

APPLICATIONS OF THE MELLIN-PERRON FORMULA IN NUMBER THEORY

by

Marko Ragnar Riedel

B. Sc. (Computer Science) University of British Columbia

A THESIS SUBMITTED IN CONFORMITY WITH THE REQUIREMENTS  
FOR THE DEGREE OF  
MASTER OF SCIENCE

GRADUATE DEPARTMENT OF COMPUTER SCIENCE  
UNIVERSITY OF TORONTO

August 1996

© Copyright by Marko Ragnar Riedel, 1996

## Abstract

### Applications of the Mellin-Perron Formula in Number Theory

Marko Ragnar Riedel

A thesis submitted in conformity with the requirements

for the degree of Master of Science

Graduate Department of Computer Science

University of Toronto

1996

In a 1995 paper, P. Flajolet describes how to evaluate harmonic sums by the Mellin transform. We use his method to obtain an exact formula for consecutive approximations of the area of a fractal ornament delineated by three alternating Koch curves, and the average order of the number of lattice points inside a paraboloid. We define *multiplicative self-similarity*, i.e. a criterion for the existence of a Fourier series expansion of the solution to certain linear recurrences.

P. Flajolet's method replaces an earlier, more complex method developed by H. Delange. This thesis applies P. Flajolet's method to results previously proved by H. Delange's, i.e. the evaluation of alternating digital sums, digital sums in periodic Cantor bases, asymptotic results for digital sums with arbitrary base/weight function combinations, and the error term of a sum related to the number of integers representable as a sum of three squares.

## Table of Contents

<b>Abstract</b>	<b>ii</b>
<b>Acknowledgement</b>	<b>vi</b>
<b>Notation and miscellany</b>	<b>vii</b>
<b>1 Introduction</b>	<b>1</b>
1.1 Generating functions . . . . .	3
1.1.1 A classical example I: integer points inside a circle . . . . .	7
1.1.2 A classical example II: binary trees . . . . .	8
1.2 Harmonic sums and the Mellin transform . . . . .	13
1.2.1 The Mellin transform . . . . .	16
1.2.2 Harmonic sums . . . . .	21
1.3 The Mellin-Perron formula . . . . .	24
1.3.1 A classical example III: the mergesort recurrence . . . . .	29
1.3.2 A fractal ornament . . . . .	37
1.3.3 Digital sums . . . . .	42
1.3.4 Counting sums of three squares . . . . .	49
1.3.5 Lattice points inside a paraboloid . . . . .	55
1.4 Notes . . . . .	59
<b>2 Analytic Number Theory</b>	<b>62</b>
2.1 Point sets in the complex plane . . . . .	63
2.2 Curves, contours and connectedness . . . . .	64
2.3 Limits, continuity, and differentiation . . . . .	65
2.4 Convergent series of analytic functions . . . . .	66
2.4.1 Analytic continuation, Laurent series and classification of singularities . . . . .	67
2.5 The Cauchy residue theorem . . . . .	71

2.6	Dirichlet series . . . . .	71
2.6.1	The Riemann and Hurwitz $\zeta$ functions . . . . .	73
2.7	The analytic version of the fundamental theorem of arithmetic . . . . .	74
2.7.1	Some useful Dirichlet generating functions . . . . .	76
2.8	Integrals of the Hurwitz $\zeta$ -function . . . . .	80
2.9	The Mellin-Perron formula . . . . .	82
2.9.1	The Mellin transform . . . . .	82
2.9.2	The Mellin-Perron formula . . . . .	85
2.9.3	The use of the Mellin-Perron formula when $m = 1$ . . . . .	86
2.10	Mellin-Perron formulae for the Hurwitz $\zeta$ -function . . . . .	87
2.11	Notes . . . . .	89
<b>3</b>	<b>The area of a fractal ornament</b>	<b>91</b>
3.1	Preliminaries . . . . .	91
3.2	A first approximation . . . . .	92
3.3	Exact analysis . . . . .	93
3.3.1	Evaluating the integral . . . . .	94
3.4	Verification . . . . .	96
3.5	Interpretation . . . . .	98
<b>4</b>	<b>Digital Sums</b>	<b>100</b>
4.1	Definitions . . . . .	100
4.2	Alternating digital sums . . . . .	102
4.2.1	Application of the Mellin-Perron formula . . . . .	102
4.2.2	Evaluating the integral . . . . .	103
4.3	Periodic weights in general . . . . .	104
4.4	Intermezzo: digital sum paradigms . . . . .	106
4.5	Digital sums relative to $\kappa$ when $\kappa(j + 1)/\kappa(j) = q(j + 1)$ is periodic . . . . .	107
4.6	Digital sums in the factorial number system . . . . .	108
4.7	The general digital sum problem . . . . .	110
4.8	Notes . . . . .	114

<b>5</b>	<b>Counting sums of three squares</b>	<b>115</b>
5.1	Preliminaries . . . . .	115
5.2	Application of the Mellin-Perron formula . . . . .	117
5.3	Notes . . . . .	119
<b>6</b>	<b>A paraboloid and the lattice points that it contains</b>	<b>120</b>
6.1	What to look for . . . . .	121
6.2	Preliminaries . . . . .	121
6.3	Application of the Mellin-Perron formula with $m = 2$ . . . . .	124
6.4	Significance of the result . . . . .	127
6.4.1	Volume plus error in the boundary surface . . . . .	127
6.4.2	Use of estimates for $r_2(n)$ . . . . .	128
6.4.3	Comparison of the two elementary methods to ours . . . . .	128
6.5	Notes . . . . .	129
	<b>Bibliography</b>	<b>130</b>
	<b>Index</b>	<b>133</b>

## Acknowledgement

I dedicate this work to my mother; what I know of imagination, joy and freedom I owe to her; and to my family, Mr. and Mrs. Norbert Ricker, and Yvonne and Shayne Konar; without their generous and kind help I would not have been able to establish myself in Canada, let alone begin an academic career.

I am grateful to my supervisor, Dr. C. W. Rackoff, for his patience and persistent questioning. I deeply respect his ability quickly to discover the essentials of a problem, and his insistence on clarity and simplicity. He has taught me always to ask first: why? rather than: how? and that a well-posed question is half the answer. There are few things more important and useful than a sound research methodology; I now have the foundation to develop one, and I do so thanks to our weekly meetings.

My second reader, Dr. V. Kumar Murty, provided insightful commentary and valuable suggestions for future research. His discussion of the paraboloid lattice problem in terms of modular forms was succinct, fascinating, and highly instructive; his unique perspective was revelatory in its informed choice of figure and ground, which differed from my own. His exposition of Beuker's proof of the irrationality of Apéry's constant had some of the exploratory, enthusiastic spirit of summer mathematics camp.

Dr. David W. Boyd and Dr. Hugh Montgomery pointed me to Mandelbrojt's gap theorem and other theorems on poles and natural boundaries of Dirichlet series.

Many people have helped me during my stay in Toronto. I cannot list everyone. Daniel Panario encouraged me to proceed with this thesis and shared the excitement of analytical combinatorics. Mikhail Soutchanski and Evguenia Ternovskaia were gracious hosts and fellow cinephiles; Mikhail's sensible advice was all the more valuable for being given so politely and unintrusively. Sherif Ghali assisted in times of financial and emotional distress; he introduced me to the music of Kurt Weill. My acquaintance with David Modjeska has greatly enriched me; if I am more able to appreciate architecture, the use of space in film, the language of Joseph Conrad, and the intricate, formalist and humanist fictions of Jorge L. Borges, he deserves the credit. Anna Wilkes née Frammartino's love of Italian opera was contagious indeed. I will never forget Massey College, if only for the wine at High Tables, that was graciously and freely given by the Senior Fellows. Narly Golestani proved a constant correspondent, and later, a comfortably polyglot companion with a convivial *legereté* of spirit to match.

Dr. Maria Klawe, Dr. James Little, Dr. David W. Kirkpatrick and Dr. N. Hutchinson of the University of British Columbia encouraged me to pursue graduate studies. Dr. Lon Rosen's lectures on complex analysis largely determined my choice of thesis topic, and were precisely as mesmerizing as his performances of the Brahms chamber music repertoire. Dr. R. Anstee's brilliant counsel ("the meaning of life is to eat high fibre") will undoubtedly serve me well for decades to come.

I am grateful to the Natural Sciences and Engineering Research Council for their generous assistance through a NSERC PGS A grant.

## Notation and miscellany

The following sets are used.

- $\mathbb{N} = \{0, 1, 2, 3, \dots\}$
- $\mathbb{Z}^+ = \{1, 2, 3, \dots\}$
- $\mathbb{Z} = \{\dots, -2, -1, 0, 1, 2, \dots\}$
- $\mathbb{R}$ . The set of reals.
- $\mathbb{R}^+ = \{x \mid x \in \mathbb{R}, x > 0\}$ . The set of positive reals.
- $\mathbb{C} = \{x + iy \mid x, y \in \mathbb{R}\}$ ,  $i^2 = -1$ . The set of complex numbers.
- $\mathbb{Q}[i] = \{a + bi \mid x, y \in \mathbb{Z}\}$ . The quadratic field of *Gaussian integers*.

We use the notation  $s = \sigma + it = re^{i\theta}$  for the complex number  $s$  and its real and imaginary parts  $\sigma$  and  $t$ . The *conjugate*  $\bar{s}$  of  $s$  is the number  $\bar{s} = \sigma - it$ . Note that  $\overline{s_1 s_2} = \bar{s}_1 \bar{s}_2$ . The *norm*, *modulus*, or *absolute value* of  $s$  is the non-negative real number  $|s| = r = +\sqrt{\sigma^2 + t^2}$ . Note that  $|n^s| = |e^{s \log n}| = |e^{\sigma \log n + it \log n}| = n^\sigma |e^{it \log n}| = n^\sigma$ .

We write  $a \mid b$ , where  $a, b \in \mathbb{Z}$  to indicate that  $a$  divides  $b$ , i.e. that there exists a  $c \in \mathbb{Z}$  such that  $ac = b$ . If no such  $c$  exists, we write  $a \nmid b$ .

We write  $(m, n) = r$  to indicate that  $r$  is the greatest common divisor of  $m$  and  $n$ ; when  $(m, n) = 1$ , we say that  $m$  and  $n$  are relatively prime. We define

$$a \bmod b = k \Leftrightarrow a \equiv k(b) \text{ and } k \in \{0, 1, \dots, b-1\} \subset \mathbb{N}.$$

The function  $\phi(n)$  is the Euler totient function;  $\phi(n)$  gives the number of  $k$  such that  $1 \leq k \leq n$  and  $(k, n) = 1$ ;  $\phi(n)$  counts the number of positive integers less than or equal to  $n$  that are relatively prime to  $n$ .

Some useful properties of Gaussian integers are:

- If  $\alpha, \beta \in \mathbb{Q}[i]$ , we write  $\alpha \mid \beta$  ( $\alpha \nmid \beta$ ) if there exists (does not exist) a  $\gamma \in \mathbb{Q}[i]$  such that  $\alpha\gamma = \beta$ .
- The units of  $\mathbb{Q}[i]$  are  $1, i, -1, -i$  or  $i^0, i^1, i^2, i^3$ .
- Two  $\alpha, \beta \in \mathbb{Q}[i]$  are *associated* if  $\alpha\gamma = \beta$  where  $\gamma$  is a unit. This is written  $\alpha \sim \beta$ .
- The eight trivial divisors of a Gaussian integer  $\alpha \in \mathbb{Q}[i]$  are its four associates and the four units, i.e.  $1, \alpha, i, i\alpha, -1, -\alpha, -i, -i\alpha$ .
- An  $\alpha \in \mathbb{Q}[i]$  is a *prime* iff  $\alpha$  is neither zero nor a unit and has only trivial divisors.
- The primes of  $\mathbb{Q}[i]$  are
  - $1 + i$
  - $\pi, \bar{\pi}$ , where  $p = \pi\bar{\pi}$  for a natural prime  $p \equiv 1(4)$
  - $p$  where  $p$  is a natural prime  $p \equiv 3(4)$ .

Note that  $2 = i^3(1 + i)^2$ .

We point out the *fundamental theorem of arithmetic*, i.e. that every  $n \in \mathbb{Z}^+$  has a unique prime factorization

$$n = \prod_{k=1}^r p_k^{u_k}$$

with  $p_k \in \mathbb{Z}^+$  prime and  $p_k < p_{k+1}$ . The fundamental theorem of arithmetic for Gaussian integers states that for  $\alpha \in \mathbb{Q}[i]$ ,  $\alpha$  not a unit or zero, we have

$$\alpha = i^l \prod_{k=1}^r \pi_k^{l_k},$$

with the  $\pi_k$  primes of  $\mathbb{Q}[i]$ . This factorization is unique in the sense that a different factorization  $\alpha = i^m \prod_{k=1}^r \rho_k^{m_k}$  can be reordered such that  $\pi_{k_1} \sim \rho_{k_2}$  and  $l_{k_1} = m_{k_2}$ . In particular, the factorization of an  $n \in \mathbb{N}$ ,  $n > 1$  in  $\mathbb{Q}[i]$  is given by

$$n = i^{3l} (1 + i)^{2l} \prod_{k=1}^r \pi_k^{l_k} \bar{\pi}_k^{l_k} \prod_{k=1}^s \rho_k^{m_k}$$

with  $\pi_k \bar{\pi}_k$  a natural prime of the form  $p \equiv 1(4)$  and  $\rho_k$  a natural prime of the form  $p \equiv 3(4)$ .

A sum  $\sum$  without an index variable is understood to range over  $\mathbb{Z}^+$ ; a product  $\prod_p$  ranges over all primes  $p$ . In the context of such products we use  $p_r$  to refer to the  $r$ th prime in the sequence of primes

$\{2, 3, 5, \dots\}$ . The sum  $\sum_{j=1}^0$  is defined to be zero, as are all empty sums; empty products are defined to be one.

The binomial coefficient  $\binom{m}{k}$  is given by

$$\binom{m}{k} = \frac{m!}{k!(m-k)!},$$

which generalizes to

$$\binom{s}{k} = \frac{s(s-1)\dots(s-(k-1))}{k!},$$

where  $s \in \mathbb{C}$ .

The binomial theorem states that

$$\sum_{k=0}^m \binom{m}{k} a^k b^{m-k} = (a+b)^m,$$

where  $m$  is an integer. This generalizes to

$$\sum_{k=0}^{\infty} \binom{s}{k} a^k b^{s-k} = (a+b)^s,$$

$s \in \mathbb{C}$ , which converges for  $|b/a| \leq 1$ .

The following two identities are easily verified by induction:

$$\sum_{k=0}^m k = \frac{1}{2}m(m+1) \quad \text{and} \quad \sum_{k=0}^m k^2 = \frac{1}{6}m(m+1)(2m+1).$$

Geometric progressions are summed by

$$\sum_{k=0}^m q^k = \frac{q^{m+1} - 1}{q - 1} \quad \text{and} \quad \sum_{k=0}^{\infty} q^k = \frac{1}{1 - q}$$

where  $|q| < 1$  in the second formula.

If

$$A(z) = \sum_{n=0}^{\infty} a_n z^n$$

then

$$[z^n]A(z) = a_n,$$

i.e. the notation  $[z^n]A(z)$  is used to extract the  $n$ th coefficient of the variable  $z$  in a power series.

The letters  $\mathcal{A}, \mathcal{B}, \mathcal{C} \dots$  denote sets; the complement of a set  $\mathcal{A}$  is denoted by  $\bar{\mathcal{A}}$ .

We use the symbol  $\sim$  to indicate *asymptotic equivalence*. We write

$$f(z) \sim g(z)$$

iff

$$\lim_{z \rightarrow z_0} \frac{f(z)}{g(z)} = 1$$

where  $f, g : \mathbb{C} \mapsto \mathbb{C}$  and  $z_0$  is a point of the extended complex plane; usually  $z_0 = 0$  or  $z_0 = \infty$ .

A sequence  $\{\phi_k(z)\}$  is an *asymptotic sequence* iff  $|\phi_{k+1}(z)/\phi_k(z)| \rightarrow 0$  as  $z \rightarrow z_0$ . We say that  $f(z)$  has an *asymptotic expansion*

$$\sum a_k \phi_k(z) \quad \text{and write} \quad f(z) \sim \sum a_k \phi_k(z)$$

iff  $\{\phi_k(z)\}$  is an asymptotic sequence,  $\{a_k\} \subset \mathbb{C}$ , and

$$\lim_{z \rightarrow z_0} \frac{|f(z) - \sum^n a_k \phi_k(z)|}{\phi_n(z)} = 0.$$

Asymptotic series are usually divergent, but they may converge. For example, if  $f(z)$  is a function that is single-valued and analytic outside some circle  $|z| > R$ , it has a convergent series expansion at  $z_0 = \infty$ , and this expansion coincides with the asymptotic one.

The series used in this thesis have  $\phi_k(z) = z^k$  when  $z_0 = 0$ , and  $\phi_k(z) = z^{-k}$  when  $z_0 = \infty$ .

## Chapter 1

### Introduction

This thesis is concerned with a specific set of applications of the Mellin-Perron formula in analytic number theory. We will treat examples from four domains:

- fractal curves, more precisely, the area of a fractal ornament delineated by three Koch-like curves,
- digital sums, the classic example of which are binary digital sums, i.e. the sum of the binary digits of zero, through one, two, etc., up to an integer  $n - 1$ ,
- a counting problem related to the number of integers representable as sums of three squares, and
- the number of lattice points inside a paraboloid.

Research in digital sums has a variety of applications in the analysis of algorithms, such as register allocation strategies, random channel networks, sorting networks, and divide-and-conquer recurrences. The other three domains are of a more theoretical nature.

The diversity of these four applications calls for a unified framework of discussion, in order that we may see why these problems are computationally similar to one another. We can construct this framework by placing them in successively larger domains. The Mellin-Perron formula is an instance of a more general problem domain known as the study of *harmonic sums*. Harmonic sums are in turn part of the vast field of *generating function mathematics*, which includes combinatorial methods that use ordinary, exponential, and probabilistic generating functions, or, in the case of multiplicative number theory, Dirichlet series. Another term for *generating function mathematics* is *analytic combinatorics*. This second term testifies to the key paradigm of the method, i.e. the study of *discrete structures*, be they trees, lattices or prime numbers, by means of *continuous methods*, be they those of real-variable analysis or the complex residue calculus. This introduction attempts to convey some of the excitement that accompanies this approach, some of its history (e.g. the history of the Mellin-Perron formula is the history of its earlier guise as a so-called “discontinuous factor”, and we will discuss the nature and

usage of this factor by earlier mathematicians), and most importantly, to present a coherent framework for the results of subsequent chapters.

Technical definitions of the terminology and precise statements of lemmata and theorems are to be found in subsequent chapters; this introduction aims to provide an overview of key principles and methods rather than discuss technicalities. An effort has been made to present the material verbally rather than symbolically wherever possible. The reader should consult the next chapter when additional information or a precise definition of a technical term is desired.

The structure of this chapter is the following. First, we discuss generating function methodology; we list frequently used classes of generating functions and their characteristics. Next, we present two examples. These are not central to this thesis, but they provide additional motivation and historical background to our work. Second, we discuss harmonic sums and the Mellin transform, which is the main mathematical tool we will use. Detailed examples are included. Finally we present an outline of the body of this thesis; we state the topic of each chapter and explain how it fits into the framework of harmonic sums in general and the Mellin-Perron formula in particular.

*How to read this document.* The continuum of potential readers of this thesis extends from the logician or computer scientist with little background in complex analysis and analytical combinatorics on the one hand, and the specialist in analytic number theory and combinatorics, on the other. The first wishes to acquire a skill; the second wants to know what contribution this thesis makes to the field. These two types of readers correspond to two modes of reading this thesis.

- The reader who is new to analytical combinatorics should read the introduction very carefully, and consult chapter two for technical details. As this is not a textbook, this reader may have to consult the papers referenced in the notes to chapters one and two for additional background material. The next four chapters are arranged roughly in order of increasing difficulty.
- The specialist can probably omit the first two chapters. It should suffice to read the abstract and proceed to the result chapters, i.e. chapters three, four, five and six. The second chapter can function as a ready reference for the theorems used in those chapters.

Every reader is situated at a different point along this continuum and will adjust his reading accordingly. An effort has been made to accommodate both types.

The author bears the sole responsibility for any errors that remain in this thesis.

### 1.1 Generating functions

A generating function encodes information about a discrete sequence  $\{a_n\}$  in a continuous function  $A(z)$ . This encoding takes the form of a series expansion; expand  $A(z)$  into a series, and the  $n$ th coefficient of the series will contain some form of  $a_n$ , i.e. the  $n$ th number in the sequence that is being generated. Different forms of series expansions correspond to different types of generating functions. Three common types of generating functions are

- *ordinary and exponential generating functions*, where

$$A(z) = \sum_{n=0}^{\infty} a_n z^n \quad \text{and} \quad A(z) = \sum_{n=0}^{\infty} \frac{a_n}{n!} z^n$$

and

- *Dirichlet generating functions*, where

$$A(z) = \sum_{n=1}^{\infty} \frac{a_n}{n^z}.$$

We should point out that the first two types are commonly known as power series, and the third type as Dirichlet series. We use the term *generating function* to emphasise their use. Ordinary and exponential generating functions are used in combinatorial enumeration; Dirichlet generating functions are used in analytic number theory. Our presentation of the first two types will be restricted to a brief overview and historical background, since they are not central to this thesis. We include them here because they are undoubtedly the most compelling reason why computer scientists would want to study generating function methods.

Generating functions are useful because there exists a *mapping* between *set-theoretical operations*, which are applied to the objects being enumerated or the factors of the numbers being studied, and various *algebraic compositions* of generating functions.

In the case of ordinary and exponential generating functions, this map takes the following form. We consider classes of combinatorial structures, i.e. a set  $\mathcal{A}$ , and a size function  $|\cdot| : \mathcal{A} \mapsto \mathbb{R}$  defined on the elements  $\alpha$  of  $\mathcal{A}$ . For example, the size of a permutation is the number of its elements; the size of a tree is the number of nodes. We use alternate combinatorial forms of the two types of generating functions; these are

$$A(z) = \sum_{\alpha \in \mathcal{A}} z^{|\alpha|} \quad \text{and} \quad A(z) = \sum_{\alpha \in \mathcal{A}} \frac{z^{|\alpha|}}{|\alpha|!}.$$

Identification of coefficients now reveals the rôle of the  $a_n$ ; they count the number of structures whose size function is equal to  $n$ . (The term *size function* should not be taken too literally; the function  $|\cdot|$  may cover a wide range of statistics for  $\mathcal{A}$ .)

The use of ordinary and exponential generating functions rests on the following intuitive extension of the size function  $|\cdot| : \mathcal{A} \mapsto \mathbb{R}$  to the  $k$ -tuples of  $\mathcal{A}^k$ ; the size  $|(\alpha_1, \alpha_2, \dots, \alpha_k)|$  of  $(\alpha_1, \alpha_2, \dots, \alpha_k)$  is  $|\alpha_1| + |\alpha_2| + \dots + |\alpha_k|$ , i.e. the size of a composite object is the sum of the sizes of its components. Set-theoretic operations now become equivalent to algebraic generating function compositions; e.g. the cartesian product  $\mathcal{B} \times \mathcal{C}$  of two sets of combinatorial objects  $\mathcal{B}$  and  $\mathcal{C}$  contains the pairs  $(\beta, \gamma) \in \mathcal{B} \times \mathcal{C}$ ; each such pair has size  $|\beta| + |\gamma|$ ; hence the number of pairs of size  $n = |\beta| + |\gamma|$  is the number of pairs of size 0 and  $n$ , 1 and  $n-1$  etc. or  $b_0c_n, b_1c_{n-1}$  etc. which is  $[z^n](B(z) \cdot C(z))$ .

We can now describe the nature of the map; all entries in the following table, except for the first, are obtained by the construction used for the cartesian product.

Operation		Ordinary generating function equivalent
Union	$\mathcal{A} = \mathcal{B} + \mathcal{C}$	$A(z) = B(z) + C(z)$
Product	$\mathcal{A} = \mathcal{B} \times \mathcal{C}$	$A(z) = B(z) \cdot C(z)$
Sequence	$\mathcal{A} = \mathfrak{S}\{\mathcal{B}\}$	$A(z) = \frac{1}{1-B(z)}$
Powerset	$\mathcal{A} = \mathfrak{P}\{\mathcal{B}\}$	$A(z) = e^{\sigma(z)}$ where $\sigma(z) = \sum_{k=1} (-1)^{k-1} \frac{B(z^k)}{k}$
Multiset	$\mathcal{A} = \mathfrak{M}\{\mathcal{B}\}$	$A(z) = e^{\sigma(z)}$ where $\sigma(z) = \sum_{k=1} \frac{B(z^k)}{k}$
Cycle	$\mathcal{A} = \mathfrak{C}\{\mathcal{B}\}$	$A(z) = e^{\sigma(z)}$ where $\sigma(z) = \sum_{k=1} \frac{\phi(k)}{k} \log \frac{1}{1-B(z^k)}$

These operations have their usual definitions, adapted for the purpose of enumerating discrete, finite combinatorial structures; for example, the sequence of a set  $\mathcal{A}$  is the set  $\{\epsilon\} + \mathcal{A} + \mathcal{A} \times \mathcal{A} + \mathcal{A} \times \mathcal{A} \times \mathcal{A} + \dots$ ; the powerset of  $\mathcal{A}$  is the set of finite subsets; multisets allow repetitions, i.e.  $\mathfrak{M}\{\mathcal{B}\} = \prod_{\beta \in \mathcal{B}} \mathfrak{S}\{\{\beta\}\}$ , and cycles are sequences up to cyclic permutations. (A single cycle of length  $n$  is equivalent to  $n$  sequences.) There are some caveats to be heeded, e.g. we define the union always to be disjoint, as in  $\mathcal{B} + \mathcal{C} = \{\epsilon_1\} \times \mathcal{B} \cup \{\epsilon_2\} \times \mathcal{C}$ ; the reader should consult the notes.

Exponential generating functions are used when each subcomponent that is counted by the size function also bears a unique label. An unlabelled combinatorial structure of size  $n$  corresponds to  $n!$  labelled combinatorial structures of the same size. The following table describes the map for exponential generating functions.

Operation		Exponential generating function equivalent
Union	$\mathcal{A} = \mathcal{B} + \mathcal{C}$	$A(z) = B(z) + C(z)$
Product	$\mathcal{A} = \mathcal{B} \times \mathcal{C}$	$A(z) = B(z) \cdot C(z)$
Sequence	$\mathcal{A} = \mathfrak{S}\{\mathcal{B}\}$	$A(z) = \frac{1}{1-B(z)}$
Set	$\mathcal{A} = \mathfrak{P}\{\mathcal{B}\}$	$A(z) = e^{B(z)}$
Cycle	$\mathcal{A} = \mathfrak{C}\{\mathcal{B}\}$	$A(z) = \log \frac{1}{1-B(z)}$

The content of these two tables is often referred to as *the folklore theorem of combinatorial enumeration*, because most of the constructions listed have been in use for a long time, and without necessarily being grouped or identified as theorems.

The use of ordinary and exponential generating functions follows this sequence of steps: first, set-theoretic relations among the objects being considered are established; second, these relations are translated into their generating function equivalents; third, these generating functions are expanded in order to compute the coefficients and hence the desired statistic on the subset of objects that have the same size. We construct the generating function in order to capture a set statistic; we expand the generating function to compute the statistic.

The case of Dirichlet generating functions is slightly different; there is a single map, given below. This map helps to evaluate Dirichlet generating functions.

Operation		Dirichlet generating function equivalent
Dirichlet convolution	$a_n = \sum_{d n} b_d c_{n/d}$	$A(z) = B(z) \cdot C(z)$

The Dirichlet convolution of two sequences is obtained by considering all possible pairs of numbers in the cartesian product of the two sequences and adding those with the same index product.

Dirichlet generating functions are perhaps the most widely used tool in multiplicative analytic number theory, although there is only one principal identity for them. The body of this thesis is concerned exclusively with applications of Dirichlet generating functions.

These generating functions form one of two factors of the Mellin transform of a harmonic sum. Harmonic sums are evaluated in the poles of the transform function. In order to locate these poles exactly, we need a closed form of the transform function. We obtain this form by evaluating the associated Dirichlet generating function. The Dirichlet series captures the so-called frequencies of the harmonic sum; its closed form locates the poles. Here we proceed from an expansion to the original function, rather than the other way around, as is the case for ordinary and exponential generating functions.

**Generating functions and complexity classes**

It is also worth pointing out that generating functions implicitly define sets of complexity classes. This subject seems to have received little attention. E.g. a result by Chomsky states that the enumeration of the elements of a given length  $n$  of regular languages  $\mathcal{L}$  corresponds precisely to the set of rational generating functions, i.e.  $|\mathcal{L} \cap \{0, 1\}^n| = [z^n]f(z)$ , where  $f(z)$  always a rational function. The folklore theorem on combinatorial enumeration, formalized by Flajolet, implicitly defines a complexity class that includes enumeration functions for partitions and trees, for example, but does not include enumeration functions for graphs or graph statistics. This thesis studies a class of problems whose Dirichlet generating functions exhibit additive or multiplicative self-similarity. Evidently we can define complexity classes of algorithms with respect to the enumeration functions of their problem instances.

**Generating functions; usage pattern and motivation**

The number of results that have been obtained by generating function methods, and the key role that generating functions have played in the work of mathematicians such as Euler, Jacobi, Riemann, Hardy and Polya, is more than sufficient justification to study their applications, as is the renewed interest this field by mathematicians working in computer science, such as DeBruijn, Knuth, Odlyzko and Flajolet, who use generating functions to investigate, for example, the distribution of properties in the average instance of a given data structure, or the average, worst and best-case performance of algorithms. Questions such as that of the average height of a binary tree when all trees with  $n$  nodes are equally likely, or the average-case complexity of sorting algorithms, such as mergesort, can all be answered by generating function methods. The results in question are often difficult to obtain by any other method; there are also cases when the use of generating functions yields an improvement in the exactness of a result.

Therefore it is hardly necessary to argue the use of generating function methods. If we nonetheless include an example, this choice is dictated by its utility in a subsequent chapter and its importance in the development of mathematics. (We will also discuss two examples that have a more immediate relation to computer science, namely the problem of counting the number of binary trees with a given number of internal nodes, and the mergesort recurrence.) In the next section we present some of the history of a classical problem, namely that of counting the number of lattice, i.e. integer points inside a circle. This example has all the features that we associate with good generating function mathematics.

- A discrete enumeration problem is encoded very straightforwardly, one might say automatically, in the series expansion of a continuous function.
- The coefficients of this series expansion are studied by means of complex-variable methods.
- The complex function involved is of such profound importance to mathematics that it links to or even sustains the development of a whole new field, i.e. that of theta functions and modular forms.

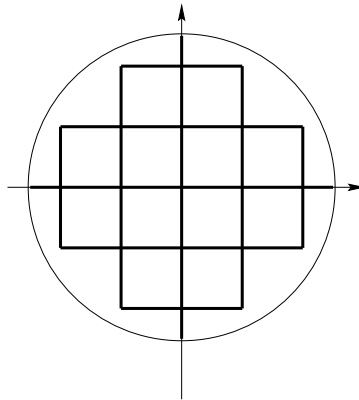
It is not reasonable nor statistically likely to expect everyday applications of generating functions in computer science, i.e. applications that are frequently concerned with partitions, permutations or trees, to have as far-reaching an importance to computer science as the study of theta functions did to mathematics, but it suggests a scope of possibility.

### 1.1.1 A classical example I: integer points inside a circle

The problem that we study is the evaluation of the number  $r_2(n)$  for natural numbers  $n$ , where  $r_2(n)$  gives the number of ways to write  $n$  as the sum  $n = x^2 + y^2$  of two squares; the numbers  $x, y$  are integers. For example,  $r_2(25) = 12$ , since  $25 = 0^2 + (\pm 5)^2$ ,  $25 = (\pm 5)^2 + 0^2$ ,  $25 = (\pm 3)^2 + (\pm 4)^2$  and  $25 = (\pm 4)^2 + (\pm 3)^2$ . The formula

$$A(x) = \sum_{n \leq x} r_2(n)$$

relates the number  $r_2(n)$  to the number of lattice points inside a circle of radius  $\sqrt{x}$ . (We set  $r_2(0) = 1$ .) This is illustrated in the diagram below.



The number of lattice points inside a circle of radius  $\sqrt{25/4}$  is

$$r_2(6) + r_2(5) + \dots + r_2(1) + r_2(0) = 0 + 8 + 4 + 0 + 4 + 4 + 1 = 21.$$

We can discuss this problem in the framework of ordinary generating functions and their associated operators. Let  $\mathcal{B}$  and  $\mathcal{C}$  be two copies of the set of integers  $\mathbb{Z}$ ; the size function is  $k \mapsto k^2$ ; there is one object of size zero, i.e. the number 0, and two objects of size  $m^2$ , i.e. the numbers  $-m$  and  $m$ . The object pair  $(\beta, \gamma)$  has size  $|\beta| + |\gamma|$ , hence the cartesian product  $\mathcal{B} \times \mathcal{C}$  contains as many objects of size  $n$  as there are ways to write  $n$  as the sum of two integer squares. Translated into generating functions, this gives

$$\theta(z) = \sum_{n=0} r_2(n)z^n = \left(1 + 2 \sum_{m=1} z^{m^2}\right)^2 = \left(\sum_{m=-\infty}^{\infty} z^{m^2}\right)^2.$$

(This problem is so simple that we could have obtained  $\theta(z)$  without the use of the operator framework.)

