  by means of a pseudo-random number generator in the range  $1 < a < n - 1$ .

1. Compute  $b := a^t \bmod n$  using the binary algorithm for exponentiation (see Algorithm 1 in Chapter 3 of the first volume). If  $b = 1$  or  $b = n - 1$  then exit ( $n$  is a strong probable prime base  $a$ ).
2. For  $j = 1$  to  $s - 1$  do: Compute  $b := b^2 \bmod n$ ; if  $(b = n - 1)$  then exit ( $n$  is a strong probable prime base  $a$ ).
3. Exit:  $n$  is composite. □

This test can be repeated several times with different pseudo-randomly chosen  $a$ . In 1980 Monier and Rabin independently showed that an integer  $n$  that passes the test as a strong probable prime is prime with probability at least  $3/4$ , so that  $m$  tests increase this probability to  $1 - 1/4^m$  [163, 178]. In fact, for large test integers  $n$ , the probability is even closer to unity. Damgard, Landrock and Pomerance showed in 1993 that if  $n$  has  $k$  bits, then this probability is greater than  $1 - k^2 4^{2-\sqrt{k}}$ , and for certain  $k$  is even higher [97]. For instance, if  $k$  has 500 bits, then this probability is greater than  $1 - 1/4^{28m}$ . Thus a 500-bit integer that passes this test even once is prime with the prohibitively safe odds—the chance of a false declaration of primality is less than one part in Avogadro’s number ( $6 \times 10^{23}$ ). If it passes the test for four pseudo-randomly chosen integers  $a$ , then the chance of false declaration of primality is less than one part in a googol ( $10^{100}$ ). Such probabilities are many orders of magnitude less than the chance that an undetected hardware or software error has occurred in the computation.

A number of more advanced probabilistic primality testing algorithms are now known. The current state-of-the-art is that such tests can determine the primality of integers with hundreds to thousands of digits. Additional details of these schemes are available in [86].



For these reasons, probabilistic primality tests are considered entirely satisfactory for practical use, even for applications, such as large interbank financial transactions, that have extremely high security requirements. Nonetheless mathematicians have long sought tests that remove this last iota of uncertainty, yielding a mathematically rigorous certificate of primality. Indeed, the question of whether there exists a “polynomial time” primality test has long stood as an important unsolved question in pure mathematics.

Thus it was with considerable elation that such an algorithm was recently discovered, by Manindra Agrawal, Neeraj Kayal and Nitin Saxena (initials AKS) of the Indian Institute of Technology in Kanpur, India [2]. Their discovery sparked worldwide interest, including a prominent report in the *New York Times* [184]. Since the initial report in August 2002, several improvements have been made. We present here a variant of the original algorithm due to Lenstra [151], as implemented by Richard Crandall and Jason Papadopoulos, who note that the implementations of Daniel Bernstein already provide significant acceleration [95].

**Algorithm 3** *Variant AKS provable primality test.*

Suppose we are given an integer  $p$  that we wish to establish as either prime or composite. This test assumes the existence of a table of primes covering integers up to roughly  $(\log_2 p)^2$ .

1. Establish that  $p$  is not a proper power, meaning that  $p \neq a^b$  for  $b > 1$ . This can be done by computing  $p^{1/c}$  for primes  $c$  up to  $\lceil \log_2(p) \rceil$ , using appropriately high-precision floating-point arithmetic. If  $p$  is a prime power, then of course it is composite.
2. Set  $v = \lceil \log_2 p \rceil$ . For integers  $r$  beginning with  $v + 1$ , test whether  $r$  is prime, and the multiplicative order of  $p$  modulo  $r$  is at least  $v$ . If not, increment  $r$  by one until these conditions are met.
3. For  $a = 1$  to  $r - 1$  test that the relation

$$(x - a)^p \equiv (x^p - a) \pmod{p, (x^r - 1)} \quad (7.2.4)$$

holds in the polynomial ring  $(\mathbb{Z}/p\mathbb{Z})[x]$ . If this condition holds for each  $a$  in the given range, then  $p$  is prime.  $\square$

Arithmetic in the polynomial ring  $(\mathbb{Z}/n\mathbb{Z})[x]$  can be accelerated by noting that polynomial multiplication is simply another instance of acyclic convolution, which can be computed using fast Fourier transforms (FFTs), as in FFT-based multiprecision multiplication (which is discussed in the next section). Proof that this algorithm produces a certificate of primality are in Lenstra's manuscript [151], and a working implementation of this scheme (which uses Crandall's PrimeKit software) is available from Crandall's website <http://www.perfsci.com>. A brief review of AKS implementation issues is available at [95].

### 7.3 Roots of Polynomials

In Section 6.2.5 of the first volume, we showed how a relatively simple scheme involving Newton iterations can be used to compute high-precision square roots and even to perform high-precision division. This Newton iteration scheme is, in fact, quite general and can be used to solve many kinds of equations, both algebraic and transcendental. One particularly useful application, frequently encountered by experimental mathematicians, is to find roots of polynomials. This is done by using a careful implementation of the well-known version of Newton's iteration

$$x_{k+1} = x_k - \frac{p(x)}{p'(x)}, \quad (7.3.5)$$

where  $p'(x)$  denotes the derivative of  $p(x)$ . As before, this scheme is most efficient if it employs a level of numeric precision that starts with ordinary double precision (16-digit) or double-double precision (32-digit) arithmetic until convergence is achieved at this level, then approximately doubles with each iteration until the final level of precision is attained. One additional iteration at the final or penultimate precision level may be needed to insure full accuracy.

Note that Newton's iteration can be performed, as written in (7.3.5), with either real or complex arithmetic, so that complex roots of polynomials (with real or complex coefficients) can be found almost as easily as real roots. Evaluation of the polynomials  $p(x)$  and  $p'(x)$  is most efficiently performed using Horner's rule: for example, the polynomial  $p(x) = p_0 + p_1x + p_2x^2 + p_3x^3 + p_4x^4 + p_5x^5$  is evaluated as  $p(x) = p_0 + x(p_1 + x(p_2 + x(p_3 + x(p_4 + xp_5))))$ .

There are two issues that arise here that do not arise with the Newton iteration schemes for division and square root. The first is the selection of the

starting value—if it is not close to the desired root then successive iterations may jump far away. If you have no idea where the roots are (or how accurate the starting value must be), then a typical strategy is to try numerous starting values, covering a wide range of likely values, and then make an inventory of the approximate roots that are found. If you are searching for complex roots, note that it is often necessary to use a two-dimensional array of starting values. These “exploratory” iterations can be done quite rapidly, since typically only a modest numeric precision is required—in almost all cases just ordinary double precision (16 digits) or double-double precision (32 digits) arithmetic will do. Once the roots have been located in this fashion, then the full-fledged Newton scheme can be used to produce their precise high-precision values.

The second issue is how to handle repeated roots. The difficulty here is that, in such cases, convergence to the root is very slow, and instabilities may throw the search far from the root. In these instances, note that we can write  $p(x) = q^2(x)r(x)$ , where  $r$  has no repeated roots (if all roots are repeated, then  $r(x) = 1$ ). Now note that  $p'(x) = 2q(x)r(x) + q^2(x)r'(x) = q(x)[2r(x) + q(x)r'(x)]$ . This means that if  $p(x)$  has repeated roots, then these roots are also roots of  $p'(x)$ , and, conversely, if  $p(x)$  and  $p'(x)$  have a common factor, then the roots of this common factor are repeated roots of  $p(x)$ . This greatest common divisor polynomial  $q(x)$  can be found by performing the Euclidean algorithm (in the ring of polynomials) on  $p(x)$  and  $p'(x)$ . The Newton iteration scheme can then be applied to find the roots of both  $q(x)$  and  $r(x)$ . It is possible, of course, that  $q(x)$  also has repeated roots, but recursive application of this scheme quickly yields all individual roots.

In the previous paragraph, we mentioned the possible need to perform the Euclidean algorithm on two polynomials, which involves polynomial multiplication and division. For modest-degree polynomials, a simple implementation of the schemes learned in high school algebra suffices—just represent the polynomials as strings of high-precision numbers. For high-degree polynomials, polynomial multiplication can be accelerated by utilizing fast Fourier transforms and a convolution scheme that is almost identical (except for release of carries) to the scheme, mentioned in Section 6.2 of the first volume, to perform high-precision multiplication. High-degree polynomial division can be accelerated by a Newton iteration scheme, similar to that mentioned above for high-precision division. See [86] for additional details on high-speed polynomial arithmetic.

We noted above that if the starting value is not quite close to the desired root, then successive Newton iterations may jump far from the root, and eventually

converge to a different root than the one desired. In general, suppose we are given a degree- $n$  polynomial  $p(x)$  with  $m$  distinct complex roots  $r_k$  (some may be repeated roots). Define the function  $Q_p(z)$  as the limit achieved by successive Newton iterations that start at the complex number  $z$ ; if no limit is achieved then set  $Q_p(z) = \infty$ . Then the  $m$  sets  $\{z : Q_p(z) = r_k\}$  for  $k = 1, 2, \dots, m$  constitute a partition of the complex plane, except for a filamentary set of measure zero that separates the  $m$  sets. In fact, each of these  $m$  sets is itself an infinite collection of disconnected components.

The collection of these Newton-Julia sets and their boundaries form pictures of striking beauty, and are actually quite useful in gaining insight on both the root structure of the original polynomial and the behavior of Newton iteration solutions. Some of the most interesting graphics of this type are color-coded plots of the function  $N_p(z)$ , which is the number of iterations required for convergence (to some accuracy  $\epsilon$ ) of Newton's iteration for  $p(x)$ , beginning at  $z$  (if the Newton iteration does not converge at  $z$ , then set  $N_p(z) = \infty$ ). A black-and-white variation of such a plot for the cubic polynomial  $p(x) = x^3 - 1$ , which displays a dot if  $N_p(z) \geq 20$ , for  $\epsilon = 10^{-30}$ , is shown in Figure 7.1. A color version is shown in the Color Supplement (Chapter 8) of the first volume.

## 7.4 Numerical Quadrature

Experimental mathematicians very frequently find it necessary to calculate definite integrals to high precision. Recall the examples given in Chapters 1 and 5 of the first volume, wherein we were able to experimentally identify certain definite integrals as analytic expressions, based only on their high-precision numerical value.

To briefly reprise one example, we were inspired by a recent problem in the *American Mathematical Monthly* [3]. We found by using one of the quadrature routines to be described below, together with a PSLQ integer relation detection program, that if  $C(a)$  is defined by

$$C(a) = \int_0^1 \frac{\arctan(\sqrt{x^2 + a^2}) dx}{\sqrt{x^2 + a^2}(x^2 + 1)}, \quad (7.4.6)$$

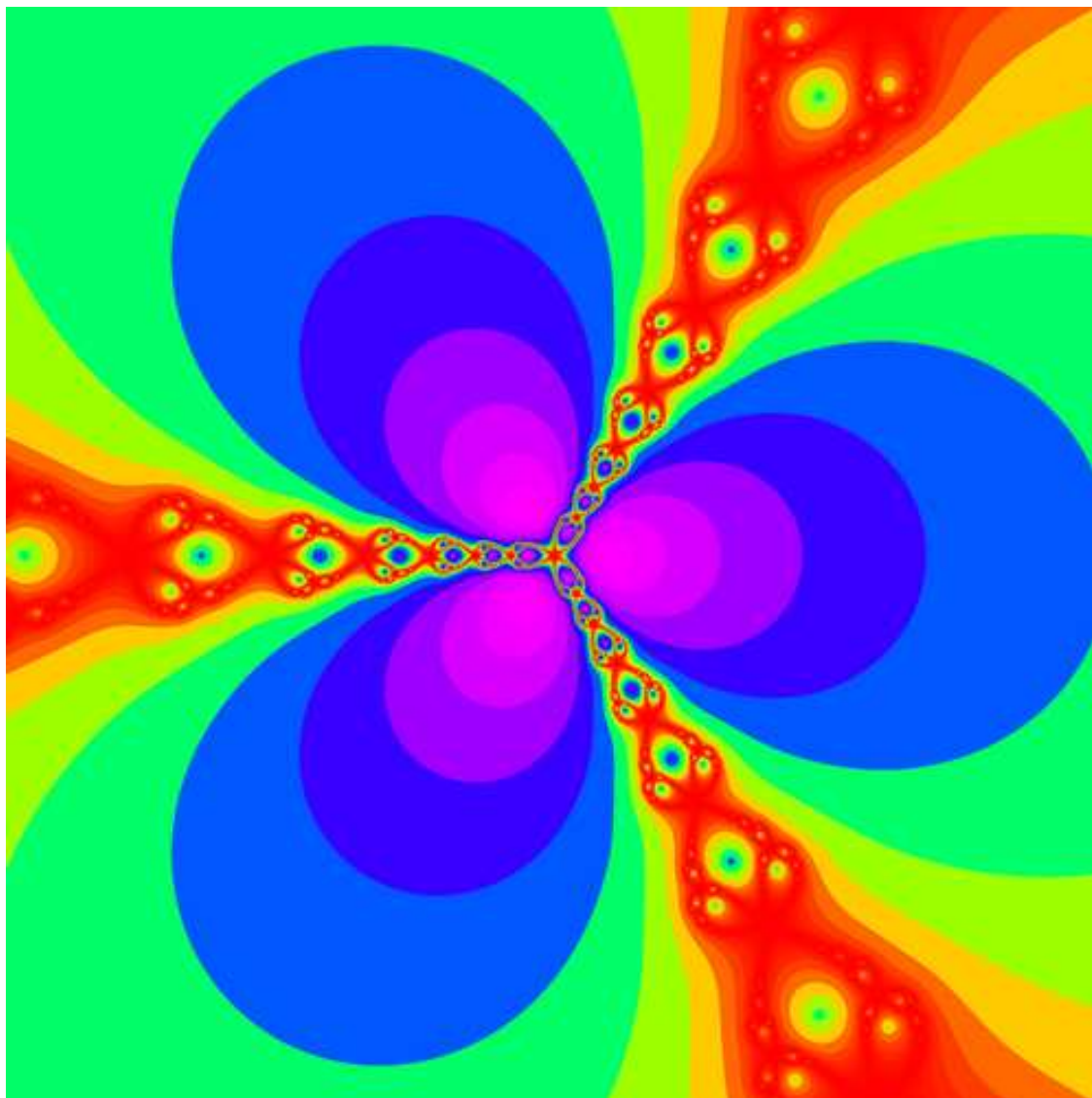


Figure 7.1: Newton-Julia set for  $p(x) = x^3 - 1$

then

$$\begin{aligned} C(0) &= \pi \log 2/8 + G/2 \\ C(1) &= \pi/4 - \pi\sqrt{2}/2 + 3\sqrt{2} \arctan(\sqrt{2})/2 \\ C(\sqrt{2}) &= 5\pi^2/96, \end{aligned} \tag{7.4.7}$$

where  $G = \sum_{k \geq 0} (-1)^k / (2k+1)^2$  is Catalan's constant. The third of these results is the result from the *Monthly*. These particular results then led to the following general result, among others:

$$\int_0^\infty \frac{\arctan(\sqrt{x^2 + a^2}) dx}{\sqrt{x^2 + a^2}(x^2 + 1)} = \frac{\pi}{2\sqrt{a^2 - 1}} \left[ 2 \arctan(\sqrt{a^2 - 1}) - \arctan(\sqrt{a^4 - 1}) \right]. \tag{7.4.8}$$

The commercial packages *Maple* and *Mathematica* both include rather good high-precision numerical quadrature facilities. However, these packages do have some limitations, and in many cases much faster performance can be achieved with custom-written programs. And in general it is beneficial to have some understanding of quadrature techniques, even if you rely on software packages to perform the actual computation.

We describe here three state-of-the-art, highly efficient techniques for numerical quadrature. You can try programming these schemes yourself, or you can refer to the C++ and Fortran-90 programs available at <http://www.expmath.info>.

### 7.4.1 Gaussian Quadrature

Our first quadrature scheme is known as Gaussian quadrature, named after the famous mathematician who first discovered this method in the 19th century. Gaussian quadrature is particularly effective for functions that are bounded, continuous and smooth on a finite interval. Unfortunately, many references only give tables of abscissas and weights, which are useless when we need multi-hundred digit accuracy. Thus we include here a fairly reasonably efficient scheme for computing these parameters to any desired accuracy (although, as we will see, the quadratic scaling of this scheme looms as an obstacle to very high-precision implementation).

Gaussian quadrature approximates the definite integral of the real function  $f(x)$  on  $(-1, 1)$  as

$$\int_{-1}^1 f(x) dx \approx \sum_{j=1}^n w_j f(x_j), \quad (7.4.9)$$

where the abscissas  $x_j$  are the zeroes of the degree- $n$  Legendre polynomial  $P_n(x)$  on  $[-1, 1]$ , and the weights  $w_j$  are given by

$$w_j = \frac{-2}{(n+1)P'_n(x_j)P_{n+1}(x_j)}, \quad (7.4.10)$$

for  $1 \leq j \leq n$  (see [14, pg. 285]). The abscissas can be computed using a Newton iteration scheme, where the starting value for  $x_j$  is given by  $\cos[\pi(j-1/4)/(n+1/2)]$  [175, pg. 125]. The Legendre polynomial function values can be computed using an  $n$ -long iteration of the recurrence  $P_0(x) = 0$ ,  $P_1(x) = 1$  and

$$(k+1)P_{k+1}(x) = (2k+1)xP_k(x) - kP_{k-1}(x),$$

for  $k \geq 2$ . The derivative is computed as  $P'_n(x) = n(xP_n(x) - P_{n-1}(x))/(x^2 - 1)$ .