We have reduced the computation of  $r_2(n)$  to the problem of computing the  $n$ th coefficient  $[z^n]\theta(z)$  of  $\theta(z)$ . A celebrated, deep result that was first discovered by Jacobi yields the answer. This result is a theta function identity, of which

$$\left(\sum_{m=-\infty}^{\infty} z^{m^2}\right)^2 = 1 + 4 \sum \frac{z^n}{1+z^{2n}} = 1 + 4 \sum \sum \left(z^{(4m-3)n} - z^{(4m-1)n}\right)$$

is a special case. The difficulty lies in proving the first equality; the second one is easily verified. Formally we have

$$\sum \frac{z^n}{1+z^{2n}} = \sum \sum_{m=1} (-1)^{m-1} z^{(2m-1)n} = \sum \sum z^{(4m-3)n} - z^{(4m-1)n}.$$

This solves the problem. We compare coefficients and conclude that  $r_2(n)$  is *the difference between the number of odd divisors of  $n$  that are congruent 1 modulo 4 and the number of odd divisors of  $n$  that are congruent 3 modulo 4*. (There are a variety of “elementary” proofs of this result, see the discussion of Theorem 6.2.1.)

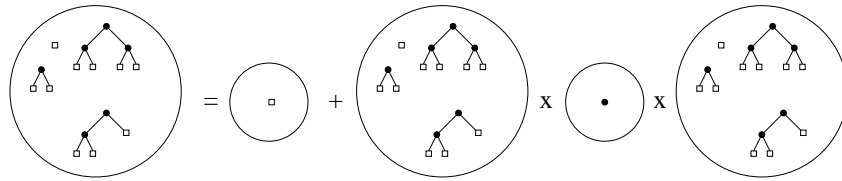
We will return to the study of lattice points in a later chapter. The problem easily generalizes to higher dimensions, and it has given rise to some very profound 20th-century mathematics; Ramanujan, Hardy and others worked in this area.

### 1.1.2 A classical example II: binary trees

There are numerous examples of the generating function technique. We have discussed the evaluation of  $r_2(n)$ . In this section we present a classic combinatorial enumeration problem related to binary trees; this problem is more immediately relevant to the computer scientist. We will also discuss the mergesort recurrence once the necessary background material has been presented.

For the student of ordinary generating functions, to count the number of binary trees with a given number of internal nodes is the equivalent of writing a “hello world” program.

Let  $\mathcal{B}$  be the set of binary trees. The size  $|\beta|$  of a tree  $\beta$  is the number of internal nodes. A binary tree is either an external node, or an internal node with two subtrees, i.e. every parent node has either zero or two children. This decomposition is shown below:



and we may write it as

$$\mathcal{B} = \mathcal{E} + \mathcal{B} \times \mathcal{N} \times \mathcal{B}.$$

(In the diagram, the set  $\mathcal{B}$  of binary trees is represented by four of its elements; the singleton sets  $\mathcal{E}$  and  $\mathcal{N}$  are represented in full.) The generating function of a size zero external node is  $E(z) = 1z^0 = 1$ ; there is a single object of size zero. The generating function of a size one internal node is  $N(z) = 1z^1 = z$ ; there is a single object of size one. Let  $B(z)$  be the generating function of the number  $b_n$  of binary trees of size  $n$ . We have

$$B(z) = E(z) + B(z)N(z)B(z) = 1 + zB(z)^2.$$

We solve this for  $B(z)$  and obtain

$$B(z) = \frac{1 \pm \sqrt{1 - 4z}}{2z}.$$

By the binomial theorem,

$$\begin{aligned} \sqrt{1 - 4z} &= \sum_{n=0}^{\infty} \binom{1/2}{n} (-4z)^n = 1 - 2z + \sum_{n=2}^{\infty} \frac{\frac{1}{2}(\frac{1}{2}-1)\dots(\frac{1}{2}-(n-1))}{1 \cdot 2 \dots n} (-4)^n z^n \\ &= 1 - 2z + \sum_{n=2}^{\infty} \frac{1(-1)(-3)\dots(3-2n) \cdot 2 \cdot 4 \dots (2n)}{n! n!} (-1)^n z^n \\ &= 1 - 2z - \sum_{n=2}^{\infty} \frac{1}{2n-1} \frac{(-1)(-3)\dots(-(2n-3))(-2n-1)}{n! n!} 2 \cdot 4 \dots (2n) (-1)^n z^n \\ &= 1 - 2z - \sum_{n=2}^{\infty} \frac{1}{2n-1} \binom{2n}{n} z^n. \end{aligned}$$

The positive root gives  $b_0 = -1$ , the negative root  $b_0 = 1$ , hence we choose the negative root:

$$\begin{aligned} B(z) &= \frac{1 - 1 + 2z + \sum_{n=2}^{\infty} \frac{1}{2n-1} \binom{2n}{n} z^n}{2z} \\ &= 1 + \sum_{n=2}^{\infty} \frac{1}{2} \frac{1}{2n-1} \binom{2n}{n} z^{n-1} = 1 + \sum_{n=1}^{\infty} \frac{1}{2} \frac{1}{2n+1} \binom{2n+2}{n+1} z^n \\ &= 1 + \sum_{n=1}^{\infty} \frac{1}{2} \frac{1}{2n+1} \frac{(2n+1)(2n+2)}{(n+1)^2} \binom{2n}{n} z^n = \sum_{n=0}^{\infty} \frac{1}{n+1} \binom{2n}{n} z^n. \end{aligned}$$

This proves that the number  $b_n$  of binary trees with  $n$  internal nodes is

$$\frac{1}{n+1} \binom{2n}{n},$$

i.e. the  $n$ th *Catalan* number.

The technique that we have demonstrated applies to a diverse group of problems. It is not always possible to evaluate the coefficients of the generating function directly. The following example is slightly more complicated and shows what kinds of difficulties can arise.

Consider the size of the *intersection* of two binary trees, which is a complexity measure over *pairs* of binary trees. Given two trees  $\beta_1$  and  $\beta_2$ , we can define their intersection recursively – if either tree is empty, the intersection is empty, otherwise the intersection is the tree whose left (right) subtree is the intersection of the two left (right) subtrees of  $\beta_1$  and  $\beta_2$ . If we overlay  $\beta_1$  with  $\beta_2$ , their intersection is the tree that includes all “double nodes.” We wish to compute the average size of the intersection of two binary trees  $\beta_1$  and  $\beta_2$  given that  $|\beta_1| + |\beta_2| = n$ , i.e. the total number of internal nodes of  $\beta_1$  and  $\beta_2$  is  $n$ . The size function for the intersection is the number of internal nodes.

Formally we have

$$\beta_1 \cap \beta_2 = \begin{cases} \epsilon & \text{if } \beta_1 = \epsilon \text{ or } \beta_2 = \epsilon \\ \beta_1^l \cap \beta_2^l \times \{\nu\} \times \beta_1^r \cap \beta_2^r & \text{otherwise,} \end{cases}$$

where  $\beta^l$  is the left and  $\beta^r$  is the right subtree of a tree  $\beta$ ,  $\epsilon$  is the empty tree and  $\nu$  is a single internal node.

This yields the following expression for the size function  $s(\beta_1, \beta_2)$  of the intersection of  $\beta_1$  and  $\beta_2$ .

$$s(\beta_1, \beta_2) = \begin{cases} 1 & \text{if } \beta_1 = \epsilon \text{ or } \beta_2 = \epsilon \\ 1 + s(\beta_1^l, \beta_2^l) + s(\beta_1^r, \beta_2^r) & \text{otherwise.} \end{cases}$$

Under a uniform probability distribution, the average value  $s(n)$  of the intersection of two trees  $\beta_1$  and  $\beta_2$  with a total of  $n$  internal nodes is the sum of  $s(\beta_1, \beta_2)$  over all pairs  $(\beta_1, \beta_2)$ , divided by the total number of such pairs, i.e.

$$s(n) = \frac{\sum_{|\beta_1|+|\beta_2|=n} s(\beta_1, \beta_2)}{\sum_{|\beta_1|+|\beta_2|=n} 1}.$$

Clearly

$$\sum_{|\beta_1|+|\beta_2|=n} 1 = [z^n] \sum_{(\beta_1, \beta_2) \in \mathcal{B}^2} z^{|\beta_1|+|\beta_2|} = [z^n] B(z)^2.$$

Define  $S(z)$  by

$$\sum_{|\beta_1|+|\beta_2|=n} s(\beta_1, \beta_2) = [z^n] \sum_{(\beta_1, \beta_2) \in \mathcal{B}^2} s(\beta_1, \beta_2) z^{|\beta_1|+|\beta_2|} = [z^n] S(z).$$

We need to compute  $S(z)$  in order to find  $s(n)$ .

Let  $\mathcal{V} = \mathcal{B} - \{\epsilon\}$ ; note that the respective generating function is  $V(z) = B(z) - 1$ . We use the following decomposition of the set of pairs of binary trees:

$$\mathcal{B}^2 = \{(\epsilon, \epsilon)\} + \{\epsilon\} \times \mathcal{V} + \mathcal{V} \times \{\epsilon\} + \mathcal{V} \times \mathcal{V}.$$

There are four cases.

- The pair of empty trees:

$$\sum_{(\beta_1, \beta_2) \in \{(\epsilon, \epsilon)\}} s(\beta_1, \beta_2) z^{|\beta_1|+|\beta_2|} = 1$$

- The left element empty and the right not empty:

$$\sum_{(\beta_1, \beta_2) \in \{\epsilon\} \times \mathcal{V}} s(\beta_1, \beta_2) z^{|\beta_1|+|\beta_2|} = \sum_{\beta_2 \in \mathcal{V}} z^{|\beta_2|} = V(z) = B(z) - 1$$

- The left element not empty and the right empty:

$$\sum_{(\beta_1, \beta_2) \in \mathcal{V} \times \{\epsilon\}} s(\beta_1, \beta_2) z^{|\beta_1|+|\beta_2|} = \sum_{\beta_1 \in \mathcal{V}} z^{|\beta_1|} = V(z) = B(z) - 1$$

- Neither element empty.

$$\sum_{(\beta_1, \beta_2) \in \mathcal{V} \times \mathcal{V}} s(\beta_1, \beta_2) z^{|\beta_1|+|\beta_2|} = \sum_{(\beta_1, \beta_2) \in \mathcal{V} \times \mathcal{V}} (1 + s(\beta_1^l, \beta_2^l) + s(\beta_1^r, \beta_2^r)) z^{|\beta_1|+|\beta_2|}.$$

The last expression takes somewhat longer evaluate. The computation is shown below.

$$\begin{aligned}
& \sum_{(\beta_1, \beta_2) \in \mathcal{V} \times \mathcal{V}} (1 + s(\beta_1^l, \beta_2^l) + s(\beta_1^r, \beta_2^r)) z^{|\beta_1| + |\beta_2|} \\
= & (B(z) - 1)^2 + \sum_{(\beta_1^l, \beta_1^r) \in \mathcal{B}^2, (\beta_2^l, \beta_2^r) \in \mathcal{B}^2} (s(\beta_1^l, \beta_2^l) + s(\beta_1^r, \beta_2^r)) z^{2 + |\beta_1^l| + |\beta_1^r| + |\beta_2^l| + |\beta_2^r|} \\
= & (B(z) - 1)^2 \\
+ & z^2 \sum_{(\beta_1^l, \beta_2^l) \in \mathcal{B}^2, (\beta_1^r, \beta_2^r) \in \mathcal{B}^2} s(\beta_1^l, \beta_2^l) z^{|\beta_1^l| + |\beta_2^l|} z^{|\beta_1^r| + |\beta_2^r|} \\
+ & z^2 \sum_{(\beta_1^l, \beta_2^l) \in \mathcal{B}^2, (\beta_1^r, \beta_2^r) \in \mathcal{B}^2} s(\beta_1^r, \beta_2^r) z^{|\beta_1^l| + |\beta_2^l|} z^{|\beta_1^r| + |\beta_2^r|} \\
= & (B(z) - 1)^2 + 2z^2 S(z) B(z)^2
\end{aligned}$$

Combining the four cases, we have

$$S(z) = 1 + B(z) - 1 + B(z) - 1 + (B(z) - 1)^2 + 2z^2 S(z) B(z)^2 = B(z)^2 + 2z^2 S(z) B(z)^2$$

or

$$S(z) = \frac{B(z)^2}{1 - 2z^2 B(z)^2}.$$

Although there is an unreasonable amount of manipulation involved, the coefficients  $[z^n]S(z)$  can be computed exactly. In practice one would use an asymptotic method to obtain the dominant term of  $[z^n]S(z)$ . The purpose of this section was to demonstrate the use and manipulation of ordinary generating functions in practical applications, such as binary trees, and we will not pursue this example. The reader should consult the notes for more information.

This concludes our brief discussion of (ordinary) generating functions. We now pass to the main subject of this introduction, i.e. harmonic sums, the Mellin transform, and the relation between harmonic sums and Dirichlet generating functions.

## 1.2 Harmonic sums and the Mellin transform

Harmonic sums are sums of the form

$$G(x) = \sum_k \lambda_k g(\mu_k x),$$

where the  $\lambda_k$  are the *amplitudes*, the  $\mu_k$  are the *frequencies* and  $g(x)$  is the *base function*. We consider harmonic sums because *we wish to evaluate  $G(x)$  at a set of particular points  $x_0, x_1, \dots$  or at all  $x \in \mathbb{R}$ .*

**Example.** The harmonic sum

$$\sum_{1 \leq k < n} \lambda_k \left(1 - \frac{k}{n}\right)^m,$$

is an instance with base function  $g(y) = (1 - y)^m$ , amplitude  $\lambda_k$  and frequency  $k$ , evaluated at  $x = 1/n$ , in other words, *a finite sum of the first  $n - 1$  numbers of a sequence times a polynomial moment of degree  $m$ .* The body of this thesis is concerned exclusively with sums of this type.

The study of harmonic sums is closely related to that of Mellin transforms. Harmonic sums are treated by the following method:

1. Compute the Mellin transform of the harmonic sum.

$$G(x) \xrightarrow{\mathfrak{M}} G^*(s)$$

2. Use the inverse Mellin transform to obtain an expression of the harmonic sum in terms of an integral of the transform function.

$$G^*(s) \xrightarrow{\mathfrak{M}^{-1}} G(x)$$

3. Use the singularities of the transform function to evaluate the integral and thus to obtain an expansion of the harmonic sum.

$$\mathfrak{M}^{-1}[G^*(s); x] = G(x) \sim \pm \sum_{\zeta \in \text{Sing}(G^*(s)x^{-s}) \cap H} \text{Res}(G^*(s)x^{-s}; s = \zeta),$$

where the sign is determined by the choice of the half-plane  $H$ . The sum in this formula is taken over the singularities of  $G^*(s)x^{-s}$ . We compute the residue at each pole in  $H$ , and add them. This operation is a consequence of the Cauchy residue theorem. (The purpose of  $H$  will be explained in the next section, as will be the presence or absence of error terms in the above formula. In practice the convergence of the sum may be stronger than merely asymptotic.)

The computation of the sum in the last formula can be made automatic by the use of the secondary map

$$\frac{A}{(s-\zeta)^{k+1}} \xrightarrow{\text{Res}(f(s)x^{-s}; s=\zeta)} A \frac{(-1)^k}{k!} x^{-\zeta} (\log x)^k.$$

It is a basic characteristic of harmonic sums that the Mellin transform  $G^*(s)$  of  $G(x)$  has the form

$$G^*(s) = \Lambda(s)g^*(s),$$

where  $g^*(s)$  is the transform of the base function and  $\Lambda(s)$  is the Dirichlet generating function of the  $\lambda_k$ , i.e.

$$\Lambda(s) = \sum \frac{\lambda_k}{k^s}.$$

This means that when the sums are evaluated by the method described above, the poles of the Dirichlet generating function and some additional terms correspond to an expression of the sum.

The following example illustrates the method that we have described.

**Example.** *Harmonic numbers.*

**Step 1: definition of the harmonic sum and computation of the appropriate Mellin transform.** Let  $\lambda_k = 1/k$ ,  $\mu_k = 1/k$  and  $g(x) = x/(1+x) = 1/(1+1/x)$ ; consider the harmonic sum

$$h(x) = \sum_k \lambda_k g(\mu_k x) = \sum_k \frac{1}{k} \frac{x/k}{1+x/k} = \sum \left( \frac{1}{k} - \frac{1}{x+k} \right).$$

This sum is of interest because

$$h(n) = \sum \left( \frac{1}{k} - \frac{1}{n+k} \right) = \sum \frac{1}{k} - \sum_{k=n+1} \frac{1}{k} = \sum_{k=1}^n \frac{1}{k} = H_n,$$

the  $n$ th harmonic number.

The principal operation of the first step in the evaluation of harmonic sums is the computation of the Mellin transform of the base function  $g(y)$  and the computation of the Dirichlet generating function

$\Lambda(s)$ . We first compute the transform of the base function. We have  $\mathfrak{M}[1/(1+x); s] = \pi/\sin(\pi s)$  (this is a standard transform) and hence

$$\mathfrak{M}\left[\frac{x}{1+x}; s\right] = -\frac{\pi}{\sin(\pi s)}.$$

Now we compute the Dirichlet generating function  $\Lambda(s)$ . We have

$$\Lambda(s) = \sum \frac{1}{k} k^s = \sum \frac{1}{k^{1-s}} = \zeta(1-s).$$

(The function  $\zeta(s)$  is described in section 2.6.1. We also point out that the above equality only holds for  $\sigma < 0$ . For the moment we omit convergence issues in order to concentrate on the procedural aspects of the example.) We conclude that the Mellin transform of  $h(x)$  is

$$-\frac{\pi}{\sin(\pi s)}\zeta(1-s).$$

**Step 2: inversion of the map.** By Mellin inversion,

$$\mathfrak{M}^{-1}\left[-\frac{\pi}{\sin(\pi s)}\zeta(1-s); x\right] = h(x).$$

This is equivalent to the inversion integral

$$\int_{c-i\infty}^{c+i\infty} \left(-\frac{\pi}{\sin(\pi s)}\zeta(1-s)\right) x^{-s} ds = h(x).$$

(The choice of  $c$  will be explained later.) This integral representation permits the computation of  $h(x)$ , because the integral can be evaluated by the Cauchy Residue theorem, i.e. it is a sum of residues of  $h^*(s)x^{-s}$ .

**Step 3: computation of the poles of the transform function and the corresponding terms in the asymptotic expansion.** We use the fact that

$$h(x) \sim - \sum_{\varsigma \in \text{Sing}(h^*(s)x^{-s}) \cap H} \text{Res}(h^*(s)x^{-s}; s = \varsigma),$$

where  $H$  is the right half-plane, chosen for an expansion at infinity. We must compute the set of poles  $\text{Sing}(h^*(s)x^{-s}) \cap H$  and map them back to the terms of the expansion of  $h(x)$ . The poles of  $h^*(s)$  in the right half-plane are at  $s = 0$ , where we have a double pole and

$$h^*(s) = \frac{1}{s^2} - \frac{\gamma}{s} + \dots$$

and at  $s = k$ ,  $k \in \mathbb{Z}^+$ , where we have

$$h^*(s) = -(-1)^k \frac{\zeta(1-k)}{s-k} + \dots$$

These poles map back to

$$-\log x - \gamma$$

and

$$-\frac{1}{2x} \quad \text{for } k=1, \quad -\frac{(-1)^k B_k}{k} \frac{1}{x^k} \quad \text{for } k \geq 2.$$

We conclude that Harmonic numbers satisfy the asymptotic expansion

$$H_n \sim \log n + \gamma + \frac{1}{2n} + \sum_{k \geq 2} \frac{(-1)^k B_k}{k} \frac{1}{n^k}.$$

This expansion is exact; it converges for  $n \geq 1$ .

The reader will have noticed the use of terms such as *half-plane* in the above discussion or perhaps have wondered why we chose the poles that we did choose when we evaluated  $H_n$ . In order to make these notions precise, we need to study the Mellin transform. We will return to the subject of harmonic sums once the basic terminology has been explained. We have deliberately chosen to outline the methodology first, and then proceed to a more detailed presentation. The object is to supply a strategy that the reader can use to tackle harmonic sums. The questions to ask are these: what is the Mellin transform of the sum and where is it defined? What are the poles of the transform function? How do they map to an asymptotic expansion? These questions define a particular harmonic sum problem; this hurdle overcome, the rest is fairly straightforward computation, admittedly of a kind that must be done carefully, e.g. if the domain of convergence of the Mellin integral is determined incorrectly, we will choose the wrong poles.

### 1.2.1 The Mellin transform

The purpose of this section is to introduce the Mellin transform and its inverse, and outline how these two transforms are computed. We will first define the Mellin transform and then illustrate the steps required to compute it. We will then define the inverse transform, and again illustrate it with a simple example.

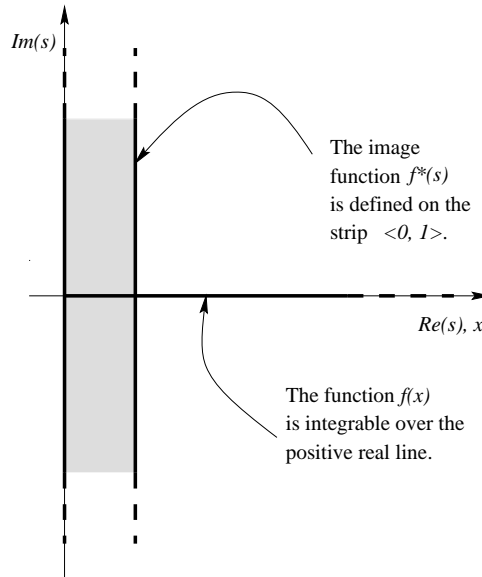
The Mellin transform maps the space of **functions that are integrable along the positive real line** to that of **complex functions** that are analytic on a *vertical strip of the complex plane*. This strip may in many cases be extended to a larger domain. The map is given by

$$\mathfrak{M}[f(x); s] = f^*(s) = \int_0^{+\infty} f(x)x^{s-1} dx.$$

**Example.** The Mellin transform of

$$f(x) = \frac{1}{1+x} \quad \text{is} \quad \mathfrak{M}[f(x); s] = f^*(s) = \frac{\pi}{\sin \pi s}$$

in the vertical strip of width one that is adjacent to, and to the right of the imaginary axis; we write  $\langle 0, 1 \rangle$  to denote this domain. The diagram shows the domains of  $f(x)$  and its image  $f^*(s)$ .



The function  $\frac{1}{1+x}$  is integrable over  $(0, +\infty)$  and its Mellin transform  $\frac{\pi}{\sin \pi s}$  is defined in the strip  $\langle 0, 1 \rangle$ .

Given a function  $f(x)$  what is the meaning of the *fundamental strip*  $\langle u, v \rangle$  where the image function  $f^*(s)$  converges? If we rewrite the Mellin integral in the form

$$\left( \int_0^1 + \int_1^{+\infty} \right) f(x) \frac{x^s}{x} dx,$$

it is evident that the integral converges iff the first integral remains bounded at zero and the second integral vanishes at infinity. This places a constraint on  $s$ . Suppose  $f(x) \in \mathcal{O}(x^u)$  as  $x$  goes to zero and  $f(x) \in \mathcal{O}(x^v)$  as  $x$  goes to infinity. The first half converges as long as the real part of  $s$  is larger than  $-u$ . The second half converges as long as the real part of  $s$  is less than  $-v$ . Each of these constraints defines a half-plane. The fundamental strip is the intersection of these two half-planes.

**Example.** Let  $f(x) = e^{-x} - 1 + x$ . The following series expansion holds as  $x$  goes to zero.

$$f(x) = \sum_{n=2}^{\infty} \frac{(-1)^n x^n}{n!}$$

Therefore

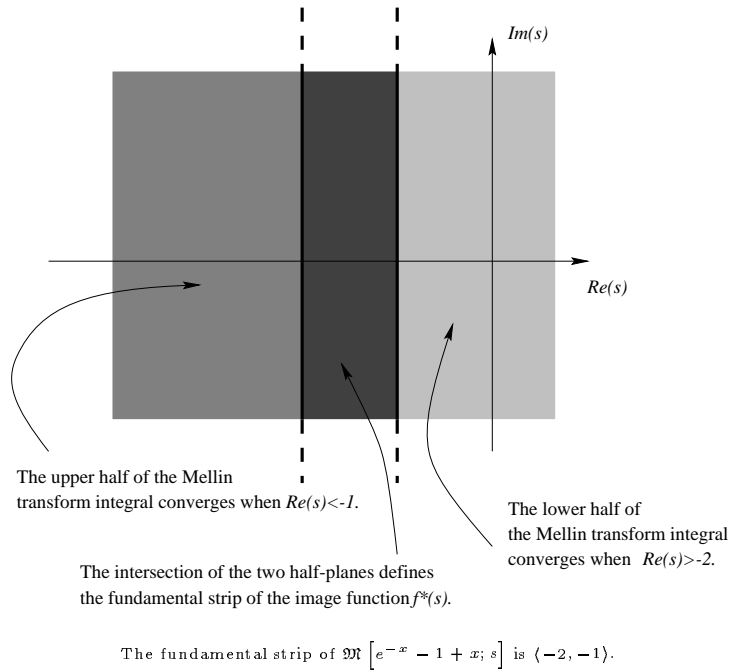
$$\int_0^1 f(x) \frac{x^s}{x} dx$$

converges when  $\sigma > -2$ , where we have set  $s = \sigma + it$ . On the other hand,  $f(x) \sim x$  as  $x$  goes to infinity.

Hence

$$\int_1^{+\infty} f(x) \frac{x^s}{x} dx$$

converges for  $\sigma < -1$ . The fundamental strip of the image function is  $\langle -2, -1 \rangle$ ; incidentally,  $f^*(s) = \Gamma(s)$ , where  $\Gamma(s)$  is the Euler gamma function. This situation is illustrated below.



The inverse map from an image function  $f^*(s)$  back to  $f(x)$  is given by the integral

$$\frac{1}{2\pi i} \int_{c-i\infty}^{c+i\infty} f^*(s) x^{-s} ds = f(x)$$

where  $c \in \mathbb{R}$  is located in the fundamental strip  $\langle \alpha, \beta \rangle$  of the Mellin integral of  $f(x)$ ; i.e. the inverse transform is given by an integral taken along any vertical line in the fundamental strip. The following example shows how to compute this transform. The reader should pay particular attention to the two

different expansions constructed for  $f(x)$ . We will compute two such expansions, one of them based on the poles of  $f^*(s)$  in the left half-plane, the other in the right half-plane. This is precisely what happens in the evaluation of harmonic sums; we shall see that the two half-planes correspond to an expansion of  $G(x)$  at zero and infinity, respectively.

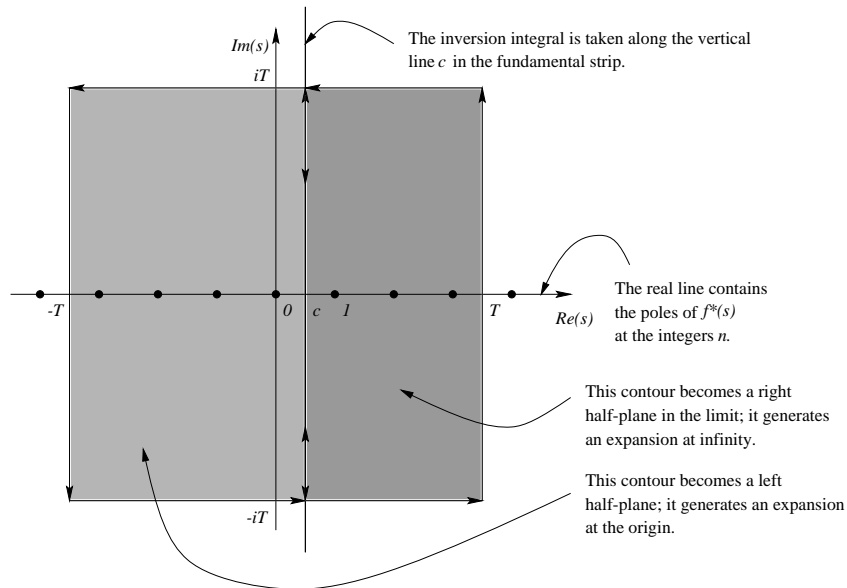
**Example.** Consider the transform pair

$$f(x) = \frac{1}{1+x} \quad \text{and} \quad f^*(s) = \frac{\pi}{\sin \pi s}.$$

The poles of  $f^*(s)$  are at those  $s$  where  $\sin \pi s$  is zero, i.e. at all integers. The inverse Mellin transform is given by

$$\frac{1}{2\pi i} \int_{c-i\infty}^{c+i\infty} \frac{\pi}{x^s \sin \pi s} ds.$$

We can evaluate this integral by the Cauchy residue theorem, taking the limit of a rectangular contour that eventually includes all of  $\langle -\infty, c \rangle$  or  $\langle c, \infty \rangle$ . In the first case the contour consists of the original integral, a horizontal line segment from  $c + iT$  to  $-T + iT$ , a vertical line segment from  $-T + iT$  to  $-T - iT$  and a horizontal line segment from  $-T - iT$  to  $c - iT$ ; in the second of the original integral, traversed from top to bottom, a segment from  $c - iT$  to  $T - iT$ , one from  $T - iT$  to  $T + iT$  and one from  $T + iT$  to  $c + iT$ . These two contours are shown below.



The two contours that are used to compute the Mellin inversion integral of  $\frac{\pi}{\sin \pi s}$ .

We must be careful to choose a sequence of increasing  $T$  that avoids the poles of  $f^*(s)$ ; these poles are at the integers and therefore a sequence of the form  $T_j = (2j + 1)/2$ ,  $j \in \mathbb{Z}$ , with  $T$  chosen midway between poles, will suffice.

The integral equals the sum of the residues inside the contour, plus the contribution from one extra vertical and the two horizontal segments. The poles  $n$  are all simple, with residue  $(-1)^n$ , as we can see from

$$\frac{\pi}{\sin \pi s} = \frac{2\pi i}{e^{i\pi s} - e^{-i\pi s}} = \frac{2\pi i e^{i\pi s}}{e^{2i\pi s} - 1}$$

and

$$\lim_{s \rightarrow n} (s - n) \frac{2\pi i e^{i\pi s}}{e^{2i\pi s} - 1} = 2\pi i (-1)^n \lim_{s \rightarrow n} \frac{1}{2i\pi e^{2i\pi s}} = (-1)^n.$$

Let  $s = \sigma + it$ . Then

$$\left| \frac{\pi}{\sin \pi s} \right| = \left| \frac{2\pi i e^{i\pi s}}{e^{2i\pi s} - 1} \right| = 2\pi \frac{e^{-\pi t}}{|e^{2i\pi s} - 1|}$$

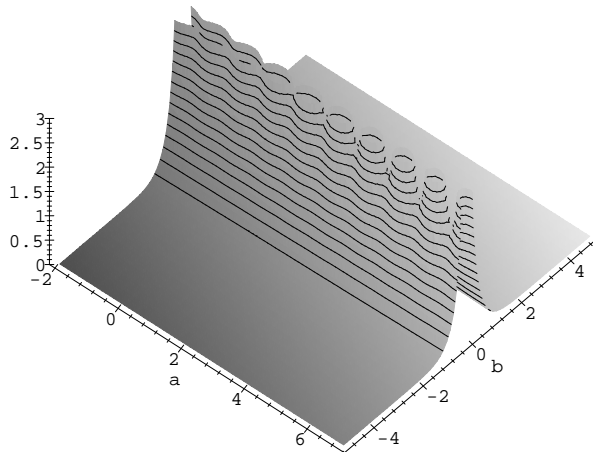
and

$$\left| \frac{\pi}{\sin \pi s} \right| = \left| \frac{2\pi i e^{-i\pi s}}{1 - e^{-2i\pi s}} \right| = 2\pi \frac{e^{\pi t}}{|1 - e^{-2i\pi s}|}.$$

We point out that

$$|e^{2\pi i\sigma - 2\pi t} - 1| \rightarrow 1 \text{ as } t \rightarrow +\infty \quad \text{and} \quad |1 - e^{-2\pi i\sigma + 2\pi t}| \rightarrow 1 \text{ as } t \rightarrow -\infty$$

because the  $e^{\pm 2\pi i\sigma}$  terms merely rotate around the unit circle.



Contour plot of  $|(1 + 1/16)^s \pi / \sin(\pi s)|$  for  $s = a + bi$  and  $-2 \leq a \leq 7$  and  $-5 \leq b \leq 5$ , showing the poles

of  $\pi / \sin(\pi s)$  and the exponential decay away from the real axis.

Hence

$$\left| \frac{\pi}{\sin \pi s} \right| \in \mathcal{O}(e^{-\pi t})$$

in the upper half plane, and

$$\left| \frac{\pi}{\sin \pi s} \right| \in \mathcal{O}(e^{+\pi t})$$

in the lower half. This shows that the contribution from the one extra vertical and the two extra horizontal segments in each contour is  $\mathcal{O}(e^{-T} x^{\pm T} / (\pm T + 1))$  and vanishes in the limit, as illustrated by the contour plot. Therefore the Mellin inversion integral

$$\frac{1}{2\pi i} \int_{c-i\infty}^{c+i\infty} \frac{\pi}{x^s \sin \pi s} ds$$

is

$$\sum_{n=0}^{\infty} (-1)^n x^n = \frac{1}{1+x} = f(x),$$

for  $|x| < 1$ , if we evaluate it in the left half-plane and

$$-\sum_{n=1}^{\infty} (-1)^n x^{-n} = \frac{1}{1+x} = f(x),$$

for  $|x| > 1$ , if we evaluate it in the right half-plane.

The reader will have noticed that the two expansions of  $x$  obtained from the left and right half-planes are complete expansions of  $f(x)$  about zero and infinity. This is a special case of the *Mapping theorem*, which states that the terms of an (asymptotic) expansion of  $f(x)$  about zero or infinity correspond precisely to poles of  $f^*(s)$  in the left and right half-plane, respectively. Therefore, (asymptotic) expansions of the original function  $f(x)$  can be obtained from the poles of its image function  $f^*(s)$ . This is the key principle in the evaluation of many types of harmonic sums, such as dyadic harmonic sums, which arise in connection with the study of random resource allocation. This thesis is concerned with *exact*, rather than asymptotic results.

### 1.2.2 Harmonic sums

The first step in the evaluation of a harmonic sum of the form

$$G(x) = \sum_k \lambda_k g(\mu_k x)$$

is the use of the *linearity and rescaling* property of the Mellin transform. This property states that *the Mellin transform of single base function term with amplitude  $\lambda$  and positive frequency  $\mu$  is the transform of the base function times the amplitude and the inverse frequency to the power  $s$* . In other words, amplitudes and frequencies factor. Let the transform of  $g(x)$  be  $g^*(s)$ , then the transform of  $\lambda g(\mu x)$  is  $\lambda \mu^{-s} g^*(s)$ . Therefore the Mellin transform associates the Dirichlet series

$$\Lambda(s) = \sum \frac{\lambda_k}{k^s}$$

to the harmonic sum  $G(x)$ , and the transform of  $G(x)$  is

$$G^*(s) = \left( \sum \frac{\lambda_k}{k^s} \right) g^*(s) = \Lambda(s) g^*(s).$$

(The technical condition for this statement to hold is that the abscissa of absolute convergence of  $\Lambda(s)$  lies within the fundamental strip of  $g^*(s)$ .)

The second step is the use of Mellin inversion. By the Mapping theorem, the poles of  $\Lambda(s)g^*(s)x^{-s}$  correspond to an (asymptotic) expansion of  $G(x)$  at zero or infinity, depending on the half-plane (left or right) selected. Hence *we can evaluate the harmonic sum  $G(x)$  by computing and adding the residues in the left or right half-plane*. (There are again technical conditions that need to be fulfilled for this statement to hold; these are summarized as follows. If the residues that contribute to the expansion of  $G(x)$  are located in a strip  $\langle \gamma, \beta \rangle$ , the transform  $g^*(s)$  of the base function must decay faster than any negative power of  $|s|$  in this strip and on its boundary – we say that  $g^*(s)$  is of *fast decrease* – and the Dirichlet series must not grow faster than some fixed power of  $|s|$  in this region – we say that  $\Lambda(s)$  is of *slow increase*.) This principle is known as the *generalized Mellin summation formula*, i.e.

$$G(x) = \sum_k \lambda_k g(\mu_k x) \sim \pm \sum_{\varsigma \in \text{Sing}(\Lambda(s)g^*(s)x^{-s}) \cap H} \text{Res}(\Lambda(s)g^*(s)x^{-s}; s = \varsigma),$$

where  $H$  is the half-plane to the left (right) of the fundamental strip and the sign positive (negative) for an asymptotic expansion about zero (infinity).

It must be pointed out that this step relies critically on the concept of *analytic continuation*. The Dirichlet generating function of the  $\lambda_k$  has a half-plane of convergence to the right of an abscissa  $\sigma_c$ . The transform  $g^*(s)$  of the base function  $g(x)$  is defined in a fundamental strip  $\langle \alpha, \beta \rangle$ . The intersection of the half-plane and the strip must be non-empty if the Mellin summation formula is to apply. In order to evaluate  $G(x)$  in terms of the poles of  $G^*(s)$  in  $H$ , we require an *analytic continuation* of  $G^*(s)$  beyond the fundamental strip, into all of  $H$ , i.e. a function that matches  $G^*(s)$  in the fundamental strip, and is meromorphic in  $H$ . Section 2.4.1 explains this concept and defines the relevant terms.

**Example.** The Dirichlet series  $\sum_{k=0} 1/4^{ks}$  converges for  $\sigma > 0$ . However, the function  $4^s/(4^s - 1)$  is meromorphic in all of  $\mathbb{C}$  with poles at  $2\pi ik/\log 4$ ,  $k \in \mathbb{Z}$  and agrees with  $\sum_{k=0} 1/4^{ks}$  on  $\sigma > 0$ . It defines the analytic continuation of the latter.

### 1.3 The Mellin-Perron formula

Recall that we motivated the study of harmonic sums by considering the example

$$\sum_{1 \leq k < n} \lambda_k \left(1 - \frac{k}{n}\right)^m.$$

This type of sum is evaluated by the Mellin-Perron formula. The Mellin-Perron formula is a specific instance of generalized Mellin summation, as discussed above. The evaluation paradigm of harmonic sums applies, i.e. we begin by computing the Mellin transform of the harmonic sum, we invert the integral and obtain an expansion in terms of the poles, and we map this expansion to an expansion of the sum. The Mellin-Perron formula encapsulates the Mellin-transform of the harmonic sum via the Mellin inversion integral. It is no longer necessary to use the Mellin transform explicitly when the Mellin-Perron formula is applied.

We will begin our discussion with some historical remarks. As pointed out, the standard derivation of the Mellin-Perron formula is by means of harmonic sums and Mellin inversion. In fact this proof can be presented in a more straightforward setting. We include this traditional proof because it shows by what process the Mellin-Perron formula might originally have been discovered; we will then present the Mellin transform framework.

The traditional proof uses the “discontinuous factor” described by the following lemma.

**Lemma.**

$$\phi(y) = \frac{1}{2\pi i} \int_{c-i\infty}^{c+i\infty} \frac{y^s}{s(s+1)\cdots(s+m)} ds = \begin{cases} \frac{1}{m!} \left(1 - \frac{1}{y}\right)^m & \text{if } 1 \leq y \\ 0 & \text{if } 0 < y \leq 1, \end{cases}$$

where  $y \in \mathbb{R}^+$ ,  $m \in \mathbb{Z}^+$  and  $c \geq 1$ .

This “discontinuous factor” appears early in the mathematical literature, in various forms and without always being identified as a theorem; for example, Landau used it repeatedly in a series of papers published from 1915 onwards.

The above equality for the discontinuous factor  $\phi(y)$  is easily verified with the Cauchy residue theorem.

**Proof.** There are two cases.

**Case 1.**  $1 \leq y$ .

The term

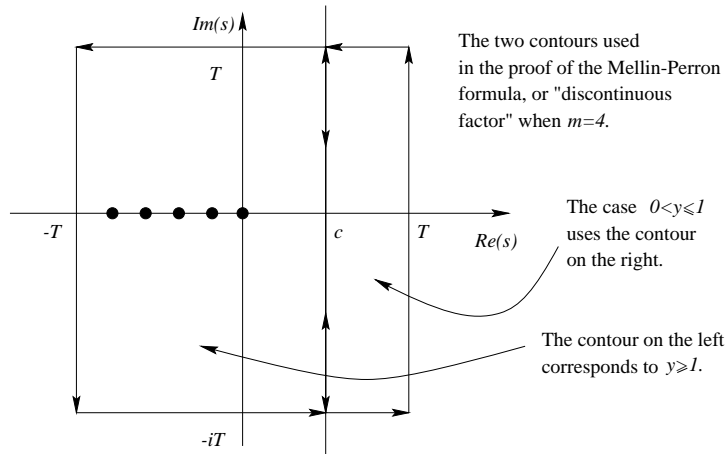
$$\frac{y^s}{s(s+1)\cdots(s+m)}$$

is meromorphic with residues

$$\frac{y^{-k}}{(-k)(-k+1)\cdots(-k+k-1)(-k+k+1)\cdots(-k+m)} = y^{-k} \frac{(-1)^k}{k!(m-k)!}$$

where  $0 \leq k \leq m$ . Therefore the sum of these residues is

$$\sum_{k=0}^m y^{-k} \frac{(-1)^k}{k!(m-k)!} = \frac{1}{m!} \sum_{k=0}^m \binom{m}{k} \left(-\frac{1}{y}\right)^k 1^{m-k} = \frac{1}{m!} \left(1 - \frac{1}{y}\right)^m.$$



Now consider the left contour shown in the diagram. The integral along the vertical segment at  $c$  in the right-half plane approaches

$$\int_{c-i\infty}^{c+i\infty} \frac{y^s}{s(s+1)\cdots(s+m)} ds$$

as  $T$  goes to infinity. Along the two horizontal segments from  $-T \pm iT$  to  $c \pm iT$ , the integrand is bounded by

$$\frac{y^\sigma}{T^m}$$

and because the term

$$\frac{1}{1+\sigma} \frac{y^{1+\sigma}}{T^m}$$

with  $\sigma = -T$ ,  $\sigma = c$  vanishes as  $T$  goes to infinity (recall that  $1 \leq y$ ), the contribution from these two segments is zero. The integrand is bounded by

$$\frac{y^{-T}}{T(T-1)\cdots(T-m)}$$

on the vertical segment in the left half-plane; hence the integral is bounded by

$$\frac{2y^{-T}}{(T-1) \cdots (T-m)}$$

and its contribution is zero also. (We used  $1 \leq y$  for the second time.)

**Case 2.**  $0 < y \leq 1$ .

Consider the contour in the right half-plane. Along the horizontal segments we may use the same bound as in the first case, with  $\sigma = c$  and  $\sigma = T$ ; hence these integrals vanish ( $0 < y \leq 1$ ). The integrand is bounded by

$$\frac{y^T}{T(T+1) \cdots (T+m)}$$

on the vertical segment in the right half-plane; its contribution is zero because  $0 < y \leq 1$ .

The principal feature of the “discontinuous factor” is that it can be used to evaluate finite sums. Suppose we have a finite sum over the indices  $k$  from 1 to  $n-1$ . Evidently  $\phi(y)$  is non-zero if  $1/y$  lies in  $(0, 1)$  and zero otherwise. We need only find a map such that the set  $\{1, \dots, n-1\}$  maps to a subrange of  $(0, 1)$  and  $\{n, n+1, \dots\}$  to a subrange of  $[1, \infty)$ . Clearly  $1/y = k/n$  is such a map. We obtain

$$\frac{1}{2\pi i} \int_{c-i\infty}^{c+i\infty} \frac{1}{k^s} \frac{n^s}{s(s+1) \cdots (s+m)} ds = \begin{cases} \frac{1}{m!} \left(1 - \frac{k}{n}\right)^m & \text{if } k < n \\ 0 & \text{if } n \leq k. \end{cases}$$

By a formal argument we finally have

$$\frac{1}{m!} \sum_{k=1}^{n-1} \lambda_k \left(1 - \frac{k}{n}\right)^m = \frac{1}{2\pi i} \int_{c-i\infty}^{c+i\infty} \left(\sum \frac{\lambda_k}{k^s}\right) \frac{n^s}{s(s+1) \cdots (s+m)} ds.$$

This is the **Mellin-Perron formula**. (In practice we need to verify the convergence of the Dirichlet series on the right; the Mellin transform framework provides the necessary apparatus.) The mathematical process described here can be summarized as follows.

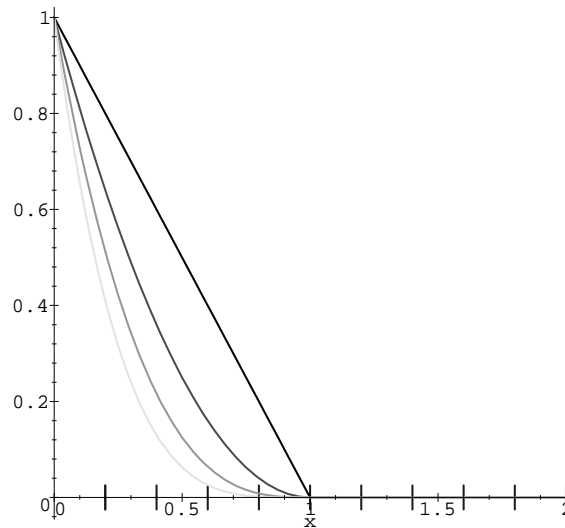
- The “discontinuous factor”  $\phi(y)$ ,  $y$  positive, is zero when  $y$  is less than or equal to one. It partitions the real line into a lower interval, where it vanishes, and an upper interval, where it is a polynomial in  $1/y$ , of degree  $m$ .
- If the terms of an infinite sum in  $k$  are multiplied by this factor and  $y = y(k)$ , only those terms with  $y(k)$  in the upper interval contribute to the sum.

- Hence  $\phi(y)$  can be used to evaluate a finite subrange of a sum of terms of the form  $\lambda_k \phi(y(k))$ , where  $y(k)$  is used to select the range.
- The result is an integral along a vertical line in the complex plane, and the Cauchy Residue theorem applies. The singularities of

$$\left( \sum \frac{\lambda_k}{k^s} \right) \frac{n^s}{s(s+1) \cdots (s+m)}$$

in the left or right half-plane can be used to compute a formula for the sum.

The Mellin-transform framework supplies a more general and more precise context for these observations. The underlying motivation stays the same. In particular, the Mellin-transform view adds two additional perspectives. One, that the Mellin-Perron formula is a specific instance of harmonic sum formulas, and hence, two, that its evaluation corresponds to Mellin inversion. The latter allows us to use the Mapping theorem; in other words, an evaluation in the left half-plane corresponds to an expansion of the harmonic sum at zero, one in the right half-plane, to an expansion at infinity. We present the Mellin-transform framework below.



Plot of  $g(x)$  for  $m \in \{1, 2, 3, 4\}$ . The vertical marks on the  $x$  axis show which  $k/n$  contribute to  $G(x)$  when  $n = 5$ .

We wish to evaluate the harmonic sum

$$\sum_{1 \leq k < n} \lambda_k \left( 1 - \frac{k}{n} \right)^m$$

where  $m, n \in \mathbb{Z}^+$ . This is equivalent to

$$\sum_1^{\infty} \lambda_k g\left(\frac{k}{n}\right)$$

where

$$g(x) = \begin{cases} (1-x)^m & \text{if } 0 < x \leq 1 \\ 0 & \text{otherwise.} \end{cases}$$

(The choice of  $g(x)$  corresponds to the “discontinuous factor”;  $g(x)$  is the polynomial that restricts the sum to the first  $n-1$  terms. The graph explains the process.) It is not difficult to see that

$$\mathfrak{M}[g(x); s] = \frac{1}{s(s+1) \cdots (s+m)}.$$

Evidently the sum

$$\sum_1^{\infty} \lambda_k g\left(\frac{k}{n}\right)$$

is a harmonic sum  $G(x)$  of the form

$$\sum_1^{\infty} \lambda_k g(kx)$$

with amplitudes  $\lambda_k$ , frequencies  $\mu_k = k$  and evaluated at  $x = 1/n$ . Therefore the transform function  $G^*(s)$  is

$$\Lambda(s) \frac{1}{s(s+1) \cdots (s+m)}.$$

By Mellin inversion we thus have

$$G(x) = \frac{1}{2\pi i} \int_{c-i\infty}^{c+i\infty} \Lambda(s) \frac{x^{-s}}{s(s+1) \cdots (s+m)} ds$$

and in particular

$$G\left(\frac{1}{n}\right) = \sum_{1 \leq k < n} \lambda_k \left(1 - \frac{k}{n}\right)^m = \frac{1}{2\pi i} \int_{c-i\infty}^{c+i\infty} \Lambda(s) \frac{n^s}{s(s+1) \cdots (s+m)} ds.$$

This is the Mellin-Perron formula.

The alert reader will undoubtedly have wondered why the Mellin inversion integral is used at all; it appears that we could have omitted this step and evaluated  $G(x)$  by the generalized Mellin summation formula. The integral is necessary because the transform  $g^*(s)$  of the base function  $g(x)$  is of order  $\mathcal{O}(|s|^{-m})$  in any vertical strip; it is not of fast decrease. (E.g. we cannot compute the cost of Karatsuba multiplication by the Mellin-Perron formula because the Dirichlet generating function-transform pair is not of fast decrease.) There will usually be an error term of some kind, which is generated by the integral at the lower (upper) end of the evaluation strip.

### 1.3.1 A classical example III: the mergesort recurrence

Until this point we have developed a series of instantiations. We began with the framework of generating functions, and discussed some principal instances, namely ordinary, exponential and Dirichlet generating functions. Harmonic sums are a specific application of Dirichlet generating functions. We discussed the Mellin-transform methods that are used to evaluate these sums. Finally, we fixed yet another parameter and restricted ourselves to harmonic sums that can be evaluated by the Mellin-Perron formula. We have now completed our descent along increasingly specific branches of mathematics. The next level of instantiation will yield leaves, i.e. the four results of this thesis, which use the Mellin-Perron formula to evaluate harmonic sums of the type

$$\sum_{1 \leq k < n} \lambda_k \left(1 - \frac{k}{n}\right)$$

and

$$\sum_{1 \leq k < n} \lambda_k \left(1 - \frac{k}{n}\right)^2.$$

Note that we restrict our attention to the cases  $m = 1$  and  $m = 2$ .

We ask the following two questions. Why does the Mellin-Perron formula apply to problems as diverse as evaluating digital sums or counting lattice points, i.e. what common characteristic do these problems have? The second question concerns the fact that three out of the four problems that we will treat lead to Fourier series expansions of the harmonic sum. The mergesort recurrence also leads to Fourier series. Therefore we ask what feature of the problem induces a Fourier series.

These two questions have simple answers. The Mellin-Perron formula applies to iterated sums of a sequence  $\{\lambda_k\}$  when the corresponding Dirichlet generating function  $\Lambda(s)$  can be analytically continued into the left half-plane, where it must have a closed form that can be used to find the poles, and hence an expansion of the sum. If there is an infinite number of poles, and they are regularly spaced, one or more Fourier series result. If the sequence  $\{\lambda_k\}$  satisfies a recurrence of the form

$$\begin{aligned} \lambda_{lm} &= r\lambda_m + e_{lm} \\ \lambda_{lm+1} &= e_{lm+1} \\ \lambda_{lm+2} &= e_{lm+2} \\ &\dots \\ \lambda_{m+l-1} &= e_{m+l-1}, \end{aligned}$$

where  $m > 1$ , and the fluctuation  $\{e_k\}$  at  $k = lm + r$ ,  $0 \leq r < l$  is independent of  $m$ , i.e. periodic,

together with a set of initial values

$$\lambda_1 = e_1, e_2, e_3 \dots e_{l-1},$$

or if it is a linear combination of such terms, then  $\Lambda(s)$  has an infinite number of regularly spaced poles. Here  $r$  is a non-zero constant and  $l$  is a positive integer larger than or equal to two. We say that the sequence  $\{\lambda_k\}$  is *multiplicatively self-similar* with scale factor  $l$  and fluctuation  $\{e_k\}$ . This terminology will be motivated later; we point out that different types of fluctuation induce various *fractal* properties of the Fourier series, where the term *fractal*<sup>1</sup> is taken to mean both self-similarity up to scaling, and the state of being “broken”, i.e. non-differentiable, at a dense set of points.

In order to make these notions more precise, we will study an example, namely the behavior of the mergesort algorithm. We will develop a framework that allows us to determine whether the Mellin-Perron formula applies to a given problem, and whether a Fourier series occurs in the result. Finally we will use this framework to give a unified presentation of the results of this thesis, and some of their history.

The problem is the following. We are given an array of integers that contains some permutation of the integers from 1 to  $n$ . All  $n!$  permutations are equally likely. We wish to sort this array in nondecreasing order. We use mergesort; i.e. we recursively sort the bottom and top halves of the array, and merge the two halves. The single-element array is already sorted. We are interested in the number of comparisons that this algorithm uses in the best, average and worst cases.

We need to be more specific about the merge step. Two sorted arrays of size  $\mu$  and  $\nu$  are merged as follows. We compare the last elements of the two source arrays; the larger of the two is removed from the source array and placed at the highest empty slot of the target array, which has size  $\mu + \nu$ . When one of the source arrays is exhausted, we copy the other one to the target array. The number of comparisons is  $\mu + \nu - S$ , where  $S$  is the number of elements left in the unexhausted source array.

The number of comparisons is given by the following recurrence

$$T(n) = T\left(\left\lfloor \frac{n}{2} \right\rfloor\right) + T\left(\left\lceil \frac{n}{2} \right\rceil\right) + e_n,$$

where it is immediate that  $e_n = n - 1$  in the worst case, and  $e_n = n - \lfloor n/2 \rfloor = \lceil n/2 \rceil$  in the best case. It is an instructive exercise to show that

$$e_n = n - \frac{\lfloor n/2 \rfloor}{\lfloor n/2 \rfloor + 1} - \frac{\lceil n/2 \rceil}{\lceil n/2 \rceil + 1}$$

---

<sup>1</sup>Lat. *fractus*, p.part. of *frangere*, to break.

in the average case. The base case is  $T(1) = 0$  for all three types. (The reader is referred to the chapter notes for more information.) The salient feature of this analysis is that the problem has been reduced to a *divide-and-conquer* recurrence of the type

$$f_n = f_{\lfloor n/2 \rfloor} + f_{\lceil n/2 \rceil} + e_n.$$

Now let  $e_0 = e_1 = f_0 = 0$  and define  $\nabla a_n = a_n - a_{n-1}$ ,  $\Delta a_n = a_{n+1} - a_n$ , for  $\{a_n\}$  any sequence. After a few simple manipulations (see section 2.9.3), we find that

$$\begin{aligned}\Delta \nabla f_{2m} &= \Delta \nabla f_m + \Delta \nabla e_{2m} \\ \Delta \nabla f_{2m+1} &= \Delta \nabla e_{2m+1}\end{aligned}$$

and  $\Delta \nabla f_1 = e_2 = \Delta \nabla e_1$ . Furthermore, the inverse of the  $\Delta \nabla$  operator is given by the *iterated sum*

$$\sum_{k=1}^{n-1} \sum_{l=1}^k \Delta \nabla f_l = \sum_{k=1}^{n-1} (n-k) \Delta \nabla f_k = f_n - n f_1.$$

We make the following two critical observations.

- The solution  $\{f_n\}$  of the mergesort recurrence, or more generally of a divide-and-conquer recurrence of the type

$$f_n = f_{\lfloor n/2 \rfloor} + f_{\lceil n/2 \rceil} + e_n$$

is an *iterated sum* of  $\Delta \nabla f_k$ .

- The sequence  $\{\Delta \nabla f_k\}$  is *multiplicatively self-similar* with scale factor  $l = 2$  and fluctuation  $\{\Delta \nabla e_k\}$ .

These observations have two consequences. First, the iterated sum

$$\sum_{k=1}^{n-1} (n-k) \Delta \nabla f_k = f_n - n f_1$$

can be rewritten as the harmonic sum

$$\sum_{k=1}^{n-1} \Delta \nabla f_k \left(1 - \frac{k}{n}\right) = \frac{1}{n} f_n - f_1,$$

and hence the Mellin-Perron formula for  $m = 1$  applies. Second, the Dirichlet generating function

$$W(s) = \sum \frac{\Delta \nabla f_k}{k^s} \quad \text{has the form} \quad W(s) = \frac{1}{1-2^{-s}} \sum \frac{\Delta \nabla e_k}{k^s}$$

or

$$W(s) = \frac{\Xi(s)}{1 - 2^{-s}} \quad \text{where} \quad \Xi(s) = \sum \frac{\Delta \nabla e_k}{k^s}$$

and hence  $W(s)$  has, in addition to the poles of  $\Xi(s)$ , a sequence of regularly spaced poles generated by

$$\frac{1}{1 - 2^{-s}}.$$

These poles are at

$$s = \chi_k = \frac{2k\pi i}{\log 2} \quad \text{where} \quad k \in \mathbb{Z}$$

and translate back into a Fourier series for  $f_n$ .

We combine these two observations into the following statement. If  $\{f_n\}$  is defined by a divide-and-conquer recurrence of the type

$$f_n = f_{\lfloor n/2 \rfloor} + f_{\lceil n/2 \rceil} + e_n$$

where  $e_n \in \mathcal{O}(n^r)$ , then  $f_n$  satisfies

$$f_n = n f_1 + \frac{n}{2\pi i} \int_{c-i\infty}^{c+i\infty} \frac{\Xi(s)n^s}{1 - 2^{-s}} \frac{ds}{s(s+1)},$$

where  $c > r + 1$ . (The condition on  $c$  ensures that the Dirichlet series for  $\Xi(s)$  converges; consult section 2.6 for more information.)

It remains to explain how the poles of the integrand translate into a Fourier series. Recall that this translation is accomplished by the secondary map

$$\frac{A}{(s - \zeta)^{\kappa+1}} \xrightarrow{\text{Res}(f(s)x^{-s}; s = \zeta)} A \frac{(-1)^\kappa}{\kappa!} x^{-\zeta} (\log x)^\kappa.$$

Suppose that  $\Xi(s)/(s(s+1))$  is analytic at  $\chi_k$ . Then  $s = \chi_k$  is a simple pole of  $(1 - 2^{-s})^{-1}$ . Hence we are treating the case

$$x = \frac{1}{n}, \quad A = \frac{\Xi(\chi_k)}{\chi_k(\chi_k + 1)} \lim_{s \rightarrow \chi_k} \frac{s - \chi_k}{1 - 2^{-s}}.$$

It follows that the terms of the Fourier series are given by

$$\frac{1}{\log 2} \frac{\Xi(\chi_k)}{\chi_k(\chi_k + 1)} n^{2k\pi i / \log 2},$$

i.e. we have a Fourier series  $F(u)$  in  $\log_2 n$ ,

$$F(\log_2 n) = \sum \frac{1}{\log 2} \frac{\Xi(\chi_k)}{\chi_k(\chi_k + 1)} e^{2k\pi i u},$$

where the sum is over all  $k$  such that  $\chi_k$  is a simple pole. The other poles of the Mellin-Perron integrand, such as the one at  $s = -1$ , must also be included in the formula for  $f_n$ , but except for the one at  $s = 0$ , they do not contribute to the Fourier series.

We point out that a multiplicative self-similarity with scale factor  $l$  will result in a Fourier series in  $\log_l n$ . The argument is precisely the same as in the case  $l = 2$  that we studied above.

We return to our case study of the mergesort recurrence. The above observations apply. It is a matter of computation, straightforward for the best and worst cases, considerably more subtle in the average case, to obtain the Fourier series and the terms that correspond to the poles of  $1/s(s+1)$  and  $\Xi(s)$ . Once these are computed, we have an exact expression for the three types of complexity, with a dominant term in  $\theta(n \log_2 n)$  and a Fourier series for the fluctuation around this term. The reader should have no difficulty understanding or re-deriving the following statement.

Let  $T(n)$ ,  $Y(n)$  and  $U(n)$  be the worst, best and average-case complexity of mergesort. Let

$$\chi_k = \frac{2k\pi i}{\log 2}.$$

Define the sequences  $\{a_k\}$ ,  $\{d_k\}$  and  $\{b_k\}$ , as follows:

	$k = 0$	$k \in \mathbb{Z} \setminus \{0\}$
$a_k$	$\frac{1}{2} - \frac{1}{\log 2}$	$\frac{1}{\log 2} \frac{1}{\chi_k(\chi_k+1)}$
$d_k$	$\log_2 \sqrt{\pi} - \frac{1}{2 \log 2} - \frac{1}{4}$	$-\frac{1}{\log 2} \frac{\zeta(\chi_k)}{\chi_k(\chi_k+1)}$
$b_k$	$\frac{1}{2} - \frac{1}{\log 2} - \frac{1}{\log 2} \sum \frac{2}{(m+1)(m+2)} \log \left( \frac{2m+1}{2m} \right)$	$\frac{1}{\log 2} \frac{1+\Psi(\chi_k)}{\chi_k(\chi_k+1)}$

where  $\zeta(s)$  is the Riemann zeta function (see also section 2.6.1) and  $\Psi(s)$  is given by

$$\Psi(s) = \sum \frac{2}{(m+1)(m+2)} \left( -\frac{1}{(2m)^s} + \frac{1}{(2m+1)^s} \right).$$

Let  $A(u)$ ,  $D(u)$  and  $B(u)$  be the Fourier series

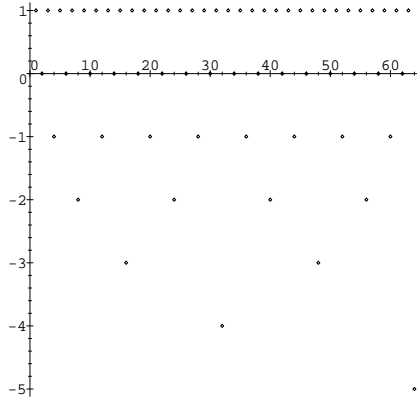
$$A(u) = \sum_{k \in \mathbb{Z}} a_k e^{2k\pi i u}, \quad D(u) = \sum_{k \in \mathbb{Z}} d_k e^{2k\pi i u}, \quad B(u) = \sum_{k \in \mathbb{Z}} b_k e^{2k\pi i u}.$$

The worst, best and average case complexities of mergesort are given by

$$\begin{aligned} T(n) &= n \log_2 n + nA(\log_2 n) + 1, \\ Y(n) &= \frac{1}{2}n \log_2 n + nD(\log_2 n), \\ U(n) &= n \log_2 n + nB(\log_2 n) + \epsilon(n), \end{aligned}$$

where  $\epsilon(n) \in \mathcal{O}(1)$ .

**Self-similarity in  $\lambda_k$  and fractal characteristics of the corresponding Fourier series**



Graph of a sequence  $\{\Delta \nabla f_k\}$  that is *multiplicatively self-similar* with scale factor  $l = 2$  and fluctuation  $\Delta \nabla e_k = -(-1)^k$ .

We wish to explain the terms *self-similarity* and *fractal*. The first question that we must answer is this. Why should a sequence  $\{\Delta \nabla f_k\}$  that satisfies the recurrence

$$\Delta \nabla f_{2m} = \Delta \nabla f_m + \Delta \nabla e_{2m} \quad \text{and} \quad \Delta \nabla f_{2m+1} = \Delta \nabla e_{2m+1}$$

be termed *multiplicatively self-similar*? We can motivate this terminology by considering the best-case behaviour of mergesort. Recall that this is the case where  $e_n = \lfloor n/2 \rfloor$  and hence  $\Delta \nabla e_k = -(-1)^k$ . The recurrence becomes

$$\Delta \nabla f_{2m} = \Delta \nabla f_m - 1 \quad \text{and} \quad \Delta \nabla f_{2m+1} = 1.$$

(Incidentally, this implies that  $\Delta \nabla f_k = 1 - v_2(k)$ , where  $v_2(k)$  is the exponent of the largest power of two that divides  $k$ . This in turn implies  $Y(n) = \sum_{m < n} v(n)$ , where  $v(n)$  is the sum of the binary digits of  $n$ . The latter identity is worth pointing out, because it suggests an alternate method to study  $Y(n)$ . We will return to this subject when we discuss digital sums.) The above recurrence induces two kinds of self-similarity, both readily apparent from on the graph shown at the beginning of this section. (In the following, we use an operator  $\sigma$  that maps a number  $k$  to a sequence  $\sigma_1(k), \sigma_2(k), \dots, \sigma_l(k)$ , where the  $\sigma_i$  are functions  $\mathbb{N} \mapsto \mathbb{N}$  or  $\mathbb{R} \mapsto \mathbb{R}$ . The significance of this operator will be explained below.)

- (The whole and its parts.) If we skip all odd  $k$ , “compress” the remaining even values by a factor of two, i.e. by  $n \mapsto n/2$ , and subtract one, we obtain the original sequence. This immediately implies the existence of regularly spaced poles, and hence, a Fourier series when an iterated sum of the sequence is evaluated by the Mellin-Perron formula. Why? In terms of the Dirichlet generating

function, to skip the odd  $k$  is to split the function into two sums and to “compress” is to express the even sum as an exponential factor times the original generating function. The difference between the two terms introduces the regularly spaced poles.