In a typical implementation of Gaussian quadrature, multiple “levels” or phases are employed, with a set of abscissas and weights pre-calculated for each level, starting with say  $n = 3$ , and then with  $n$  doubling with each level up to some maximum phase  $m$ . For many well-behaved functions, Gaussian quadrature typically exhibits quadratic convergence, in the sense that after a few initial levels, each subsequent level approximately doubles the number of correct digits. However, each level performed also approximately doubles the computation time, compared with the previous level.

We summarize this procedure as follows [175, pg. 125].

**Algorithm 4** *Gaussian quadrature.*

Initialize (compute abscissas and weights):

For  $k := 1$  to  $m$  do:

Set  $n := 3 \cdot 2^k$ ;

For  $j := 1$  to  $n/2$  do:

Set  $r := \cos[\pi(j-1/4)/(n+1/2)]$ ;

Iterate until  $r = t_5$  to within the “epsilon” of the working precision level:

```

Set  $t_1 := 1$  and  $t_2 := 0$ ;
For  $j_1 := 1$  to  $n$  do:
Set  $t_3 := t_2$  and  $t_2 = t_1$  and calculate  $t_1 := [(2j_1 - 1)rt_2 - (j_1 - 1)t_3]/j_1$ ;
enddo
Calculate  $t_4 = n(rt_1 - t_2)/(r^2 - 1)$  and then set  $t_5 := r$  and  $r := r - t_1/t_4$ ;
end iterate
Set  $x_{j,k} = r$  and  $w_{j,k} = 2/[(1 - r^2)t_4^2]$ ;
enddo; enddo

```

```

Perform quadrature for a function  $f(x)$  on  $(-1, 1)$ :
For  $k := 1$  to  $m$  (or until successive instances yield results identical to within
the working precision) do:
Set  $n := 3 \cdot 2^k$  and set  $s := 0$ ;
For  $j = 1$  to  $n/2$  do:
Set  $s := s + w_{j,k}[f(-x_{j,k}) + f(x_{j,k})]$ ;
enddo
Set  $S_k := s$ ;
enddo

```

□

The Newton iteration scheme given in lines 4–11 above can be accelerated by utilizing a dynamic level of precision, as in other Newton-based algorithms. This means that the first few iterations are performed with modest precision, say double precision or double-double accuracy. Once convergence has been achieved at this level, additional iterations are performed with a level of precision that nearly doubles with each iteration, until the final level of precision has been achieved.

Error bounds are known for Gaussian quadrature, but since they rely on bounds for the  $n$ -th derivatives of  $f(t)$ , they are not very useful in this context, since we seek schemes that work for arbitrary functions, where such information is generally not available. See Section 7.4.4 below for some alternative means to estimate errors and determine when results are sufficiently accurate.

## 7.4.2 Error Function Quadrature

The second scheme we will discuss here is known as “error function” or “erf” quadrature. While error function quadrature is not as efficient as Gaussian



quadrature for continuous, bounded, well-behaved functions on finite intervals, it often produces highly accurate results even for functions with (integrable) singularities or vertical derivatives at one or both endpoints of the interval. In contrast, Gaussian quadrature typically performs very poorly in such instances.

The error function quadrature scheme and the tanh-sinh scheme to be described in the next section are based on the Euler-Maclaurin summation formula, which can be stated as follows [14, pg. 280]. Let  $m \geq 0$  and  $n \geq 1$  be integers, and define  $h = (b - a)/n$  and  $x_j = a + jh$  for  $0 \leq j \leq n$ . Further assume that the function  $f(x)$  is at least  $(2m + 2)$ -times continuously differentiable on  $[a, b]$ . Then

$$\int_a^b f(x) dx = h \sum_{j=0}^n f(x_j) - \frac{h}{2} (f(a) + f(b)) - \sum_{i=1}^m \frac{h^{2i} B_{2i}}{(2i)!} (f^{(2i-1)}(b) - f^{(2i-1)}(a)) - E, \quad (7.4.11)$$

where  $B_{2i}$  denote the Bernoulli numbers, and

$$E = \frac{h^{2m+2} (b - a) B_{2m+2} f^{2m+2}(\xi)}{(2m + 2)!}, \quad (7.4.12)$$

for some  $\xi \in (a, b)$ .

In the circumstance where the function  $f(x)$  and all of its derivatives are zero at the endpoints  $a$  and  $b$ , the second and third terms of the Euler-Maclaurin formula are zero. Thus the error in a simple step-function approximation to the integral, with interval  $h$ , is simply  $E$ . But since  $E$  is then less than a constant times  $h^{2m+2}/(2m + 2)!$ , for any  $m$ , we conclude that the error goes to zero more rapidly than any power of  $h$ . In the case of a function defined on  $(-\infty, \infty)$ , the Euler-Maclaurin summation formula still applies to the resulting doubly infinite sum approximation, provided as before that the function and all of its derivatives tend to zero for large positive and negative arguments.

This principle is utilized in the error function and tanh-sinh quadrature scheme (see the next subsection) by transforming the integral of  $f(x)$  on a finite interval, which we will take to be  $(-1, 1)$  for convenience, to an integral on  $(-\infty, \infty)$  using the change of variable  $x = g(t)$ . Here  $g(x)$  is some monotonic function with the property that  $g(x) \rightarrow 1$  as  $x \rightarrow \infty$ , and  $g(x) \rightarrow -1$

as  $x \rightarrow -\infty$ , and also with the property that  $g'(x)$  and all higher derivatives rapidly approach zero for large arguments. In this case we can write, for  $h > 0$ ,

$$\int_{-1}^1 f(x) dx = \int_{-\infty}^{\infty} f(g(t))g'(t) dt = h \sum_{-\infty}^{\infty} w_j f(x_j), \quad (7.4.13)$$

where  $x_j = g(hj)$  and  $w_j = g'(hj)$ . If the convergence of  $g'(t)$  and its derivatives to zero is sufficiently rapid for large  $|t|$ , then even in cases where  $f(x)$  has a vertical derivative or an integrable singularity at one or both endpoints, the resulting integrand  $f(g(t))g'(t)$  will be a smooth bell-shaped function for which the Euler-Maclaurin summation formula applies, as described above. In such cases we have that the error in the above approximation decreases faster than any power of  $h$ . The summation above is typically carried out to limits  $(-N, N)$ , beyond which the terms of the summand are less than the “epsilon” of the multiprecision arithmetic being used.

The error function integration scheme uses the function  $g(t) = \operatorname{erf}(t)$  and  $g'(t) = (2/\sqrt{\pi})e^{-t^2}$ . Note that  $g'(t)$  is merely the bell-shaped probability density function, which is well-known to converge rapidly to zero, together with all of its derivatives, for large arguments. The error function  $\operatorname{erf}(x)$  can be computed to high precision as  $1 - \operatorname{erfc}(x)$ , using the following formula given by Crandall [89, pg. 85] (who in turn attributes it to a 1968 paper by Chiarella and Reichel):

$$\operatorname{erfc}(t) = \frac{e^{-t^2} \alpha t}{\pi} \left( \frac{1}{t^2} + 2 \sum_{k \geq 1} \frac{e^{-k^2 \alpha^2}}{k^2 \alpha^2 + t^2} \right) + \frac{2}{1 - e^{2\pi t/\alpha}} + E, \quad (7.4.14)$$

where  $|E| < e^{-\pi^2/\alpha^2}$ . The parameter  $\alpha > 0$  here is chosen small enough to ensure that the error  $E$  is sufficiently small. We summarize this scheme with the following algorithm statement. Here  $n_p$  is the precision level in digits, and  $\epsilon$  is the “epsilon” level, which is typically  $10^{-n_p}$ .

**Algorithm 5** *Error function complement [erfc] evaluation.*

Initialize:

Set  $\alpha := \pi/\sqrt{n_p \log(10)}$ , and set  $n_t := n_p \log(10)/\pi$ .

Set  $t_2 := e^{-\alpha^2}$ ,  $t_3 := t_2^2$ , and  $t_4 := 1$ .

For  $k := 1$  to  $n_t$  do: set  $t_4 := t_2 \cdot t_4$ ,  $E_k := t_4$ ,  $t_2 := t_2 \cdot t_3$ ; enddo.

Evaluation of function, with argument  $x$ :

Set  $t_1 := 0$ ,  $t_2 := x^2$ ,  $t_3 := e^{-t_2}$  and  $t_4 := \epsilon/(1000 \cdot t_3)$ .

For  $k := 1$  to  $n_t$  do: set  $t_5 := E_k/(k^2\alpha^2 + t_2)$  and  $t_1 := t_1 + t_5$ .

If  $|t_5| < t_4$  then exit do; enddo.

Set  $\operatorname{erfc}(x) := t_3\alpha x/\pi \cdot (1/t_2 + 2t_1) + 2/(1 - e^{2\pi x/\alpha})$ . □

We now state the algorithm for error function quadrature. As with the Gaussian scheme,  $m$  levels or phases of abscissas and weights are pre-computed in the error function scheme. Then we perform the computation, increasing the level by one (each of which approximately doubles the computation, compared to the previous level), until an acceptable level of estimated accuracy is obtained (see Section 7.4.4). In the following,  $\epsilon$  is the “epsilon” level of the multiprecision arithmetic being used.

**Algorithm 6** *Error function quadrature.*

Initialize:

Set  $h := 2^{2-m}$ .

For  $k := 0$  to  $20 \cdot 2^m$  do:

Set  $t := kh$ ,  $x_k := 1 - \operatorname{erfc}(t)$  and  $w_k := 2/\sqrt{\pi} \cdot e^{-t^2}$ .

If  $|x_k - 1| < \epsilon$  then exit do; enddo.

Set  $n_t = k$  (the value of  $k$  at exit).

Perform quadrature for a function  $f(x)$  on  $(-1, 1)$ :

Set  $S := 0$  and  $h := 4$ .

For  $k := 1$  to  $m$  (or until successive values of  $S$  are identical to within  $\epsilon$ ) do:

$h := h/2$ .

For  $i := 0$  to  $n_t$  step  $2^{m-k}$  do:

If  $(\operatorname{mod}(i, 2^{m-k+1}) \neq 0$  or  $k = 1)$  then

If  $i = 0$  then  $S := S + w_0 f(0)$  else  $S := S + w_i(f(-x_i) + f(x_i))$  endif.

endif; enddo; endo.

Result =  $hS$ . □

### 7.4.3 Tanh-Sinh Quadrature

The third scheme we will discuss here is known informally as “tanh-sinh” quadrature. It is not well-known, but based on the authors’ experience, it deserves to

be taken seriously because of its speed, robustness and ease of implementation. Like the error function quadrature scheme, it is often successful in producing high-precision quadrature values even for functions with (integrable) singularities or vertical derivatives at endpoints. It was first introduced by Hidetosi Takahasi and Masatake Mori [198, 164].

The tanh-sinh scheme is very similar to the error function scheme, in that it is based on the Euler Maclaurin summation formula. The difference here is that it employs the transformation  $x = \tanh(\pi/2 \cdot \sinh t)$ , where  $\sinh t = (e^t - e^{-t})/2$ ,  $\cosh t = (e^t + e^{-t})/2$  and  $\tanh t = \sinh t / \cosh t$ . As before, this transformation converts an integral on  $(-1, 1)$  to an integral on the entire real line, which then can be approximated by means of a simple step-function summation. In this case, by differentiating the transformation, we obtain the abscissas  $x_k$  and the weights  $w_k$  as

$$\begin{aligned} x_j &= \tanh[\pi/2 \cdot \sinh(jh)] \\ w_j &= \frac{\pi/2 \cdot \cosh(jh)}{\cosh^2[\pi/2 \cdot \sinh(jh)]}. \end{aligned} \quad (7.4.15)$$

Note that these functions involved here are compound exponential, so for example the weights  $w_j$  converge very rapidly to zero. As a result, the tanh-sinh quadrature scheme is often even more effective than the error function scheme in dealing with singularities at endpoints.

**Algorithm 7** *tanh-sinh quadrature.*

Initialize: Set  $h := 2^{-m}$ .

For  $k := 0$  to  $20 \cdot 2^m$  do:

Set  $t := kh$ ,  $x_k := \tanh(\pi/2 \cdot \sinh t)$  and  $w_k := \pi/2 \cdot \cosh t / \cosh^2(\pi/2 \cdot \sinh t)$ ;

If  $|x_k - 1| < \epsilon$  then exit do; enddo.

Set  $n_t = k$  (the value of  $k$  at exit).

Perform quadrature for a function  $f(x)$  on  $(-1, 1)$ :

Set  $S := 0$  and  $h := 1$ .

For  $k := 1$  to  $m$  (or until successive values of  $S$  are identical to within  $\epsilon$ ) do:

$h := h/2$ .

For  $i := 0$  to  $n_t$  step  $2^{m-k}$  do:

If  $(\text{mod}(i, 2^{m-k+1}) \neq 0$  or  $k = 1)$  then

If  $i = 0$  then  $S := S + w_0 f(0)$  else  $S := S + w_i(f(-x_i) + f(x_i))$  endif.  
 endif; enddo; endo.  
 Result =  $hS$ . □

#### 7.4.4 Practical Considerations for Quadrature

Each of the schemes described above have assumed a function of one variable defined and continuous on the interval  $(-1, 1)$ . Integrals on other finite intervals  $(a, b)$  can be found by applying a linear change of variable:

$$\int_a^b f(t) dt = \frac{b-a}{2} \int_{-1}^1 f\left(\frac{b+a}{2} + \frac{b-a}{2}x\right) dx. \quad (7.4.16)$$

Note also that integrable functions on an infinite interval can, in a similar manner, be reduced to an integral on a finite interval, for example:

$$\int_0^\infty f(t) dt = \int_0^1 [f(x) + f(1/x)/x^2] dx. \quad (7.4.17)$$

Integrals of functions with singularities (such as “corners” or step discontinuities) within the integration interval (i.e., not at the endpoints) should be broken into separate integrals.

The above algorithm statements each suggest increasing the level of the quadrature (the value of  $k$ ) until two successive levels give the same value of  $S$ , to within some tolerance  $\epsilon$ . While this is certainly a reliable termination test, it is often possible to stop the calculation earlier, with significant savings in run time, by means of making reasonable projections of the current error level. In this regard, the authors have found the following scheme to be fairly reliable: Let  $S_1$ ,  $S_2$ , and  $S_3$  be the value of  $S$  at the current level, the previous level, and two levels back, respectively. Then set  $D_1 := \log_{10} |S_1 - S_2|$ ,  $D_2 := \log_{10} |S_1 - S_3|$ , and  $D_3 := \log_{10} \epsilon - 1$ . Now we can estimate the error  $E$  at level  $k > 2$  as  $10^{D_4}$ , where  $D_4 = \min(0, \max(D_1^2/D_2, 2D_1, D_3))$ . These estimation calculations may be performed using ordinary double precision arithmetic.

All three quadrature schemes have been implemented in C++ and Fortran-90 programs available at <http://www.expmath.info>.

### 7.4.5 2-D and 3-D Quadrature

The error function and tanh-sinh quadrature schemes can be easily generalized to perform two-dimensional (2-D) and three-dimensional (3-D) quadrature. Run times are typically many times higher than with one-dimensional (1-D) integrals. However, if one is content with say 32-digit or 64-digit results (by using double-double or quad-double arithmetic, respectively), then many two-variable functions can be integrated in reasonable run time (say a few minutes). One advantage that these schemes have is that they are very well suited to parallel processing. Thus even several-hundred digit values can be obtained for 2-D and 3-D integrals if one can utilize a highly parallel computer, such as a “Beowulf” cluster. One can even envision harnessing many computers on a geographically distributed grid for such a task, although the authors are not aware of any such attempts, as of this date.

One sample computation of this sort, performed by one of the present authors, produced the following evaluation:

$$\int_{-1}^1 \int_{-1}^1 \frac{dx dy}{\sqrt{1+x^2+y^2}} = 4 \log(2 + \sqrt{3}) - \frac{2\pi}{3}. \quad (7.4.18)$$

## 7.5 Infinite Series Summation

We have already seen numerous examples in previous chapters of mathematical constants defined by infinite series. In experimental mathematics work, it is usually necessary to evaluate such constants to say several hundred digit accuracy. The commercial software packages *Maple* and *Mathematica* include quite good facilities for the numerical evaluation of series. However, as with numerical quadrature, these packages do have limitations, and in some cases better results can be obtained using custom-written computer code. In addition, even if one relies exclusively on these commercial packages, it is useful to have some idea of the sorts of operations that are being performed by such software.

Happily, in many cases of interest to the experimental mathematician, infinite series converge sufficiently rapidly that they can be numerically evaluated to high precision by simply evaluating the series directly as written, stopping the summation when the individual terms are smaller than the “epsilon” of the

multiprecision arithmetic system being used. All of the BBP-type formulas, for instance, are of this category. But other types of infinite series formulas present considerable difficulties for high precision evaluation. Two simple examples are Gregory's series for  $\pi/4$  and a similar series for Catalan's constant:

$$\begin{aligned}\pi &= 1 - 1/3 + 1/5 - 1/7 + \cdots \\ G &= 1 - 1/3^2 + 1/5^2 - 1/7^2 + \cdots .\end{aligned}\tag{7.5.19}$$

We describe here one technique that is useful in many such circumstances. In fact, we have already been introduced to it in an earlier section of this chapter: It is the Euler-Maclaurin summation formula. The Euler-Maclaurin formula can be written in somewhat different form than before, as follows [14, page 282]. Let  $m \geq 0$  be an integer, and assume that the function  $f(x)$  is at least  $(2m+2)$ -times continuously differentiable on  $[a, \infty)$ , and that  $f(x)$  and all of its derivatives approach zero for large  $x$ . Then

$$\sum_{j=a}^{\infty} f(j) = \int_a^{\infty} f(x) dx + \frac{1}{2}f(a) - \sum_{i=1}^m \frac{B_{2i}}{(2i)!} f^{(2i-1)}(a) + E, \tag{7.5.20}$$

where  $B_{2i}$  denote the Bernoulli numbers, and

$$E = \frac{B_{2m+2} f^{(2m+2)}(\xi)}{(2m+2)!}, \tag{7.5.21}$$

for some  $\xi \in (a, \infty)$ .

This formula is not effective as written. The strategy is instead to evaluate a series manually for several hundred or several thousand terms, then to use the Euler-Maclaurin formula to evaluate the tail. Before giving an example, we need to describe how to calculate the Bernoulli numbers  $B_{2k}$ , which are required here. The simplest way to compute them is to recall that [1, page 807]

$$\zeta(2k) = \frac{(2\pi)^{2k} |B_{2k}|}{2(2k)!}, \tag{7.5.22}$$

which can be rewritten as

$$\frac{B_{2k}}{(2k)!} = \frac{2(-1)^{k+1} \zeta(2k)}{(2\pi)^{2k}}. \tag{7.5.23}$$

The Riemann zeta function at real arguments  $s$  can in turn be computed using the formula [64]

$$\zeta(s) = \frac{-1}{2^n(1-2^{1-s})} \sum_{j=0}^{2n-1} \frac{e_j}{(j+1)^s} + E_n(s), \quad (7.5.24)$$

where

$$e_j = (-1)^j \left( \sum_{k=0}^{j-n} \frac{n!}{k!(n-k)!} - 2^n \right) \quad (7.5.25)$$

(the summation is zero when its index range is null) and  $|E_n(s)| < 1/(8^n|1-2^{1-s}|)$ . This scheme is encapsulated in the following algorithm.

**Algorithm 8** *Zeta function evaluation.*

Initialize: Set  $n = P/3$ , where  $P$  is the precision level in bits, and set  $t_1 := -2^n$ ,  $t_2 := 0$ ,  $S := 0$ , and  $I = 1$ .

For  $j := 0$  to  $2n - 1$  do: If  $j < n$  then  $t_2 := 0$  elseif  $j = n$  then  $t_2 := 1$  else  $t_2 := t_2 \cdot (2n - j + 1)/(j - n)$  endif.

Set  $t_1 := t_1 + t_2$ ,  $S := S + I \cdot t_1/(j+1)^s$  and  $I := -I$ ; enddo.

Return  $\zeta(s) := -S/[2^n \cdot (1 - 2^{1-s})]$ . □

A more advanced method to compute the zeta function in the particular case of interest here, where we need the zeta function evaluated at all even integer arguments up to some level  $m$ , is described in [16].

We will illustrate the above by calculating Catalan's constant using the Euler-



Maclaurin formula. We can write

$$\begin{aligned}
 G &= (1 - 1/3^2) + (1/5^2 - 1/7^2) + (1/9^2 - 1/11^2) + \cdots \\
 &= 8 \sum_{k=0}^{\infty} \frac{2k+1}{(4k+1)^2(4k+3)^2} \\
 &= 8 \sum_{k=0}^n \frac{2k+1}{(4k+1)^2(4k+3)^2} + 8 \sum_{k=n+1}^{\infty} \frac{2k+1}{(4k+1)^2(4k+3)^2} \\
 &= 8 \sum_{k=0}^n \frac{2k+1}{(4k+1)^2(4k+3)^2} + 8 \int_{n+1}^{\infty} f(x) dx + 4f(n+1) \\
 &\quad - 8 \sum_{i=1}^m \frac{B_{2i}}{(2i)!} f^{(2i-1)}(n+1) + 8E,
 \end{aligned} \tag{7.5.26}$$

where  $f(x) = (2x+1)/[(4x+1)^2(4x+3)^2]$  and  $|E| < 3/(2\pi)^{2m+2}$ . Using  $m = 20$  and  $n = 1000$  in this formula, we obtain a value of  $G$  correct to 114 decimal digits. We presented the above scheme for Catalan's constant because it is illustrative of the Euler-Maclaurin method. However serious computation of Catalan's constant can be done more efficiently using the Boole summation formula (see (16)), the recently discovered BBP-type formula (given in Table 3.5 of the first volume), Ramanujan's formula (given in Item 7 of the Chapter 6 in the first volume), or Bradley's formula (also given in Item 7 of Chapter 6 in the first volume).

One less-than-ideal feature of the Euler-Maclaurin approach is that high-order derivatives are required. In many cases of interest, successive derivatives satisfy a fairly simple recursion and can thus be easily computed with an ordinary hand-written computer program. In other cases, these derivative are sufficiently complicated that such calculations are more conveniently performed in a symbolic computing environment such as *Mathematica* or *Maple*. In a few applications of this approach, a combination of symbolic computation and custom-written numerical computation is required to produce results in reasonable runtime[18].

### 7.5.1 Computation of Multiple Zeta Constants

As we saw in Chapter 3, one class of mathematical constants that has been of particular interest to experimental mathematicians in the past few years are

multiple zeta constants. Research in this arena has been facilitated by the discovery of methods that permit these constants to high precision. While Euler-Maclaurin-based schemes can be used (and in fact were used) in these studies, they are limited to 2-order sums. We present here an algorithm that permits even high-order sums to be evaluated to hundreds or thousands of digit accuracy. We will limit our discussion here to multiple zeta constants of the form

$$\zeta(s_1, s_2, \dots, s_n) = \sum_{n_1 > n_2 > \dots > n_k} \frac{1}{n_1^{s_1} n_2^{s_2} \dots n_k^{s_k}}, \quad (7.5.27)$$

for positive integers  $s_k$  and  $n_k$ , although in general technique we describe here has somewhat broader applicability.

This scheme is as follows [36]. For  $1 \leq j \leq m$ , define the numeric strings

$$a_j = \{s_j + 2, \{1\}^{r_j}, s_{j+1}, \{1\}^{r_{j+1}}, \dots, s_m + 2, \{1\}^{r_m}\} \quad (7.5.28)$$

$$b_j = \{r_j + 2, \{1\}^{s_j}, r_{j-1}, \{1\}^{s_{j-1}}, \dots, r_1 + 2, \{1\}^{s_1}\}, \quad (7.5.29)$$

where by the notation  $\{1\}^n$  we mean  $n$  repetitions of 1. For convenience, we will define  $a_{m+1}$  and  $b_0$  to be the empty string. Define

$$\delta(s_1, s_2, \dots, s_k) = \prod_{j=1}^k \left[ \sum_{\nu_j=1}^{\infty} 2^{-\nu_j} \left( \sum_{i=j}^k \nu_i \right)^{-s_j} \right]. \quad (7.5.30)$$

Then we have

$$\begin{aligned} \zeta(a_m) = & \sum_{j=1}^m \left[ \sum_{t=0}^{s_j+1} \delta(s_j + 2 - t, \{1\}^{r_j}, a_{j+1}) \delta(\{1\}^t, b_{j-1}) \right. \\ & \left. + \sum_{u=1}^{r_j} \delta(\{1\}^u, a_{j+1}) \delta(r_j + 2 - u, \{1\}^{s_j}, b_{j-1}) \right] + \delta(b_m). \end{aligned} \quad (7.5.31)$$

See the discussion in Chapter 3, as well as [36] and [87] for further details. An on-line tool that implements this procedure is available at <http://www.cecm.sfu.ca/projects/ezface+>. This procedure has also been implemented as part of the Experimental Mathematician's Toolkit, available at <http://www.expmath.info>.

## 7.6 Commentary and Additional Examples

1. **Primes in  $\pi$ .**  $\pi_1 = 3, \pi_2 = 31, \pi_6 = 314159$ , and  $\pi_{38}$  are the only prime initial strings in the first 500 digits of  $\pi$ . The next probable prime string is  $\pi_{16208}$ . This suspected prime would be a superb candidate for the new AKS-class of rigorous, deterministic algorithms discussed in Section 7.2. For more information on primes in  $\pi$ , see <http://mathpages.com/home/kmath184/kmath184.htm>.

2. **Adaptation of Putnam problem 1993–A5.** Consider the integral

$$J(a, b) = \int_a^b \frac{(x^2 - x)^2}{(x^3 - 3x + 1)^2} dx.$$

Show that

$$J(-100, -10) + J\left(\frac{1}{101}, \frac{1}{11}\right) + J\left(\frac{101}{100}, \frac{11}{10}\right)$$

is rational. Hint: Numerically observe that the given expression is rational. Then consider the changes of variables  $x \rightarrow 1 - 1/x$  and  $x \rightarrow 1/(1 - x)$ , respectively, to move the second and third integrals to the interval  $[-100, -10]$ . Answer: 11131110/107634259.

3. **Berkeley problem 5.10.26.** Evaluate

$$\mathcal{I} = \int_0^{2\pi} \frac{\cos^2(3t)}{5 - 4 \cos(2t)} dt.$$

Answer:  $3\pi/8$ . Hint:  $\mathcal{I} = \frac{1}{2} \operatorname{Re} \left( i \int_{|z|=1} (z^3 + 1)/(2z^2 - 5z + 2) dz \right) = -\pi \operatorname{Res} \left( \frac{z^3 + 1}{(2z^2 - 5z + 2)}, z = \frac{1}{2} \right)$ .

4. **Berkeley problem 5.10.27.** Evaluate

$$\mathcal{I} = \int_{|z|=1} \frac{\cos^3(z)}{z^3} dz$$

taken counter-clockwise. Answer:  $3\pi i$ . Hint:  $2\pi i \operatorname{Res}(\cos^3(z)/z^3, z = 0)$ .

5. **Berkeley problem 5.11.4.** Evaluate

$$\int_0^{\infty} \frac{\sin^2(t)}{t^2} dt.$$

Answer:  $\pi/2$ . Hint: Integrate by parts or use residues.

6. **Berkeley problem 5.11.5.** Evaluate

$$\int_{-\infty}^{\infty} \frac{\sin^3(t)}{t^3} dt \quad \left( = \frac{3\pi}{4} \right).$$

Hint:  $4 \sin^3(z) = \operatorname{Re} 3 \exp(iz) - \exp(3iz)$ . Use Cauchy's theorem on a contour  $C$  which consists of semicircles of radius  $\varepsilon \rightarrow 0$  and  $R \rightarrow \infty$  along with the intervals  $[-R, -\varepsilon]$  and  $[\varepsilon, R]$ .

7. **Berkeley problem 5.11.8.** Show

$$\int_0^{\infty} \frac{\sin(x)}{x(x^2 + a^2)} dx = \frac{\pi(1 - e^{-a})}{2a^2}.$$

8. **Berkeley problem 5.11.14.** Evaluate

$$\int_0^{\infty} \frac{1 + x^2}{1 + x^4} dx \quad \left( = \frac{\pi}{\sqrt{2}} \right).$$

Hint: Again use residues or consider  $\int_0^t (1 + x^2)/(1 + x^4) dx$ .

9. **Berkeley problem 5.11.22.** Show

$$\int_0^{\infty} \frac{x}{\sinh(x)} dx = \frac{\pi^2}{4}.$$

Then determine

$$\int_0^t \frac{x}{\sinh(x)} dx.$$

10. **Berkeley problem 5.11.25.** Show

$$\int_0^{\infty} \frac{\log^2(x)}{1 + x^2} dx = \frac{\pi^3}{8}.$$

11. Evaluate the following integrals, by numerically computing them and then trying to recognize the answers, either by using the Inverse Symbolic Calculator at <http://www.cecm.sfu.ca/projects/ISC/>, or else by using a PSLQ facility, such as that built into the Experimental Mathematician's Toolkit, available at <http://www.expmath.info>.

These examples are taken from Gradshteyn and Ryzhik's reference [125]. All of the answers are simple one- or few-term expressions involving familiar mathematical constants such as  $\pi$ ,  $e$ ,  $\sqrt{2}$ ,  $\sqrt{3}$ ,  $\log 2$ ,  $\zeta(3)$ ,  $G$  (Catalan's constant) and  $\gamma$  (Euler's constant). We recognize that many of these can be evaluated analytically using symbolic computing software (depending on the available versions). The intent here is to provide exercises for numerical quadrature and constant recognition facilities.

$$(a) \quad \int_0^1 \frac{x^2 dx}{(1+x^4)\sqrt{1-x^4}} \quad (7.6.32)$$

$$(b) \quad \int_0^\infty x e^{-x} \sqrt{1-e^{-2x}} dx \quad (7.6.33)$$

$$(c) \quad \int_0^\infty \frac{x^2 dx}{\sqrt{e^x-1}} \quad (7.6.34)$$

$$(d) \quad \int_0^{\pi/4} x \tan x dx \quad (7.6.35)$$

$$(e) \quad \int_0^{\pi/2} \frac{x^2 dx}{1-\cos x} \quad (7.6.36)$$

$$(f) \quad \int_0^{\pi/4} (\pi/4 - x \tan x) \tan x dx \quad (7.6.37)$$

$$(g) \quad \int_0^{\pi/2} \frac{x^2 dx}{\sin^2 x} \quad (7.6.38)$$

$$(h) \quad \int_0^{\pi/2} \log^2(\cos x) dx \quad (7.6.39)$$

$$(i) \quad \int_0^1 \frac{\log^2 x \, dx}{x^2 + x + 1} \quad (7.6.40)$$

$$(j) \quad \int_0^1 \frac{\log(1 + x^2) \, dx}{x^2} \quad (7.6.41)$$

$$(k) \quad \int_0^\infty \frac{\log(1 + x^3) \, dx}{1 - x + x^2} \quad (7.6.42)$$

$$(l) \quad \int_0^\infty \frac{\log x \, dx}{\cosh^2 x} \quad (7.6.43)$$

$$(m) \quad \int_0^1 \frac{\arctan x}{x\sqrt{1-x^2}} \quad (7.6.44)$$

$$(n) \quad \int_0^{\pi/2} \sqrt{\tan t} \, dt \quad (7.6.45)$$

Answers: (a)  $\pi/8$ , (b)  $\pi(1+2\log 2)/8$ , (c)  $4\pi(\log^2 2 + \pi^2/12)$ , (d)  $(\pi \log 2)/8 + G/2$ , (e)  $-\pi^2/4 + \pi \log 2 + 4G$ , (f)  $(\log 2)/2 + \pi^2/32 - \pi/4 + (\pi \log 2)/8$ , (g)  $\pi \log 2$ , (h)  $\pi/2(\log^2 2 + \pi^2/12)$ , (i)  $8\pi^3/(81\sqrt{3})$ , (j)  $\pi/2 - \log 2$ , (k)  $2(\pi \log 3)/\sqrt{3}$ , (l)  $\log \pi - 2 \log 2 - \gamma$ , (m)  $[\pi \log(1 + \sqrt{2})]/2$ , (n)  $\pi\sqrt{2}/2$ .

12. Evaluate the following infinite series, by numerically computing them and then trying to recognize the answers, either by using the Inverse Symbolic Calculator at <http://www.cecm.sfu.ca/projects/ISC/>, or else by using a PSLQ facility, such as that built into the Experimental Mathematician's Toolkit, available at <http://www.expmath.info>.