- (Generating the sequence.) If we start with the one-element sequence  $\{1\}$  and append a copy, subtracting one from the last element of the copy, we get  $\{1, 0\}$ . If we iterate this operator, we obtain successively

$$\{1\}, \quad \{1, 0\}, \quad \{1, 0, 1, -1\}, \quad \{1, 0, 1, -1, 1, 0, 1, -2\},$$

etc., i.e. length  $2^r$  prefixes of the sequence. This can be summarized by the initial-value/operator pair

$$1; \quad k \mapsto 1, k - 1.$$

It is not difficult to see that how the two kinds of similarity can be obtained from one another. Both correspond to multiplicative self-similarities where the fluctuation  $\{e_k\}$  at  $k = lm + r$ ,  $0 \leq r < l$  is independent of  $m$ , i.e. periodic. Consider the operator  $\sigma$  and the initial value  $f_1$ . If  $\sigma$  has the property that  $\sigma_1(f_1) = f_1$ , then it generates the sequence  $\{f_n\}$

$$\{f_1\}, \quad \{\sigma_1(f_1), \dots, \sigma_l(f_1)\}, \quad \{\sigma_1(\sigma_1(f_1)), \dots, \sigma_l(\sigma_1(f_1)), \dots, \sigma_1(\sigma_l(f_1)), \dots, \sigma_l(\sigma_l(f_1))\}$$

etc., or

$$\{f_1\}, \quad \{f_1, \dots, \sigma_l(f_1)\}, \quad \{f_1, \dots, \sigma_l(f_1), \dots, \sigma_1(\sigma_l(f_1)), \dots, \sigma_l(\sigma_l(f_1))\}$$

etc. On the other hand, a sequence  $\{f_n\}$  that is multiplicatively self-similar with scale factor  $l$  and  $\{e_k\}$  periodic may be written as

$$f_1, e_2, \dots, e_{l-1}, rf_1 + e_l, e_{l+1}, \dots, e_{2l-1}, rf_2 + e_l, e_{l+1}, \dots, e_{2l-1}, \dots$$

It follows that the initial-value/operator pair

$$f_1; \quad f_1 \mapsto f_1, e_2, \dots, e_{l-1}, rf_1 + e_l; \quad k \mapsto e_{l+1}, e_{l+2}, \dots, e_{2l-1}, rk + e_l$$

generates  $\{f_n\}$ , where we have selected the range  $l$  to  $2l - 1$  to represent the period of  $\{e_k\}$ . Note that this imposes a constraint on  $e_{l+1}$  in the case when there exists an  $n > 1$  such that  $f_n = f_1$ . If such an  $n$  exists, we can construct the generator representation of  $\{f_n\}$  by means of an operator  $\sigma$  iff  $e_{l+1} = f_1$ .

The second issue that we must address is the meaning of *fractal* when applied to the Fourier series rather than to the sequence  $\{\lambda_k\}$  itself. In this context the term refers to a continuous function, which is not differentiable at a dense set of points. We now clarify this remark. The dense set of points in question is always the set of  $l$ -adic rationals  $p/l^r$ , e.g. when the scale factor is  $l = 2$ , we study the differentiability of the Fourier series  $F(u)$  at the dyadic rationals  $p/2^r$ . These numbers form a dense set. If we can show that  $F(u)$  is continuous and non-differentiable at the  $l$ -adic rationals, we have proved that  $F(u)$  is fractal in the function-theoretic sense. The first half is usually easy, because the exponential terms of the series are continuous in  $u$ ; therefore it only remains to check the integers. The second part is more difficult; we must show that the limit

$$\lim_{h \rightarrow 0} \frac{F(u+h) - F(u-h)}{2h}$$

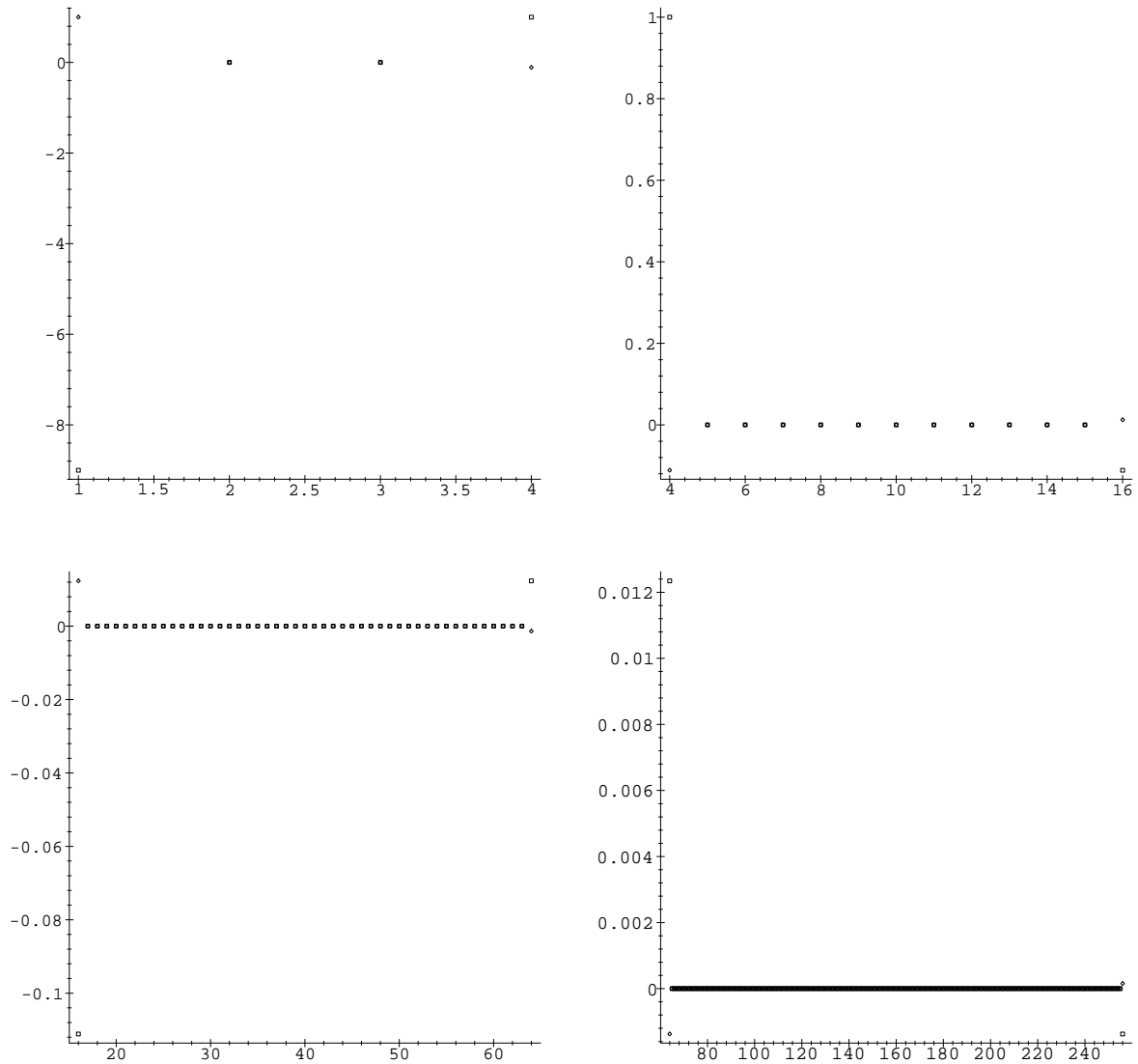
does not exist for  $u$  a dyadic rational. It is sometimes possible to prove a stronger result, namely that  $F(u)$  is nowhere differentiable. In those cases (we will encounter one in a digital sum shortly), we let  $u \pm h$  approach  $u$  along the respective set of rationals. As an additional criterion, we note that a constant error term  $\Xi(s)$  implies that  $F(u)$  is not a fractal function.

Recall that  $B(u)$  is the Fourier series for the average-case behavior of mergesort. We can show that  $B(u)$  is continuous, and that it has a cusp (read: is non-differentiable) at any dyadic rational  $p/2^r$ . Hence we are justified in describing the average-case fluctuation as *fractal*.

### Format of the remainder of this chapter

We will introduce and discuss the four specific applications of the Mellin-Perron formula that we study in this thesis. Recall that the two criteria for applicability of the Mellin-Perron formula are, first, that we must be concerned with an iterated sum of a sequence  $\{\lambda_k\}$ , and second, that the sequence  $\{\lambda_k\}$  should be multiplicatively self-similar. Therefore we will present each application by first describing the problem, then the sum involved, and finally, we will verify that the underlying  $\{\lambda_k\}$  is multiplicatively self-similar. We also comment on the fractal nature of the Fourier series obtained. We will conclude each section with a historical perspective, where such perspectives are relevant. The notes to this chapter document the historical material.

1.3.2 A fractal ornament



Graph of the two sequences that combine linearly into  $\{\Delta \nabla a_k\}$ . The scale factor is  $l = 4$  and the fluctuation is  $\Delta \nabla e_k = 0$ .

The first application of the Mellin-Perron formula that we consider is the fractal ornament pictured and described in section 3.1. This ornament is obtained by starting with a equilateral triangle with unit side length. We then proceed to replace the sides of the triangle while moving around the triangle in a counterclockwise fashion. Every line segment is replaced by four smaller ones of equal length. The outer two of the four are collinear with the original segment. They are the two segments that remain when we remove the middle third of the original. The middle third is replaced by two line segments such that the

replaced segment and the two new ones form an equilateral triangle. The number of turns taken around the circumference determines whether this new, smaller triangle lies inside or outside of the previous figure. We alternate between inward and outward; the first turn is inward.

We wish to compute the area of the ornament after  $n - 1$  segments have been replaced. Let  $a_n$  be this area. A few straightforward manipulations show that

$$\begin{aligned}\Delta\nabla a_{4^r} &= \frac{\sqrt{3}}{4} \frac{1}{9} \left( \left(-\frac{1}{9}\right)^{r-1} - \left(-\frac{1}{9}\right)^r \right) & (r \geq 0) \\ \Delta\nabla a_n &= 0 & (\text{otherwise}).\end{aligned}$$

This identity implies that  $a_n$  is

- an iterated sum of  $\Delta\nabla a_n$ , and
- that  $\Delta\nabla a_n$  is multiplicatively self-similar with scale factor  $l = 4$ .

The last statement perhaps requires some explanation. The above equation for  $\Delta\nabla a_n$  shows that it is a linear combination of two sequences  $\{\alpha_k\}$  with different initial values  $\alpha_1$ . The graphs of these two sequences are included at the beginning of this section.

It is easy to verify that the recurrence for  $\{\alpha_k\}$  is

$$\alpha_{4m} = -\frac{1}{9}\alpha_m, \quad \alpha_{4m+1} = 0, \quad \alpha_{4m+2} = 0, \quad \alpha_{4m+3} = 0,$$

with the two initial values  $\alpha_1 = -9$  and  $\alpha_1 = 1$ , respectively. Alternatively we can use the characterizations

$$-9; \quad -9 \mapsto -9, 0, 0, 1; \quad k \mapsto 0, 0, 0, -\frac{1}{9}k$$

and

$$1; \quad 1 \mapsto 1, 0, 0, -\frac{1}{9}; \quad k \mapsto 0, 0, 0, -\frac{1}{9}k,$$

which are obtained from our earlier translation rule for the two kinds of self-similarity. Hence  $\Delta\nabla a_k$  is indeed multiplicatively self-similar, as claimed. Because the scale factor is  $l = 4$ , we can expect a Fourier series in  $\log_4 n$ .

Will the Fourier series be a fractal function? The Dirichlet generating function  $A(s)$  of  $\Delta\nabla a_k$  is

$$A(s) = -\frac{\sqrt{3}}{4} 10 \frac{4^s}{9 \cdot 4^s + 1},$$

hence the error term  $\Xi(s)$  is constant.  $F(u)$  will not be a fractal function.

## History and background

The problem of computing the area of a fractal ornament at each step in its construction appears to be new. It is one of two problems in this thesis (the other one is study of integer lattice points inside a paraboloid) that were discovered by the following procedure. First, we know that iterated sums of multiplicatively self-similar sequences can be treated by the Mellin-Perron formula. *Ergo* we ask, second, where do these sums occur, such that the underlying sequence has a Dirichlet generating function with easy-to-locate poles. Thus the object is to find a sequence that also has a meaningful, and ideally interesting concrete interpretation – we build the problem to fit the solution.

There is nonetheless one useful question that we may ask, i.e. what is it about a fractal ornament that makes the area of successive approximations be the right kind of sum? In our example, the boundary of the ornament is clearly isomorphic to a tree, the branch factor being four. At every step, we replace a segment by four new ones, its children. Therefore we can account for all the segments introduced and replaced in the process by drawing a 4-tree, with each node bearing the coordinates of the endpoints of the segments. It so happens that the area added to, or subtracted from the ornament at every step stays constant at each level of the tree. The area at step  $n$  is the sum of these additions and subtractions. If they are constant, however, their difference will be zero except when we change from one level to the next. Moreover, this change is effected by scaling – we scale the current triangle by a constant factor. These are precisely the characteristics of a double sum of a multiplicatively self-similar sequence.

Although there is no background literature on computing the area enclosed by successive approximations to a fractal curve, there is of course extensive material on fractal curves themselves. The curve used in our example is inspired by the Koch curve. We informally sketch some of the characteristics of these two curves; we confine ourselves to *construction by similarities* as it applies to curves, and to *dimension*.

We will need some terminology. We begin with the basic observation that there are four different ways to map the oriented plane vector  $AB$  to  $CD$ , where we permit any combination of scaling, translation and reflection. There are two possibilities to map the end points by scaling and translating, i.e.  $\{\{A, C\}, \{B, D\}\}$  and  $\{\{A, D\}, \{B, C\}\}$ , which are combined with no reflection, or a reflection over  $CD$ . These four *similarities* may all be written as

$$F(x) = \rho Mx + b,$$

where  $\rho \in \mathbb{R}^+$ ,  $M$  is an orthonormal matrix and  $b$  a translation vector.

We use similarities in the plane in order to define self-similar curves. We start with  $N + 1$  distinct points in the plane,  $N \geq 2$ :

$$A = A_1, A_2, \dots, A_{N+1} = B,$$

such that the distance between pairs  $(A_i, A_{i+1})$  is less than the distance between  $A$  and  $B$ , and  $N$  similarities

$$F_i(AB) = A_i A_{i+1}.$$

We can prove that there exists a unique curve  $\Gamma$  such that

$$\Gamma = \bigcup_{i=1}^N F_i(\Gamma),$$

and we say that the polygon  $P_1$  on the vertices  $A_1, A_2, \dots, A_{N+1}$  is the *generator* of  $\Gamma$ . Informally we may take  $\Gamma$  to be the limit curve of the successive approximations

$$P_{k+1} = \bigcup_{i=1}^N F_i(P_k).$$

This iterative model also allows us to show the existence of a continuous parameterization of  $\Gamma$  such that  $\Gamma = \gamma(I)$ , where  $I = [0, 1]$ .

We have the following *simplicity criterion*, which permits us to test whether a given generator  $P_1$  leads to a self-intersecting curve. This criterion states that if there exists a closed bounded set  $D$  of non-null area such that  $F_i(D) \subset D$  for all  $i \in [1, N]$ , and that  $F_i(D)$  and  $F_{i+1}(D)$  only intersect in a single point,  $A_{i+1}$ , and that all other pairs  $F_i(D)$  and  $F_j(D)$  do not intersect at all, then  $\Gamma$  is simple.

The question of the *dimension* of a plane curve is answered by considerations that use an object known for better or for worse as the *Minkowski sausage* of  $\Gamma$ . This is the set  $\Gamma(\epsilon)$  of all points of a distance at most  $\epsilon$  from  $\Gamma$ , i.e.

$$\Gamma(\epsilon) = \bigcup_{(x,y) \in \Gamma} \{(x', y') \mid |(x', y') - (x, y)| \leq \epsilon\}.$$

We consider the area of  $\Gamma(\epsilon)$ , denoted  $\mathcal{A}(\Gamma(\epsilon))$ . For example, if  $\Gamma$  is a straight line, this area is  $2\epsilon L(\Gamma) + \pi\epsilon^2$ ; if  $\Gamma$  includes all the points of a unit square with side length  $a$ , it is  $a^2 + 4a\epsilon + \pi\epsilon^2$ ; if  $\Gamma$  is a single point it is  $\pi\epsilon^2$ . We compute the limit

$$\frac{\log \mathcal{A}(\Gamma(\epsilon))}{\log \epsilon},$$

and for the three cases get respectively,

$$\lim_{\epsilon \rightarrow 0} \frac{(2L(\Gamma) + 2\pi\epsilon)/(2\epsilon L(\Gamma) + \pi\epsilon^2)}{1/\epsilon} = 1, \quad \lim_{\epsilon \rightarrow 0} \frac{a^2}{\log \epsilon} = 0, \quad \text{and} \quad \lim_{\epsilon \rightarrow 0} \frac{\log \pi}{\log \epsilon} + 2 = 2.$$

When we take into account that the dimension of a point should be zero, that of a line one, and that of a square two, these examples motivate the following definition. The *upper and lower dimensions*  $\Delta(\Gamma)$  and  $\delta(\Gamma)$  of a curve  $\Gamma$  are

$$\Delta(\Gamma) = \limsup_{\epsilon \rightarrow 0} \left( 2 - \frac{\log \mathcal{A}(\Gamma(\epsilon))}{\log \epsilon} \right) \quad \text{and} \quad \delta(\Gamma) = \liminf_{\epsilon \rightarrow 0} \left( 2 - \frac{\log \mathcal{A}(\Gamma(\epsilon))}{\log \epsilon} \right).$$

There exists a charming theorem concerning the dimension of curves constructed by similarities. Recall that to every similarity  $F_i$  of the generator  $P_1$  we associate a number  $\rho_i$ , the *similarity ratio*. By definition of  $P_1$  we have  $\sum_{i=1}^N \rho_i \geq 1$ . If the  $F_i$  also fulfill the simplicity criterion, we have  $\sum_{i=1}^N \rho_i^2 \leq 1$ . We now use the fact that the function  $x \mapsto \sum_{i=1}^N \rho_i^x$  is strictly decreasing and continuous in  $[1, 2]$  to conclude that there exists a unique solution  $e$  of the equation  $\sum_{i=1}^N \rho_i^e = 1$  in  $[1, 2]$ , the *similarity exponent* of  $G$ . Bouligand. The theorem announced above states that

$$\sigma = \Delta(\Gamma) = \delta(\Gamma).$$

Armed with these definitions, we study two examples. The first is the Koch curve, which served as the inspiration to the fractal ornament problem. The second is the curve of which three copies actually delineate the ornament. (The reader may wish to skip ahead to section 3.1 for additional illustrations.)

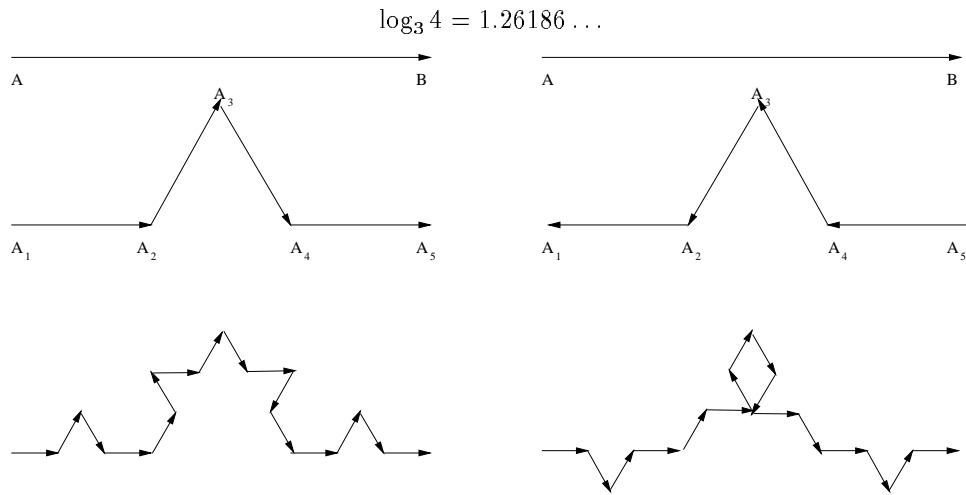
**Example.** The diagram at the end of this section illustrates the difference between the two curves. The Koch curve is on the left. The three rows show the start segment  $AB$ , the generator  $P_1$  and the first approximation  $P_2$ . The novelty of our area computation consists in the fact that we enumerate the approximations segment by segment, and not layer by layer.

The similarities  $F_1, F_2, F_3$  and  $F_4$  that define each curve are completely defined by (1) the orientation shown on the segment  $AB$  and the segments  $A_i A_{i+1}$  and (2) the fact that the curve always extends to the left of each oriented segment.

Clearly both curves submit to the similarity exponent equation

$$4 \left( \frac{1}{3} \right)^\sigma = 1$$

and hence they have dimension



Generators of the Koch curve, and the curve that delineates the fractal ornament.

The reader who wishes to learn more about self-similarity in curves, or fractals in general should consult the chapter notes.

### 1.3.3 Digital sums

The chapter on digital sums examines a variety of digital sum problems. We will study a representative example. The other digital sums that we evaluate are perhaps more involved, but the problem and the solution paradigm remain the same.

The canonical digital sum problem studies the sum of digits in the binary representation of a positive integer  $n$ . For example, if we take  $n = 19$ , which we may write as  $n = (19)_{10}$ , in order to indicate that base 10 is used, then we have  $n = (10011)_2$  and the sum of digits in binary  $v(n)$  evaluates to  $v(19) = 1 + 0 + 0 + 1 + 1 = 3$ . Binary digital sums have been studied extensively, and we will present some of this history later. For the moment, it suffices to note that the binary digital sum problem exhibits the first and most obvious characteristic of all digital sum problems: we wish to study a function  $v(n)$  that is a sum of the digits of  $n$  represented in some base  $q$ .

The next step is to ask what variations of this problem exist, and whether we can say anything of interest about them. One fairly natural generalization of the problem, always subject to the constraint that  $v(n)$  should be a sum of digits, is to associate a weight function with the digits of  $n$ . There are different kinds of weight functions, of course; the function may depend on the position of the respective digit, on its value, or both. Consider alternating digital sums: taking  $q = 2$ , we have  $v(19) = (1)1 +$

$(-1)0 + (1)0 + (-1)1 + (1)1 = 1$ . Alternating digital sums use the weight function  $w(j) = (-1)^j$ , where  $j$  is the position in  $(n)_q$ , starting with  $j = 0$  at the right. Digital sum problems may also be generalized by considering the representation of  $n$  in a Cantor base  $\kappa$ , rather than the more familiar base- $q$  representation, where each position has weight  $q^j$ . Cantor bases include ordinary integer bases as a special case.

There is an additional consideration. We ask how the function  $v(n)$  should be studied. The object of any study is the discovery of new properties of the object being studied; the property that concerns us here is *the rate of growth* of  $v(n)$ . It is not difficult to see that  $v(n)$  behaves quite irregularly; it shares this characteristic with many famous functions of number theory that depend, like  $v(n)$ , on the multiplicative structure, i.e. prime factorization, of  $n$ . Therefore we study the sum of the first  $n - 1$   $v(k)$  rather than  $v(n)$ , i.e. we replace  $v(n)$  by a function that is smoother than  $v(n)$  itself. This function is

$$\frac{1}{n} \sum_{k=1}^{n-1} v(k).$$

Let us return to the problem of alternating digital sums. In keeping with the framework of this section, we wish to show that the sum

$$\frac{1}{n} \sum_{k=1}^{n-1} v(k)$$

is an iterated sum of a sequence that is multiplicatively self-similar, where we have fixed  $v(n)$  to be the alternating digital sum of  $n$  in the integer base  $q$ .

It is not difficult to see that

$$\nabla v(n) = (-1)^{v_q(n)} - (q - 1) (v_q(n) \bmod 2).$$

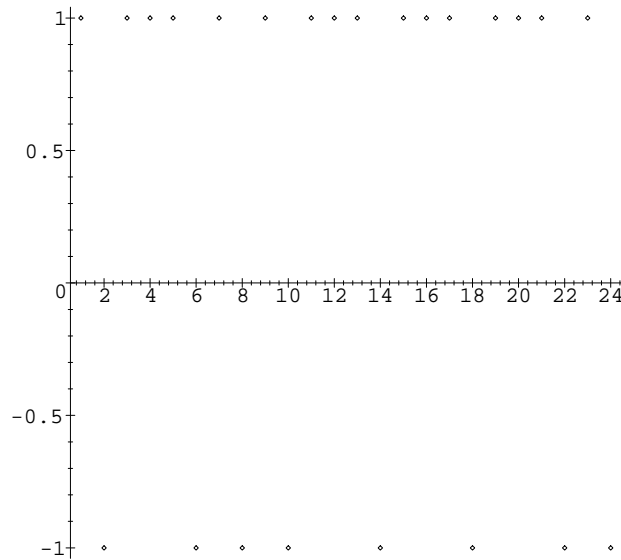
The computation is described in section 4.2.1; it might be instructive to verify this formula without looking ahead. In any case, it yields the first part of the claim immediately. Because we are evaluating a sum over a range of  $v(n)$  (from 1 to  $n - 1$ ), that sum is a double iterated sum of  $\nabla v(n)$ . It remains to check that  $\nabla v(n)$  is multiplicatively self-similar. We may rewrite the identity for  $\nabla v(n)$  as follows:

$$\nabla v(n) = (-1)^{v_q(n)} + (q - 1) \frac{1}{2} \left( (-1)^{v_q(n)} - 1 \right).$$

It follows that  $\nabla v(n)$  is a combination of a term in  $(-1)^{v_q(n)}$ , and a constant term.

The function  $(-1)^{v_q(n)}$  is multiplicatively self-similar with scale factor  $q^2$  and an error term that depends on  $r = n \bmod q^2$ ; it is zero when  $r$  is zero, one when  $q$  does not divide  $r$ , and minus one otherwise. The recurrence for  $(-1)^{v_q(n)}$  is (let  $0 < r < q^2$ )

$$\begin{aligned} (-1)^{v_q(q^2 m)} &= (-1)^{v_q(m)} \\ (-1)^{v_q(q^2 m+r)} &= \begin{cases} 1 & \text{if } q \nmid r, \\ -1 & \text{if } q \mid r. \end{cases} \end{aligned}$$



Graph of the sequence  $\{(-1)^{v_2(k)}\}$ , which is *multiplicatively self-similar* with scale factor  $l = 4$  and fluctuation

$$\{e_k\} = ((0, 1, 2, 3) \mapsto (0, 1, -1, 1)) (k \bmod 4).$$

The translation into an alternate representation by a generator  $\sigma$  is

$$1; \quad 1 \mapsto 1, -1, 1, 1; \quad -1 \mapsto 1, -1, 1, -1,$$

where we have taken  $q = 2$ . The same sequence is plotted in the graph.

We now ask whether the Fourier series for alternating digital sums will be a fractal function. The answer is yes; although we do not include a proof here, we can expect it to be entirely like the case of ordinary digital sums. The method used there is outlined in the next section.

### History and background

The history of research into the properties of digital sums is best exemplified by the developments that led to the Fourier expansion of

$$\frac{1}{n} \sum_{k=1}^{n-1} v(k)$$

where  $v(k)$  is the sum of digits of  $n$  written in base  $q$ . There are three phases to this history; these phases are first, investigation by elementary methods, second, application of a Delange-type method (in fact the problem occasioned the discovery of this method) and third, the use of the Mellin-Perron formula to obtain an exact, complete solution of the problem.

The first phase consists of results that evaluate the dominant term of the sum, and attempt to bound the error term. The evaluation of the dominant term is in fact perfectly straightforward. It rests on the following intuitive observation. The rightmost digit of  $n$  cycles through 0 to  $q - 1$ . So does the next digit, but in steps of  $q$  zeros, followed by  $q$  ones etc. The next digit also follows this cycle, this time with step size  $q^2$ . The sum of a full cycle of digits is  $0 + 1 + \dots + q - 1 = \frac{1}{2}(q - 1)q$ . The number of cycles up to  $n$  is  $\sim n/q$  for the rightmost digit. It is  $\sim n/q^2$  for the next one, where each of these contains  $q$  full cycles, etc. This gives a total of  $n(1/q + (1/q^2)q + \dots (1/q^l)q^{l-1})$  where  $l = \lfloor \log_q n \rfloor$ . Therefore the dominant term of the sum is

$$\sim \frac{1}{n} \frac{n}{q} l \frac{1}{2}(q - 1)q \sim \frac{q - 1}{2} \log_q n.$$

The dominant term is not very interesting. It is the error term that matters. In hindsight we know that this term generates the Fourier series, while the simple approximation that we used to find the dominant term is already exact.

Bush is credited with the above approximation. He showed in 1940 that

$$\frac{1}{x} \sum_{k \leq x} v(k) \sim \frac{q - 1}{2} \log_q x.$$

Bellman and Shapiro were the first to give a bound on the error term; they showed in 1948 that

$$\left| \frac{1}{x} \sum_{k \leq x} v(k) - \frac{q - 1}{2} \log_q x \right| \in \mathcal{O}(\log \log x).$$

Mirsky improved this to  $\mathcal{O}(1)$  in 1949; Drazin and Griffith studied the error term in a 1952 paper. Finally Trollope found an explicit expression of the error term in 1968. Actually he treated only the case  $q = 2$ , but claimed that his method, which was fairly complicated even for  $q = 2$ , would generalize to

$q \in \mathbb{Z}^+, q \geq 2$ . The result that he proved is this. Suppose  $2^l \leq n < 2^{l+1}$ , with  $l \in \mathbb{Z}^+$ , i.e.  $l = \lfloor \log_2 n \rfloor$ . Define  $x$  by  $n = 2^l(1+x)$ , which implies  $0 \leq x < 1$ . Then we have

$$\frac{1}{n} \sum_{k=1}^{n-1} v(k) = \frac{1}{2} \log_2 n - \frac{2^{l-1}}{n} (2f(x) + (1+x) \log_2(1+x) - 2x)$$

where

$$f(x) = \sum_{r=0}^{\infty} \frac{1}{2^r} g(2^r x)$$

and  $g(x)$  is a function with period 1 such that

$$g(x) = \begin{cases} \frac{x}{2} & \text{if } 0 \leq x < \frac{1}{2} \\ \frac{1-x}{2} & \text{if } \frac{1}{2} \leq x < 1. \end{cases}$$

Knowing as we do that the error term is “periodic in  $\log_2 n$ ”, i.e. given by a function of period 1 evaluated at  $\log_2 n$ , we recognize the key contribution of this result, which is to take the relative position  $x$  of  $n$  between  $2^l$  and  $2^{l+1}$ , and express the error in terms of this position.

The second phase is marked by a single 1975 paper, that of Hubert Delange *sur la fonction sommatoire de la fonction “somme des chiffres”*. This paper is remarkable for several reasons. Delange solved the problem for any base  $q$ . Trollope claimed that his method could treat any  $q$ , but had not actually given a demonstration. Second, Delange showed that the error term could be expressed by a periodic function of period 1 evaluated at  $\log_2 n$ , and computed the Fourier series expansion of this function. He was the first to recognize its fractal nature; he showed that it is nowhere differentiable. The expansion of sums of multiplicatively self-similar series often includes a fractal fluctuation when there is a non-zero error term; therefore, Delange discovered a truly characteristic feature of the problem. Third, and perhaps most importantly, Delange’s paper implicitly gave a versatile procedure for treating digital sums and other problems of a similar type. Kirschenhofer and Tichy extended this method to Cantor-bases in 1985; we present some of their results in our digital sum chapter. Osbaldestin and Shiu used the method in 1988 to compute an error term related to sums of three squares, which will be described in the next section, and in one of the results chapters. Careful study of the papers that use a *Delange-type method* indicates that there seems to be an element of the straightforward and procedural rather than of innovation in its application.

We proceed to discuss Delange’s result and the characteristics of a Delange-type method. Recall that we seek to evaluate the sum of the sum of digits  $v(k)$  of the first  $n-1$  integers  $k$  written in base  $q$ .

Delange proved that

$$\frac{1}{n} \sum_{k=1}^{n-1} v(k) = \frac{q-1}{2} \log_q n + F(\log_q n).$$

The function  $F(u)$  is defined as follows. We let

$$g(u) = \int_0^u \left( \lfloor qt \rfloor - q \lfloor t \rfloor - \frac{q-1}{2} \right) dt \quad \text{and} \quad h(u) = \sum_{r=0}^{\infty} q^{-r} g(q^r u)$$

and set

$$F(u) = \frac{q-1}{2} (1 + \lfloor u \rfloor - u) + q^{1+\lfloor u \rfloor - u} h\left(q^{-(1+\lfloor u \rfloor - u)}\right).$$

The function  $F(u)$  is periodic with period 1 and continuous for all  $u \in \mathbb{R}$ . Delange also proved that  $F(u)$  has the Fourier series expansion

$$F(u) = \sum_{k \in \mathbb{Z}} c_k e^{2k\pi i u}$$

where

$$c_0 = \frac{q-1}{2} \frac{\log 2\pi - 1}{\log q} - \frac{q+1}{4} \quad \text{and} \quad c_k = i \frac{q-1}{2k\pi} \left(1 + \frac{2k\pi i}{\log q}\right)^{-1} \zeta\left(\frac{2k\pi i}{\log q}\right).$$

(Incidentally, the reader should recognize the  $1/s(s+1)$  term from the Mellin-Perron integrand.)