These examples have been provided to the authors by Gregory and David Chudnovsky of the Institute for Mathematics and Supercomputing at Brooklyn Polytechnic College. All of the answers are simple one- or few-term expressions involving familiar mathematical constants such as  $\pi$ ,  $e$ ,  $\sqrt{2}$ ,  $\sqrt{3}$ ,

$\log 2$ ,  $\zeta(3)$ ,  $G$  (Catalan's constant) and  $\gamma$  (Euler's constant).

$$(a) \quad \sum_0^{\infty} \frac{50n - 6}{2^n \binom{3n}{n}} \quad (7.6.46)$$

$$(b) \quad \sum_0^{\infty} \frac{2^{n+1}}{\binom{2n}{n}} \quad (7.6.47)$$

$$(c) \quad \sum_0^{\infty} \frac{12n2^{2n}}{\binom{4n}{2n}} \quad (7.6.48)$$

$$(d) \quad \sum_0^{\infty} \frac{(4n)!(1 + 8n)}{4^{4n}n!^4} \quad (7.6.49)$$

$$(e) \quad \sum_0^{\infty} \frac{(4n)!(19 + 280n)}{4^{4n}n!^4 99^{2n+1}} \quad (7.6.50)$$

$$(f) \quad \sum_0^{\infty} \frac{(2n)!(3n)!4^n(4 + 33n)}{n!^5 108^n 125^n} \quad (7.6.51)$$

$$(g) \quad \sum_0^{\infty} \frac{(-27)^n(90n + 177)}{16^n \binom{3n}{n}} \quad (7.6.52)$$

$$(h) \quad \sum_0^{\infty} \frac{275n - 158}{2^n \binom{3n}{n}} \quad (7.6.53)$$

$$(i) \quad \sum_0^{\infty} \frac{8^n(520 + 6240n - 430n^2)}{\binom{4n}{n}} \quad (7.6.54)$$

$$(j) \quad \sum_0^{\infty} \frac{\binom{2n}{n}}{n^2 4^n} \quad (7.6.55)$$

$$(k) \quad \sum_0^{\infty} \frac{(-1)^n}{n^3 2^n \binom{2n}{n}} \quad (7.6.56)$$

$$(l) \quad \sum_0^{\infty} \frac{8^n(338 - 245n)}{3^n \binom{3n}{n}} \quad (7.6.57)$$

$$(m) \quad \sum_1^{\infty} \frac{(-9)^n \binom{2n}{n}}{6n^2 64^n} - \sum_1^{\infty} \frac{3^n \binom{2n}{n}}{n^2 16^n} \quad (7.6.58)$$

Answers: (a)  $\pi$ , (b)  $\pi + 4$ , (c)  $3\pi + 8$ , (d)  $2/(\pi\sqrt{3})$ , (e)  $2/(\pi\sqrt{11})$ , (f)  $15\sqrt{3}/(2\pi)$ , (g)  $120 - 64\log 2$ , (i)  $45\pi - 1164$ , (j)  $\pi^2/6 - 2\log^2 2$ , (k)  $\log^3 2/6 - \zeta(3)/4$ , (l)  $162 - 6\pi\sqrt{3} - 18\log 3$ , (m)  $\pi^2/18 + \log^2 2 - \log^3 3/6$ .

13. **Pisot and Salem numbers.** A real algebraic integer  $\alpha > 1$  is a *Pisot number* (resp. *Salem number*) if all other roots of its monic polynomial lie inside (respectively, inside or on) the unit disc. An algebraic number  $\alpha$  is a Pisot number if and only if, for  $n \in \mathbb{N}$ , the fractional parts  $\{n\alpha\} \rightarrow 0$  as  $n \rightarrow \infty$ . The smallest Pisot number is the largest root of  $z^3 - z - 1$  and is approximately  $1.3247179\dots$ . The smallest Salem number is *conjectured* to be the largest root of *Lehmer's polynomial*

$$p(z) = z^{10} + z^9 - z^7 - z^6 - z^5 - z^4 - z^3 + z + 1,$$

which has two real roots  $0.850137130927042\dots$  and  $1.176280818259917\dots$ . The conjecture is called *Lehmer's problem (1933)* [65].

- (a) Show that for a Pisot number the fractional parts  $\{n\alpha\} \rightarrow 0$  as  $n \rightarrow \infty$ . Hint: use Newton's formula for

$$\sigma_n = \sum_{p(\beta)=0} \beta^n,$$

where  $p$  is the polynomial associated with  $\alpha$ .

- (b) Computationally establish Lehmer's conjecture up to as high a degree as possible (this has been done to about degree 40).

14. **Halley's method.** Suppose  $f$  is a suitably smooth real function and  $N \geq 1$  is integer. Consider the iteration starting at an initial value  $y_0 = y$  and iterating

$$y_{n+1} = y_n + \frac{(1/f)^{(N)}(y)}{(1/f)^{(N+1)}(y)} \Big|_{y=y_n}.$$

For  $N = 0$  this is Newton's method to find a zero of  $f$ . For  $N = 1$ , this is a locally cubically convergent method due to the astronomer and mathematician Halley. Show in general that the method converges of order  $N + 2$  for the initial guess close enough to a (non-degenerate) zero of  $f$ .



15. **Bender's continued fraction for the Bernoulli numbers.** The Bernoulli numbers as exact rational numbers can be generated from a continued fraction, due to Bender, which makes a nice equally divergent counterpart to Exercise 56 of Chapter 7 of the first volume. Formally,

$$\mathcal{S}(a) = 1 + \frac{b_0 a^2}{1 + \frac{b_1 a^2}{1 + \frac{b_2 a^2}{1 + \frac{b_3 a^2}{\ddots}}}}$$

$$= \sum_{n=0}^{\infty} B_{2n} a^{2n},$$

with  $b_0 = 1/6$  and

$$b_n = \frac{n(n+1)^2(n+2)}{4(2n+1)(2n+3)},$$

for  $n > 0$ .

- (a) While the series does not converge in any obvious sense, it gives a very satisfactory symbolic expansion. Decide whether  $\mathcal{S}(a)$  coincides with the subsequent three equivalent expressions:

$$\begin{aligned} \mathcal{S}(a) &\stackrel{?}{=} a \int_0^{\infty} t \coth\left(\frac{ta}{2}\right) e^{-t} dt \\ &= a \int_0^1 \frac{\log(s)}{s^a - 1} ds - \frac{a}{2} = 1 + \frac{a}{2} + a \sum_{n=1}^{\infty} \frac{1}{(an+1)^2}. \end{aligned}$$

- (b) Evaluate  $\mathcal{S}(4)$ ,  $\mathcal{S}(1/4)$ ,  $\mathcal{S}(3)$ , and  $\mathcal{S}(1/3)$ .

- (c) Show that  $2n \int_0^{\infty} x^{2n-1} e^{-\pi x} \operatorname{cosech}(\pi x) dx = |B_{2n}|$  for  $n = 1, 2, \dots$ . (What are the even moments?) Deduce, for all real  $a$ , that

$$\begin{aligned} \Psi'(1+a^{-1})/a + a/2 &= 1 + \frac{a}{2} + a \sum_{n=1}^{\infty} \frac{1}{(an+1)^2} \\ &= \int_0^{\infty} \coth\left(\frac{\pi x}{a}\right) \frac{2x}{(1+x^2)^2} dx, \end{aligned}$$

and that

$$\sum_{n=1}^N B_{2n} a^{2n} = 4a \int_0^\infty \frac{ya \left(1 - (-1)^N (ya)^{2N} \{1 + N(1 + (ya)^2)\}\right)}{(e^{2\pi y} - 1)(1 + (ya)^2)^2} dy,$$

for  $N > 0$ .

(d) In consequence,

$$\left| \int_0^\infty \coth\left(\frac{\pi x}{a}\right) \frac{2x}{(1+x^2)^2} dx - \sum_{n=0}^{N-1} B_{2n} a^{2n} \right| < |B_{2N}| a^{2N},$$

and we obtain a genuine asymptotic expansion.

(e) Show, for  $a > 1$ , that

$$\begin{aligned} \mathcal{S}(a) &= \frac{a(3+2a+a^2)}{2(1+a)^2} - \sum_{n=1}^{\infty} \frac{n}{(-a)^n} (\zeta(n+1) - 1) \\ &= \frac{1}{2a} \pi^2 \csc^2\left(\frac{\pi}{a}\right) - 2 \frac{a^2}{(1-a^2)^2} - 2 \sum_{n=1}^{\infty} \frac{n}{a^{2n}} (\zeta(2n+1) - 1). \end{aligned}$$

For what  $a$  are these valid equalities?

(f) Show that

$$\mathcal{T}(a) = \int_0^\infty \tanh(ax) e^{-x} dx = \int_0^1 \frac{1-y^{2a}}{1+y^{2a}} dy$$

generates the continued fraction

$$\frac{a}{1 + \frac{\frac{a}{2a^2}}{1 + \frac{\frac{a}{6a^2}}{1 + \frac{\frac{a}{12a^2}}{1 + \frac{\frac{a}{20a^2}}{1 + \frac{\frac{a}{\dots}}{\dots}}}}}} = - \sum_{n \geq 1} T_{2n+1} a^{2n-1}.$$

Here the  $T_n$  are, as before, tangent numbers.

(g) Provide a justification for the last identity, along the lines given above for the Bernoulli numbers.

(h) Show

$$\frac{\mathcal{T}(a) + 1}{2} = F\left(\frac{1}{2a}, 1; \frac{1}{2a} + 1; -1\right)$$

is of the form for Gauss's continued fraction.

16. **The Boole summation formula.** Boole summation is a counterpart of Euler-Maclaurin summation especially fitted to alternating series. Based upon Euler numbers, rather than Bernoulli numbers, it is described in detail in [37], and in the revised version of Abramowitz and Stegun available at <http://dlmf.nist.gov/>. We recall the *Euler polynomials* are defined by

$$\frac{2e^{tx}}{e^t + 1} = \sum_{n=0}^{\infty} E_n(x) \frac{t^n}{n!}$$

for  $|t| < \pi$ , and let the *periodic Euler function* be defined by  $\overline{E}_n(x) = E_n(x)$  for  $0 < x \leq 1$  and  $\overline{E}_n(x+1) = -\overline{E}_n(x)$  for all  $x$ . Then the Euler numbers are recovered as  $E_n = 2^n E_n(\frac{1}{2})$ .

The version of Boole summation we need tells us that if  $f$  has  $m$  derivatives on  $t \geq x$ , and  $f^{(k)}(t) \rightarrow 0$  as  $t \rightarrow \infty$  for  $0 \leq k \leq m$  then

$$\sum_{n=0}^{\infty} (-1)^n f(x+h+n) = \sum_{k=0}^m \frac{E_k(h)}{2k!} f^{(k)}(x) + R_m \quad (7.6.59)$$

where

$$R_m = \frac{1}{2(m-1)!} \int_0^{\infty} \overline{E}_{m-1}(h-t) f^{(m)}(x+t) dt.$$

(a) Use (7.6.59) to prove the formulas for the “errors” in Gregory series for  $\pi$  and the classical series for  $\log 2$  given in Section 2.2 of the first volume.

(b) Apply the same technique to Catalan's constant and to  $\zeta(3)$ .

17. **Value Recycling,  $\zeta$  and  $\Gamma$ .** This material is culled from [53]. The notion of *recycling* is that previously calculated  $\zeta$ -values—or initialization tables

of those calculations—are re-used to aid in the extraction of other  $\zeta$ -values, or that many  $\zeta$ -values are somehow simultaneously determined, and so on. So by value recycling, we intend that the computation of a collection of  $\zeta$ -values is more efficient than establishment of independent values. We record

$$\pi t \cot(\pi t) = -2 \sum_{m=0}^{\infty} \zeta(2m) t^{2m}, \quad (7.6.60)$$

and the *incomplete gamma function*, given (at least for  $\operatorname{Re}(z) > 0$ ) by:

$$\Gamma(a, z) = \int_z^{\infty} t^{a-1} e^{-t} dt = \frac{2z^a e^{-z}}{\Gamma(1-a)} \int_0^{\infty} \frac{t^{1-2a} e^{-t^2}}{t^2 + z} dt,$$

where the integral representation is valid for  $\operatorname{Re}(a) < 1$ . Evaluation of  $\Gamma(a, z)$  is not as problematic as it may seem; many computer systems have suitable machinery. We note the special cases  $\Gamma(s, 0) = \Gamma(s)$  and  $\Gamma(1, z) = e^{-z}$ , and the recursion

$$a\Gamma(a, z) = \Gamma(a+1, z) - z^a e^{-z}. \quad (7.6.61)$$

Also,

$$\begin{aligned} \zeta(s)\Gamma\left(\frac{1}{2}s\right) &= \frac{\pi^{s/2}}{s(s-1)} + \sum_{n=1}^{\infty} n^{-s}\Gamma\left(\frac{1}{2}s, \pi n^2\right) \\ &\quad + \pi^{s-1/2} \sum_{n=1}^{\infty} n^{s-1}\Gamma\left(\frac{1}{2}(1-s), \pi n^2\right), \end{aligned} \quad (7.6.62)$$

$$\begin{aligned} \zeta(s)\Gamma(s) &= -\frac{\lambda^s}{2s} + \frac{\lambda^{s-1}}{s-1} + \sum_{n=0}^{\infty} n^{-s}\Gamma(s, \lambda n) \\ &\quad - 2\lambda^{s-1} \sum_{n=1}^{\infty} \left(\frac{\lambda}{2\pi i}\right)^{2n} \frac{\zeta(2n)}{2n+s-1}, \end{aligned} \quad (7.6.63)$$

- (a) One can use either of (7.6.62) or (7.6.63) to efficiently evaluate zeta at each of  $N$  arguments  $\{s, s+2, s+4, \dots, s+2(N-1)\}$  for any complex  $s$ . This approach is fruitful for obtaining a set of zeta values at odd positive integers, for example. The idea is to exploit the recursion relation (7.6.61) for the incomplete gamma function and,

when  $N$  is sufficiently large, effectively remove incomplete gamma evaluations via  $\Gamma(\{s/2\}, x), \Gamma(\{(1-s)/2\}, x)$  where  $\{z\}$  here denotes the fractional part of  $z$ , over a collection of  $x$ -values, and then use the above recursion either backwards or forwards to rapidly evaluate series terms for the whole set of desired zeta values.

Given the initial  $\Gamma(\{s/2\}, x)$  evaluations, further effort is sharply reduced. When  $\{s + 2k\}$  are odd integers, precomputation involve only  $\Gamma(0, x)$  and  $\Gamma(1/2, x)$  values—known classically as exponential-integral and error-function values. Reference [89] contains explicit pseudocode for a recycling evaluation of  $\zeta(3), \zeta(5), \dots, \zeta(L)$  via (7.6.62), and one initializes error function and exponential-integral values by:

$$\begin{aligned} \{\Gamma(1/2, \pi n^2) : n \in [1, \lfloor D \rfloor]\}, & \quad (7.6.64) \\ \{\Gamma(0, \pi n^2) : n \in [1, \lfloor D \rfloor]\}, & \end{aligned}$$

when  $D$  decimal digits precision is ultimately desired for each  $\zeta$  value. The notion of “recycling” takes its purest form in this method, for the incomplete-gamma evaluations above are reused for every  $\zeta(\text{odd})$ .

- (b) A second approach, relevant for *even* integer arguments, involves a method of series inversion used by J. P. Buhler for numerical analysis on Fermat’s Last Theorem and on the Vandiver conjecture [70, 71]. This uses a generating function for Bernoulli numbers, and invokes Newton’s method for series inversion of the key elementary functions. To obtain values at even positive integers, one may use an expansion related to (7.6.60). One has:

$$\frac{\sinh(2\pi\sqrt{t})}{4\pi\sqrt{t}} \frac{2\pi^2 t}{\cosh(2\pi\sqrt{t}) - 1} = - \sum_{n=0}^{\infty} (-1)^n \zeta(2n) t^n,$$

which we have written this way for Algorithm 9. We have split the left-hand side into two series, each in  $t$ : one of the form  $(\sinh \sqrt{z})/\sqrt{z}$  the other like  $(\cosh \sqrt{z} - 1)/z$ . The idea is to invert the latter series via a fast polynomial inversion algorithm (Newton method). Using  $t$  as a place-holder, one reads off the  $\zeta$ -values as coefficients in a final polynomial. In Algorithm 9, we assume that  $\zeta(2), \zeta(4), \dots, \zeta(2N - 2)$  are desired. The polynomial arithmetic is

most efficient when truncation of large polynomials occurs as needed. We denote by  $q(t) \bmod t^k$  truncation of polynomial  $q$  through  $t^{k-1}$ . Also, polynomial multiplication operation is signified by “\*”.

**Algorithm 9** *Recycling scheme for  $\zeta(0), \zeta(2), \zeta(4), \dots, \zeta(2(N-1))$ .*

- (1) [Denominator setup] Create the polynomial  $f(t) = (\cosh(2\pi\sqrt{t}) - 1)/(2\pi^2t)$ , through degree  $N$  (i.e., through power  $t^N$  inclusive);
- (2) [Newton polynomial inversion, to obtain  $g := f^{-1}$ ] Set  $p = g = 1$ ;  
 while( $p < \deg(f)$ ) do begin  
      $p = \max(2p, \deg(f))$ ;  $h = f \bmod t^p$ ;  $g = (g + g * (1 - h * g)) \bmod t^p$ ; end;
- (3) [Numerator setup] Create the polynomial  $k(t) = \sinh(2\pi\sqrt{t})/(4\pi\sqrt{t})$ , through degree  $N$ ;  
 $g = g * k \bmod t^{2N-1}$ ; for  $n \in [0, 2N-2]$ , read off  $\zeta(2n)$  as  $-(-1)^n$  times the coefficient of  $t^n$  in polynomial  $g(t)$ .  $\square$

In step (1) the polynomial can have floating point or symbolic coefficients with their respective powers of  $\pi$  and so on. If used symbolically, the  $\zeta$  values of the indicated finite set are exact, through  $\zeta(2N-2)$  inclusive. The method has been used, numerically so that fast Fourier transform methods may also be applied, to calculate the relevant  $\zeta$ -values for high-precision values of the Khintchine constant [16].

If memory storage is an issue, there is a powerful technique called *multisectioning*, whereby one calculates all the  $\zeta(2k)$  for  $k$  lying in some congruence class (mod 4, 8 or 16 say), using limited memory for that calculation, then moving on to the next congruence class, and so on. Observe that, by looking only at even-indexed Bernoulli numbers in the previous algorithm, we have effectively multisectioned by 2 already. To multisection by 4, observe:

$$\frac{x \cosh x \sin x \pm x \cos x \sinh x}{\sinh x \sin x} = 2 \sum_{n \in S^\pm} \frac{B_n}{n!} (2x)^n,$$

where the sectioned sets are  $S^+ = \{0, 4, 8, 12, \dots\}$  and  $S^- = \{2, 6, 10, 14, \dots\}$ . The key is that the denominator  $(\sinh x \sin x)$  is, perhaps

surprisingly,  $x^2$  times a series in  $x^4$ , namely we have the attractive series

$$\sinh x \sin x = \sum_{n \in S^-} (-1)^{(n-2)/4} 2^{n/2} \frac{x^n}{n!}, \quad (7.6.65)$$

so that the key Newton inversion of a polynomial approximant to the denominator only has *one-fourth* the terms of the standard Bernoulli denominator  $(e^x - 1)$ . Thus, reduced memory is used to establish a congruence class of Bernoulli indices, then that memory is reused for the next congruence class, and so on. Thus, these methods function well in parallel or serial environments.

Multisectioning was used by Buhler and colleagues—as high as level-16 sections—to verify Fermat’s Last Theorem to exponent 8,000,000. They desired Bernoulli numbers modulo primes, and so employed integer arithmetic. A detailed analysis of multisectioning is to be found in Kevin Hare’s 1999 MSc. thesis [134].

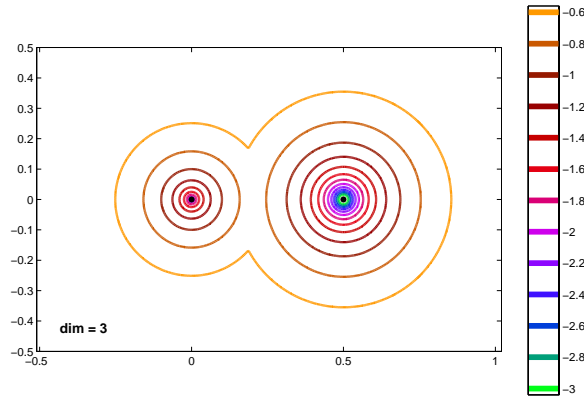
- (c) A third approach is to contemplate continued fraction representations that yield  $\zeta$ -values. For example, the well known fraction for  $\sqrt{z} \coth \sqrt{z}$  gives:

$$\frac{\pi^2 z}{3 + \frac{\pi^2 z}{5 + \frac{\pi^2 z}{7 + \dots}}} = 2 \sum_{n=1}^{\infty} (-1)^{n-1} \zeta(2n) z^n.$$

This is advantageous if one already has an efficient continued fraction engine. As a final alternative for fast evaluation at even positive integer arguments, there is an interesting approach due to Plouffe and Fee, in which the Von-Staudt-Clausen formula for the fractional part of  $B_n$  is invoked, then asymptotic techniques are used to determine the integer part. In this way the number  $B_{200,000}$  has been calculated in exact, rational form.

18. **Pseudospectra.** A fine example of the changing nature of numerical mathematics is afforded by the use of *pseudospectra*. One considers the  $\varepsilon$ -pseudospectrum of a complex  $n \times n$  matrix  $A$  for  $\varepsilon > 0$

$$\Lambda_\varepsilon(A) = \{z \in \mathbb{C} : z \in \Lambda(B), \text{ for some } \|B - A\| \leq \varepsilon\},$$

Figure 7.2: The pseudospectra of matrix  $A(0.5)$ 

which consists of the eigenvalues of all nearby matrices (in operator norm), and coincides with the spectrum of  $A$ ,  $\Lambda(a)$ , for  $\varepsilon = 0$ . This is much more stable than the spectrum and can be graphed very informatively for various values of  $\varepsilon$  as illustrated in Figure 7.2 for the matrix  $A(0.5)$ . Here,

$$A(t) = \begin{bmatrix} t & -1/4 & 0 \\ 0 & t & 0 \\ 0 & 0 & t - 1/2 \end{bmatrix} \quad A^{-1}(t) = \begin{bmatrix} t^{-1} & 1/4t^{-2} & 0 \\ 0 & t^{-1} & 0 \\ 0 & 0 & 2(2t-1)^{-1} \end{bmatrix}.$$

Thus,  $A(t)$  has eigenvalues  $t, t, t - 1/2$ , and the inverse exists for  $t \neq 0, 1/2$ . Try to correlate the graphic information with the behavior of  $A^{-1}(t)$ .

See <http://web.comlab.ox.ac.uk/projects/pseudospectra/> for a host of tools and examples, from which the picture (Figure 7.3) of a 200 dimension Toeplitz matrix with “symbol”  $t^{-5} - t$  is taken.

Similarly, the  $\varepsilon$ -spectral radius of a matrix,  $\alpha_\varepsilon(A) = \{\operatorname{Re} z : z \in \Lambda_\varepsilon(A)\}$ , is more robust than the spectral radius ( $\varepsilon = 0$ ), and is important in engineering control problems. The stability of the pseudospectrum is, in large, explained by its variational formulation.

19. **Square-full numbers and one bits.** A *square-full number*, also known as squareful or powerful number, is a natural number all of whose prime



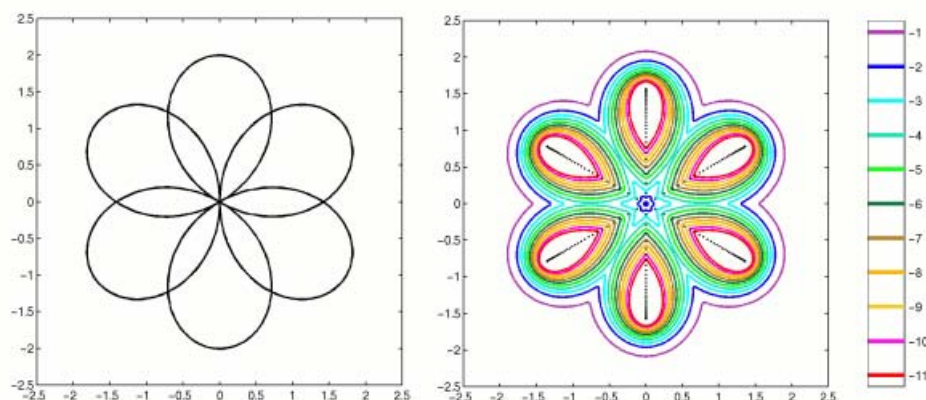


Figure 7.3: Daisy matrix pseudospectra

factors occur to multiplicity greater than one. In [17] it is shown that the number of 1's in the first  $N$  bits of the theta function value  $\sigma = \left(\sum_{n \geq 0} 2^{-n^2}\right)^2 = (\theta_3(1/2)/2 + 1/2)^2$  behaves like

$$C_0 \frac{N}{\sqrt{\log N}}, \text{ where } C_0 = \frac{16L^3}{\pi^2} \sum_g \frac{b(\tau(g))}{\psi(g)}, \quad (7.6.66)$$

and  $g$  runs over the square-full integers divisible solely by primes that are congruent to 1 (mod 4). Here  $b$  denotes the number of bits,  $\tau$  the number of divisors,  $\psi(g) = g \prod_{p|g} (1 + 1/p)$ , and  $L$  is the *Landau constant*

$$L = \left( \frac{1}{2} \prod_{p \equiv 3 \pmod{4}} \left(1 - \frac{1}{p^2}\right)^{-1} \right)^{1/2} = 0.764223653 \dots$$

- (a) Compute the predicted density (7.6.66) of ones in the binary expansion of  $\sigma$ , for  $N = 10^n$ ,  $n = 1, \dots, 8$ .
  - (b) Compare the empirical density for the same values of  $N$ .
20. **Converting series into products.** Use the ideas in Exercise 8 of Chapter 3 to produce two algorithms for converting series into products of the form in that exercise; and compare the efficiency of the two methods.

21. **Self-reference and Grelling’s paradox of heterologicality.** One of the classic linguistic (or predicate) paradoxes—which like Russell’s naive set theory paradox of all non-self containing sets—prefigures Gödel’s theorem and Chaitin’s work is as follows. Say that a word or property is *homological* if it describes itself and *heterological* otherwise. Thus, “short” is short and, arguably, “ugly” is ugly and so both are homological. By contrast “long” is not long and so is heterological. A problem occurs when we try to decide if “heterological” is a heterological property. This is usually dealt with by restricting the range of predicates or the class of sets. It would be amusing if we could show that Khintchine’s constant did not respect Khintchine’s constant.
22. **Progress and silence.** “It’s generally the way with progress that it looks much greater than it really is” is the epigraph that Ludwig Wittgenstein (1889–1951) (“whereof one cannot speak, thereof one must be silent”) had wished for an unrealized joint publication of *Tractatus Logico-Philosophicus* (1922) and *Philosophical Investigations* (1953). This suggests the two volumes are not irreconcilable as often described. Wittgenstein was one of the influential members of the Vienna Circle from which Gödel took many ideas.
23. **Chaitin on randomness.** It seems apropos to end with Greg Chaitin’s views in “The Creative Life: Science vs Art,” an article available at the URL <http://www.cs.umaine.edu/~chaitin/cdg.html>.

The message is that mathematics is quasi-empirical, that mathematics is not the same as physics, not an empirical science, but I think it’s more akin to an empirical science than mathematicians would like to admit.

Mathematicians normally think that they possess absolute truth. They read God’s thoughts. They have absolute certainty and all the rest of us have doubts. Even the best physics is uncertain, it is tentative. Newtonian science was replaced by relativity theory, and then—wrong!—quantum mechanics showed that relativity theory is incorrect. But mathematicians like to think that mathematics is forever, that it is eternal. Well, there is an element of that. Certainly a mathematical proof gives more

certainty than an argument in physics or than experimental evidence, but mathematics is not certain. This is the real message of Gödel's famous incompleteness theorem and of Turing's work on uncomputability.

You see, with Gödel and Turing the notion that mathematics has limitations seems very shocking and surprising. But my theory just measures mathematical information. Once you measure mathematical information you see that any mathematical theory can only have a finite amount of information. But the world of mathematics has an infinite amount of information. Therefore it is natural that any given mathematical theory is limited, the same way that as physics progresses you need new laws of physics. Mathematicians like to think that they know all the laws. My work suggests that mathematicians also have to add new axioms, simply because there is an infinite amount of mathematical information. This is very controversial. I think mathematicians, in general, hate my ideas. Physicists love my ideas because I am saying that mathematics has some of the uncertainties and some of the characteristics of physics. Another aspect of my work is that I found randomness in the foundations of mathematics. Mathematicians either don't understand that assertion or else it is a nightmare for them . . .



# Bibliography

- [1] Milton Abramowitz and Irene A. Stegun. *Handbook of Mathematical Functions*. Dover Publications, New York, 1970.
- [2] Manindra Agrawal, Neeraj Kayal, and Nitin Saxena. PRIMES is in P. <http://www.cse.iitk.ac.in/news/primality.pdf>, 2002.
- [3] Zafar Ahmed. Definitely An Integral. *American Mathematical Monthly*, 109:670–671, 2002.
- [4] D. Albers, G. Alexanderson, and C. Reid. *More Mathematical People*. Harcourt Brace Jovanovich, Orlando, 1990.
- [5] W. R. Alford, A. Granville, and C. Pomerance. There Are Infinitely Many Charnichael Numbers. *Annals of Mathematics*, 140:1–20, 1994.
- [6] G. Almkvist and A. Granville. Borwein and Bradley’s Apery-like Formulae for  $\zeta(4n+3)$ . *Experimental Mathematics*, 8:197–204, 1999.
- [7] G. Almkvist, C. Krattenthaler, and J. Petersson. Some New Formulas for  $\pi$ . *Manuscript*, 2001.
- [8] Gerg Almkvist and Herbert S. Wilf. On the Coefficients in the Hardy-Ramanujan-Rademacher Formula for  $p(n)$ . *Journal of Number Theory*, 50:329–334, 1995.
- [9] George E. Andrews. *The Theory of Partitions*. Addison-Wesley, Boston, 1976.
- [10] George E. Andrews. *q-Series: Their Development and Application in Analysis, Number Theory, Combinatorics, Physics and Computer Algebra*. C.B.M.S. Regional Conference Series in Math, No. 66, American Mathematical Society, Providence, 1986.

- [11] George E. Andrews, Richard Askey, and Ranjan Roy. *Special Functions*. Cambridge University Press, Cambridge, 1999.
- [12] B. H. Arnold. A Topological Proof of the Fundamental Theorem of Algebra. *American Mathematical Monthly*, 56:465–466, 1949.
- [13] B. H. Arnold and Ivan Niven. A Correction. *American Mathematical Monthly*, 58:104, 1951.
- [14] Kendall E. Atkinson. *An Introduction to Numerical Analysis*. John Wiley and Sons, Hoboken, NJ, 1989.
- [15] C. Bachoc, R. Coulangeon, and G. Nebe. Designs in Grassmannian Spaces and Lattices. *Journal of Algebraic Combinatorics*, 16:5–19, 2002.
- [16] David H. Bailey, Jonathan M. Borwein, and Richard E. Crandall. On the Khintchine Constant. *Mathematics of Computation*, 66:417–431, 1997.
- [17] David H. Bailey, Jonathan M. Borwein, Richard E. Crandall, and Carl Pomerance. On the Binary Expansions of Algebraic Numbers. <http://crd.lbl.gov/~dhbailey/dhbpapers/algebraic.pdf>, 2003.
- [18] David H. Bailey, Jonathan M. Borwein, and Roland Girgensohn. Experimental Evaluation of Euler Sums. *Experimental Mathematics*, 3:17–30, 1994.
- [19] David H. Bailey and David J. Broadhurst. A Seventeenth-Order Polylogarithm Ladder. <http://crd.lbl.gov/~dhbailey/dhbpapers/ladder.pdf>, 1999.
- [20] Alexander Barvinok. *A Course in Convexity*. American Mathematical Society, 2002.
- [21] P. Bateman. On the Representation of a Number as the Sum of Three Squares. *Transactions of the AMS*, 71:70–101, 1951.
- [22] E. Bedocchi. Nota ad una congettura sui numeri primi. *Rivista di Matematica della Università di Parma*, 11:229–236, 1985.
- [23] Michael Berry. Why Are Special Functions Special? *Physics Today*, 54(4):11–12, 2001.
- [24] Frits Beukers. A Rational Approach to Pi. *Nieuw Archief voor Wiskunde*, 5:372–379, 2000.
- [25] Rajendra Bhatia. *Fourier Series*. Hindustan Book Agency, Delhi, India, 1993.

- [26] Patrick Billingsley. *Probability and Measure*. John Wiley, Hoboken, NJ, 1986.
- [27] L.M. Blumenthal. *Theory and Applications of Distance Geometry*. Chelsea, New York, 1970.
- [28] Harold P. Boas. Majorant Series. *Journal of the Korean Mathematical Society*, 37:321–337, 2000.
- [29] D. Borwein, J. M. Borwein, P. B. Borwein, and R. Girgensohn. Giuga’s Conjecture on Primality. *American Mathematical Monthly*, 103:40–50, 1996.
- [30] D. Borwein, J. M. Borwein, and R. Girgensohn. Explicit Evaluation of Euler Sums. *Proceedings of the Edinburgh Mathematical Society*, 38:273–294, 1995.
- [31] D. Borwein, J. M. Borwein, and C. Pinner. Convergence of Madelung-Like Lattice Sums. *Transactions of the American Mathematical Society*, 350:3131–3167, 1998.
- [32] D. Borwein, J. M. Borwein, and K. Taylor. Convergence of Lattice Sums and Madelung’s Constant. *Journal of Mathematical Physics*, 26:2999–3009, 1985.
- [33] David Borwein and Jonathan M. Borwein. Some Remarkable Properties of Sinc and Related Integrals. *The Ramanujan Journal*, 5:73–90, 2001.
- [34] David Borwein, Jonathan M. Borwein, and Bernard A. Mares Jr. Multi-Variable Sinc Integrals and Volumes of Polyhedra. <http://www.cecm.sfu.ca/preprints/2001pp.html>, 2001.
- [35] J. M. Borwein and G. De Barra. Nested Radicals. *American Mathematical Monthly*, 98:735–739, 1991.
- [36] J. M. Borwein, D. M. Bradley, D. J. Broadhurst, and P. Lisoněk. Special Values of Multidimensional Polylogarithms. *Transactions of the American Mathematical Society*, 353:907–941, 2001.
- [37] Jonathan Borwein, Peter Borwein, and K. Dilcher. Pi, Euler Numbers and Asymptotic Expansions. *American Mathematical Monthly*, 96:681–687, 1989.
- [38] Jonathan Borwein and Kwok-Kwong (Stephen) Choi. On Dirichlet Series for Sums of Squares. *Ramanujan Journal*, to appear.
- [39] Jonathan Borwein, Richard Crandall, and Greg Fee. On Ramanujan’s AGM fraction. <http://www.cecm.sfu.ca/preprints/2002pp.html>, 2003.