There are three steps to the method: the first is to construct an integral representation of the  $j$ th digit  $d_j(n)$  in the base- $q$  representation of  $n$ . This representation is

$$d_j(n) = \int_n^{n+1} \left( \left\lfloor \frac{u}{q^j} \right\rfloor - q \left\lfloor \frac{u}{q^{j+1}} \right\rfloor \right) du$$

and its proof is elementary in the sense that it does not use any mathematics other than the definition of digit  $d_j(n)$ . The second step uses the fact that we can interchange summations in the double sum over  $k$  and the digits and powers of  $q$  in the base- $q$  representation of  $k$ . This interchange results in an integral from 0 to  $n$  of the integrand shown in the formula for  $d_j(n)$ . This integral is then expressed in terms of  $g(u)$  (a simple transformation, immediate when we compare the integrands in the sum integral of  $d_j(n)$  and in  $g(u)$ ). The rest is algebraic manipulation. When the second step is completed, we have the formula for  $F(u)$  in terms of  $h(u)$  and  $g(u)$  that was given above. It remains to compute the Fourier series. In order to compute the  $\{c_k\}$  in

$$F(u) = \sum_{k \in \mathbb{Z}} c_k e^{2k\pi i u},$$

we start with the basic formula

$$c_k = \int_0^1 F(u) e^{-2k\pi i u} du.$$

If we expand  $F(u)$  in terms of  $g(u)$ , as given by the formula, we arrive at the integral

$$\int_{1/q}^{\infty} \frac{g(u)}{u^{s+1}} du,$$

which is absolutely convergent for  $\operatorname{Re}(s) > 0$ . This integral is treated with

$$\int_1^{\infty} \frac{\lfloor u \rfloor}{u^{s+1}} du = \frac{1}{s} \zeta(s),$$

and we have the Fourier coefficients.

We claimed earlier that the method of Delange can be used in an algorithmic, procedural manner. This is best illustrated with an example. In 1985, Kirschenhofer, Prodinger and Tichy studied digital sums with different weight functions and/or in Cantor bases  $\kappa$ . They examined the asymptotic behavior of such sums and gave a Fourier series expansion of alternating digital sums. In a Cantor base  $\kappa$ , two sequences,  $\{q(j)\}$  and  $\{\kappa(j)\}$ , where  $\kappa(j) = \prod_1^j q(k)$ , take the place of  $q$  and the powers  $q^j$  of  $q$  in an ordinary base. The procedure remains exactly the same, as do many of the intermediate integrals and formulas. For example, we use

$$d_j(n) = \int_n^{n+1} \left( \left\lfloor \frac{u}{\kappa(j)} \right\rfloor - q \left\lfloor \frac{u}{\kappa(j+1)} \right\rfloor \right) du$$

for the  $j$ th digit, and we define a family of functions

$$g_j(u) = \int_0^u \left( \lfloor q(j)t \rfloor - q(j) \lfloor t \rfloor - \frac{q(j)-1}{2} \right) dt;$$

the Fourier coefficients are again treated with

$$\int_1^{\infty} \frac{\lfloor u \rfloor}{u^{s+1}} du = \frac{1}{s} \zeta(s).$$

As pointed out earlier, Delange was also the first to study the fractal nature of the series  $F(u)$ . He showed that  $F(u)$  is continuous and nowhere differentiable, which is a stronger result than non-differentiability at a dense set of points. He did this by reducing the everywhere non-differentiability of  $F(u)$  to that of  $h(u)$  in  $(0, 1)$ . Given a  $\theta \in (0, 1)$  he shows that the sequence

$$\frac{h(x'_k) - h(x_k)}{x'_k - x_k}$$

with  $x_k \leq \theta < x'_k$  and  $x_k, x'_k \rightarrow \theta$  in the dense set of  $q$ -adic rationals does not have a limit. (In fact  $x_k$  is the length- $k$  prefix of the  $q$ -adic expansion of  $\theta$  and  $x'_k$  its successor, obtained by incrementing the last digit.)

The third phase is one of refinement and extension rather than novelty. The principal contribution is by Philippe Flajolet, who discovered in 1994 that digital sums can be expressed as harmonic sums and evaluated by the Mellin-Perron formula. Given the body of classical knowledge about Mellin transforms, Dirichlet series and the Riemann zeta function (the most well-known Dirichlet series), Flajolet's paper effectively reduces the problem to an instructive type of minor exercise. The ideas that motivate the computation are familiar to any mathematician who works in analytic number theory, and the computation itself takes only a few lines. Nonetheless if it had not been for the detour via Delange's method, who introduced the idea of Fourier series, the classical number theory approach might not have been discovered. Its outlines and simplicity appear sharper and more visible in retrospect than they were before Delange suggested where to focus.

*Contribution of this thesis.* The application of the Mellin-Perron formula to digital sum problems faces two chief obstacles: we must be able to locate the poles of the associated Dirichlet generating function, and to continue it analytically into the left half-plane. We study alternating digital sums, digital sums with periodic weights, and digital sums where the quotient sequence  $\kappa(j+1)/\kappa(j)$  of the Cantor base  $\kappa$  is periodic, and we show that all of these can be treated by the Mellin-Perron formula. We compute the Fourier series for alternating digital sums. Finally, we study factorial digital sums, where the Dirichlet generating function cannot be continued into the left half-plane. We use the Mellin-Perron formula to extract information about the asymptotically dominant term of this and similar digital sums. Our work both extends P. Flajolet's method to a wider class of digital sums, and supplies new proofs of some existing results concerning the asymptotic behavior of digital sums.

### 1.3.4 Counting sums of three squares

The study of integers that are representable as sums of squares is a classical problem in number theory. We have already seen some of its history when we discussed the function  $r_2(n)$ , i.e. the number of integer solutions  $(x, y)$  of  $n = x^2 + y^2$ . The list of mathematicians who have worked on this problem and its generalizations to  $r_k(n)$  reads like a compendium of greats: Euler, Gauss, Jacobi, Eisenstein, Ramanujan, Hardy and others. Here we consider the special case  $k = 3$ , and we ask not how many solutions there are of  $n = x_1^2 + x_2^2 + x_3^2$  (this is  $r_3(n)$ ), but rather for which integers  $n$  solution triples  $(x_1, x_2, x_3)$  exist (i.e. whether  $r_3(n)$  is zero or not). As in the case of digital sums, we study a sum of  $(0 \mapsto 0; 1, 2, \dots \mapsto 1)(r_3(n))$ , rather than the term  $(0 \mapsto 0; 1, 2, \dots \mapsto 1)(r_3(n))$  itself, which exhibits highly irregular behavior. Given  $N$ , we wish to count the number of positive integers  $n < N$  that are

representable as a sum of three squares. We will review some of the history of this problem in the next section.

The principal tool of the investigation is a theorem by Gauss, which states that  $n$  is representable as a sum of three squares if and only if it does not have the form  $n = 4^l(8k + 7)$ , where  $l, k \in \mathbb{Z}^+$ . For example,  $10 = 3^2 + 1^2 + 0^2$  and  $14 = 3^2 + 2^2 + 1^2$ ; no such representation exists for  $n = 15$ . If we let  $Q$  be the set of integers  $n \in \mathbb{Z}^+$  representable as sums of three squares including 0, and take  $k(n)$  to be the characteristic function of  $\bar{Q}$ , i.e.  $k = (Q \mapsto 0; \bar{Q} \mapsto 1)$ , we may write

$$Q(N) = \sum_{n \in Q, 0 < n \leq N} 1 = N - \sum_{0 < n \leq N} k(n).$$

Define  $\Delta(N)$  as follows:

$$Q(N) = \frac{5}{6}N + \Delta(N),$$

i.e.

$$\Delta(N) = \frac{1}{6}N - \sum_{0 < n \leq N} k(n)$$

and let  $\Delta(0) = 0$ . (The choice of terms and the definition of the error function are historically motivated, as shall be explained later.) We consider the measure

$$\frac{1}{N} \sum_{0 \leq n < N} \Delta(n)$$

which is

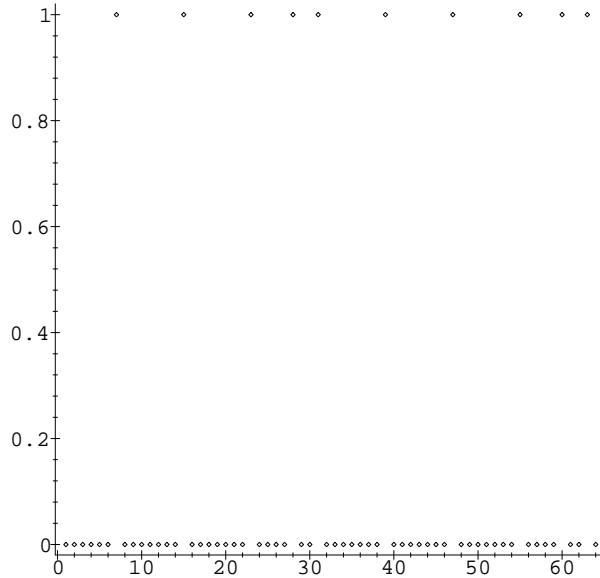
$$\frac{1}{12}N - \frac{1}{12} - \frac{1}{N} \sum_{n=1}^{N-1} \sum_{l=1}^n k(l).$$

The latter reduces the problem to the evaluation of

$$\sum_{n=1}^{N-1} \sum_{l=1}^n k(l).$$

We recall our criterion for the applicability of the Mellin-Perron formula and the existence of a Fourier series expansion: we must be concerned with an iterated sum of a multiplicatively self-similar sequence. It is immediate that we have an iterated sum; this follows from the identity above. We must check that  $k(n)$  is multiplicatively self-similar. Indeed the scale factor is  $l = 4$  and the error term zero except when  $n$  is congruent 7 modulo 8. The recurrence for  $k(n)$  is

$$\begin{aligned} k(8m) &= k(2m) & k(8m+4) &= k(2m+1) \\ k(8m+1) &= 0 & k(8m+5) &= 0 \\ k(8m+2) &= 0 & k(8m+6) &= 0 \\ k(8m+3) &= 0 & k(8m+7) &= 1. \end{aligned}$$



Graph of the sequence  $\{k(n)\}$ , which is *multiplicatively self-similar* with scale factor  $l = 4$

and fluctuation  $\{e_k\} = ((0, \dots, 6) \mapsto 0; 7 \mapsto 1) (k \pmod{4})$ .

This recurrence requires some commentary. At first glance it might seem that it does not have the canonical form of the recurrence for multiplicative self-similarity, or perhaps that it has scale factor  $l = 8$ , rather than  $l = 4$ , as claimed. We discussed two equivalent criteria for self-similarity; the first, that the sequence contain a copy of itself, obtained when we omit all  $n$  that are not exact multiples of the scale factor, and second, that it can be generated by the iterative application of an operator  $\sigma$  of the form  $\sigma_1(k), \sigma_2(k), \dots, \sigma_l(k)$ . It is the first description that indicates most readily the existence of regularly spaced poles in the Dirichlet generating function, and hence a Fourier series in the solution. When we consider the recurrence for  $k(n)$ , it is immediate that the scale factor is  $l = 4$  and not  $l = 8$ ; when we skip all  $n$  that are not multiples of 4, we obtain the original sequence. Furthermore, this operation generates poles of the form

$$\frac{2\pi ik}{\log 4}.$$

The structure of the recurrence also indicates that we must look for a pair of operators  $\sigma'$  and  $\sigma''$  as the generators of  $k(n)$ , one to generate the bottom range, shown in the left column, and one to generate the top range, shown on the right. This alternate characterization is

$$0, 0; \quad k, l \mapsto 0, 0, 0, k, 0, 0, 1, l.$$

The question of the Fourier series being fractal is open. The fact that the error term  $\Xi(s)$  is not constant, and the form of the coefficients obtained strongly suggest that it is.

### History and background

As indicated, the study of integers  $n$  representable as sums of three squares goes back to Gauss, whose theorem on the form of such  $n$  we use in this thesis. In 1908, Edmund Landau was the first to note that the set  $Q$  has asymptotic density  $5/6$  and introduce the error term  $\Delta(N)$ . In fact his work implies that

$$\Delta(N) \ll \log N \quad \text{as } n \rightarrow \infty.$$

M. C. Chakrabarti refined this result in 1940. He proved that

$$\frac{1}{6} \leq \Delta(N) < \frac{1}{3} \log_2 N + \frac{1}{2}$$

where the bounds are sharp, i.e. the lower bound is assumed and the upper approached arbitrarily closely. He also showed that the values of  $\Delta(N)/\log_2 N$  are dense in the interval  $(0, 1/3)$ .

Chakrabarti based his result on the following formula for  $\Delta(N)$ . Let the binary digits of  $N$  be given by

$$(N)_2 = \sum_{j=0}^r d_j 2^j$$

where  $0 \leq d_j < 2$  and  $d_r = 1$ . Chakrabarti showed that

$$\Delta(N) = \frac{2}{3} \sum_{j=0}^{2j \leq r} d_{2j} + \frac{1}{3} \sum_{j=1}^{2j-1 \leq r} d_{2j-1} - \sum_{j=1}^{2j \leq r} (0, 1, 2 \mapsto 0; 3 \mapsto 1)(d_{2j} + d_{2j-1} + d_{2j-2}) - \frac{1}{2} d_0;$$

i.e. he represented  $\Delta(N)$  by a digital sum. P. Shiu noticed in 1988 that this formula suggests that  $\Delta(N)$  should be  $3/16 \log_2 N$  on average. To see this, note that there will be  $\sim \log_2 N$  digits in the binary representation of  $N$ . The first sum counts the digits at even positions, which are on average

$$\sim \frac{1}{N} \left( \frac{1}{2} N \frac{1}{2} \log_2 N \right) = \frac{1}{4} \log_2 N.$$

The same argument applies to the second sum, again giving

$$\sim \frac{1}{4} \log_2 N.$$

The third sum counts 111 strings, i.e. one out of eight possibilities  $\{0,1\}^3$ , at even positions in  $(N)_2$ , giving an average of

$$\sim \frac{1}{N} \left( \frac{1}{8} N \frac{1}{2} \log_2 N \right) = \frac{1}{16} \log_2 N.$$

The average of the last term is of a lower order than  $\log_2 N$ , namely

$$\frac{1}{N} \frac{1}{2} N = \frac{1}{2}.$$

Therefore, this informal argument does indeed suggest that the average of  $\Delta(N)$  is

$$\sim \left( \frac{2}{3} \frac{1}{4} + \frac{1}{3} \frac{1}{4} - \frac{1}{16} \right) \log_2 N = \frac{3}{16} \log_2 N.$$

P. Shiu was able to prove his conjecture in a 1988 paper; he found that Chakrabarti's digital sum for  $\Delta(N)$  did not suffice to obtain the result, and used a base-4 sum instead. He showed that

$$\left| \frac{1}{N} \sum_{n \leq N} \Delta(n) - \frac{3}{16} \log_2 N \right| \in \mathcal{O}(1)$$

as  $N \rightarrow \infty$  and that for  $\epsilon \in \mathbb{R}^+$ , the number of  $n \leq N$  such that

$$\left| \Delta(n) - \frac{3}{16} \log_2 n \right| > (\log n)^{\frac{1}{2} + \epsilon}$$

is  $o(N)$ .

Suppose that

$$(n)_4 = \sum_{r \geq 0} d_r(n) 4^r$$

where  $0 \leq d_r(n) < 4$ . P. Shiu's base-4 digital sum for  $\Delta(N)$  is

$$\Delta(n) = \frac{1}{6} d_0(n) + \sum_{r \geq 1} \gamma(a_r(n), a_{r-1}(n))$$

where  $\gamma(a, a')$  is given by the following table of values.

	$a = 0$	$a = 1$	$a = 2$	$a = 3$
$a' = 0, 1, 2$	0	$\frac{2}{3}$	$\frac{1}{3}$	1
$a' = 3$	0	$-\frac{1}{3}$	$\frac{1}{3}$	0

A.H. Osbaldestin and Shiu subsequently used this base-4 digital sum to compute the Fourier series for  $\frac{1}{N} \sum_{n < N} \Delta(n)$ . In their 1989 paper, they show that there exists a periodic function  $F(u)$  with period 1 such that for  $N \geq 1$ ,

$$\frac{1}{N} \sum_{0 \leq n < N} \Delta(n) = \frac{3}{8}L + F(L) + \frac{\delta(N)}{N} \quad \text{where } L = \log_4 N \quad \text{and } \delta(N) = \begin{cases} \frac{1}{8} & N \text{ odd,} \\ 0 & N \text{ even.} \end{cases}$$

The function  $F(u)$  is a Fourier series

$$\sum_{k \in \mathbb{Z}} c_k e^{2\pi i k u}$$

with coefficients

$$c_0 = -\frac{31}{48} - \frac{3}{8 \log 4} - \frac{1}{\log 4} (\zeta'(0, 7/16) + \zeta'(0, 15/16))$$

and

$$c_k = -\frac{1}{2\pi i k} \left(1 + \frac{2\pi i k}{\log 4}\right)^{-1} \left(\zeta\left(\frac{2\pi i k}{\log 4}, \frac{7}{16}\right) + \zeta\left(\frac{2\pi i k}{\log 4}, \frac{15}{16}\right)\right), \quad k \neq 0.$$

They use the method of Delange. In order to show how their proof adapts Delange's treatment of base- $q$  digital sums to P. Shiu's special base-4 digital sum, we will retrace the key steps of the argument. We recall that Delange begins by constructing an integral formula for the  $j$ th digit in the base- $q$  representation of  $n$ . So do Osbaldestin and Shiu, except that  $\gamma(a_r(n), a_{r-1}(n))$  now takes the place of  $a_r(n)$ . Their integral formula is

$$\gamma(a_r(n), a_{r-1}(n)) = \int_n^{n+1} \beta\left(\frac{t}{4^{r+1}}\right) dt.$$

The reader is encouraged to compare this formula with the earlier ones of Delange and Kirschenhofer. The function  $\beta(t)$  is a piecewise constant function with period 1. The next step is to find an analogue of the earlier  $g(u)$  and  $g_j(u)$ . This is the function  $\alpha(u)$ , which is piecewise linear, has period 1 and is given by

$$\alpha(u) = \int_0^u \left(\beta(t) - \frac{3}{8}\right) dt.$$

With these definitions, an interchange of summations and a few simple algebraic manipulations yield the formula for in terms of  $F(u)$ . The formula for the Fourier series expansion eventually leads to an integral of the form

$$\int_{1/4}^{\infty} \frac{\alpha(u)}{u^{s+1}} du$$

(compare with the corresponding integral in  $g(u)$ ), and this integral can be expressed in terms of the Hurwitz  $\zeta$ -function at  $a = 7/16$  and  $a = 15/16$ . The Fourier coefficients then follow.

*Contribution of this thesis.* We give a straightforward, short proof of the Osbaldestin-Shiu result by the Mellin-Perron formula. The new proof is to the old one as Flajolet's evaluation of binary digital sums is to Delange's.

### 1.3.5 Lattice points inside a paraboloid

The last of the four case studies that make up the body of this thesis concerns the number of lattice points inside an upside-down paraboloid in three-space. A verbal description of this object might be as follows. We start with the right half of the parabola  $z = x^2$ , in the  $xz$ -plane, which is positioned in three-space such that  $x$  corresponds to width and  $z$  to height; the values along the  $x$  axis increase from left to right, so that the right half of  $z = x^2$  corresponds to  $x \geq 0$ . Now we turn this half-parabola upside down, making it into  $z = -x^2$ . Finally, we shift it upwards along the  $z$  axis by an integer distance  $n$ , obtaining  $z = n - x^2$ . The paraboloid  $P_n$  is the surface of rotation that results when we rotate  $z = n - x^2$  around the  $z$  axis in three-space. Actually  $P_n$  is restricted to the part of the paraboloid that lies above and in the  $xy$ -plane. It looks like a dome. Its intersection with the  $xy$ -plane is the circle  $x^2 + y^2 = n$ , its intersection with any plane that contains the  $z$  axis is an upside-down parabola, e.g. its intersection with the  $xz$  plane is  $z = n - x^2$ , with the  $yz$  plane  $z = n - y^2$ . The introduction to chapter 6 contains a picture of this paraboloid and its equation, which is

$$\{(x, y, z) \mid x^2 + y^2 = n - z, z \in [0, n]\}.$$

The paraboloid  $P_n$  contains some number of lattice points, i.e. points  $(a, b, c)$  with integer coordinates  $a, b, c$ . We define a weight function on these lattice points – if  $(a, b, c)$  lies in the interior of  $P_n$ , it has weight unity, if it lies on the paraboloid itself, it has weight a half. The sum of these weights over all lattice points inside and on  $P_n$  is the number  $V_n$ . A smoothness consideration such as that used in the digital sum case, and with sums of three squares, leads us to study the measure

$$\frac{1}{N} \sum_{n=0}^{N-1} V_n.$$

This measure does not share all characteristics of the previous three examples. These are the similarities and differences.

- The area  $a_n$  of the fractal ornament, digital sums and the error term  $\Delta(n)$  of the sum-of-three-squares function are *double iterated sums*. In a sense, so is  $\frac{1}{N} \sum_{n=0}^{N-1} V_n$ . To be exact, the problem reduces to two iterated sums, one triple, the other double:

$$\frac{1}{N} \sum_{n=0}^{N-1} \sum_{k=1}^n \sum_{l=1}^{k-1} r_2(l) + \frac{1}{2N} \sum_{n=0}^{N-1} \sum_{k=1}^n r_2(k).$$

Here  $r_2(k)$  is the number of integer solutions of  $n = x^2 + y^2$ , a function that we discussed earlier. It is this particular combination of iterated sums that permits us to apply the Mellin-Perron formula for  $m = 2$ , because it reduces to

$$\frac{1}{2} N \sum_{n=1}^{N-1} \left(1 - \frac{n}{N}\right)^2 r_2(n).$$

The reader will recognize the above as the type of harmonic sum that is evaluated with the Mellin-Perron formula.

- Prior to this example, we studied iterated sums of sequences that had multiplicative self-similarities. These similarities were useful for two reasons: *they result in closed forms of the respective Dirichlet generating function*, which permits us to locate the poles of the Mellin integrand and hence evaluate the sum. Furthermore, *the specific kind of multiplicative self-similarity that we have studied always generates one or more column of regularly spaced poles along a vertical line in the right-half plane*. These poles sum to a Fourier series in the expression of the sum.

The function  $r_2(k)$  does not exhibit multiplicative self-similarity. However, the principal requirement for the existence of an explicit solution by the Mellin-Perron formula is that we know the location of the poles of the Dirichlet generating function of the amplitudes, i.e. in this case of

$$\sum \frac{r_2(n)}{n^s}.$$

In fact we do. We can locate the poles and study their residues via the alternate representation

$$\sum \frac{r_2(n)}{n^s} = \frac{1}{4^{s-1}} \zeta(s) \left( \zeta\left(s, \frac{1}{4}\right) - \zeta\left(s, \frac{3}{4}\right) \right).$$

This function has a single pole at  $s = 1$ , and it is simple. It contributes the dominant term in the evaluation of the sum.

Owing to the fact that the Dirichlet generating function only has a single pole, the question of the “Fourier series” being fractal does not apply to this problem. The solution only has two terms, the dominant one and an error term; hence there is no series and no possibility of its being fractal.

## History and background

Among the four problems treated in this thesis, this one is the second to have been constructed in reverse, i.e. knowing the kinds of sums that can be evaluated by the Mellin-Perron formula, we looked for and found a sum that has a meaningful interpretation as a lattice point problem. (The first problem to be constructed like this was the one about the fractal ornament.) It is perhaps useful to describe how one obtains the particular choice of weights and surface from the characteristics of possible sums. We begin with the fact that  $r_2(n)$  has a Dirichlet generating function that can be expressed in terms of zeta functions. This is classic knowledge; we discussed some of the relevant history earlier. By a theorem of Whittaker and Watson, we know that the maximum order of the Hurwitz zeta function along horizontal line segments inside and parallel to  $[0, 1]$  is  $\mathcal{O}(t \log t)$ , with  $t$  the imaginary part of  $s$  in  $\zeta(s)$ , and going to infinity along vertical lines. We also know that the Mellin-Perron formula generates an integral along a vertical line in the right-half plane, and that this integral is evaluated by shifting it into the left half-plane and taking residues into account; if we wish to have an exact formula, we must ensure that the contribution along the connecting horizontal segments vanishes. The Dirichlet generating function of  $r_2(n)$  is a function in  $\zeta(s)^2$ , and hence  $\mathcal{O}(t^2 \log^2 t)$  on the relevant horizontal line segments. The Mellin-Perron formula includes a factor  $1/s(s+1) \dots (s+m)$  for an  $m+1$ -times iterated sum. This leads to the first key observation: if the  $\zeta$  terms along the vertical segments are to vanish, we need  $m \geq 2$ .

Therefore we ask what iterated sum of  $r_2(n)$  is of interest and has  $m \geq 2$ ? The lattice points inside a circle are a sum ( $m = 0$ ) of  $r_2(n)$ , those inside a paraboloid a double sum ( $m = 1$ ), and hence, their average is a triple sum with  $m = 2$ . *Ergo*, the average fulfills the requirement.

There is an additional constraint, however. If we visualize a triple sum by mapping iteration to dimension and single terms to dots, we see that the case  $m = 2$  corresponds to a half square pyramid of dots, obtained from the full one by slicing it along one of the diagonals of the square at the base. On the other hand, the Mellin-Perron formula for  $m = 2$  counts a full pyramid. Therefore, if our sum is to be amenable to the formula, it must include the missing half. Actually the dots are discrete and hence the half isn't an exact half, but rather the predecessor in the sequence of pyramids. In fact we need a sum that corresponds to predecessor, successor addition of half-pyramid sums. The predecessor is the interior of the successor. When we add the two, the interior points get counted twice, and the surface points once. This leads us to conclude that we require a weight function that assigns one to interior points, half to surface points, and zero to exterior points, which function also happens to be an intuitive

choice as a volume measure.

The paraboloid problem is like the fractal ornament problem in that there is no direct antecedent in the literature, but there is a vast number of auxiliary research results. This research is the study of *diophantine equations*, and *lattice points*. The two are synonymous because diophantine equations are polynomial equations in several variables, and are studied with respect to the number and existence of integer solutions. These equations also define areas, surfaces or more generally, bodies in  $n$ -space, and it is natural to ask what points with integer coordinates are located inside these bodies. Those points are of course the integer solutions of the respective diophantine equation.

The most basic and central result in this field is *Minkowski's lattice point theorem*. A subset  $M \subset \mathbb{R}^k$  of  $k$ -space is convex if for any two points  $\gamma, \lambda \in M$  the line segment that connects  $\gamma$  to  $\lambda$  is a subset of  $M$ . It is symmetric about the origin if  $\gamma \in M$  implies  $-\gamma \in M$ . Minkowski's lattice point theorem states that convex, measurable  $M$  that are symmetric about the origin and have volume greater than  $2^k$  contain at least one lattice point other than the origin. This bound is sharp, as the  $k$ -cube centered at  $O$  and of side length 2 shows.

The classic problem in lattice point theory is to compute the number of lattice points inside a  $k$ -ball of radius  $\sqrt{t}$ . We denote by  $A_k(t)$  the number of such points. We have already pointed out that this number is equivalent to counting the solutions of

$$x_1^2 + x_2^2 + \dots + x_k^2 \leq t,$$

which is in turn equivalent to counting and adding the number of solutions of

$$x_1^2 + x_2^2 + \dots + x_k^2 = 0, 1, 2 \dots [t].$$

Hence

$$A_k(t) = 1 + \sum_{n \leq t} r_k(n),$$

where  $r_k(n)$  counts the solutions of

$$x_1^2 + x_2^2 + \dots + x_k^2 = n.$$

We saw earlier that the evaluation of  $r_2(n)$  and other  $r_k(n)$  has given rise to entire branches of number theory, and a great deal of exciting mathematics. This problem is also remarkable for the fact that functions like  $r_2(n)$  have been studied since antiquity, e.g. by Pythagoras, and that we nonetheless find Hardy and Ramanujan working on the general even and odd  $r_k(n)$  during the first third of this century.

We conclude this brief review with an elementary result concerning  $A_k(t)$ . It is hoped that the reader is sufficiently intrigued to continue independently along this beautiful branch of mathematics.

The following formula describes the volume  $V_k(t)$  of a  $k$ -ball of radius  $\sqrt{t}$ . We have

$$V_k(t) = \frac{\pi^{k/2} t^{k/2}}{\Gamma(k/2 + 1)}.$$

We can prove that

$$|A_k(t) - V_k(t)| \in \mathcal{O}\left(t^{k/2-1/2}\right),$$

or equivalently, that the number of lattice points inside a  $k$ -ball is equal to the volume plus an error term in the order of the boundary surface. This result is part of a general pattern, according to which we can count lattice points by computing the volume and assuming an error term in the boundary surface. E.g. by inspection, this also holds for the sum of  $V_n$  of our paraboloid.

#### 1.4 Notes

For a textbook-style introduction to combinatorial enumeration and generating functions, consult [FS93]. Principles and paradigms for ordinary generating functions are discussed on [FS93, p. 1-5], for exponential generating functions on [FS93, p. 43-57].

The discussion of integer lattice points inside a circle and their relation to theta functions is taken from [BMP55a, p. 182-189], which incidentally also includes an elementary presentation of generating function methods; see [BMP55a, p. 228-243]. The Bateman manuscript project remains an invaluable resource for formulas and survey articles in a variety of fields. Elliptic functions in general and theta functions in particular are discussed on [BMP55b, p. 295-380]; the reader who wishes to learn more about the theta function identity used to evaluate  $r_2(n)$  should check [BMP55b, p. 355-360].

The binary tree example is discussed on [FS93, p. 23-24]. The presentation of the tree intersection statistic is from [CDM91, p. 187-189], which also includes an overview of generating function methods and lists some essential theorems from complex analysis.

The sections on Mellin transforms and harmonic sums are entirely based on [FGD95]. I have tried to present the results of that paper in an accessible form, with the target audience being computer scientists rather than mathematicians; many of the examples were expanded to make them somewhat easier to read. The harmonic sum evaluation paradigm of section 1.2 is given in Theorem 5 on [FGD95, p. 26]. The example concerning Harmonic numbers is on [FGD95, p. 27].

I will now list references for the detailed presentation of the Mellin transform and its relation to harmonic sums, i.e. for the content of sections 1.2.1 and 1.2.2. The definition and properties of the Mellin transform can be found on [FGD95, p. 9-11]. A useful dictionary of Mellin transforms is on [FGD95, p. 13]. The Mellin inversion theorem is on [FGD95, p. 14]. The transform of  $(1+x)^{-1}$  is discussed on [FGD95, p. 18], that of  $e^{-x} - 1 + x$  on [FGD95, p. 13, 19]. The formal statement of the Mapping theorem is on [FGD95, p. 16], Theorem 3, and [FGD95, p. 19], Theorem 4. The definition of harmonic sums, the evaluation theorem, and the Mellin summation formula are from [FGD95, p. 23-27].

My discussion of Landau's use of the Mellin-Perron formula via the "discontinuous factor", i.e. the material on the history of the Mellin-Perron formula, is based on a two-line footnote of his, found in a 1915 paper, which is reprinted on [Lan62, p. 11-29]. The footnote is on [Lan62, p. 20]. (Incidentally, the mergesort paper by Flajolet and Golin presents the "discontinuous factor" proof of the case  $m = 1$  on [FG94, p. 679-680].) The second derivation of the Mellin-Perron formula, this time in the context of harmonic sums, is generally based on [FGD95] but also uses observations from [FGK<sup>+</sup>94, p. 295-297].

A complete probabilistic analysis of the best, average and worst cases of mergesort is given on [FG94, p. 674-679]. This paper also includes a list of references for textbook-style discussions of the mergesort algorithm. The solution paradigm and formulae for divide-and-conquer recurrences is based on [FG94, p. 679-680], and the treatment of divide-and-conquer recurrences by the Mellin-Perron formula on [FG94, p. 681]. The three theorems for the worst, best and average case behavior of mergesort are on [FG94, p. 682, 684, 686]. Our definition of the term *fractal* as it applies to functions uses remarks from [FG94, p. 690] and [Del75, p. 33].

The material on curves constructed by similarities is from [Tri95, p. 177-194], a text that is notable for its lively style of presentation and the diversity of the material covered. Fractal dimension as it applies to curves is discussed on [Tri95, p. 115-123].

The section on the history of digital sum problems is substantially based on the survey in [Del75]. Bush's first result is in [Bus40]; Drazin and Griffith's in [DG52], Trollope's in [Tro68]. Delange's influential paper is [Del75]; the proof of the main theorem is on [Del75, p. 33-35], while the Fourier series expansion is computed on [Del75, p. 43-47]. The early work of Kirschenhofer and Tichy on this problem is in [KPT85]; P. Flajolet's paper is [FGK<sup>+</sup>94]. This last paper includes an extensive bibliography and a useful survey of computer science applications of digital sums.

The introduction to the sum-of-three squares problem uses the historical review in [Shi88]. The

relevant classical papers are Landau's, [Lan08] and Chakrabarti's, [Cha40]. P. Shiu's digital base-4 sum for the error term is derived on [Shi88, p. 205-207]. Osbaldestin and Shiu's application of Delange's method to the error term is on [OS89, p. 371-373].

The auxiliary material on lattice points and diophantine equations is from [Fri82]. Minkowski's lattice point theorem is proved on [Fri82, p. 1-3]; the approximation of the number of lattice points by the volume of the solid plus an error term on the order of its boundary is on [Fri82, p. 3-6]; a more general version may be found on [Kra81, p. 180-182]. Fricker also discusses the Gauss-Jacobi theorem on sums of three squares, see [Fri82, p. 20-25], and includes two proofs of the formula for  $r_2(n)$ , Gauss's elementary one and another by prime factorization of Gaussian integers, parts of which we use in our result chapter, where more specific references are given. For the two proofs, see [Fri82, p. 9-15].

## Chapter 2

### Analytic Number Theory

This chapter presents those definitions from analytic number theory that will be used in the remainder of this thesis. The material is from classical analysis and we will present the main results without proof. When we require a customized lemma we will usually sketch the proof. We will treat a result in greater depth if it helps elucidate the material of subsequent chapters. Those minor lemmata and corollaries that appear without a reference are specific to this thesis.