- [40] Jonathan M. Borwein. Problem 10281. *American Mathematical Monthly*, 100:76–77, 1993.
- [41] Jonathan M. Borwein. Revivals. Problem 10281. *American Mathematical Monthly*, 103:911–912, 1996.
- [42] Jonathan M. Borwein. Solutions to Problem 10281. *American Mathematical Monthly*, 103:181–183, 1996.
- [43] Jonathan M. Borwein and David H. Bailey. *Mathematics by Experiment: Plausible Reasoning in the 21st century*. A K Peters Ltd, Natick, MA, 2003.
- [44] Jonathan M. Borwein and Peter B. Borwein. *Pi and the AGM: A Study in Analytic Number Theory and Computational Complexity*. CMS Series of Monographs and Advanced books in Mathematics, John Wiley, Hoboken, NJ, 1987.
- [45] Jonathan M. Borwein and Peter B. Borwein. A Remarkable Cubic Mean Iteration. In *Lecture Notes in Mathematics, vol. 1435: Proceedings of Computational Methods and Function Theory*, pages 27–31. Springer-Verlag, Heidelberg, 1990.
- [46] Jonathan M. Borwein and Peter B. Borwein. A Cubic Counterpart of Jacobi’s Identity and the AGM. *Transactions of the American Mathematical Society*, 323:691–701, 1991.
- [47] Jonathan M. Borwein and Peter B. Borwein. Strange Series Evaluations and High Precision Fraud. *American Mathematical Monthly*, 99:622–640, 1992.
- [48] Jonathan M. Borwein and Peter B. Borwein. On the Generating Function of the Integer Part:  $[na+b]$ . *Journal of Number Theory*, 43:293–318, 1993.
- [49] Jonathan M. Borwein, Peter B. Borwein, and Frank G. Garvan. Some Cubic Identities of Ramanujan. *Transactions of the American Mathematical Society*, 343:35–47, 1994.
- [50] Jonathan M. Borwein, Peter B. Borwein, Roland Girgensohn, and Sheldon Parnes. Experimental Mathematical Investigation of Decimal and Continued Fraction Expansions of Selected Constants. *Manuscript*, 1995.
- [51] Jonathan M. Borwein, David M. Bradley, and David J. Broadhurst. Evaluations of  $k$ -fold Euler/Zagier Sums: A Compendium of Results for



- Arbitrary  $k$ . *The Wilf Festschrift, Electronic Journal of Combinatorics*, 4(2), 1997.
- [52] Jonathan M. Borwein, David M. Bradley, David J. Broadhurst, and Petr Lisoněk. Combinatorial Aspects of Multiple Zeta Values. *Electronic Journal of Combinatorics*, 5, R38, 1998.
- [53] Jonathan M. Borwein, David M. Bradley, and Richard E. Crandall. Computational Strategies for the Riemann Zeta Function. *Journal of Computational and Applied Mathematics*, 121:247–296, 2000.
- [54] Jonathan M. Borwein, David J. Broadhurst, and Joel Kamnitzer. Central Binomial Sums, Multiple Clausen Values and Zeta Values. *Experimental Mathematics*, 10:25–41, 2001.
- [55] Jonathan M. Borwein and Kwok-Kwong Stephen Choi. On the Representations of  $xy + xz + yx$ . *Experimental Mathematics*, 9:153–158, 2000.
- [56] Jonathan M. Borwein, Kwok-Kwong Stephen Choi, and Wilfrid Pigulla. Continued Fractions as Accelerations of Series. <http://www.cecm.sfu.ca/preprints/2003pp.html>, 2003.
- [57] Jonathan M. Borwein and Roland Girgensohn. Addition Theorems and Binary Expansions. *Canadian Journal of Mathematics*, 47:262–273, 1995.
- [58] Jonathan M. Borwein and Roland Girgensohn. Solution to AMM Problem 10901. *American Mathematical Monthly*, 108, 2001.
- [59] Jonathan M. Borwein and Roland Girgensohn. Evaluations of Binomial Series. <http://www.cecm.sfu.ca/preprints/2002pp.html>, 2002.
- [60] Jonathan M. Borwein and Adrian S. Lewis. Partially-Finite Convex Programming in  $L^1$ : Entropy Maximization. *SIAM Journal on Optimization*, 3:248–267, 1993.
- [61] Jonathan M. Borwein and Adrian S. Lewis. *Convex Analysis and Nonlinear Optimization. Theory and Examples*. CMS books in Mathematics, Springer-Verlag, vol. 3, Heidelberg, 2000.
- [62] Jonathan M. Borwein and Mark A. Limber. Maple as a High Precision Calculator. *Maple News Letter*, 8:39–44, 1992.

- [63] Jonathan M. Borwein, Petr Lisoněk, and John A. Macdonald. Arithmetic-Geometric Means Revisited. *MapleTech*, pages 20–27, 1997.
- [64] Peter Borwein. An Efficient Algorithm for the Riemann Zeta Function. <http://www.cecm.sfu.ca/preprints/1995pp.html>, 1995.
- [65] Peter B. Borwein. *Computational Excursions in Analysis and Number Theory*. Springer-Verlag, CMS Books in Mathematics, Heidelberg, 2002.
- [66] Peter B. Borwein and Christopher Pinner. Polynomials with  $(0, +1, -1)$  Coefficients and Roots Close to a Given Point. *Canadian Journal of Mathematics*, 49:887–915, 1997.
- [67] Douglas Bowman and David M. Bradley. Multiple Polylogarithms: A Brief Survey. volume 291 of *Contemporary Mathematics*, pages 71–92. American Mathematical Society, 2001.
- [68] Douglas Bowman and David M. Bradley. On Multiple Polylogarithms and Certain Generalizations Possessing Periodic Argument Lists. <http://germain.umemat.maine.edu/faculty/bradley/papers/pub.html>, 2003.
- [69] Douglas Bowman and David M. Bradley. Resolution of Some Open Problems Concerning Multiple Zeta Evaluations of Arbitrary Depth. *Compositio Mathematica*, to appear.
- [70] J. Buhler, R. Crandall, R. Ernvall, and T. Metsänkylä. Irregular Primes to Four Million. *Mathematics of Computation*, 61:151–153, 1993.
- [71] J. Buhler, R. Crandall, and R. W. Sompolski. Irregular Primes to One Million. *Mathematics of Computation*, 59:717–722, 1992.
- [72] P.L. Butzer and R.J. Nessel. *Fourier Analysis and Approximation*. Birkhäuser, Basel, 1971.
- [73] A. R. Calderbank, R. H. Hardin, E. M. Rains, P. W. Shor, and N. J. A. Sloane. A Group-Theoretic Framework for the Construction of Packings in Grassmannian Spaces. *Journal of Algebraic Combinatorics*, 9:129–140, 1999.
- [74] E. Rodney Canfield. Engel’s Inequality for Bell Numbers. *Journal of Combinatorial Theory, Series A*, 72:184–187, 1995.
- [75] Julien Cassaigne and Steven R. Finch. A Class of 1-Additive Sequences and Quadratic Recurrences. *Experimental Mathematics*, 4:49–60, 1995.

- [76] K. Chandrasekharen. *Arithmetical Functions*. Springer-Verlag, Heidelberg, 1970.
- [77] Man-Duen Choi. Tricks or Treats with the Hilbert Matrix. *American Mathematical Monthly*, 90:301–312, 1983.
- [78] J. H. Conway, R. H. Hardin, and N. J. A. Sloane. Packing Lines, Planes, etc.: Packings in Grassmannian Space. *Experimental Mathematics*, 5:139–159, 1996.
- [79] J. H. Conway and N. J. A. Sloane. *Sphere Packings, Lattices and Groups*. Springer-Verlag, Heidelberg, 1999.
- [80] Shaun Cooper. Sums of Five, Seven and Nine Squares. *Ramanujan Journal*, 6:469–490, 2002.
- [81] Robert M. Corless, Gaston H. Gonnet, David E. G. Hare, David J. Jeffrey, and Donald E. Knuth. On the Lambert W function. *Advances in Computational Mathematics*, 5:329–359, 1996.
- [82] Robert M. Corless and David J. Jeffrey. Graphing Elementary Riemann Surfaces. *ACM SIGSAM Bulletin: Communications in Computer Algebra*, 32:11–17, 1998.
- [83] Robert M. Corless, David J. Jeffrey, and Donald E. Knuth. A Sequence of Series for the Lambert W function. pages 197–204, 1997.
- [84] David A. Cox, John B. Little, and Donal O'Shea. *Using Algebraic Geometry*. Springer-Verlag, Heidelberg, 1998.
- [85] David A. Cox, Donal O'Shea, and John B. Little. *Ideals, Varieties and Algorithms*. Springer-Verlag, Heidelberg, 1996.
- [86] Richard Crandall and Carl Pomerance. *Prime Numbers: A Computational Perspective*. Springer-Verlag, Heidelberg, 2001.
- [87] Richard E. Crandall. Fast Evaluation of Multiple Zeta Sums. *Mathematics of Computation*, 67:1163–1172, 1996.
- [88] Richard E. Crandall. On the Quantum Zeta Function. *Journal of Physics A: Mathematical and General*, 29:6795–6816, 1996.
- [89] Richard E. Crandall. *Topics in Advanced Scientific Computation*. Springer-Verlag, Heidelberg, 1996.

- [90] Richard E. Crandall. New Representations for the Madelung Constant. *Experimental Mathematics*, 8:367–379, 1999.
- [91] Richard E. Crandall and Joseph P. Buhler. Elementary Function Expansions for Madelung Constants. *Journal of Physics A: Mathematics and General*, 20:5497–5510, 1987.
- [92] Richard E. Crandall and Joseph P. Buhler. On the Convergence Problem for Lattice Sums. *Journal of Physics A: Mathematics and General*, 23:2523–2528, 1990.
- [93] Richard E. Crandall and J. F. Delord. The Potential within a Crystal Lattice. *Journal of Physics A: Mathematics and General*, 20:2279–2292, 1987.
- [94] Richard E. Crandall, Karl Dilcher, and Carl Pomerance. A Search for Wieferich and Wilson Primes. *Mathematics of Computation*, 66:433–449, 1997.
- [95] Richard E. Crandall and Jason S. Papadopoulos. On the Implementation of AKS-Class Primality Tests.  
<http://developer.apple.com/hardware/ve/acgresearch.html>, 2003.
- [96] D. P. Dalzell. On  $22/7$  and  $355/113$ . *Eureka*, 34:10–13, 1971.
- [97] I. Damgard, P. Landrock, and C. Pomerance. Average Case Error Estimates for the Strong Probable Prime Test. *Mathematics of Computation*, 61:177–194, 1993.
- [98] Ingrid Daubechies and Jeffrey C. Lagarias. Two-Scale Difference Equations. I. Existence and Global Regularity of Solutions. *SIAM Journal of Mathematical Analysis*, 22:1338–1410, 1991.
- [99] Ingrid Daubechies and Jeffrey C. Lagarias. Two-Scale Difference Equations. II. Local Regularity, Infinite Products of Matrices, and Fractals. *SIAM Journal of Mathematical Analysis*, 23:1031–1079, 1992.
- [100] P. Delsarte, J.-M. Goethals, and J. J. Seidel. Spherical Codes and Designs. *Geometriae Dedicata*, 6:363–388, 1977.
- [101] John E. Dennis Jr. and Victoria Torczon. Direct Search Methods on Parallel Machines. *SIAM Journal of Optimization*, 1(1):448–474, 1991.
- [102] Keith Devlin. *Mathematics: The Science of Patterns: The Search for Order in Life, Mind and the Universe*. W. H. Freeman, New York, 1997.

- [103] Leonard Eugene Dickson. *History of the Theory of Numbers*, volume 1–3. Chelsea Publishing (1952), Carnegie Institute (1991), American Mathematical Society, Providence, 1999.
- [104] Tomslav Doslic and Darko Velian. Calculus Proofs of Some Combinatorial Identities. *Matheamtical Inequalities and Applications*, 6:197–209, 2003.
- [105] John Duncan and Colin M. McGregor. Carleman’s Inequality. *American Mathematical Monthly*, 110:03:424–430, 2003.
- [106] William Dunham. *The Mathematical Universe*. John Wiley, Hoboken, NJ, 1994.
- [107] William Dunham. *Euler: The Master of Us All*. Mathematical Association of America (Dolciani Mathematical Expositions), Washington, 1999.
- [108] Freeman Dyson. Review of Nature’s Numbers by Ian Stewart. *American Mathematical Monthly*, 103:612, 1996.
- [109] Harold M. Edwards. *Riemann’s Zeta Function*. Academic Press, New York, 1974.
- [110] Noam D. Elkies. On the Sums  $\sum_{k=-\infty}^{\infty} (4k + 1)^{-n}$ . *American Mathematical Monthly*, 110:561–573, 2003.
- [111] Paul Erdős. On a Family of Symmetric Bernoulli Convolutions. *Transactions of the American Mathematical Society*, 61:974–976, 1939.
- [112] Philippe Flajolet and Bruno Salvy. Euler Sums and Contour Integral Representations. *Experimental Mathematics*, 7:15–34, 1997.
- [113] M. K. Fort Jr. Some Properties of Continuous Functions. *American Mathematical Monthly*, 59:372–375, 1952.
- [114] L. Gårding. An Inequality for Hyperbolic Polynomials. *Journal of Mathematics and Mechanics*, 8:957–965, 1959.
- [115] Adriano Garsia. Arithmetic Properties of Bernoulli Convolutions. *Transactions of the American Mathematical Society*, 102:409–432, 1962.
- [116] Roland Girgensohn. Functional Equations and Nowhere Differentiable Functions. *Aequationes Mathematicae*, 46:243–256, 1993.

- [117] Roland Girgensohn. Nowhere Differentiable Solutions of a System of Functional Equations. *Aequationes Mathematicae*, 47:89–99, 1994.
- [118] Roland Girgensohn. A Unified View of Classical Summation Kernels. *Resultate der Mathematik (Results in Mathematics)*, 38:48–57, 2000.
- [119] G. Giuga. Su una presumibile proprietà caratteristica dei numeri primi. *Rendiconti Scienze Matematiche e Applicazioni A*, 83:511–528, 1950.
- [120] J.-M. Goethals and J. J. Seidel. Spherical designs. volume 34, pages 255–272. American Mathematical Society, 1979.
- [121] J.-M. Goethals and J. J. Seidel. Cubature Formulae, Polytopes and Spherical Designs. pages 203–218. Springer-Verlag, Heidelberg, 1981.
- [122] J.-M. Goethals and J. J. Seidel. The Football. *Nieuw Archief Wisk*, 29:50–58, 1981.
- [123] Daniel Goldston and Cem Yildirim. Small Gaps Between Primes. <http://www.aimath.org/goldston.tech>, 2003.
- [124] R. William Gosper. Unpublished Research Announcement. *Manuscript*, 1974.
- [125] I. S. Gradshteyn and I. M. Ryzhik. *Table of Integrals, Series, and Products, Fifth Edition*. Academic Press, New York, 1994.
- [126] Ronald L. Graham, Donald E. Knuth, and Oren Patashnik. *Concrete Mathematics*. Addison-Wesley, Boston, 1994.
- [127] Francois Guénard and Henri Lemberg. *La Méthode Expérimentale en Mathématiques*. Scopos, Springer-Verlag, Heidelberg, 2001.
- [128] O. Guler. Hyperbolic Polynomials and Interior Point Methods for Convex Programming. *Mathematics of Operations Research*, 22:350–377, 1997.
- [129] Richard K. Guy. Aliquot Sequences. In *Unsolved Problems in Number Theory*, pages 60–62. Springer-Verlag, Heidelberg, 1994.
- [130] R. H. Hardin and N. J. A. Sloane. New Spherical 4-Designs. *Discrete Mathematics*, 106/107:255–264, 1992.
- [131] R. H. Hardin and N. J. A. Sloane. McLaren’s Improved Snub Cube and Other New Spherical Designs in Three Dimensions. *Discrete Computational Geometry*, 15:429–441, 1996.