We recall the following definitions from real analysis.

**Definition 2.0.1** *We write*

$$\lim_{x \rightarrow x_0^+} f(x) = y \quad \left( \lim_{x \rightarrow x_0^-} f(x) = y \right)$$

*if for all  $\epsilon \in \mathbb{R}^+$  there exists a  $\delta \in \mathbb{R}^+$  such that*

$$x - x_0 < \delta \quad (x_0 - x < \delta)$$

*implies*

$$|f(x) - y| < \epsilon.$$

*We use the notation*

$$\lim_{x \rightarrow x_0^+} f(x) = f(x_0^+) \quad \text{and} \quad \lim_{x \rightarrow x_0^-} f(x) = f(x_0^-).$$

This definition permits a distinction between the limiting value of a function  $f(x)$  as we approach  $x$  from above and from below. We will use such limits when we discuss Mellin inversion integrals of transforms of discontinuous functions. For example,  $|0^+|/0^+ = 1$  and  $|0^-|/0^- = -1$ .

**Definition 2.0.2** *Let  $\{a_n\}$  be a sequence of real numbers. We write*

$$\limsup_{n \rightarrow \infty} a_n = \alpha \quad \left( \liminf_{n \rightarrow \infty} a_n = \alpha \right)$$

*if*

1.  $\forall \epsilon \in \mathbb{R}^+ \exists N \in \mathbb{Z}^+$  such that  $n \geq N$  implies  $a_n < \alpha + \epsilon$  ( $a_n > \alpha - \epsilon$ ), and
2.  $\forall \epsilon \in \mathbb{R}^+, j \in \mathbb{Z}^+ \exists k \in \mathbb{Z}^+, k \geq j$  such that  $a_n > \alpha - \epsilon$  ( $a_n < \alpha + \epsilon$ ).

If the sequence is not bounded from above (below), we define

$$\limsup_{n \rightarrow \infty} a_n = +\infty \quad \left( \liminf_{n \rightarrow \infty} a_n = -\infty \right).$$

If  $\lim_{n \rightarrow \infty} a_n = -\infty$  ( $\lim_{n \rightarrow \infty} a_n = +\infty$ ), we define

$$\limsup_{n \rightarrow \infty} a_n = -\infty \quad \left( \liminf_{n \rightarrow \infty} a_n = +\infty \right).$$

We may paraphrase this definition as follows. *The limit superior (inferior) of a sequence is the largest (smallest) number that can be obtained as a limit of one of its subsequences.*

**Definition 2.0.3** Ascending sequences of positive reals. Let  $\Lambda = \{\lambda_n\} \subset \mathbb{R}^+$  be an ascending sequence of positive reals:  $0 < \lambda_1 < \lambda_2 < \dots$ . The **upper density** of  $\Lambda$  is the quantity

$$D = \limsup_{n \rightarrow \infty} \frac{n}{\lambda_n}.$$

The **step**  $h$  of  $\lambda$  is

$$h = \liminf_{n \rightarrow \infty} (\lambda_{n+1} - \lambda_n).$$

**Example.** The sequence  $\lambda_n = n \log q$ ,  $q \in \mathbb{N}$ ,  $q > 1$  has upper density  $D = 1/\log q$  and step  $h = \log q$ . The sequence  $\lambda_n = \log(n+1)! = \sum_{k=1}^n \log(k+1)$  has upper density  $D = 0$  and step  $h = +\infty$ . (Recall that  $\sum_{k=1}^n \log(k+1) \sim n \log n$ .)

## 2.1 Point sets in the complex plane

The complex plane  $\mathbb{C}$  is a linear vector space over the reals; i.e.  $\mathbb{C} = \{(x, y) \mid x, y \in \mathbb{R}\}$ , the set of ordered pairs  $(x, y)$  of reals, along with the standard definitions of equality, addition and multiplication.

**Definition 2.1.1** (Point sets.)

**point set** A point set is any collection  $S \subseteq \mathbb{C}$  of points in the complex plane. The complement of  $\bar{S}$  of  $S$  is the set  $\{z \mid z \notin S\} = \mathbb{C} \setminus S$ .

**$\epsilon$ -neighborhood** An  $\epsilon$ -neighborhood  $N_\epsilon(z_0)$  is the set  $\{z \mid |z - z_0| < \epsilon\}$ , i.e. a disk centered at  $z_0$ , with radius  $\epsilon$ , and the boundary not included in the set. A deleted neighborhood is one with its center removed, i.e.  $N_\epsilon(z_0) \setminus \{z_0\}$ . Where it is clear from the context, we may use  $N_\epsilon$  in place of  $N_\epsilon(z_0)$ .

**limit point** The point  $z_0$  is a limit point of the point set  $S$  if every  $\epsilon$ -neighborhood of  $z_0$  contains at least one point of  $S$  different from  $z_0$ , i.e. we can shrink the disk at  $z_0$  but will always retain at least one other point of  $S$  inside it. Note that  $z_0$  need not be in  $S$ . Alternatively, we may say that  $z_0$  is a limit point of  $S$  if every deleted neighborhood of  $z_0$  contains some point of  $S$ , i.e.  $\forall \epsilon > 0 \exists z \in N_\epsilon(z_0) \setminus \{z_0\} : z \in S$ .

**interior point** The point  $z_0$  is an interior point of  $S$  if some  $\epsilon$ -neighborhood of  $z_0$  contains only points in  $S$ , i.e.  $\exists \epsilon > 0 : N_\epsilon(z_0) \subseteq S$ .

**open point set**  $S$  is open if it contains only interior points:  $\forall z_0 \in S \exists \epsilon > 0 : N_\epsilon(z_0) \subseteq S$ .

**closed point set**  $S$  is closed if it contains all of its limit points or has no limit points.

**boundary points** The point  $z_0$  is a boundary point if every  $\epsilon$ -neighborhood of  $z_0$  contains both points in  $S$  and  $\bar{S}$ , i.e.  $\forall \epsilon > 0 \exists z_1 \in S, z_2 \in \bar{S} : z_1, z_2 \in N_\epsilon(z_0)$ . Note that  $z_0$  need not be in  $S$ .

## 2.2 Curves, contours and connectedness

**Definition 2.2.1** (Curves and contours.) A **simple (closed) curve** is a set of points defined by a parametric equation  $z(t) = x(t) + iy(t)$ ,  $0 \leq t \leq 1$  where  $x(t)$  and  $y(t)$  are continuous real-valued functions such that  $t_1 \neq t_2 \Rightarrow z(t_1) \neq z(t_2)$  ( $z(t_1) = z(t_2) \iff (t_1, t_2) \in \{(0, 1), (1, 0)\}$ ). A **simple (closed) smooth curve** is a simple (closed) curve with parameterization  $z(t) = x(t) + iy(t)$  where  $x(t)$  and  $y(t)$  have continuous derivatives  $x'(t)$  and  $y'(t)$  and  $x'(t)^2 + y'(t)^2 \neq 0$ . A **simple (closed) contour** is like a simple (closed) smooth curve, but only requires piecewise continuity in the derivatives.

The smoothness requirement in the above definition ensures that  $z(t)$  has a tangent for  $t \in [0, 1]$ .

**Definition 2.2.2** (Domains.) A **domain** is a nonempty open connected point set. A point set  $S$  is **connected** if every pair of points in  $S$  can be joined by a simple curve lying entirely in  $S$ . A domain  $D$  is **simply connected** if every closed curve in  $D$  can be shrunk to a point without leaving  $D$ .

We will occasionally use *region* as a synonym of *domain*.

### 2.3 Limits, continuity, and differentiation

**Definition 2.3.1** (Limits.) For  $w = f(z)$  defined on some deleted  $r$ -neighborhood  $N_r(z_0) \setminus \{z_0\}$  of  $z_0$ ,  $\lim_{z \rightarrow z_0} f(z) = a$  means that for all positive  $\epsilon$  there exists a positive  $\delta$  such that for all  $z$  in the deleted neighborhood and the disk  $\{z \mid |z - z_0| < \delta\}$ ,  $|f(z) - a| < \epsilon$ ; i.e.  $\forall \epsilon > 0 \exists \delta > 0 : z \in N_r(z_0) \setminus \{z_0\}, z \in N_\delta(z_0) \Rightarrow |f(z) - a| < \epsilon$ .

**Definition 2.3.2** (Continuity.) A function  $w = f(z)$ , defined in  $S$ , is **continuous** at  $z_0 \in S$  if  $f(z_0) \neq \infty$  and  $\lim_{z \rightarrow z_0} f(z) = f(z_0)$ ;  $f(z)$  is **uniformly continuous** on  $S$  if it is continuous in  $S$  and for every positive  $\epsilon$ , there exists a  $\delta$  independent of  $z_0$  such that for all  $z$  in the disk  $\{z \mid |z - z_0| < \delta\}$ ,  $|f(z) - f(z_0)| < \epsilon$ ; i.e.  $\forall \epsilon > 0 \exists \delta > 0 : z \in S, z_0 \in S, |z - z_0| < \delta \Rightarrow |f(z) - f(z_0)| < \epsilon$ .

The emphasis in the definition of uniform continuity is on the fact that  $\delta$  is independent of  $\epsilon$ ; for any  $\epsilon$  the difference between any two image points can be made arbitrarily small, as long as the two pre-images are within a distance of  $\delta$  of each other, anywhere in  $S$ .

**Theorem 2.3.1** If  $f(z)$  is continuous on the closed set  $S$ , it is uniformly continuous in  $S$ .

**Definition 2.3.3** (Differentiation.) The function  $w = f(z)$ , defined in some  $\epsilon$ -neighborhood of  $z_0$ , is **differentiable** at  $z_0$  if  $f(z_0) \neq \infty$  and

$$\lim_{z \rightarrow z_0} \frac{f(z) - f(z_0)}{z - z_0}$$

exists. This limit is denoted by  $f'(z_0)$ .

If  $w = f(z)$  is defined in some  $\epsilon$ -neighborhood of  $z_0$  then  $f$  is **analytic at**  $z_0$  if it is differentiable in some  $\epsilon$ -neighborhood of  $z_0$ ;  $f(z)$  is **analytic on** a domain  $S$  if it is analytic at every  $z_0 \in S$ .

The following lemma relates differentiability and continuity.

**Lemma 2.3.1** If  $f'(z_0)$  exists,  $f(z)$  is continuous at  $z_0$ .

## 2.4 Convergent series of analytic functions

**Definition 2.4.1** A sequence  $\{z_n\}$  of complex numbers **converges** to a complex number  $z_0$  if for all  $\epsilon > 0$ , there is a  $N \in \mathbb{Z}^+$  such that  $n \geq N$  implies that  $|z_n - z_0| < \epsilon$ .

An infinite series  $\sum a_k$  is said to **converge** to  $S$  and we write  $\sum a_k = S$  if the sequence  $s_n = \sum_1^n a_k$  of partial sums converges to  $S$ . A complex series is said to **converge absolutely** if  $\sum |a_k|$  converges.

**Definition 2.4.2** Suppose  $f_n : A \mapsto \mathbb{C}$  is a sequence of functions defined on the point set  $A$ . The sequence  $\{f_n\}$  **converges pointwise** if for each  $z \in A$ ,  $\{f_n(z)\}$  converges. It **converges uniformly** to a function  $f$  if for all  $\epsilon > 0$ , there is a  $N \in \mathbb{Z}^+$  such that  $n \geq N$  implies that  $|f_n(z) - f(z)| < \epsilon$  for all  $z \in A$ .

A series  $\sum g_k(z)$  **converges pointwise** if the corresponding partial sums  $s_n(z) = \sum_{k=1}^n g_k(z)$  converge pointwise. A series  $\sum g_k(z)$  **converges uniformly** if  $s_n(z)$  converges uniformly.

**Lemma 2.4.1** Cauchy criterion. The sequence  $\{f_n(z)\}$  converges uniformly on  $A$  iff for all  $\epsilon > 0$ , there is a  $N \in \mathbb{Z}^+$  such that  $n \geq N$  implies that  $|f_n(z) - f_{n+r}(z)| < \epsilon$  for all  $z \in A$  and  $r \in \mathbb{Z}^+$ .

A series  $\sum g_k(z)$  converges uniformly on  $A$  iff for all  $\epsilon > 0$ , there is a  $N \in \mathbb{Z}^+$  such that  $n \geq N$  implies that

$$\left| \sum_{k=n+1}^{n+r} g_k(z) \right| < \epsilon$$

for all  $z \in A$  and  $r \in \mathbb{Z}^+$ .

**Lemma 2.4.2** Weierstrass  $M$ -test. Let  $\{g_n\}$  be a sequence of functions defined on a set  $A \subset \mathbb{C}$ . Suppose there exist  $\{M_n\} \subset \mathbb{R}$ , where  $M_n \geq 0$  such that  $|g_n(z)| \leq M_n$  for all  $z \in A$  and  $\sum M_n$  converges. Then  $\sum g_n(z)$  converges absolutely and uniformly on  $A$ .

**Theorem 2.4.1** Let  $A$  be a region in  $\mathbb{C}$  and let  $\{f_n\}, \{g_k\}$  be sequences of **analytic** functions defined on  $A$ . If  $f_n$  converges uniformly to  $F$  on every closed disk contained in  $A$ , then  $f$  is analytic. If  $g(z) = \sum g_k(z)$  converges uniformly on every closed disk in  $A$ , then  $g$  is analytic on  $A$  and  $g'(z) = \sum g'_k(z)$  pointwise on  $A$  and uniformly on every closed disk contained in  $A$ .

### 2.4.1 Analytic continuation, Laurent series and classification of singularities

The use of Laurent expansions and residues is central to this thesis. The Laurent expansion of a function  $f$  generalizes the concept of a Taylor expansion. Unlike Taylor expansions, Laurent expansions apply not only to points where  $f$  is analytic, but also to points where it fails to be analytic, i.e. to *singularities*. In order to discuss the Laurent expansion, we must first define the terms *analytic continuation*, and *singularity*. These are important in their own right; the reader will recall from the introduction that the evaluation of harmonic sums by the Mellin transform requires that we obtain an analytic continuation of the Dirichlet series in the amplitudes to the left or right of its fundamental strip.

**Definition 2.4.3** *An analytic function element  $(f, D)$  is an analytic function  $f(z)$  along with its domain of definition  $D$ . A function element  $(f_2, D_2)$  is a **direct analytic continuation** of another element  $(f_1, D_1)$  if  $D = D_2 \cap D_1 \neq \emptyset$  and  $f_1 \equiv f_2$  in  $D$ . A **complete analytic function** is the collection of all possible analytic function elements  $(f, D)$  starting with a given element  $(f_0, D_0)$  such that a chain of direct continuations exists between  $(f, D)$  and  $(f_0, D_0)$ . A **singularity** of a complete analytic function is a limit point of a domain of one or more elements that is not itself in the domain of any element.*

Note that a complete analytic function may be multi-valued; the standard example is the complex logarithm. We will briefly discuss this example, because it demonstrates one of the most common techniques for analytic continuation used in actual applications. This technique is analytic continuation by power series along curves; the domains  $D$  of the function elements are disks. Suppose  $f(z)$  is analytic in a neighborhood  $U$  of  $z_0$ , and we wish to obtain its analytic continuation to a point  $z_1$  not in  $U$ . Let  $\gamma$  be a curve that joins  $z_0$  to  $z_1$ . If we can expand  $f$  into a power series of radius  $\rho$  that such that  $\{z \mid |z - z_0| < \rho\}$  includes more of  $\gamma$  than  $U$ , this series defines an analytic continuation of  $f$  along  $\gamma$ . We can repeat this process with the goal of eventually reaching  $z_1$ . We will reach  $z_1$  if the successive radii of convergence do not shrink to zero. There are theorems to decide when the resulting function will be single valued; if we continue a function  $f$  around a simple closed curve, we cannot return to the starting point with a different value unless there is a singularity of  $f$  inside the curve; alternatively, continuation along two different curves with the same start and end points leads to the same value of  $f$  at the end point, unless there is a singularity of  $f$  between the two curves.

**Example.** We study the complex logarithm. We will construct an instance where analytic continuation around a singularity results in a different value on return to the starting point. The initial function

element  $(f_0, D_0)$  has  $f_0(z) = \log z$  and  $D_0 = \{z \mid |z - 1| < 1\}$ , where  $f_0(z)$  is the branch of the logarithm defined by

$$\sum_{k=0}^{\infty} \frac{(-1)^k (z-1)^{k+1}}{k}, \quad |z-1| < 1,$$

i.e. with  $\log 1 = 0$ . One way to continue  $f_0(z)$  along  $\gamma = \{z \mid |z| = 1\}$ , the interior of which includes  $z = 0$ , is via the following nine function elements (here  $0 \leq m \leq 8$ ).

$$f_m(z) = m \frac{i\pi}{4} + \sum_{k=0}^{\infty} \frac{(-1)^k (z - e^{mi\pi/4})^{k+1}}{k e^{mi\pi/4}}, \quad D_m = \{z \mid |z - e^{mi\pi/4}| < 1\}.$$

The function element  $(f_m, D_m)$  is obtained from  $(f_{m-1}, D_{m-1})$  by noting that the center  $z_m$  of the disk  $D_m$  lies inside  $D_{m-1}$ . Hence we can compute the Taylor expansion of  $f_{m-1}$  at  $z_m$  in terms of the expansion at  $z_{m-1}$ . This new expansion has radius of convergence  $\rho = 1$  and defines the next function element in the analytic continuation of  $(f_{m-1}, D_{m-1})$ .

The function elements  $(f_0, D_0)$  and  $(f_8, D_8)$  clearly cover the same domain  $\{z \mid |z - 1| < 1\}$ , but their values at the points of this domain differ by  $2\pi i$ . It is natural to ask whether there exists a paradigm to describe this behavior. This thesis is not concerned with multivalued analytic continuation. Nonetheless we remark in passing that such a paradigm does exist. It rests on a striking idea by Riemann, the so-called *Riemann surface* of the complete analytic function  $f$ . We construct this surface so that  $f$  is single-valued on it. We envision it as situated “over” the complex plane  $\mathbb{C}$ . In fact it is embedded in  $\mathbb{C}^2$ , i.e. a four-dimensional space, but there are cases when we can use the single dimension  $z$  (“height”) of three-space to capture useful information about the shape of the surface in  $\mathbb{C}^2$ . The function elements  $(f, D)$  of the complete analytic function  $f$  are situated “over”  $D$ . They are patches of the surface. If two function elements are direct continuations of one another, their respective domains overlap; the corresponding patches on the surface overlap also. If two function elements have the same domain, but give different values for  $f$ , the corresponding two patches do not overlap. If the difference in values is one-dimensional, we can indicate it by situating the two patches at different heights over  $\mathbb{C}$ . If the image  $f(z)$  of a point  $z \in \mathbb{C}$  has cardinality  $n = |f(z)| \geq 1$ , we map  $z$  to  $n$  different points on the Riemann surface, one for each value in the image set. These points are said to lie on different *sheets* of the surface. The sheets of the Riemann surface are copies of  $\mathbb{C}$  connected such that we can continuously pass from one sheet to another along suitably chosen curves and in this way obtain all the values of the complete analytic function  $f(z)$ .

Riemann surface of  $\log z$ ; 5 sheets of the surface are shown.

The Riemann surface of  $\log z$  is a surface that can be visualized in three-space. This is because the difference between the values “above” a specific  $z$  is an imaginary constant ( $2\pi i$ ), and can therefore be mapped to the single three-space dimension “height”.

**Theorem 2.4.2** Laurent expansion. *Let  $r_1, r_2 \in \mathbb{R}$ ,  $r_1 \geq 0$ ,  $r_2 > 0$  and  $z_0 \in \mathbb{C}$ . Define the annulus  $A = \{z \in \mathbb{C} \mid r_1 < |z - z_0| < r_2\}$ . The combinations  $r_1 = 0$  (deleted neighborhood) or  $r_2 = \infty$  (open complement of a disk relative to  $\mathbb{C}$ ) or both ( $\mathbb{C} \setminus 0$ ) are permitted. Let  $f$  be analytic on  $A$ . There exist  $b_n \in \mathbb{C}$  such that we may write*

$$f(z) = \sum_{n=-\infty}^{\infty} b_n (z - z_0)^n$$

and this series converges absolutely on  $A$  and uniformly on any closed annulus  $B \subset A$  of the form  $B_{\rho_1, \rho_2} = \{z \in \mathbb{C} \mid \rho_1 \leq |z - z_0| \leq \rho_2\}$  where  $r_1 < \rho_1 < \rho_2 < r_2$ . For  $\gamma$  any circle  $\{z_0 + re^{i\theta} \mid 0 \leq \theta \leq 2\pi\}$  around  $z_0$  with radius  $r$  and  $r_1 < r < r_2$  the coefficients are given by

$$b_n = \frac{1}{2\pi i} \int_{\gamma} \frac{f(\zeta)}{(\zeta - z_0)^{n+1}} d\zeta.$$

This expansion is the **Laurent expansion** of  $f$  around  $z_0$  in  $A$  and it is unique.

A singularity of  $f(z)$  in a region  $A$  can be classified according to the Laurent expansion about the singularity. We are concerned particularly with expansions in deleted neighborhoods, i.e. in annuli with  $r_1 = 0$ .

**Definition 2.4.4** Classification of singularities. Let  $f$  be analytic in a region  $A$  that contains a deleted neighborhood  $N_\epsilon(z_0) \setminus \{z_0\}$ , and let  $f$  fail to be analytic at  $z_0$ . We say that  $z_0$  is an **isolated singularity**. Let  $\{b_n\}$  be the coefficients of the Laurent expansion of  $f$  in the annulus  $N_\epsilon(z_0) \setminus \{z_0\}$ .

- **Principal part.** The expansion

$$\sum_{n=-\infty}^{-1} b_n (z - z_0)^n$$

is the principal part of  $f$  at  $z_0$ .

- **Pole.** An isolated singularity is a pole if the principal part has only a finite number of non-zero coefficients.
- **Pole of order  $k$ .** The point  $z_0$  is a pole of order  $k$  if the principal part has the form

$$\sum_{n=-k}^{-1} b_n (z - z_0)^n.$$

- **Simple pole.** A simple pole is a pole of order 1.
- **Essential singularity.** If the number of zero coefficients in the principal part is finite,  $z_0$  is an essential singularity.
- **Residue.** The coefficient  $b_{-1}$  is the residue of  $f$  at  $z_0$ . We write  $b_{-1} = \text{Res}(f(z); z = z_0)$ .
- **Removable singularity.** The point  $z_0$  is a removable singularity if all the coefficients of the principal part are zero.
- **Meromorphic functions.** A function is meromorphic in a region  $A$  if it is analytic in  $A$  with the exception of poles. A function  $f$  is meromorphic if it is meromorphic in  $\mathbb{C}$ .

We use  $\text{Sing}(f(z))$  to denote the set of finite singularities of  $f$ .

It is important that we be able to compute the residues of a function  $f$  at its isolated singularities; e.g., the Mellin-Perron formula produces an integral that can be evaluated or estimated with the residue theorem.

**Lemma 2.4.3** Computation of residues. Let  $f$  have an isolated singularity at  $z_0$  and let  $k \in \mathbb{N}$  be the smallest integer such that  $\lim_{s \rightarrow z_0} (z - z_0)^k f(z)$  exists. Then  $f(z)$  has a pole of order  $k$  at  $z_0$  and if we let  $\phi(z) = (z - z_0)^k f(z)$ , then  $\phi$  can be defined uniquely at  $z_0$  so that  $\phi$  is analytic at  $z_0$  and

$$\text{Res}(f(z); z = z_0) = \frac{\phi^{(k-1)}(z_0)}{(k-1)!}.$$

## 2.5 The Cauchy residue theorem

**Definition 2.5.1** Let  $\gamma$  be a closed curve in  $\mathbb{C}$  and  $z_0$  be a point not on  $\gamma$ . Then the **index** or **winding number** of  $\gamma$  with respect to  $z_0$  is defined by

$$I(\gamma, z_0) = \frac{1}{2\pi i} \int_{\gamma} \frac{dz}{z - z_0}.$$

The curve  $\gamma$  winds around  $z_0$   $I(\gamma, z_0)$  times.

**Theorem 2.5.1** (Cauchy residue theorem.) Let  $A$  be a region and let  $z_1, z_2, \dots, z_n$  be distinct points in  $A$ . Let  $f$  be analytic on  $A \setminus \{z_1, z_2, \dots, z_n\}$ . Let  $\gamma$  be a closed curve in  $A$  homotopic to, i.e. smoothly shrinkable to, a point in  $A$ . Assume none of  $z_1, z_2, \dots, z_n$  lie on  $\gamma$ . We have

$$\frac{1}{2\pi i} \int_{\gamma} f(z) dz = \sum_{i=1}^n \text{Res}(f(z); z = z_i) I(\gamma, z_i).$$

We will usually apply this theorem with  $I(\gamma, z_i) = 1$ .

## 2.6 Dirichlet series

The sequences considered in this section are of the form  $\{a_n\}_{n \geq 1}$ ,  $a_n \in \mathbb{C}$ ; i.e.  $\{a_n\}_{n \geq 1}$  is an arithmetical function.

**Definition 2.6.1** (Dirichlet Series.) The Dirichlet series associated to  $\{a_n\}_{n \geq 1}$ ,  $a_n \in \mathbb{C}$  is

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

The function  $A(s)$  is known as the Dirichlet generating function of  $\{a_n\}_{n \geq 1}$ . The generalized Dirichlet series  $(A, \Lambda)$ , with  $\Lambda$  as in Definition 2.0.3 and  $A = \{a_n\}_{n \geq 1}$  is given by

$$\sum_{n=1}^{\infty} a_n e^{-\lambda_n s}.$$

We will be concerned mostly with the first kind.

**Theorem 2.6.1** (Abcissae of convergence.) If  $\sum |a_n n^{-s}|$  does not converge for all  $s$  or diverge for all  $s$ , then there exists  $\sigma_a \in \mathbb{R}$  such that  $A(s) = \sum a_n n^{-s}$  converges absolutely if  $\sigma > \sigma_a$  but does not converge absolutely if  $\sigma < \sigma_a$ ;  $\sigma_a$  is the abscissa of absolute convergence of  $A(s)$ .

If  $A(s) = \sum a_n n^{-s}$  does not converge everywhere or diverge everywhere, then there exists  $\sigma_c \in \mathbb{R}$  such that  $A(s)$  converges if  $\sigma > \sigma_c$  but does not converge if  $\sigma < \sigma_c$ ;  $\sigma_c$  is the abscissa of convergence of  $A(s)$ .

The following remarkable theorem is due to S. Mandelbrojt.

**Theorem 2.6.2** *Let  $\Lambda$  be of positive step  $h$ ; let  $D$  be its upper density. Let  $f(s) = (A, \Lambda)$  with  $\sigma_a$  finite. For  $\alpha > 0$ ,  $\beta \geq 0$  there exists a continuous function  $A(\alpha, \beta)$  with  $A(\alpha, 0) = 0$  such that for all  $t_0 \in \mathbb{R}$   $f(s)$  has a singular point in the rectangle  $\{\sigma + it \mid \sigma_a - A(h, D) \leq \sigma \leq \sigma_a, |t - t_0| \leq \pi D\}$ . One such function  $A(\alpha, \beta)$  is*

$$A(\alpha, \beta) = \begin{cases} \pi\beta - (3 \log(\alpha\beta) - \frac{17}{2})\beta & \text{when } \beta > 0 \\ 0 & \text{when } \beta = 0. \end{cases}$$

**Example.** For the Dirichlet series  $\sum 2^{-ks}$ ,  $\sigma_a = 0$ ,  $h = \log 2$  and with  $D = 1/\log 2$  the height of the rectangle, vertically centered at  $t_0$ , becomes  $2\pi/\log 2$ . Indeed  $f(s) = -1 + 2^s/(2^s - 1)$  is meromorphic in all of  $\mathbb{C}$  with poles at  $2\pi ik/\log 2$ ,  $k \in \mathbb{Z}$ .

Theorem 2.6.2 has the following immediate corollary.

**Corollary 2.6.1** (Theorem of Fabry-Pólya.) *If  $h > 0$ ,  $D = 0$  and  $\sigma_a$  is finite, then every point on the abscissa of absolute convergence is a singular point of  $f$ ; i.e.  $\sigma_a$  is a natural boundary of  $f$ .*

**Example.** The Dirichlet series  $A(s) = \sum 1/(k+1)!$  has  $D = 0$  and  $h = +\infty$ , hence the line  $\sigma = 0$  is a natural boundary of  $f(s)$ , and  $f(s)$  has no analytic continuation into the left half-plane.

The following theorem lists analytic properties of Dirichlet series.

**Theorem 2.6.3** *The Dirichlet series*

$$A(s) = \sum_{n=1}^{\infty} \frac{a_n}{n^s}$$

*is analytic in its half-plane of convergence  $\sigma > \sigma_c$ . Its derivative  $A'(s)$  is represented in this half-plane by the series*

$$A'(s) = \sum_{n=1}^{\infty} \frac{a_n \log n}{n^s}.$$

*$A'(s)$  has the same abscissa of convergence  $\sigma_c$  and abscissa of absolute convergence  $\sigma_a$  as  $A(s)$ .*

### 2.6.1 The Riemann and Hurwitz $\zeta$ functions

**Definition 2.6.2** *The Hurwitz zeta function is given by*

$$\zeta(s, a) = \sum_{n=0}^{\infty} \frac{1}{(n+a)^s}.$$

*The Riemann zeta function is the function*

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

*Both series define analytic functions for  $\sigma > 1$ .*

The Riemann  $\zeta$  function is probably the most famous of all Dirichlet series. This is because it can be used to study the distribution of primes. The relation between the Riemann  $\zeta$  function and the sequence of primes will be explained in the next section. The following theorems list various properties of the  $\zeta$  function.

**Theorem 2.6.4** *The equality*

$$\pi^{-s/2} \Gamma\left(\frac{s}{2}\right) \zeta(s) = \frac{1}{s(s-1)} + \int_0^{\infty} \left(x^{s/2-1} + x^{-s/2-1/2}\right) \omega(x) dx,$$

*where  $\omega(x) = \sum e^{-\pi n^2 x}$ , holds for  $\sigma > 1$ .*

The term

$$\frac{1}{s(s-1)} + \int_0^{\infty} \left(x^{s/2-1} + x^{-s/2-1/2}\right) \omega(x) dx$$

is meromorphic and provides the analytic continuation of  $\zeta(s)$  to all of  $\mathbb{C}$ .

**Theorem 2.6.5** *The Riemann  $\zeta$  function is meromorphic with a single pole at  $s = 1$ . This pole is simple and  $\text{Res}(\zeta(s); s = 1) = 1$ . Riemann's functional equation holds:*

$$\pi^{-s/2} \Gamma\left(\frac{s}{2}\right) \zeta(s) = \pi^{-(1-s)/2} \Gamma\left(\frac{1-s}{2}\right) \zeta(1-s).$$

We list some special values of  $\zeta(s, a)$ .

**Theorem 2.6.6** *We have*

$$\zeta(0, a) = \frac{1}{2} - a \quad \text{and} \quad \zeta'(0, a) = \log \Gamma(a) - \frac{1}{2} \log(2\pi)$$

*where  $\Gamma(a)$  is the Euler gamma function.*

The Bernoulli polynomials, defined below, are used to compute the values of  $\zeta(-m, a)$  where  $m \in \mathbb{N}$ .

**Definition 2.6.3** *The set of Bernoulli polynomials  $\{B_n(x)\}$  is defined by the following relation.*

$$\frac{ze^{xz}}{e^z - 1} = \sum_{n=0}^{\infty} B_n(x) \frac{z^n}{n!}.$$

The first four Bernoulli polynomials are

$$B_0(x) = 1, \quad B_1(x) = x - \frac{1}{2}, \quad B_2(x) = x^2 - x + \frac{1}{6}, \quad B_3(x) = x^3 - \frac{3}{2}x^2 + \frac{1}{2}x.$$

**Theorem 2.6.7** *Let  $m \in \mathbb{N}$ .*

$$\zeta(-m, a) = -\frac{B_{m+1}(a)}{m+1}$$

We recall some growth properties of  $\zeta(s, a)$ .

**Theorem 2.6.8** (Whittaker-Watson.) *Let  $s = \sigma + it$  and  $\delta \in (0, 1/2)$ . The following set of relations describes the growth of  $\zeta(s, a)$  in  $\langle -\delta, \infty \rangle$ .*

$$\zeta(s, a) \in \begin{cases} \mathcal{O}(|t|^{1/2} \log |t|) & \text{if } s \in \langle -\delta, \delta \rangle \\ \mathcal{O}(|t|^{1/2}) & \text{if } s \in \langle \delta, 1 - \delta \rangle \\ \mathcal{O}(|t|^{1-\sigma} \log |t|) & \text{if } s \in \langle 1 - \delta, 1 + \delta \rangle \\ \mathcal{O}(1) & \text{if } s \in \langle 1 + \delta, \infty \rangle \end{cases}$$

## 2.7 The analytic version of the fundamental theorem of arithmetic

**Definition 2.7.1** (Dirichlet product.) *Let*

$$c_n = \sum_{t|n} a_t b_{n/t};$$

$\{c_n\}_{n \geq 1}$  *is the Dirichlet product of  $\{a_n\}_{n \geq 1}$  and  $\{b_n\}_{n \geq 1}$  and is written*

$$c_n = a_n * b_n.$$

*This product is also referred to as the Dirichlet convolution of  $\{a_n\}_{n \geq 1}$  and  $\{b_n\}_{n \geq 1}$ .*

**Theorem 2.7.1** (Dirichlet products and Dirichlet series.) *Let*

$$A(s) = \sum \frac{a_n}{n^s}, \text{ and } B(s) = \sum \frac{b_n}{n^s}$$

*with abscissae of absolute convergence  $a$  and  $b$ . Let  $c_n = a_n * b_n$ . Then*

$$C(s) = \sum \frac{c_n}{n^s}$$

*converges absolutely with abscissa  $c = \max\{a, b\}$  and*

$$C(s) = A(s)B(s).$$

The proof of this theorem uses the following computation.

$$\sum \frac{c_n}{n^s} = \sum \frac{1}{n^s} a_n * b_n = \sum \frac{1}{n^s} \sum_{t|n} a_t b_{n/t} = \sum \frac{1}{n^s} \sum_{kl=n} a_k b_l = \sum_{k \geq 1} \sum_{l \geq 1} \frac{a_k b_l}{(kl)^s}.$$

**Definition 2.7.2** (Multiplicative and completely multiplicative arithmetical functions.) *An arithmetical function  $\{a_n\}_{n \geq 1}$  is multiplicative if  $a_{n_1 n_2} = a_{n_1} a_{n_2}$  when  $(n_1, n_2) = 1$ ;  $\{a_n\}_{n \geq 1}$  is completely multiplicative if  $a_{n_1 n_2} = a_{n_1} a_{n_2}$  for all  $n_1, n_2$ .*

Note that this definition implies  $a_1 = 1$  unless  $\{a_n\}_{n \geq 1}$  vanishes everywhere;  $\exists n : a_n \neq 0$  and hence  $a_n = a_1 a_n$  or  $a_1 = 1$ .

**Theorem 2.7.2** (Analytic version of the fundamental theorem of arithmetic.) *Let  $\{a_n\}_{n \geq 1}$  be multiplicative; let  $A(s)$  be its Dirichlet series with abscissa of absolute convergence  $a$ . We have*

$$\sum \frac{a_n}{n^s} = \prod_p \sum_{v=0}^{\infty} \frac{a_p^v}{p^{vs}}$$

*when  $\sigma > a$ . If  $\{a_n\}_{n \geq 1}$  is completely multiplicative, this simplifies to*

$$\sum \frac{a_n}{n^s} = \prod_p \frac{1}{1 - a_p p^{-s}}.$$

*The term on the right is known as the Euler product of  $\{a_n\}_{n \geq 1}$ .*

**Proof.** Let

$$A_p(s) = \sum_{v=0}^{\infty} \frac{a_p^v}{p^{vs}}$$

and consider

$$\prod_{p \leq p_r} A_p(s) = \sum_{v_1=0}^{\infty} \cdots \sum_{v_r=0}^{\infty} \frac{a_p^{v_1} \cdots a_p^{v_r}}{p^{v_1 s} \cdots p^{v_r s}} = \sum_r \frac{a_n}{n^s}$$

where the product ranges over the first  $r$  primes and the sum includes those  $n$  with prime factors  $\leq p_r$ .

We have

$$\lim_{r \rightarrow \infty} \left| \sum \frac{a_n}{n^s} - \sum_r \frac{a_n}{n^s} \right| = \lim_{r \rightarrow \infty} \left| \sum_{\exists \rho > r: p_\rho | n} \frac{a_n}{n^s} \right| < \lim_{r \rightarrow \infty} \left| \sum_{n=p_r+1} \frac{a_n}{n^s} \right| \leq \lim_{r \rightarrow \infty} \sum_{n=p_r+1} \frac{|a_n|}{n^s} = 0.$$

We have used the absolute convergence of  $A(s)$  in the last step. We conclude that

$$\lim_{r \rightarrow \infty} \sum_r \frac{a_n}{n^s} = \sum \frac{a_n}{n^s}.$$

It remains to verify that  $\prod_{p \leq p_r} A_p(s)$  converges. Recall that  $1 + x < e^x = 1 + x + \sum_{n=2}^{\infty} \frac{x^n}{n!}$  and hence  $\log(1+x) < x$  for  $x \in \mathbb{R}, x > -1$ . We use this inequality to obtain

$$\left| \log \prod_{p \leq p_r} A_p(s) \right| \leq \sum_{p \leq p_r} |\log A_p(s)| = \sum_{p \leq p_r} \left| \log \left( 1 + \sum_{v=1}^{\infty} \frac{a_p^v}{p^{vs}} \right) \right| \leq \sum_{p \leq p_r} \sum_{v=1}^{\infty} \frac{|a_p^v|}{p^{vs}}.$$

All the partial sums are bounded and hence the series on the left and the product both converge. We have

$$\sum \frac{a_n}{n^s} = \prod_p A_p(s).$$

This is the desired result.  $\blacksquare$

We thus have

$$\sum_{n=1}^{\infty} \frac{1}{n^s} = \prod_p \frac{1}{1-p^{-s}}.$$

This relation was already known to Euler.

### 2.7.1 Some useful Dirichlet generating functions

The examples in this section have been selected to illustrate Theorem 2.7.2, and demonstrate additional techniques for the evaluation of Dirichlet generating functions. We will use them later, when we evaluate digital sums.

**Definition 2.7.3** Let  $\{\kappa(j)\}_{j \geq 0}$  be a sequence of positive integers such that  $\kappa(0) = 1$  and  $\kappa(j) \mid \kappa(j+1)$ ,  $\kappa(j) < \kappa(j+1)$ , for  $j \geq 0$ . Then the function  $v_\kappa : \mathbb{Z}^+ \mapsto \mathbb{N}$  is defined as

$$v_\kappa(n) = \max\{j \mid j \in \mathbb{N}, \kappa(j) \mid n\}.$$

The special case  $\kappa(j) = q^j$  where  $q \geq 2$ , is denoted by  $v_q(n)$ ;

$$v_q(n) = \max\{v \mid v \in \mathbb{N}, q^v \mid n\}.$$

The function  $v_q(n)$  gives the highest power of  $q$  that divides  $n$ .

**Definition 2.7.4** The function  $\kappa^{-1}(k)$  is defined to be an integer-valued inverse of  $\kappa(j)$ .

$$\kappa^{-1}(k) = j \Leftrightarrow \kappa(j-1) < k \leq \kappa(j)$$

When  $\kappa(j) = q^j$  where  $q \geq 2$ ,

$$\kappa^{-1}(k) = \lceil \log_q k \rceil.$$

We evaluate several Dirichlet generating functions that contain  $v_\kappa$  and  $v_q$ .

**Example.** By definition of  $v_q(n)$  the function  $(-1)^{v_q(n)}$  is completely multiplicative when  $q$  is prime, and multiplicative when  $q$  is a prime power. We can evaluate the Dirichlet generating function of  $(-1)^{v_q(n)}$  by Theorem 2.7.2 when  $q$  is prime. We have

$$\begin{aligned} \sum_{n \geq 1} \frac{(-1)^{v_q(n)}}{n^s} &= \prod_p \frac{1}{1 - (-1)^{v_q(p)} p^{-s}} = \frac{1}{1 + q^{-s}} \prod_{p \neq q} \frac{1}{1 - p^{-s}} \\ &= \frac{1 - q^{-s}}{1 + q^{-s}} \prod_p \frac{1}{1 - p^{-s}} = \zeta(s) \frac{q^s - 1}{q^s + 1} = \zeta(s) \left(1 - 2 \frac{1}{q^s + 1}\right). \end{aligned}$$

In fact this result holds for composite  $q$  as well.

$$\begin{aligned} \sum_{n \geq 1} \frac{(-1)^{v_q(n)}}{n^s} &= \sum_{k \geq 1} \frac{(-1)^{v_q(kq)}}{(kq)^s} + \sum_{r=1}^{q-1} \sum_{k \geq 0} \frac{(-1)^{v_q(kq+r)}}{(kq+r)^s} = \sum_{k \geq 1} \frac{1}{q^s} \frac{(-1)^{v_q(k)+1}}{k^s} + \sum_{r=1}^{q-1} \sum_{k \geq 0} \frac{1}{(kq+r)^s} \\ &= -\frac{1}{q^s} \sum_{k \geq 1} \frac{(-1)^{v_q(k)}}{k^s} + \sum_{k \geq 1} \frac{1}{k^s} - \sum_{k \geq 1} \frac{1}{(kq)^s} = -\frac{1}{q^s} \sum_{k \geq 1} \frac{(-1)^{v_q(k)}}{k^s} + \zeta(s) \left(1 - \frac{1}{q^s}\right) \end{aligned}$$

Further manipulation yields

$$\begin{aligned} \left(1 + \frac{1}{q^s}\right) \sum_{k \geq 1} \frac{(-1)^{v_q(k)}}{k^s} &= \zeta(s) \left(1 - \frac{1}{q^s}\right) \\ \sum_{k \geq 1} \frac{(-1)^{v_q(k)}}{k^s} &= \zeta(s) \frac{q^s - 1}{q^s + 1} = \zeta(s) \left(1 - 2 \frac{1}{q^s + 1}\right). \end{aligned}$$

**Example.** We consider the Dirichlet generating function of the following term:

$$v_q(n) \bmod 2.$$

This function can be evaluated in two ways. The first of these uses the relation

$$v_q(n) \bmod 2 = \frac{1}{2} \left( 1 - (-1)^{v_q(n)} \right).$$

It follows that

$$\sum_{v_q(n) \equiv 1(2)} \frac{1}{n^s} = \frac{1}{2} \zeta(s) - \frac{1}{2} \zeta(s) \left( 1 - 2 \frac{1}{q^s + 1} \right) = \zeta(s) \frac{1}{q^s + 1}.$$

The second approach is more general; it proceeds directly from the definition of  $v_q(n) \bmod 2$ .

$$\begin{aligned} \sum_{v_q(n) \equiv 1(2)} \frac{1}{n^s} &= \sum_{k=0} \sum_{v_q(n)=2k+1} \frac{1}{n^s} = \sum_{k=0} \sum_{q \nmid m} \frac{1}{(q^{2k+1} m)^s} \\ &= \frac{1}{q^s} \sum_{k=0} \frac{1}{q^{2ks}} \sum_{q \nmid m} \frac{1}{m^s} = \frac{1}{q^s} \left( \sum_m \frac{1}{m^s} - \sum_{q|m} \frac{1}{m^s} \right) \sum_{k=0} \frac{1}{q^{2ks}} \\ &= \frac{1}{q^s} \frac{1}{1 - q^{-2s}} \zeta(s) \left( 1 - \frac{1}{q^s} \right) = \zeta(s) \frac{q^s - 1}{(1 - q^{-2s}) q^{2s}} = \zeta(s) \frac{1}{q^s + 1}. \end{aligned}$$

The half-plane of convergence of the last two functions is obtained by a trivial comparison with  $\zeta(s)$ ; hence the respective computations hold for  $\sigma > 1$ .

**Lemma 2.7.1** *Let  $\{\kappa(j)\}_{j \geq 0}$  be a sequence as in Definition 2.7.3. Let  $m \geq 2$  and  $0 \leq r < m$ ,  $m, r \in \mathbb{Z}^+$ .*

*Then*

$$\sum_{v_{\kappa}(n) \equiv r(m)} \frac{1}{n^s} = \zeta(s) \sum_{k=0} \left( \frac{1}{\kappa(mk+r)^s} - \frac{1}{\kappa(mk+r+1)^s} \right)$$

*with  $\sigma > 1$ . When  $\kappa(j) = q^j$ , this simplifies to*

$$\zeta(s) q^{(m-r-1)s} \prod_{v=1}^{m-1} (q^s - \omega_m^v)^{-1},$$

*where  $\omega_m = e^{2\pi i/m}$  is the  $m$ th primitive root of unity.*

**Proof.** The method used in the previous example ( $m = 2$  and  $r = 1$ ) can be used in the general case as well.

$$\begin{aligned} \sum_{v_{\kappa}(n) \equiv r(m)} \frac{1}{n^s} &= \sum_{k=0} \sum_{v_{\kappa}(n)=mk+r} \frac{1}{n^s} = \sum_{k=0} \sum_{\substack{\kappa(mk+r+1) \\ \kappa(mk+r)} \nmid l} \frac{1}{(\kappa(mk+r)l)^s} \\ &= \sum_{k=0} \frac{1}{\kappa(mk+r)^s} \left( \sum_l - \sum_{\substack{\kappa(mk+r+1) \\ \kappa(mk+r)} \mid l} \right) \frac{1}{l^s} = \zeta(s) \sum_{k=0} \frac{1}{\kappa(mk+r)^s} \left( 1 - \frac{\kappa(mk+r)^s}{\kappa(mk+r+1)^s} \right) \end{aligned}$$

The special case  $\kappa(j) = q^j$  gives  $\kappa(mk + r) = (q^m)^k q^r$  and hence

$$\sum_{k=0}^{\infty} \frac{1}{\kappa(mk + r)^s} = \frac{1}{q^{rs}} \sum_{k=0}^{\infty} \frac{1}{(q^{ms})^k} = \frac{1}{q^{rs}} \frac{q^{ms}}{q^{ms} - 1}.$$

With

$$q^{ms} - 1 = \prod_{v=0}^{m-1} (q^s - \omega_m^v) \quad \text{and} \quad \frac{1}{q^{rs}} - \frac{1}{q^{(r+1)s}} = \frac{q^s - 1}{q^{(r+1)s}}$$

we have the result.  $\blacksquare$

**Example.** The third and last example in this series is the Dirichlet generating function of  $v_q(n)$  itself.

The computation is straightforward.

$$\sum \frac{v_q(n)}{n^s} = \sum_{q|n} \frac{v_q(n)}{n^s} = \sum \frac{v_q(qm)}{(qm)^s} = \frac{1}{q^s} \sum \frac{1 + v_q(m)}{m^s} = \frac{1}{q^s} \left( \zeta(s) + \sum \frac{v_q(m)}{m^s} \right).$$

This gives

$$\sum \frac{v_q(n)}{n^s} = \left( 1 - \frac{1}{q^s} \right)^{-1} \frac{\zeta(s)}{q^s} = \frac{\zeta(s)}{q^s - 1}.$$

The previous example is a special case of the following lemma.

**Lemma 2.7.2** *Let  $\{\kappa(j)\}_{j \geq 0}$  be a sequence as in Definition 2.7.3 and consider a function  $t : \mathbb{N} \mapsto \mathbb{C}$ .*

*Then*

$$\sum \frac{1}{n^s} \sum_{j=1}^{v_\kappa(n)} t(j) = \zeta(s) \sum_{j=1}^{\infty} \frac{t(j)}{\kappa(j)^s};$$

*in particular,*

$$\sum \frac{v_\kappa(n)}{n^s} = \zeta(s) \sum_{j=1}^{\infty} \frac{1}{\kappa(j)^s} \quad \text{and} \quad \sum \frac{v_\kappa(n)(v_\kappa(n) + 1)}{2n^s} = \zeta(s) \sum_{j=1}^{\infty} \frac{j}{\kappa(j)^s}$$

*with  $\sigma > 1$ .*

**Proof.** It is an instructive exercise to adapt the technique employed in the evaluation of  $\sum \frac{v_q(n)}{n^s}$  to the above lemma; indeed this yields a proof. We will use a restricted Dirichlet convolution to establish the result.

$$\sum \frac{1}{n^s} \sum_{j=1}^{v_\kappa(n)} t(j) = \sum \frac{1}{n^s} \sum_{\kappa(j)|n, j>0} t(j) = \sum_{n=m\kappa(j), j>0} \sum_{\kappa(j)|n, j>0} \frac{t(j)}{m^s \kappa(j)^s} = \sum_m \sum_{j=1}^{\infty} \frac{t(j)}{m^s \kappa(j)^s} = \zeta(s) \sum_{j=1}^{\infty} \frac{t(j)}{\kappa(j)^s}$$

The key step is the use of  $\sum_{n=m\kappa(j), j>0}$ . This is a Dirichlet convolution with the divisors restricted to

$\{\kappa(j)\}_{j>0}$ . The two particular instances are obtained from  $v_\kappa(n) = \sum_{j=1}^{v_\kappa(n)} 1$  and  $\frac{1}{2} v_\kappa(n)(v_\kappa(n) + 1) =$

$\sum_{j=1}^{v_\kappa(n)} j$ . We need to verify that the latter two converge in  $\sigma > 1$ . By definition of  $v_\kappa(n)$ ,  $v_\kappa(n) \leq \log_2 n$ .

Hence  $\sigma > 2$  and  $\sigma > 3$  would suffice. In fact Theorem 2.6.3 shows that  $\sigma > 1$  for both series.  $\blacksquare$

## 2.8 Integrals of the Hurwitz $\zeta$ -function

The integrals of the generalized  $\zeta$ -function that we will encounter in the remainder of this thesis are all of a similar type. This section provides a *shifting lemma* that makes it possible to evaluate those integrals.

We will use the growth estimates of Theorem 2.6.8 for  $\zeta(s, a)$  to prove the following lemma.

**Lemma 2.8.1** (Shifting lemma.) *Let  $\alpha \in (-1, 0)$  and let  $c > 1$ . Let  $\{T_j\}_{j \geq 1} \subset \mathbb{R}^+$  be an increasing sequence of real numbers such that  $\lim_{j \rightarrow \infty} T_j = \infty$  and  $T_j > T_0$ ,  $T_0$  a fixed constant. Suppose  $\Phi(s)$  is a function that is analytic on all  $R_j = \langle \alpha, c \rangle \cap \{s = \sigma + it \mid |t| \leq T_j\}$  except for a finite number of singularities, and that there exists a constant  $M \in \mathbb{R}^+$  such that  $|\Phi(s)| < M$  independently of  $j$  on the boundary  $\delta R_j$  of  $R_j$ . Let  $S' = \text{Sing}(\Phi(s)\zeta(s, a)/(s(s+1))) \cap \langle \alpha, c \rangle$ . Under these conditions*

$$\frac{1}{2\pi i} \int_{\alpha - i\infty}^{\alpha + i\infty} \Phi(s) \frac{\zeta(s, a)}{s(s+1)} ds = \frac{1}{2\pi i} \int_{c - i\infty}^{c + i\infty} \Phi(s) \frac{\zeta(s, a)}{s(s+1)} ds - \sum_{\varsigma \in S'} \text{Res} \left( \Phi(s) \frac{\zeta(s, a)}{s(s+1)}; s = \varsigma \right)$$

This is a shifting lemma because it tells us that the integral along a line parallel to the imaginary axis and situated at  $c > 1$  maybe shifted to  $\alpha \in (-1, 0)$ , taking into account the residues of the integrand.

**Proof.** We use the following rectangular contour, where  $T = T_j > T_0$ , and the contour is traversed counterclockwise.

$$\begin{aligned} \Gamma_1 &= \{c + it \mid |t| \leq T\} & \Gamma_2 &= \{\sigma + iT \mid \alpha \leq \sigma \leq c\} \\ \Gamma_3 &= \{\alpha + it \mid |t| \leq T\} & \Gamma_4 &= \{\sigma - iT \mid \alpha \leq \sigma \leq c\} \end{aligned}$$

We apply the Cauchy residue theorem:

$$\frac{1}{2\pi i} \int_{\Gamma_1 \cup \Gamma_2 \cup \Gamma_3 \cup \Gamma_4} \frac{\Phi(s)\zeta(s, a)}{s(s+1)} ds = \sum_{\varsigma \in S'} \text{Res} \left( \frac{\Phi(s)\zeta(s, a)}{s(s+1)}; s = \varsigma \right)$$

or

$$\begin{aligned} \frac{1}{2\pi i} \int_{-\Gamma_3} \frac{\Phi(s)\zeta(s, a)}{s(s+1)} ds &= \frac{1}{2\pi i} \int_{\Gamma_1} \frac{\Phi(s)\zeta(s, a)}{s(s+1)} ds + \frac{1}{2\pi i} \int_{\Gamma_2 \cup \Gamma_4} \frac{\Phi(s)\zeta(s, a)}{s(s+1)} ds \\ &\quad - \sum_{\varsigma \in S'} \text{Res} \left( \frac{\Phi(s)\zeta(s, a)}{s(s+1)}; s = \varsigma \right). \end{aligned}$$

We have the result if we can show that (a) the integrals along  $\Gamma_1$  and  $\Gamma_3$  converge and (b) the integrals along  $\Gamma_2$  and  $\Gamma_4$  vanish as  $T = T_j \rightarrow \infty$ .

We verify (a) with a comparison test. Let  $s = c + it$ ,  $ds = idt$ .

$$\begin{aligned} \left| \int_{c-iT}^{c+iT} \frac{\Phi(s)\zeta(s, a)}{s(s+1)} ds \right| &\leq \left( \int_{-T}^{-T_0} + \int_{-T_0}^{T_0} + \int_{T_0}^T \right) \left| \frac{\Phi(c+it)\zeta(c+it)}{s(s+1)} i \right| dt \\ &< C + c_4 M \left( \int_{-T}^{-T_0} + \int_{T_0}^T \right) \frac{1}{\sqrt{c^2 + t^2} \sqrt{(c+1)^2 + t^2}} dt \\ &< C + 2c_4 M \int_{T_0}^T \frac{1}{t^2} dt = C + 2c_4 M \left[ \frac{1}{t} \right]_T^{T_0} = C_1 - \frac{2c_4 M}{T} \end{aligned}$$

The case for  $\Gamma_3$  is similar. This time we use the first rather than the fourth case of Theorem 2.6.8. Let  $s = \alpha + it$ ,  $ds = idt$ .

$$\begin{aligned} \left| \int_{\alpha-iT}^{\alpha+iT} \frac{\Phi(s)\zeta(s, a)}{s(s+1)} ds \right| &\leq \left( \int_{-T}^{-T_0} + \int_{-T_0}^{T_0} + \int_{T_0}^T \right) \left| \frac{\Phi(\alpha+it)\zeta(\alpha+it)}{s(s+1)} i \right| dt \\ &< C + c_1 M \left( \int_{-T}^{-T_0} + \int_{T_0}^T \right) \frac{|t|^{1/2} \log |t|}{\sqrt{\alpha^2 + t^2} \sqrt{(\alpha+1)^2 + t^2}} dt \\ &< C + 2c_1 M \int_{T_0}^T \frac{t^{1/2} \log t}{t^2} dt \end{aligned}$$

The last integral converges, as does, therefore, the integral along  $\Gamma_3$ .

In order to establish (b), we consider integrals along the four types of intervals listed in Theorem 2.6.8. Take  $s = \sigma \pm iT$ ,  $ds = d\sigma$ ,  $\sigma \in [1 - \delta, 1 + \delta]$ ,  $\delta \in (0, 1/2)$ . (This is the only one of the four types we will treat; the other three can be estimated in the same way.) We have

$$\begin{aligned} \left| \int_{1-\delta \pm iT}^{1+\delta \pm iT} \frac{\Phi(s)\zeta(s, a)}{s(s+1)} ds \right| &= \left| \int_{1-\delta}^{1+\delta} \frac{\Phi(\sigma \pm iT)\zeta(\sigma \pm iT, a)}{s(s+1)} d\sigma \right| \\ &\leq \frac{M}{T^2} \int_{1-\delta}^{1+\delta} |\zeta(\sigma \pm iT)| d\sigma \leq \frac{c_3 M}{T^2} \int_{1-\delta}^{1+\delta} |\pm T|^{1-\sigma} \log |\pm T| d\sigma \\ &< \frac{c_3 M \log T}{T^2} \left[ \frac{1}{\log T} T^{1-\sigma} \right]_{1+\delta}^{1-\delta} = \frac{c_3 M (T^\delta - T^{-\delta})}{T^2}. \end{aligned}$$

Hence the integrals along  $\Gamma_{2,4} \cap \langle 1 - \delta, 1 + \delta \rangle$  vanish as  $T \rightarrow \infty$ ; the same estimate works for the remaining three types. Clearly there always exists an appropriate decomposition of  $[\alpha, c]$  into a sequence of intervals that can be treated as above, e.g.  $[-1/4, 3/2] = [-1/4, 1/4] [1/4, 1 - 1/4] [1 - 1/4, 1 + 1/4] [1 + 1/4, 3/2]$ . This shows that the integrals along the horizontal segments  $\Gamma_2$  and  $\Gamma_4$  vanish as claimed, and concludes the proof of the lemma. ■

There is a straightforward generalization of this lemma.

**Lemma 2.8.2** (Generalized shifting lemma.) *Let  $m \in \mathbb{Z}^+$  and  $a_1, a_2 \dots a_n \in (0, 1]$ . Let  $\alpha \in (-1, 0)$  and let  $c > 1$ . Let  $\{T_j\}_{j \geq 1} \subset \mathbb{R}^+$  be an increasing sequence of real numbers such that  $\lim_{j \rightarrow \infty} T_j = \infty$  and  $T_j > T_0$ ,  $T_0$  a fixed constant. Suppose  $\Phi(s)$  is a function that is analytic on all  $R_j = \langle \alpha, c \rangle \cap \{s = \sigma + it \mid |t| \leq T_j\}$  except for a finite number of singularities, and that there exists a constant  $M \in \mathbb{R}^+$  such that  $|\Phi(s)| < M$  independently of  $j$  on the boundary  $\delta R_j$  of  $R_j$ . Let  $S' = \text{Sing}(\Phi(s) \prod_{n=1}^m \zeta(s, a_n) / (s(s+1) \dots (s+m))) \cap \langle \alpha, c \rangle$ . Under these conditions*

$$\begin{aligned} \frac{1}{2\pi i} \int_{\alpha-i\infty}^{\alpha+i\infty} \Phi(s) \frac{\prod_{n=1}^m \zeta(s, a_n)}{s(s+1) \dots (s+m)} ds &= \frac{1}{2\pi i} \int_{c-i\infty}^{c+i\infty} \Phi(s) \frac{\prod_{n=1}^m \zeta(s, a_n)}{s(s+1) \dots (s+m)} ds \\ &- \sum_{\zeta \in S'} \text{Res} \left( \Phi(s) \frac{\prod_{n=1}^m \zeta(s, a_n)}{s(s+1) \dots (s+m)}; s = \zeta \right) \end{aligned}$$

**Proof.** The argument is the same as in the special case  $m = 1$ . We need to verify that the integrals along  $\Gamma_{1,3}$  and  $\Gamma_{2,4}$  converge and vanish, respectively. This is immediate when we consider that

- for  $\Gamma_1$ , we have the bound

$$C_1 - \frac{2c_4 M}{T^m},$$

- for  $\Gamma_3$ , the bound

$$C + 2c_1 M \int_{T_0}^T \frac{(\log t)^m}{t^{m/2+1}} dt,$$

which converges,

- and for  $\Gamma_{2,4}$ , the bound

$$\frac{c_3 M (\log T)^{m-1} (T^{m\delta} - T^{-m\delta})}{T^{m+1}},$$

again for the type  $[1 - \delta, 1 + \delta]$ . ■

## 2.9 The Mellin-Perron formula

This section contains the definitions and lemmata that are required to state the Mellin-Perron formula and define its domain of application. We will sketch the proof of the Mellin-Perron formula, as its use often requires a more than superficial appreciation of the method.

### 2.9.1 The Mellin transform

**Definition 2.9.1** (Open strip.) *The open strip of complex numbers  $\langle \alpha, \beta \rangle$  is the set  $\{s = \sigma + it \mid \alpha < \sigma < \beta\}$ .*

**Definition 2.9.2** (Mellin transform.) *Let  $f(x)$  be locally Lebesgue integrable over  $(0, +\infty)$ . The Mellin transform of  $f(x)$  is defined by*

$$\mathfrak{M}[f(x); s] = f^*(s) = \int_0^{+\infty} f(x)x^{s-1}dx.$$

*The fundamental strip is the largest open strip where the integral converges.*

**Lemma 2.9.1** *The conditions*

$$f(x)_{x \rightarrow 0+} \in \mathcal{O}(x^u), \quad f(x)_{x \rightarrow +\infty} \in \mathcal{O}(x^v),$$

*when  $u > v$ , guarantee that  $f^*(x)$  exists in the strip  $\langle -u, -v \rangle$ .*

We apply this lemma to a family of Heaviside-like step functions.

**Definition 2.9.3** *Let*

$$H_0(x) = \begin{cases} 1 & \text{if } x \in [0, 1], \\ 0 & \text{if } x > 1 \end{cases}$$

*be defined on  $[0, +\infty)$  and let*

$$H_m(x) = (1-x)^m H_0(x) \quad \text{when } m \in \mathbb{Z}^+.$$

Note that  $H_0(x)$  has a discontinuity at  $x = 1$ ; we have  $\lim_{x \rightarrow 1-} H_0(x) = 1$  and  $\lim_{x \rightarrow 1+} H_0(x) = 0$ . Note also that  $\lim_{x \rightarrow 1-} H_m(x) = \lim_{x \rightarrow 1+} H_m(x) = 0$  when  $m \in \mathbb{Z}^+$ ;  $H_m(x)$  is continuous at  $x = 1$ .

**Lemma 2.9.2** *The Mellin transform  $H_m^*(x)$  of  $H_m(x)$ , where  $m \in \mathbb{N}$ , exists in  $\langle 0, +\infty \rangle$  and is given by*

$$H_m^*(x) = \frac{m!}{s(s+1)\dots(s+m)}.$$

We have  $H_m(x)_{x \rightarrow 0+} \in \mathcal{O}(1)$  and  $H_m(x)_{x \rightarrow +\infty} \in \mathcal{O}(x^{-b})$  for any  $b > 0$  and for  $m \in \mathbb{N}$ , hence  $H_m^*(x)$  exists in  $\langle 0, +\infty \rangle$ . Note that

$$H_0^*(x) = \int_0^1 x^{s-1} dx = \frac{1}{s} [x^s]_0^1 = \frac{1}{s}.$$

We also have

$$\begin{aligned}
H_m^*(s) &= \int_0^1 H_m(x) x^{s-1} dx \\
&= \int_0^1 H_{m-1}(x) x^{s-1} dx - \int_0^1 H_{m-1}(x) x^s dx \\
&= H_{m-1}^*(s) - \int_0^1 \frac{(1-x)^m}{m} s x^{s-1} dx \\
&= H_{m-1}^*(s) - \frac{s}{m} H_m^*(s).
\end{aligned}$$

This gives

$$H_m^*(s) = \frac{m}{s+m} H_{m-1}^*(s)$$

for  $m \in \mathbb{Z}^+$  and the lemma follows.  $\blacksquare$

We will be concerned with the linearity and the rescaling property of the Mellin transform.

**Theorem 2.9.1** (Linearity and rescaling.) *Let  $\mathcal{K} \subset \mathbb{Z}$  be a finite set of integers; let  $\mu_k, \lambda_k \in \mathbb{R}^+$ . Let the fundamental strip of  $\mathfrak{M}[f(x); s]$  be  $\langle \alpha, \beta \rangle$ . We have*

$$\mathfrak{M} \left[ \sum_k \lambda_k f(\mu_k x); s \right] = \left( \sum_k \frac{\lambda_k}{\mu_k^s} \right) \mathfrak{M}[f(x); s],$$

where  $s \in \langle \alpha, \beta \rangle$ .

Let  $y = \mu_k x$  and  $dy = \mu_k dx$ . Note that

$$\begin{aligned}
\int_0^\infty \left( \sum_k \lambda_k f(\mu_k x) \right) x^{s-1} dx &= \sum_k \lambda_k \int_0^\infty f(\mu_k x) x^{s-1} dx \\
&= \sum_k \lambda_k \int_0^\infty f(y) y^{s-1} \frac{dy}{\mu_k^s} = \left( \sum_k \frac{\lambda_k}{\mu_k^s} \right) \int_0^\infty f(y) y^{s-1} dy.
\end{aligned}$$

We were able to exchange the integral with the summation because  $\mathcal{K}$  is finite. It can be shown that this operation extends to infinite  $\mathcal{K}$  as long as  $\sum_k \lambda_k / \mu_k^s$  converges absolutely. The extended property holds in the intersection of the half-plane of convergence of  $\sum_k \lambda_k / \mu_k^s$  and the fundamental strip  $\langle \alpha, \beta \rangle$  of  $f(x)$ .

**Definition 2.9.4** (Inverse Mellin transform.)

1. (Lebesgue integration.)

Let  $f(x)$  be integrable with fundamental strip  $\langle \alpha, \beta \rangle$ . If  $c \in (\alpha, \beta)$  and  $f^*(c+it)$  is integrable, then

$$\frac{1}{2\pi i} \int_{c-i\infty}^{c+i\infty} f^*(s) x^{-s} ds = f(x)$$

almost everywhere. If  $f(x)$  is continuous, the equality holds everywhere on  $(0, +\infty)$ .

2. (Riemann integration.)

Let  $f(x)$  be locally integrable with fundamental strip  $\langle \alpha, \beta \rangle$  and be of bounded variation in a neighborhood of  $x_0$ . Then

$$\lim_{T \rightarrow \infty} \frac{1}{2\pi i} \int_{c-iT}^{c+iT} f^*(s) x^{-s} ds \Big|_{x_0} = \frac{f(x_0^+) + f(x_0^-)}{2}$$

for  $c \in (\alpha, \beta)$ .