- [132] Godfrey H. Hardy. *Ramanujan*. Chelsea, New York, 1978.
- [133] Godfrey H. Hardy and Edward M. Wright. *An Introduction to the Theory of Numbers*. Oxford University Press, Oxford, 1985.
- [134] Kevin Hare. Multisectioning, Rational Poly-exponential Functions and Parallel Computation. <http://www.cecm.sfu.ca/preprints/1999pp.html>, 1999.
- [135] John Herivel. *Joseph Fourier: The Man and the Physicist*. Oxford University Press, Oxford, 1992.
- [136] J. Herman, R. Kučera, and J. Šimša. *Equations and Inequalities: Elementary Problems and Theorems in Number Theory*, volume 1. CMS Books, Springer-Verlag, Heidelberg, 2000.
- [137] J. Holt. Mistaken Identity Theory: Why Scientists Always Pick the Wrong Man. *Lingua Franca*, page 60, March 2000.
- [138] R. Hooke and T. A. Jeeves. Direct Search Solution of Numerical and Statistical Problems. *Journal of the Association for Computing Machinery*, 8:212–229, 1961.
- [139] John H. Hubbard, Jean Marie McDill, Anne Noonburg, and Beverly H. West. A New Look at the Airy Equation with Fences and Funnels. In *The Organic Mathematics Project Proceedings*, volume 20, Ottawa, 1997. Canadian Mathematical Society.
- [140] Borge Jessen and Aurel Wintner. Distribution Functions and the Riemann Zeta Function. *Transactions of the American Mathematical Society*, 38:48–88, 1935.
- [141] G. S. Joyce and I. J. Zucker. Evaluation of the Watson integral and Associated Logarithmic Integral for the  $d$ -Dimensional Hypercubic Lattice. *Journal of Physics A*, 34:7349–7354, 2001.
- [142] G. S. Joyce and I. J. Zucker. On the Evaluation of Generalized Watson Integrals. *Proceedings of the AMS*, to appear.
- [143] Anatolij A. Karatsuba. *Basic Analytic Number Theory*. Springer-Verlag, Heidelberg, 1993.
- [144] Yitzhak Katznelson. *An Introduction to Harmonic Analysis*. Wiley, Hoboken, NJ, 1968.

- [145] Akio Kawauchi. *A Survey of Knot Theory*. Birkhäuser Verlag, Basel, 1996.
- [146] R. Kershner and A. Wintner. On Symmetric Bernoulli Convolutions. *American Journal of Mathematics*, 57:541–548, 1935.
- [147] Donald E. Knuth. *The Art of Computer Programming*, volume 2. Addison-Wesley, Boston, 1998.
- [148] Pavel P. Korovkin. On Convergence of Linear Positive Operators in the Space of Continuous Functions. *Doklady Akademii Nauk. Rossijskaya Akademiya Nauk*, 90:961–964, 1953.
- [149] George Lamb. Problem 11000. *American Mathematical Monthly*, 110:240, 2003.
- [150] Andrew Lasota and Michael C. Mackey. *Chaos, Fractals and Noise: Stochastic Aspects of Dynamics*. Springer-Verlag, Heidelberg, 1994.
- [151] H. W. Lenstra Jr. Primality Testing with Cyclotomic Rings. *preprint*, 2002.
- [152] A. S. Lewis, P. A. Parrilo, and M. V. Ramana. The Lax Conjecture Is True. *Proceedings of the American Mathematical Society*, to appear.
- [153] A. S. Lewis and H. S. Sendov. Self-Concordant Barriers for Hyperbolic Means. *Mathematical Programming*, 91, Ser. A:1–10, 2001.
- [154] Peter Lisoněk and Robert Israel. Metric Invariants of Tetrahedra via Polynomial Elimination. *Proceedings of the 2000 International Symposium on Symbolic and Algebraic Manipulation*, pages 217–219, 2000.
- [155] John E. Littlewood. *Littlewood's Miscellany*. Cambridge University Press, Cambridge, second edition, edited by Béla Bollobás edition, 1988.
- [156] I. G. Macdonald. *Symmetric Functions and Hall Polynomials*. Oxford University Press, Oxford, 1995.
- [157] K. Mahler. Arithmetische Eigenschaften der Lösungen einer Klasse von Funktionalgleichungen. *Mathematik Annalen*, 101:342–366, 1929.
- [158] George Marsaglia and John C. Marsaglia. A New Derivation of Stirling's Approximation to  $n!$ . *American Mathematical Monthly*, 97:826–829, 1990.
- [159] R. Daniel Mauldin and Karoly Simon. The Equivalence of Some Bernoulli Convolutions to Lebesgue Measure. *Proceedings of the American Mathematical Society*, 126:2733–2736, 1998.



- [160] M. Mazur. Problem 10717. *American Mathematical Monthly*, 106:167, 1999.
- [161] K. I. M. McKinnon. Convergence of the nelder-mead simplex method to a nonstationary point. *SIAM Journal on Optimization*, 9:148–158, 1998.
- [162] J. Milnor. On the Betti Numbers of Real Algebraic Varieties. *Proceedings of the American Mathematical Society*, 15:275–280, 1964.
- [163] L. Monier. Evaluation and Comparison of Two Efficient Probabilistic Primalty Testing Algorithms. *Theoretical Computer Science*, 12:97–108, 1980.
- [164] M. Mori. Developments in the Double Exponential Formula for Numerical Integration. In *Proceedings of the International Congress of Mathematicians*, pages 1585–1594, Heidelberg, 1991. Springer-Verlag.
- [165] David Mumford, Caroline Series, and David Wright. *Indra's Pearls: The Vision of Felix Klein*. Cambridge University Press, Cambridge, 2002.
- [166] J. A. Nelder and R. Mead. A simplex method for function minimization. *The Computer Journal*, 7:308–313, 1965.
- [167] Ivan Niven, Herbert S. Zuckerman, and Hugh L. Montgomery. *An Introduction to the Theory of Numbers*. John Wiley, Hoboken, NJ, 1991.
- [168] Jorge Nocedal and Stephen J. Wright. *Numerical Optimization*. Springer-Verlag, Heidelberg, 1999.
- [169] Yasuo Ohno. A Generalization of the Duality and Sum Formulas on the Multiple Zeta Values. *Journal of Number Theory*, 74:39–43, 1999.
- [170] Frank W. J. Olver. *Asymptotics and Special Functions*. Academic Press, New York, 1974.
- [171] Yuval Peres, Wilhelm Schlag, and Boris Solomyak. Sixty Years of Bernoulli Convolutions. *Manuscript*, 2001.
- [172] Yuval Peres and Boris Solomyak. Absolute Continuity of Bernoulli Convolutions, a Simple Proof. *Mathematical Research Letters*, 3:231–239, 1996.
- [173] Kenneth A. Perko. On the Classifications of Knots. *Proceedings of the American Mathematical Society*, 45:262–266, 1974.
- [174] Oskar Perron. *Die Lehre von den Kettenbrüchen*. Chelsea, New York, 1950.

- [175] William H. Press, Brian P. Flannery, Saul A. Teukolsky, and William T. Vetterling. *Numerical Recipes*. Cambridge University Press, Cambridge, 1986.
- [176] P. Pritchard, A. Moran, and A. Thyssen. Twenty-Two Primes in Arithmetic Progression. *Mathematics of Computation*, 64:1337–1339, 1995.
- [177] A.P. Prudnikov, Yu.A. Brychkov, and O.I. Marichev. *Integrals and Series* (Volumes 1 and 2). Taylor and Francis, London, 1986.
- [178] M. Rabin. Probabilistic Algorithms for Testing Primality. *Journal of Number Theory*, 12:128–138, 1980.
- [179] Bruce Reznick. *Sums of Even Powers of Real Linear Forms*. American Mathematical Society, 1992.
- [180] Bruce Reznick. Some Constructions of Spherical 5-Designs. *Linear Algebra and Its Applications*, 226/228:163–196, 1995.
- [181] Paulo Ribenboim. *The Little Book of Big Primes*. Springer-Verlag, Heidelberg, 1991.
- [182] Bernhard Riemann. Über die Anzahl der Primzahlen unter einer Gegebenen Grösse. In Hermann Weyl, editor, *Das Kontinuum und Andere Monographien*. Chelsea Publishing Co., New York, 1972.
- [183] John Riordan. *Combinatorial Identities*. John Wiley, Hoboken, NJ, 1968.
- [184] Sara Robinson. New Method Said to Solve Key Problem In Math. *New York Times*, Aug. 8 2002.
- [185] Dale Rolfsen. *Knots and Links*. Publish or Perish, Inc., Houston, 1976.
- [186] Raphael Salem. A Remarkable Class of Algebraic Integers, Proof of a Conjecture by Vijayaraghavan. *Duke Mathematical Journal*, 11:103–108, 1944.
- [187] Byron Schmuland. Random Harmonic Series. *American Mathematical Monthly*, 110:407–416, 2003.
- [188] P. D. Seymour and T. Zaslavsky. Averaging Sets: A Generalization of Mean Values and Spherical Designs. *Advances in Mathematics*, 52:213–240, 1984.
- [189] P. W. Shor and N. J. A. Sloane. A Family of Optimal Packings in Grassmannian Manifolds. *Journal of Algebraic Combinatorics*, 7:157–163, 1998.

- [190] Carl L. Siegel. Über riemanns nachlaß zur analytischen Zahlentheorie. *Quellen Studien zur Geschichte der Math. Astron. und Pys.*, 2:45–80, 1932.
- [191] L. E. Sigler. *Fibonacci's Liber Abaci: A Translation into Modern English of Leonardo Pisano's Book of Calculation*. Springer-Verlag, Heidelberg, 2002.
- [192] N. J. A. Sloane, R. H. Hardin, and P. Cara. Spherical Designs in Four Dimensions. In *Proceedings of the Workshop on Information and Coding*. IEEE Press, New York, 2003.
- [193] N. J. A. Sloane, R. H. Hardin, and P. Cara. Spherical Designs in Three and Four Dimensions. *preprint*, 2003.
- [194] N. J. A. Sloane and Simon Plouffe. *The Encyclopedia of Integer Sequences*. Academic Press, New York, 1995.
- [195] Boris Solomyak. On the Random Series  $\sum \pm \lambda^i$  (an Erdős Problem). *Annals of Mathematics*, 142:611–625, 1995.
- [196] Richard Stanley. *Enumerative Combinatorics*, volume 1 and 2. Cambridge University Press, Cambridge, 1999.
- [197] Karl R. Stromberg. *An Introduction to Classical Real Analysis*. Wadsworth, Thompson Learning, Florence, KY, 1981.
- [198] Hidetosi Takahasi and Masatake Mori. Double Exponential Formulas for Numerical Integration. *Publications of RIMS, Kyoto University*, 9:721–741, 1974.
- [199] Gary Taubes. The (Political) Science of Salt. *Science*, pages 898–907, Aug. 1998.
- [200] Radu Theodorescu. Problem 10738. *American Mathematical Monthly*, 106:471, 1999.
- [201] Edward C. Titchmarsh. *The Theory of the Riemann Zeta-Function*. Oxford University Press, 1951.
- [202] Virginia Torczon. *Multi-Directional Search: A Direct Search Algorithm for Parallel Machines*. PhD thesis, Rice University, May 1989.
- [203] Virginia Torczon. On the convergence of pattern search algorithms. *SIAM Journal on Optimization*, 7:1–25, Feb. 1997.

- [204] Michal Trott. Visualization of Riemann Surfaces of Algebraic functions. *Mathematica in Education and Research*, 6(4):15–36, 1997.
- [205] Mitsuru Utchiyama. Proofs of the Korovkin Theorems via Inequalities. *American Mathematical Monthly*, 110:334–336, 2003.
- [206] Alfred J. van der Poorten and Xuan Chuong Tran. Quasi-Elliptic Integrals and Periodic Continued Fractions. *Monatshefte fr Mathematik*, 131:155–169, 2000.
- [207] H. S. Wall. *Analytic Theory of Continued Fractions*. Chelsea, New York, 1948.
- [208] George N. Watson. The Final Problem: An Account of the Mock Theta Functions. *Journal of the London Mathematical Society*, 11.
- [209] Aurel Wintner. On Convergent Poisson Convolutions. *American Journal of Mathematics*, 57:827–838, 1935.
- [210] Benjamin Yandell. *The Honors Class: Hilbert's Problems and Their Solvers*. A K Peters, Natick, MA, 2001.
- [211] Don Zagier. The First 50 Million Prime Numbers. *The Mathematical Intelligencer*, 0:7–19, 1977.
- [212] D. Zeilberger. Identities in Search of Identity. *Theoretical Computer Science*, 117:23–38, 1993.
- [213] Antoni Zygmund. *Trigonometric Series*. Cambridge University Press, Cambridge, 1959.

# Index

- Abel, Niels Henrik, 19, 56
  - Abel summation, 97, 123
- Adamchik, Victor, iii
- additive partitions, *see* partitions
- additive sequences, 267
- affine variety, *see* variety
- Affleck, Ian, 42
- AGM iteration
  - complex, 235
- Agoh's conjecture, 252
- Agrawal, Manindra, 333
- Ahmed, Zafar, 19
- Airy, George
  - Airy function, 68
  - Airy integral, 67
- Aliquot sequences, 268
- amicable numbers, 267–268
- Andrew's convolution, 208, 230
- Apéry, Roger, 149
- Archimedes, 3, 314
- Argument principle, 302
- arithmetic-geometric mean (AGM),
  - 209, 290–292
  - $AG_2$ , 318
  - generalized, 289–295
- Aronson's sequence, 65
- Artjuhov, Mikhail, 332
  
- Banach contraction theorem, 297–298
- barrier functions, 276–279
- Bauschke, Heinz, iii
  
- BBP-type formulas, 55, 347, 349
- Beatty's theorem, 66
- Bell numbers, 47, 65
  - Bell polynomials, 65
- Bender, Carl, 357
- Benson's formula, 229
- Bernoulli convolution, 102–106
  - Bernoulli measure, 102
- Bernoulli numbers, 34, 145, **147**, 148,
  - 252, 341, 347, 357, 359
  - continued fractions, 357
- Bernoulli, Daniel, 79
- Bernstein, Daniel, 333
- Bernstein, Sergi
  - Bernstein operators, 114
  - Bernstein polynomials, 140–141
- Berry, Michael, 1Q, 48
- Bertrand's postulate, 187
- Bescovitch, A.S., 286
- Bessel functions, 40, 130, 184, 221
  - modified, 219
- beta function, 136
  - derivatives, 158–159
  - digamma representation, 159
  - Riemann zeta function, 186
- Beukers, Frits, 3
- Bézier curves, 140
- Bézout's theorem, 262
- Binet's mean value theorem, 135
- binomial series, 316
  - evaluating, 23–32

- Bohr, Harald, 321  
 Boltzmann-Shannon entropy, 299, 301  
 Bolzano, Bernhard, 121  
 Boole summation formula, 34, 349, 359  
 Borel, Emile, 73  
 Borwein, Peter, iii  
 bounded variation, 83  
 Bradley's formula, 349  
 Bradley, David, iii  
 Broadhurst, David, 55, 156, 160, 177  
 Brouncker, William, 35  
 Brouwer, Luitzen Egbertus Jan,  
   325–328  
     Brouwer's fixed point theorem,  
       326  
 Buchberger's algorithm, 257  
 Canfield, E. Rodney, 47  
 Cantor, Georg, 80  
   Cantor function, 103  
 Carathéodory, Constantin, 201  
 Carleman's inequality, 313  
 Carmichael numbers, **246**, 248, 265  
 Cassaigne, Julien, 267  
 Casteljau's algorithm, 140  
 Catalan, Eugène  
   Catalan numbers, 59, 232  
   Catalan zeta function, **153**, 218  
   Catalan's constant, 23, 32, 117,  
     186, 218, 338, 347, 359  
     evaluating, 348  
 Cauchy, Augustin-Louis  
   Cauchy convolution, 204, 213  
   Cauchy-Lindelöf theorem, 239,  
     303–305  
 Cayley, Arthur, 208  
   Cayley-Hamilton theorem, 226  
   Cayley-Menger determinant, 254,  
     **255**  
   Césaro summation, 54, 83, 96, 123  
 Chaitin, Gregory, iii, 366, 367Q  
 Champernowne number, 53  
 Chandrasekharan, K., 124  
 Chang, Jen, iii  
 Chebyshev polynomial, 51  
 Chevyshev, Pafnuty, 56  
 Chudnovsky, David and Gregory, iii,  
   354  
 Clausen, Thomas, 148  
   Clausen functions, 31  
   Clausen's product, 56, 139  
   multiple Clausen values, 178  
 Cohen, Henri, 268  
 computer algebra packages  
   bugs, 109  
   integral evaluation, 19–23  
   product evaluation, 5–8  
 conjugate function, *see* Fenchel-Young  
   conjugate  
 continued fractions, 13–17, 32–41  
   hypergeometric functions, 35  
   numerical accuracy, 13  
 contraction map, 297  
 Corless, Robert, iii  
 Crandall, Richard, iii, 253, 331, 334,  
   342  
   Madelung's constant, 221, 229  
 d'Alembert, Jean, 79  
 DAD problems, 300, 324–325  
 Daubechies, Ingrid, 103  
 David Parker's DPGraph, 280  
 de la Vallée Poussin, 266  
 Dedekind  $\eta$ -function, 206  
 Deligne words, 178  
 Devlin, Keith, 202Q  
 diagonalization algorithm, 300  
 difference equations, 103

- digamma function, 71
- Dijk paths, 231
- dilogarithms, 55
- Dirichlet, Lejeune, 79, 266
  - Dirichlet class number, 155
  - Dirichlet kernel, 115, 119
  - Dirichlet operator, 115
  - Dirichlet series, 154
    - primitive L-series, 153
    - primitive L-series, 188
- doubly stochastic matrix, 300, 324
  - doubly stochastic pattern, 300
- Drinfeld-Deligne conjecture, 157
- Dunham, William, 74
- Dyson, Freeman, 143Q
  
- Ehrhardt polynomial, 282
- eigenvector, 299
- Einstein, Albert, 99
- Ekeland variational principle, 295
  - Banach contraction, 297, 298
- Elkies, Noam D., 189
- elliptic integrals, 129, 131, 209, 313–314
  - mean iteration, 311
- Eratosthenes
  - sieve of Eratosthenes, 331
- Erdős, Paul, 104, 151, 207
- error function, 116
- Euclidean algorithm, 335
- Euler, Leonhard, 74
  - Basel problem, 75
  - continued fractions, 34–35, 40
  - Euler function, 359
  - Euler numbers, 33, 153
  - Euler polynomials, 359
  - Euler product, 146
  - Euler sums, 155, 157
    - double Euler sums, 164, 167, 224
    - unit Euler sum, 168
- Euler's beta integral, 311
- Euler's integral, 314
- Euler's pentagonal number theorem, 203
- Euler's totient function, 60, 270
- Euler-Maclaurin summation
  - formula, 33, 183, 341–342, 344, 347
  - infinite series, 347–349
  - MZVs, 349–350
- zeta function, 147
- Eureka* (journal), 3
  
- Faber, Georg, 126
  - Faber-Schauder basis, *see* Schauder basis
- face-centered cubic (FCC), 132
- fast Fourier transform (FFT), 94, 334, 335
  - convolutions, 94
- Fateman, Richard, iii
- Fatou's condition, 98
- Fee, Greg, iii
- Fejér's series, 99
- Fejér, Lipót
  - Fejér operator, 115
  - Fejér's sum, 94
  - Fejér's theorem, 93, 94
- Fenchel, Werner
  - Fenchel duality, 134
  - Fenchel-Young conjugate
    - Lambert  $W$  function, 309
- Ferguson, Helaman, iii
- Fermat, Pierre de, 267
  - Fermat primes, 253
  - Fermat's little theorem, 245
- Ferrer diagram, 202
- Feynman, Richard, 329Q

- Fibonacci, Leonardo, 221Q, 243–244
  - Fibonacci numbers, 237
  - Fibonacci sequence, 221–224, 243
- Finch, Steven, iii, 267
- fixed point, 297
  - Banach contraction theorem, 297
- Fourier series, **80**, 80–86
  - approximate identity, 95
  - Bernoulli convolution, 102–106
  - convergence, 82
    - $L^1(\mathbb{T})$ , 84
    - pointwise, 82, 84
    - uniform, 82
  - convolution theorems, 93–94
  - Hamburger moment problem, 101
  - Jordan test, 83, 98
  - kernels
    - Blackman kernel, 125
    - $(C, \alpha)$ -kernel, 125
    - de la Vallée-Poussin kernel, 124
    - definition, 95
    - Dirichlet kernel, 97, 115, 119
    - Fejér kernel, 83, 87, 96, 115, 124
    - Fejér-Korovkin kernel, 125, 141
    - Gauss kernel, 87
    - Jackson kernel, 125, 141
    - Poisson kernel, 141
  - Parseval's equation, 85
  - pathologies, 118
  - periodization, 90
  - Poisson summation formula, 90
- Fourier transform, **87**, 87–89
  - Jordan's theorem, 89
  - $L^1(\mathbb{R})$ , 87
  - $L^2(\mathbb{R})$ , 88, **88**
  - Riemann-Lebesgue lemma, 87
  - uniqueness theorem, 87
- Fourier, Joseph, 77Q, 78–80
- Fubini's theorem, 20, 110, 116
- Fundamental Theorem of Algebra, 327–328
- Galway, William, iii
- Gamma function, 74, 129, 141, 217, 316–317
  - $AG_2$ , 318
  - elliptic integrals, 213
  - identities, 181
  - incomplete, 360
  - infinite products, 5–8
  - recycling, 359
  - Riemann zeta function, 145
  - Selberg integral, 311
  - sinc integral, 135
  - Zagier conjecture, 176
- Garsia, A., 105
- Gauss, Carl Friedrich, 220
  - arithmetic-geometric mean, 291
  - continued fractions, 35–40
  - Gaussian quadrature, 338
- Gersgorin circles, 238
- Gibbs, J. Willard
  - Gibbs phenomenon, 98–99
- Giuga, G., 245
  - Giuga number, 247
  - Giuga numbers, 248
  - Giuga sequence, 251, **251**, 252
  - Giuga sequences, 268
  - Giuga's prime number conjecture, 245
- Gödel, Kurt, 366
- Goldston, Daniel, 285
- Google, 47
- Gosper, Bill, 69, 151
- Graham, Ronald, iii
- Granville, Andrew, iii, 286
- Grassman manifolds, 264



- Gregory, James  
   Gregory series, 359  
   Gregory's series, **33**, 37, 347  
 Grelling's paradox of heterologicality, 366  
 Gröbner basis, 254, 257–262  
   grevlex basis, 258  
  
 Haar basis, 126  
 Hadamard, Jacques, 326  
   Hadamard inequality, 63  
 Haenel, Christoph, iii  
 Halley, Edmond, 356  
 Hamburger moment problem, 101  
 Hankel, Hermann  
   Hankel loop integral, 317  
   Hankel matrix, 273  
 Hardy, G. H., 121, 206, 212, 219, 285, 286  
 Hata, 4  
 Heron's formula, 255  
 heterological property, 366  
 Hida, Yozo, iii  
 Hilbert, David, iQ, 146  
   Hilbert matrix, 273  
   Hilbert's theorem, 281  
 Hjab, Omar, 43  
 Hoch and Stern information measure, 301  
 Hölder convolution, 171–172  
 homological property, 366  
 Hooke and Jeeves algorithm, *see*  
   pattern search algorithm  
 Horner's rule, 334  
 Hurwitz, Adolf, 272  
 hyperbolic polynomials, 277  
 hypergeometric function, 41, 56, 129, 232, 314  
  
 arithmetic-geometric mean, 291, 292  
 contiguity relation, 35  
 continued fractions, 35–40  
 Euler's integral, 314  
 Gamma function, 315  
   multiple, 314  
 Zagier conjecture, 174, 176  
  
 ideal, 257  
 integer polytope, 282  
 integer relation algorithms, 8, 26, 31, 149, 150, 152, 155–157, 161, 178, 224  
 intuitionism, 326  
 inverse problems, 280  
 isoperimetric inequality, 128  
  
 Jacobi, Carl  
   Jacobi theta-functions, 214  
   Jacobi's quartic identity, 232  
   Jacobi's triple product, 204, 207  
   Jacobian theta functions, **207**, 207–213  
 Jeffrey, David, iii  
 Jensen's inequality, 45  
 Johnson, Samuel, 48  
 Jordan, Camille  
   Jordan test, 83, 98  
   Jordan theorem  
     sinc integrals, 110  
   Jordan's theorem, 89  
 Joyce, G. S., 130  
 Joyce, Jeff, iii  
  
 Kayal, Neeraj, 333  
 Kershner, R., 103  
 Khintchine constant, 362, 366  
 Kirchhoff's law, 315  
 Klein, Felix, 237

- Klein, James D., 73  
 knot theory, 198  
 Knuth, Donald E., 17, 122  
 Koecher, Max, 150  
 Kolmogoroff's 0-1-law, 102  
 Korelt, 265  
 Korovkin, Pavel, 97, 113  
     complex theorem, 140  
     three function theorems, 113–115,  
         139–140  
 Korselt's criterion, 246  
 Kreimer, Dirk, 160  
 Kummer-type polylogarithms, *see*  
     polylogarithms
- l'Hôpital's rule, 117  
 Lagarias, Jeffrey C., 103  
 Lagrange, Joseph-Louis, 78, 215  
 Laguerre, Edmond, 213  
 Lambert  $W$  function, 18, 239, 301,  
     305–310  
     convex conjugates, 309  
     Maclaurin series, 320  
     Riemann surfaces, 307–308  
     Stirling's formula, 306  
 Lambert series, **213**  
 Landau constant, 365  
 Laplace, Pierre-Simon  
     Laplace transform, 217  
     Laplace's method, 317  
 Laurent series, 303  
 Lax conjecture, 279  
 Lebesgue, Henri, 80, 118  
     Lebesgue constants, 96, 97  
     Lebesgue's condition, 98  
 Legendre, Adrien-Marie  
     Legendre polynomial, 339  
     Legendre's relation, 313  
     Legendre-Jacobi symbol, 154, 186,  
         218  
 Lehmer, Derrick, 253Q, 268  
     Lehmer's conjecture, 253  
     Lehmer's polynomial, 356  
     Lehmer's problem, 356  
 Lenstra, Hendrik, 333, 334  
 level set, compact, 296  
 Lewis, Adrian, iii  
 Li, Xiaoye, iii  
 Liouville, Joseph, 314  
 Lisonek, Petr, iii  
 Lisoněk, Petr, 177, 254  
 Littlewood, John Edensor, 276,  
     285–287  
 LLL, 197  
 Locke, John, 289Q  
 log-concavity, 59  
 log-convex sequence, 46  
 Lucas, Edouard, 222  
     Lucas numbers, 237  
     Lucas sequence, 222  
     Lucas-Carmichael numbers, *see*  
         Carmichael numbers  
 Luke, Russell, iii
- Macdonald, John A., 293  
 Machin formula, 37  
 Macklem, Mason, iii  
 MacMahon, Percy, 202  
 Madelung's constant, 216, 219–221,  
     229, 253  
 Magma, 255  
 Mahler, Kurt, 3  
     Mahler continued fraction, 66  
     Mahler's generating function, 65  
 Maier, Helmut, 285  
 Mares, Bernard, 113

- math software, *see* computer algebra packages
- MathSciNet, 47
- maximum entropy methods, 299–301
- maximum principle, 302
- mean iteration
  - $\frac{2}{3}$ -power mean, 311
  - logarithmic mean, 311
- Mellin transform, 128, 135, **144**, 181, 184, 217
- Menshow, D. E., 120
- Mersenne, Marin, 267
- Michelson, Albert, 99
- Milnor, John, 262
- Möbius, August
  - Möbius function, 213
  - Möbius inversion theorem, 208, 213–216
  - sums of squares, 214–216
- modular equations, 209
- moment, of a function, 101
- Moore, Gordon
  - Moore’s Law, i
- Mori, Masatake, 344
- Morin, Mathew, iii
- Motzkin, Theodore, 281
  - Motzkin numbers, 60, 65
- multidirectional search methods, 321
- multiple Clausen values, *see* Clausen
- multiple zeta constants, 350
- multiple zeta values, *see* MZV
- multisectioning, 362
- Mumford et al
  - symmetry and chaos, 235
- Mumford, David, iii, 237Q
- MZV, **155**, 157, 349–350
  - duality, 167–172
  - evaluating, 187
  - reducibility, 157–161
- Nash equilibrium theorem, 327
- neg-entropy, *see* Hoch and Stern information measure
- Newton’s method, 336
- Newton, Isaac
  - Newton iteration, 334, **334**, 336, 340
  - Newton’s lemma, 59
  - Newton’s method, 334, 356
  - Newton-Julia sets, 336
- nonexpansive map, 298
- nonnegative polynomials, 280
- normal set of primes, **248**
- North, Joseph Roy, 33
- nuclear magnetic resonance (NMR), 301
- Odlyzko, Andrew, iii
- Ohno, Yasuo
  - Ohno’s duality theorem, 191
  - Ohno-Zagier generating function, 192
- orthogonal wavelet, 126
- Papadopoulos, Jason, 333
- Parseval equation, 85, 88, 117
- partial fraction decomposition, 42
- partitions, 201, 227
- pattern search algorithm, 263, 321, 322, 324
- Pell’s equation, 213
- pentagonal numbers, 203
- perfect numbers, 266
- Perko, Kenneth, 199
- Perron, Oskar, 40
  - Perron’s continued fractions, 41
- Peters, Klaus, iii
- pi

- continued fraction, 2
- estimates, 2
- formulas, 23, 29, 49
- pictorial representations, 3
- Pick's theorem, 282
- Pisot number, **104**, 275, 356
- Plancherel transform, *see* Fourier transform
- Plouffe, Simon, 152
- Pochhammer function, 18
- Poincaré, J. Henri, 326
- Poisson kernel, 96, 124
- Poisson summation formula, 89–93, 209
  - experimental version, 92, 95
- polyhedron space, 137
- polylogarithms, 24, 149, 169, 173, 186
  - Eulerian polylogarithms, 157
  - Kummer-type polylogarithms, 179
  - multi-dimensional, 30, 167
    - extension, 189
- polynomial arithmetic, 335
- Pomerance, Carl, 331
- prime numbers, 330–334
  - primality tests, 331–334
    - AKS provable primality test, 333–334
    - sieve of Eratosthenes, 331
    - strong probable prime test, 332
  - prime counting function  $\pi(x)$ , 330
- principal branch, 306
- principle of the excluded middle, 326
- Prouhet-Tarry-Escott problem, 269
- pseudospectrum, **363**, 364
- PSLQ, 142, 152, 177, 178, 336
  - binomial sums, 31
  - Riemann zeta function, 161
  - zeta function, 149
- quadrature, 336–346
  - error function quadrature, 340–343
  - Gaussian quadrature, 338–340
  - higher dimensions, 346
  - tanh-sinh quadrature, 343–345
- quasi-elliptic integrals, 55
- Rademacher, Hans, 206
- Ramanujan, Srinivasa, 47, 206, 210, 212, 234, 286
  - additive partitions, 204–206
  - continued fractions, 69–71, 73
  - formulas, 349
  - zeta function, 151–152
- Rauhut, Holger, 251
- Rayleigh quotient, 298–299
- Regiomontanus, Johann, 221Q
- Reidemeister moves, 198
- replicative function, 122
- residue, 303
- Riemann zeta function, 18, 31, 32, 48, 84, **143**, 198, 239, 270–272, 330, 347, 348
  - alternating, **144**
  - beta function, 159
  - Catalan zeta function, **153**, 218
  - evaluating, 180, 181, 183, 187
    - even positive integers, 147, 148
    - odd positive integers, 149, 151
  - identities, 190
  - MZV
    - duality, 167–172
    - recycling, 359
    - reflection formula, **145**
- Riemann, Georg Bernhard, 80, 121, 181
  - Riemann Hypothesis, **146**
  - Riemann hypothesis
    - generalized, 254

- Riemann integral, 217
- Riemann sum, 99
- Riemann surfaces, 307–308
- Riemann-Lebesgue lemma, 85, 87, 104, 127
- Riemann-Siegel Formula, 181–183
- Riesz-Fischer theorem, 85
- Rolle' theorem, 59
- Rosenbrock function, 322
- Rouché's theorem, 302
- RSA encryption, 330, 331
- Rudin-Shapiro polynomials, 276
- Rudolph, Daniel, iii
- Salem, Raphaël, 105
  - Salem number, 356
- Salvy, Bruno, 318
- Saxena, Nitin, 333
- Scharein, Rob, iii
- Schauder, Juliusz, 126
  - Schauder basis, 126
- Schilling equation, 134
- Schmerl, James, 267
- Schur function, **283**, 284, 285
- Selberg integral, 311
- self map, 297
- Sendov, Hristo, iii
- Serrano, Luis, iii
- Siegel, Carl, 183
- Sigler, L. E., 243
- sinc integral, 106, **109**, 113, 135–138
  - iterated, 111, 138
  - multi-variable, 137–138
  - sinc space, 137
- singular value, 210
- Sloane, Neil, iii, 262–265
- Solomyak, Boris, 105
- Soundararajan, K., 286
- spectral radius, **364**
- spherical  $t$ -design, 262
- Spiegel, Eugene, 267
- square-full number, 364
- Stembridge, John, 285
- Stieltjes, Thomas Jan, 40
- Stigler's law of eponymy, 66, 326
  - Stigler, George and Stephen, 66
- Stirling, James
  - Stirling numbers, 59
    - second kind, 28
  - Stirling's formula, 18
    - Lambert  $W$  function, 306
- Stone-Weierstrass theorem, 114, 141
- strictly convex
  - power function, 316
- Sturm's theorem, 259
- symmetric functions, **283**
- Takahasi, Hidetosi, 344
- Thabit ibn Kurrah, 267
- Theodorescu, Radu, 46
- theta function, 217
  - Fibonacci sums, 223–224
- theta transform, 92, 141, 145, 184
- Titchmarsh, Edward, 183
- Torczon, Virginia, 322, 324
- torus knots, 198
- Turing, Alan, 367
- variety, 257
- Vertesi, Janet, iii
- von Neumann minimax theorem, 327
- von Staudt, Karl, 148
- von-Staudt-Clausen formula, 363
- Watson, G. N., 146Q
- Watson, G.N.
  - Watson integrals, 129–134
    - generalized, 132–133
    - logarithmic, 220

- three-dimensional, 129–131
- wavelet basis, 126
- Weierstrass, Karl, 121
- Weiss, Asia, iii
- Wilf-Zeilberger algorithm, 329–330
  - Gosper's telescoping algorithm, 151
- Wilker's inequality, 67
- Wilson's theorem, 269
- Wintner, A., 103, 104
- Wittgenstein, Ludwig, 366
  
- Yildirim, Cel, 285
- Young diagram, 284, 285
  
- Zagier, Don, 156, 160, 245Q
  - Zagier conjecture, 157, **175**
    - extensions, 194–197
    - generalizations, 176
    - proof, 173–176
- zeta function, *see* Riemann zeta function
- Zucker, I. J., 130
- Zucker, John, iii