Of course if  $\lim_{x \rightarrow x_0^+} f(x) = \lim_{x \rightarrow x_0^-} f(x)$  then

$$\frac{f(x_0^+) + f(x_0^-)}{2} = f(x_0).$$

### 2.9.2 The Mellin-Perron formula

**Theorem 2.9.2** (Mellin-Perron formula.) *Let  $c \in \mathbb{R}^+$  lie in the half-plane of absolute convergence of  $\sum_k \lambda_k/k^s$ . Then we have*

$$\frac{1}{m!} \sum_{1 \leq k < n} \lambda_k \left(1 - \frac{k}{n}\right)^m = \frac{1}{2\pi i} \int_{c-i\infty}^{c+i\infty} \left(\sum_{k \geq 1} \frac{\lambda_k}{k^s}\right) n^s \frac{ds}{s(s+1)\dots(s+m)}$$

for  $m \in \mathbb{Z}^+$ . We have

$$\sum_{1 \leq k < n} \lambda_k + \frac{\lambda_n}{2} = \frac{1}{2\pi i} \int_{c-i\infty}^{c+i\infty} \left(\sum_{k \geq 1} \frac{\lambda_k}{k^s}\right) n^s \frac{ds}{s}$$

when  $m = 0$ .

This theorem is a straightforward application of Mellin inversion.

**Proof.** Let  $F(x) = \sum_k \lambda_k f(\mu_k x)$  and use the rescaling property to obtain

$$\mathfrak{M}[F(x); s] = F^*(s) = \left(\sum_k \frac{\lambda_k}{\mu_k^s}\right) f^*(s).$$

Consider Riemann-integrable  $f(x)$  and apply the Mellin inversion formula.

$$\sum_k \lambda_k \frac{f(\mu_k x^+) + f(\mu_k x^-)}{2} = \lim_{T \rightarrow \infty} \frac{1}{2\pi i} \int_{c-iT}^{c+iT} \left(\sum_k \frac{\lambda_k}{\mu_k^s}\right) f^*(s) x^{-s} ds$$

Let  $f(x) = H_m(x)$ ,  $m \in \mathbb{N}$  and let  $\mu_k = k$ . Recall that the fundamental strip of  $H_m(x)$  is  $\langle 0, \infty \rangle$ ; let  $x = 1/n$ . This gives

$$\begin{aligned} \sum_k \lambda_k \frac{f(\mu_k x^+) + f(\mu_k x^-)}{2} &= \sum_k \lambda_k \frac{H_m(\frac{k}{n^-}) + H_m(\frac{k}{n^+})}{2} \\ &= \sum_{1 \leq k < n} \lambda_k \frac{(1 - \frac{k}{n^-})^m + (1 - \frac{k}{n^+})^m}{2} + \lambda_n \frac{H_m(1^+) + H_m(1^-)}{2} \\ &= \sum_{1 \leq k < n} \lambda_k \left(1 - \frac{k}{n}\right)^m + \lambda_n \frac{H_m(1^+) + H_m(1^-)}{2}. \end{aligned}$$

Note that

$$\lambda_n \frac{H_m(1^+) + H_m(1^-)}{2} = \begin{cases} \lambda_n/2 & \text{if } m = 0 \\ 0 & \text{if } m \in \mathbb{Z}^+. \end{cases}$$

Continuing the substitution, we have

$$\begin{aligned} \lim_{T \rightarrow \infty} \frac{1}{2\pi i} \int_{c-iT}^{c+iT} \left( \sum_k \frac{\lambda_k}{\mu_k^s} \right) f^*(s) x^{-s} ds &= \lim_{T \rightarrow \infty} \frac{1}{2\pi i} \int_{c-iT}^{c+iT} \left( \sum_k \frac{\lambda_k}{k^s} \right) \frac{m!}{s(s+1)\dots(s+m)} n^s ds \\ &= \frac{m!}{2\pi i} \int_{c-i\infty}^{c+i\infty} \left( \sum_k \frac{\lambda_k}{k^s} \right) n^s \frac{ds}{s(s+1)\dots(s+m)} \end{aligned}$$

This concludes the proof. Because the fundamental strip of  $H_m(x)$  is  $\langle 0, \infty \rangle$ , the choice of  $c > 0$  is determined by the half-plane of convergence of  $\sum_k \lambda_k/k^s$  only. ■

### 2.9.3 The use of the Mellin-Perron formula when $m = 1$

The lemma below summarizes the usage pattern of the Mellin-Perron formula when  $m = 1$ .

**Lemma 2.9.3** *Let  $\{a_n\}_{n \geq 1}$  be an arithmetical function; let  $a_0 = 0$  and let  $\{b_n\}_{n \geq 1} = \{\Delta \nabla a_n\}_{n \geq 1}$ . If  $B(s) = \sum b_n/n^s$  is the Dirichlet generating function of  $\{b_n\}_{n \geq 1}$ , we have*

$$\sum_{k=1}^{n-1} \sum_{l=1}^k b_l = a_n - na_1 = \frac{n}{2\pi i} \int_{c-i\infty}^{c+i\infty} B(s) n^s \frac{ds}{s(s+1)}$$

where  $c$  is in the half-plane of convergence of  $B(s)$ .

To see this, note first that

$$\frac{1}{1!} \sum_{k=1}^{n-1} b_k \left(1 - \frac{k}{n}\right)^1 = \frac{1}{2\pi i} \int_{c-i\infty}^{c+i\infty} B(s) n^s \frac{ds}{s(s+1)}$$

or

$$\sum_{k=1}^{n-1} b_k(n-k) = \frac{n}{2\pi i} \int_{c-i\infty}^{c+i\infty} B(s)n^s \frac{ds}{s(s+1)}.$$

We select  $m = 1$  because the iterated sum on the left cancels the operator  $\Delta\nabla$ .

$$\begin{aligned} \sum_{k=1}^{n-1} b_k(n-k) &= \sum_{k=1}^{n-1} \sum_{l=1}^k b_l = \sum_{k=1}^{n-1} \sum_{l=1}^k \Delta\nabla a_l = \sum_{k=1}^{n-1} \sum_{l=1}^k (\nabla a_{l+1} - \nabla a_l) \\ &= \sum_{k=1}^{n-1} (\nabla a_{k+1} - \nabla a_1) = \sum_{k=1}^{n-1} (a_{k+1} - a_k) - (n-1)\nabla a_1 \\ &= a_n - a_1 - (n-1)a_1 = a_n - na_1 \end{aligned}$$

## 2.10 Mellin-Perron formulae for the Hurwitz $\zeta$ -function

This section presents two Mellin-Perron formulae for the generalized  $\zeta$ -function. They will be used in later chapters. We include them here because their respective derivations hardly differ from that of the standard Mellin-Perron formula.

We apply the Mellin inversion theorem to  $F(x) = \sum_k \lambda_k f(\mu_k x)$  with  $x = r/n$ ,  $r, n \in \mathbb{Z}^+$ ,  $\mu_k = k+a$ ,  $\lambda_k = 1$ ,  $a \in \mathbb{R}$ ,  $a \in (0, 1]$ ,  $f(x) = H_1(x) = (1-x)H_0(x)$ . As we require  $\mu_k \in \mathbb{R}^+$  we take  $k \in \mathbb{N}$ . We have

$$F(x) = \sum_{k \in \mathbb{N}} \left(1 - (k+a)\frac{r}{n}\right) H_0\left((k+a)\frac{r}{n}\right)$$

and

$$F^*(s) = \left(\sum_{k \in \mathbb{N}} \frac{1}{(k+a)^s}\right) f^*(s) = \frac{\zeta(s, a)}{s(s+1)}$$

where  $\sigma > 1$ . We need to evaluate  $F(x)$ .  $H_0(x)$  vanishes outside of  $[0, 1)$ , hence we require  $0 \leq (k+a)r/n < 1$  or  $k < n/r - a$ . Let  $\mathbb{N}(u) = \{\nu < u \mid \nu \in \mathbb{N}\}$  where  $u \in \mathbb{R}^+$ . We have

$$F(x) = \sum_{k \in \mathbb{N}(n/r-a)} \left(1 - (k+a)\frac{r}{n}\right).$$

With these settings the Mellin inversion formula yields the following theorem.

**Theorem 2.10.1** *Let  $c > 1$ .*

$$\sum_{k \in \mathbb{N}(n/r-a)} \left(1 - (k+a)\frac{r}{n}\right) = \frac{1}{2\pi i} \int_{c-i\infty}^{c+i\infty} \frac{1}{r^s} \zeta(s, a) \frac{n^s}{s(s+1)} ds$$

This theorem has several useful corollaries. The first of these is obtained by setting  $r = 1$ . Let  $\alpha \in (-1, 0)$ .

**Corollary 2.10.1** *Let  $n \in \mathbb{N}$ .*

$$\frac{1}{2\pi i} \int_{\alpha-i\infty}^{\alpha+i\infty} \zeta(s, a) \frac{n^s}{s(s+1)} ds = 0$$

Let  $c = 1$ . The set of poles of  $\zeta(s, a)n^s/(s(s+1))$  in  $\langle \alpha, c \rangle$  is  $\{1, 0\}$ . We apply the shifting lemma with  $\Phi(s) = n^s$  and  $T_j = j$ . Because  $|n^s| = n^\sigma$  we can take  $M = n^c$ .

$$\begin{aligned} \frac{1}{2\pi i} \int_{\alpha-i\infty}^{\alpha+i\infty} \zeta(s, a) \frac{n^s}{s(s+1)} ds &= \frac{1}{2\pi i} \int_{c-i\infty}^{c+i\infty} \zeta(s, a) \frac{n^s}{s(s+1)} \\ &\quad - \operatorname{Res} \left( \zeta(s, a) \frac{n^s}{s(s+1)}; s=1 \right) - \operatorname{Res} \left( \zeta(s, a) \frac{n^s}{s(s+1)}; s=0 \right) \\ &= \sum_{0 \leq k < n} \left( 1 - (k+a) \frac{1}{n} \right) - \frac{n}{2} - \zeta(0, a) \\ &= n - n \frac{a}{n} - \frac{1}{n} \frac{1}{2} (n-1)n - \frac{n}{2} - \zeta(0, a) = \frac{1}{2} - a - \zeta(0, a) = 0 \end{aligned}$$

The second corollary results from taking  $r = 4$ .

**Corollary 2.10.2** *Let  $n \in \mathbb{N}$ . The value of*

$$\frac{1}{2\pi i} \int_{\alpha-i\infty}^{\alpha+i\infty} \frac{1}{4^s} \zeta(s, a) \frac{n^s}{s(s+1)} ds$$

*is given by the following table.*

	$n = 4m$	$n = 4m + 1$	$n = 4m + 2$	$n = 4m + 3$
$0 < a \leq \frac{1}{4}$	0	$-\frac{1}{n} (3a + \frac{3}{8})$	$-\frac{1}{n} (2a - \frac{1}{2})$	$-\frac{1}{n} (a - \frac{3}{8})$
$\frac{1}{4} < a \leq \frac{1}{2}$	0	$\frac{1}{n} (a - \frac{5}{8})$	$-\frac{1}{n} (2a - \frac{1}{2})$	$-\frac{1}{n} (a - \frac{3}{8})$
$\frac{1}{2} < a \leq \frac{3}{4}$	0	$\frac{1}{n} (a - \frac{5}{8})$	$\frac{1}{n} (2a - \frac{3}{2})$	$-\frac{1}{n} (a - \frac{3}{8})$
$\frac{3}{4} < a \leq 1$	0	$\frac{1}{n} (a - \frac{5}{8})$	$\frac{1}{n} (2a - \frac{3}{2})$	$\frac{1}{n} (3a - \frac{21}{8})$

We let  $c = 1$  as before and consider the poles of  $\zeta(s, a)n^s/(4^s s(s+1))$  in  $\langle \alpha, c \rangle$ , which are at 1 and 0.

We apply the shifting lemma with  $\Phi(s) = (n/4)^s$ ,  $T_j = j$  and take  $M = (n/4)^c$ .

$$\begin{aligned} \frac{1}{2\pi i} \int_{\alpha-i\infty}^{\alpha+i\infty} \zeta(s, a) \frac{n^s}{4^s s(s+1)} ds &= \frac{1}{2\pi i} \int_{c-i\infty}^{c+i\infty} \zeta(s, a) \frac{n^s}{4^s s(s+1)} \\ &\quad - \operatorname{Res} \left( \zeta(s, a) \frac{n^s}{4^s s(s+1)}; s=1 \right) - \operatorname{Res} \left( \zeta(s, a) \frac{n^s}{4^s s(s+1)}; s=0 \right) \\ &= \sum_{k \in \mathbb{N}(n/4-a)} \left( 1 - (k+a) \frac{4}{n} \right) - \frac{n}{8} - \zeta(0, a) = \epsilon(n, a) - \frac{n}{8} - \zeta(0, a) \end{aligned}$$

Suppose  $n = 4m + m_1$  where  $m_1 \in \{0, 1, 2, 3\}$ . We have  $n/4 - a = \lfloor n/4 \rfloor + m_1/4 - a$ . If  $m_1/4 < a$ , the sum over  $\mathbb{N}(n/4 - a)$  ranges from 0 to  $\lfloor n/4 \rfloor - 1$ . If  $m_1/4 \geq a$  the sum includes  $\lfloor n/4 \rfloor$ . We have two cases.

$$\epsilon(n, a) = \begin{cases} \lfloor \frac{n}{4} \rfloor - a \frac{4}{n} \lfloor \frac{n}{4} \rfloor - \frac{2}{n} (\lfloor \frac{n}{4} \rfloor - 1) \lfloor \frac{n}{4} \rfloor & \text{if } \frac{m_1}{4} < a \\ \lfloor \frac{n}{4} \rfloor + 1 - a \frac{4}{n} (\lfloor \frac{n}{4} \rfloor + 1) - \frac{2}{n} (\lfloor \frac{n}{4} \rfloor + 1) \lfloor \frac{n}{4} \rfloor & \text{if } \frac{m_1}{4} \geq a \end{cases}$$

We note that  $\lfloor n/4 \rfloor = (n - m_1)/4$  and  $\lfloor n/4 \rfloor 4/n = 1 - m_1/n$ . Hence the two terms evaluate to

$$\frac{1}{8}n + \frac{1}{2} - a + \frac{1}{n} \left( am_1 - \frac{1}{2}m_1 - \frac{1}{8}m_1^2 \right)$$

and

$$\frac{1}{8}n + \frac{1}{2} - a + \frac{1}{n} \left( a(m_1 - 4) + \frac{1}{2}m_1 - \frac{1}{8}m_1^2 \right).$$

We conclude that

$$\frac{1}{2\pi i} \int_{\alpha - i\infty}^{\alpha + i\infty} \zeta(s, a) \frac{n^s}{4^s s(s+1)} ds = \epsilon(n, a) - \frac{1}{8}n - \frac{1}{2} + a.$$

This gives the tabled values when  $\epsilon(n, a)$  is evaluated according to  $m_1$  and  $a$ .

**Example.** We can use this corollary to verify the following relation.

$$\frac{1}{2\pi i} \int_{\alpha - i\infty}^{\alpha + i\infty} (\zeta(s, 7/16) + \zeta(s, 15/16)) \frac{n^s}{4^s s(s+1)} ds = \begin{cases} 0 & \text{if } n \equiv 0 \pmod{2} \\ \frac{1}{8}n & \text{if } n \equiv 1 \pmod{2} \end{cases}$$

We have  $7/16 \in (1/4, 1/2]$  and  $15/16 \in (3/4, 1]$ . Hence we need only add the second and fourth rows of the table, with  $a$  set to  $7/16$  and  $15/16$  respectively. This result will be useful in a later chapter.

## 2.11 Notes

I consulted [Cla82] as an introduction to elementary real analysis; the preliminaries of this chapter are from [Cla82, p. 167].

There are many texts on basic complex analysis. The definitions pertaining to point sets, complex limits, and analyticity are from [Det84, p. 13-17, 27-39]; analytic continuation is discussed on [Det84, p. 152-162] and [Mar87, p. 397-411]. Convergent series of analytic functions are discussed on [Mar87, p. 206-213]. The material on Laurent series is from [Mar87, p. 246-252, 266-272] and [Det84, p. 163-170]. The winding number is discussed on [Mar87, p. 165]; the residue theorem is given on [Mar87, p. 280].

The material on Dirichlet series will be found in any good text on number theory. I have used the presentation in [Apo86, p. 224-248]. This text also includes a detailed technical proof of the Mellin-Perron formula for  $m = 0$ . I also consulted [Kra81, p. 86-87] for the proof of the analytic version of the fundamental theorem of algebra. Theorem 2.6.2 is from [Man72, p. 1, 2, 7, 102-104].

An introduction to the Riemann  $\zeta$  function is found on [Lan93, p. 415-421]. This includes a presentation of the functional equation; the reader may also wish to consult [Kar92, p. 9-11] or [WW15, p. 262-265]. The formula for  $\zeta(m, a)$  with  $m$  a negative integer is proved on [BMP55c, p. 24-27, 35-37] or [WW15, p. 260-262], for example. The growth estimates for  $\zeta(s, a)$  are from [WW15, p. 269-270].

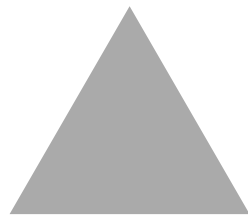
[FGK<sup>+</sup>94] note and use the shifting lemma. A statement of Corollary 2.10.1 for the case  $a = 1$  can be found on [FGK<sup>+</sup>94, p. 297].

The presentation of the Mellin-Perron formula is from [FGK<sup>+</sup>94, p. 295-297]. The note on the use of the Mellin transform when  $m = 1$  is based on observations found on [FG94, p. 680-681]. The background material on Mellin transforms is from [FGD95, p. 9-14].

## Chapter 3

### The area of a fractal ornament

We will consider the area of a fractal ornament in this chapter. It will be apparent that our discussion readily generalizes to other self-similar fractal ornaments. (Fractal ornament statistics such as area are a form of tree statistics.) It is not difficult to compute the limit of the area of a fractal ornament. Our goal will be to obtain a formula for the area in terms of  $n$ , where  $A_n$  is the  $n$ th ornament, as defined below. This question is somewhat more involved.



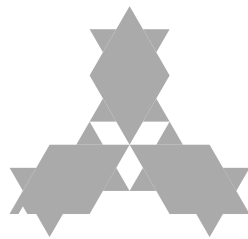
$A_1$ ; level 0 complete



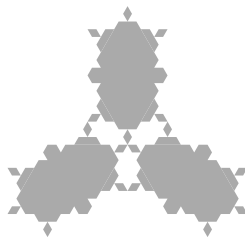
$A_3$



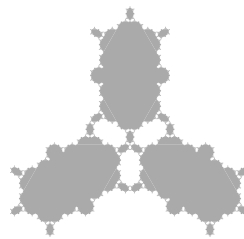
$A_5$



$A_{17}$



$A_{64}$ ; level 3 complete



$A_{4096}$ ; level 6 complete

### 3.1 Preliminaries

Six approximations of the ornament  $A$  are shown in the table above. The production rule is simple. The first ornament  $A_1$  is a triangle. Subsequent  $A_k$  are constructed by iterating over the segments that delineate the ornament, starting at the lower left and moving in a counter-clockwise fashion. A segment of length  $l$  is replaced by four new ones of length  $l/3$ . The first and last are incident on the terminal

vertices of the original. The middle two connect the first to the last. The middle vertex alternately falls outside or inside the ornament, according to the depth in the segment tree. This replacement rule is the familiar one used to generate the Koch curve.

We will consider the following question. *Given  $n$ , what is  $a_n$ , the area of  $A_n$ ?* It is important that we clarify why it should warrant attention. The construction of successive ornaments proceeds by iterating over the levels at depth  $r$ . It is easy to compute  $r$  given  $n$ ; once we have  $r$ , we can compute the offset into the current level and hence the area. In fact we will carry out this computation to verify our later results. Our purpose will be to find a function that captures the behavior of  $a_n$  and does not use approximations such as the integer logarithm. This work is in the spirit of Delange's and more recently, Flajolet's analysis of the sum-of-digits function, where it does not suffice to unroll the recursion for selected values of  $n$  and generalize to all of  $Z^+$ .

We will prove the following theorem.

**Theorem 3.1.1** *Let  $a_n$  be the area of the  $n$ th approximation to the fractal ornament described above.*

*We have*

$$a_n = \frac{\sqrt{3}}{4} \frac{10}{13} - \frac{\sqrt{3}}{4} \frac{10}{9 \log 4} n \sum_{k \in Z} \frac{n^{\rho + \chi_k}}{(\rho + \chi_k)(\rho + \chi_k + 1)}$$

where  $\rho = \frac{i\pi - \log 9}{\log 4}$  and  $\chi_k = \frac{2\pi i}{\log 4} k$ .

### 3.2 A first approximation

What follows provides a traditional, straightforward analysis of the problem, i.e. of its restriction to specific values of  $n$ . We will find these results useful in the remainder of this chapter.

It is convenient to define  $A_0$  as the empty ornament;  $a_0 = 0$ . The next approximation  $A_1$  is a triangle; all of its sides have unit length;  $a_1 = \frac{\sqrt{3}}{4}$ . We say that level  $r$  is filled when the three underlying trees with branch factor four are filled to depth  $r$ . Hence level 0 is completed with  $A_1$  (three trees of depth 0), level 1 is completed with  $A_4$  (three trees of depth 1) etc. Level  $r$  consists of  $3 \cdot 4^{r-1}$  segments (leaves). We conclude that

- level 0 is completed after 1 step ( $A_1$ )
- level  $r \geq 1$  is completed after  $1 + \sum_{k=1}^r 3 \cdot 4^{k-1} = 1 + 3 \sum_{k=0}^{r-1} 4^k = 1 + 3 \frac{4^r - 1}{4 - 1} = 4^r$  steps ( $A_{4^r}$ ).

The segment length of the triangles that are being subtracted and added decreases from  $l$  to  $l/3$  between levels and the area from  $T$  to  $T/9$ . Therefore the area of  $A_{4^r}$ ,  $r \geq 1$  is

$$\begin{aligned} a_{4^r} &= \frac{\sqrt{3}}{4} \left( 1 - 3\frac{1}{9} + 3\frac{4}{9^2} - 3\frac{4^2}{9^3} \dots (-1)^r \frac{4^{r-1}}{9^r} \right) = \frac{\sqrt{3}}{4} - \frac{\sqrt{3}}{4} \frac{3}{9} \left( 1 - \frac{4}{9} + \frac{4^2}{9^2} \dots (-1)^r \frac{4^{r-1}}{9^{r-1}} \right) \\ &= \frac{\sqrt{3}}{4} \left( 1 - \frac{3}{9} \frac{1 - (-4/9)^r}{13/9} \right) = \frac{\sqrt{3}}{4} \left( \frac{10}{13} + \frac{3}{13} \left( -\frac{4}{9} \right)^r \right) \end{aligned}$$

The limiting area  $A$  of  $\{A_{4^r}\}$ ,  $r \geq 1$  is  $\frac{\sqrt{3}}{2} \frac{5}{13}$ .

### 3.3 Exact analysis

We will consider  $\Delta \nabla a_n$  in order to compute  $a_n$ . What is  $a_n - a_{n-1} = \nabla a_n$ ? The difference between consecutive values of  $a_n$  depends on the current level. If  $n$  is on level  $r$ , the change in area is given by

$$(-1)^r \frac{1}{9^r} \frac{\sqrt{3}}{4} = \left( -\frac{1}{9} \right)^r \frac{\sqrt{3}}{4}.$$

How do we compute  $r$  given  $n$ ? Level  $r$  extends from  $4^{r-1} + 1$  to  $4^r$ . Hence  $r = \lfloor \log_4(n-1) \rfloor + 1$ .

$$\nabla a_n = a_n - a_{n-1} = \left( -\frac{1}{9} \right)^{\lfloor \log_4(n-1) \rfloor + 1} \frac{\sqrt{3}}{4} = -\frac{\sqrt{3}}{4} \frac{1}{9} \left( -\frac{1}{9} \right)^{\lfloor \log_4(n-1) \rfloor} \quad (n \geq 2)$$

$$\nabla a_1 = a_1 - a_0 = \frac{\sqrt{3}}{4}$$

We proceed to  $\Delta \nabla a_n$ .

$$\begin{aligned} \Delta \nabla a_n &= \nabla a_{n+1} - \nabla a_n = -\frac{\sqrt{3}}{4} \frac{1}{9} \left( \left( -\frac{1}{9} \right)^{\lfloor \log_4 n \rfloor} - \left( -\frac{1}{9} \right)^{\lfloor \log_4(n-1) \rfloor} \right) \\ &= \frac{\sqrt{3}}{4} \frac{1}{9} \left( \left( -\frac{1}{9} \right)^{\lfloor \log_4(n-1) \rfloor} - \left( -\frac{1}{9} \right)^{\lfloor \log_4 n \rfloor} \right) \quad (n \geq 2) \\ \Delta \nabla a_1 &= \nabla a_2 - \nabla a_1 = -\frac{1}{9} \frac{\sqrt{3}}{4} - \frac{\sqrt{3}}{4} = -\frac{10}{9} \frac{\sqrt{3}}{4} \end{aligned}$$

It is best to stop and interpret these equations.

$$\begin{aligned} \Delta \nabla a_{4^r} &= \frac{\sqrt{3}}{4} \frac{1}{9} \left( \left( -\frac{1}{9} \right)^{r-1} - \left( -\frac{1}{9} \right)^r \right) \quad (r \geq 1) \\ \Delta \nabla a_1 &= -\frac{10}{9} \frac{\sqrt{3}}{4} \\ \Delta \nabla a_n &= 0 \quad (\text{otherwise}). \end{aligned}$$

We will evaluate  $a_n$  by means of the Mellin-Perron formula for  $m = 1$ . This requires that we study the Dirichlet generating function of  $\Delta\nabla a_n$ .

$$\begin{aligned}
A(s) &= \sum_{n \geq 1} \frac{\Delta\nabla a_n}{n^s} = \Delta\nabla a_1 + \sum_{n \geq 2} \frac{\Delta\nabla a_n}{n^s} \\
&= \Delta\nabla a_1 + \frac{\sqrt{3}}{4} \frac{1}{9} \sum_{r \geq 1} \left( \left( -\frac{1}{9} \right)^{r-1} - \left( -\frac{1}{9} \right)^r \right) \frac{1}{(4^r)^s} \\
&= \Delta\nabla a_1 + \frac{\sqrt{3}}{4} \frac{1}{9} \left( -9 \sum_{r \geq 1} \left( -\frac{1}{9} \right)^r \frac{1}{(4^r)^s} - \sum_{r \geq 1} \left( -\frac{1}{9} \right)^r \frac{1}{(4^r)^s} \right) \\
&= \Delta\nabla a_1 - \frac{\sqrt{3}}{4} \frac{10}{9} \sum_{r \geq 1} \left( -\frac{1}{9 \cdot 4^s} \right)^r = \Delta\nabla a_1 - \frac{\sqrt{3}}{4} \frac{10}{9} \left( \sum_{r \geq 0} \left( -\frac{1}{9 \cdot 4^s} \right)^r - 1 \right)
\end{aligned}$$

We must determine the half-plane of convergence  $\{s = \sigma + it \mid \sigma > c\}$ , where  $c > 0$ , of  $A(s)$  if we wish to use the Mellin-Perron formula. We require  $|\frac{1}{9 \cdot 4^s}| < 1$ , i.e.  $\frac{1}{9} < 4^\sigma$  or  $\log_4 \frac{1}{9} < \sigma$ . Hence  $c = 1$  will suffice. We continue the evaluation of  $A(s)$ .

$$\begin{aligned}
A(s) &= -\frac{10}{9} \frac{\sqrt{3}}{4} + \frac{\sqrt{3}}{4} \frac{10}{9} - \frac{\sqrt{3}}{4} \frac{10}{9} \frac{1}{1 - \left( -\frac{1}{9 \cdot 4^s} \right)} \\
&= -\frac{\sqrt{3}}{4} \frac{10}{9} \frac{9 \cdot 4^s}{9 \cdot 4^s + 1} = -\frac{\sqrt{3}}{4} 10 \frac{4^s}{9 \cdot 4^s + 1}
\end{aligned}$$

The Mellin-Perron formula for  $m = 1$  tells us that (Lemma 2.9.3)

$$a_n - na_1 = \frac{n}{2\pi i} \int_{1-i\infty}^{1+i\infty} A(s) n^s \frac{ds}{s(s+1)} \text{ or } a_n = \frac{\sqrt{3}}{4} n + \frac{n}{2\pi i} \int_{1-i\infty}^{1+i\infty} A(s) n^s \frac{ds}{s(s+1)}.$$

### 3.3.1 Evaluating the integral

Let

$$B(s) = \frac{4^s}{9 \cdot 4^s + 1}.$$

The integral is evaluated with the Cauchy Residue Theorem. We will use the following rectangular contour, traversed counterclockwise.

$$\begin{aligned}
\Gamma_1 &= \{1 + it \mid |t| \leq R\} & \Gamma_2 &= \{\sigma + iR \mid -R \leq \sigma \leq 1\} \\
\Gamma_3 &= \{-R + it \mid |t| \leq R\} & \Gamma_4 &= \{\sigma - iR \mid -R \leq \sigma \leq 1\}
\end{aligned}$$

We take  $R = \frac{2j\pi}{\log 4} > 1$  so that  $\Gamma_1 \cup \Gamma_2 \cup \Gamma_3 \cup \Gamma_4$  includes the poles at  $s = 0$ ,  $s = -1$  and  $s = \rho + \chi_k$ , where  $\rho = \frac{i\pi - \log 9}{\log 4}$  and  $\chi_k = \frac{2\pi i}{\log 4}k$ ,  $|\rho + \log_4 9 + \chi_k| < R$ . What are the residues at these poles? All the poles are simple ones.

$$\begin{aligned} \operatorname{Res} \left( \frac{B(s)n^s}{s(s+1)}; s=0 \right) &= \lim_{s \rightarrow 0} \frac{4^s}{1+9 \cdot 4^s} \frac{n^s}{s+1} = \frac{1}{10} \\ \operatorname{Res} \left( \frac{B(s)n^s}{s(s+1)}; s=-1 \right) &= \lim_{s \rightarrow -1} \frac{4^s}{1+9 \cdot 4^s} \frac{n^s}{s} = \frac{1}{4+9} \frac{1}{-n} = -\frac{1}{13n} \\ \operatorname{Res} \left( \frac{B(s)n^s}{s(s+1)}; s=\rho+\chi_k \right) &= \lim_{s \rightarrow \rho+\chi_k} 4^s \frac{s - (\rho + \chi_k)}{1+9 \cdot 4^s} \frac{n^s}{s(s+1)} \\ &= -\frac{1}{9} \frac{n^{\rho+\chi_k}}{(\rho + \chi_k)(\rho + \chi_k + 1)} \lim_{s \rightarrow \rho+\chi_k} \frac{1}{9 \log 4 \cdot 4^s} \\ &= -\frac{1}{9} \frac{n^{\rho+\chi_k}}{(\rho + \chi_k)(\rho + \chi_k + 1)} \frac{1}{-9 \log 4 \cdot \frac{1}{9}} = \frac{1}{9 \log 4} \frac{n^{\rho+\chi_k}}{(\rho + \chi_k)(\rho + \chi_k + 1)} \end{aligned}$$

We have

$$\int_{\Gamma_1 \cup \Gamma_2 \cup \Gamma_3 \cup \Gamma_4} B(s)n^s \frac{ds}{s(s+1)} = 2\pi i \left( \frac{1}{10} - \frac{1}{13n} + \frac{1}{9 \log 4} \sum_{k \in \mathbb{Z}} \frac{n^{\rho+\chi_k}}{(\rho + \chi_k)(\rho + \chi_k + 1)} \right).$$

Note that  $\lim_{R \rightarrow \infty} \int_{\Gamma_1} B(s)n^s \frac{ds}{s(s+1)} = \int_{1-i\infty}^{1+i\infty} B(s)n^s \frac{ds}{s(s+1)}$ .

If we can show that  $\lim_{R \rightarrow \infty} \int_{\Gamma_{2,3,4}} B(s)n^s \frac{ds}{s(s+1)} = 0$ , we will have evaluated  $a_n$ .

- $\Gamma_2$ ;  $s = \sigma + iR$ ;  $ds = d\sigma$

$$\begin{aligned} \left| \int_{-R}^1 \frac{4^s}{1+9 \cdot 4^s} \frac{n^s}{s(s+1)} d\sigma \right| &\leq \int_{-R}^1 \left| \frac{4^s}{1+9 \cdot 4^s} \frac{n^s}{s(s+1)} \right| d\sigma \\ &= \int_{-R}^1 \frac{(4n)^\sigma}{R^2} \frac{1}{|1+9 \cdot 4^\sigma e^{2j\pi i}|} d\sigma \\ &\leq \frac{1}{R^2 (1+9 \cdot 4^{-R})} \int_{-R}^1 (4n)^\sigma d\sigma \\ &= \frac{1}{R^2 (1+9 \cdot 4^{-R}) \log 4n} (4n - (4n)^{-R}) \end{aligned}$$

This is the desired result. The integral along  $\Gamma_2$  vanishes as  $R$  goes to infinity.

- $\Gamma_3$ ;  $s = -R + it$ ;  $ds = idt$

$$\begin{aligned} \left| \int_{-R}^R \frac{4^s}{1+9 \cdot 4^s} \frac{n^s}{s(s+1)} idt \right| &\leq \int_{-R}^R \left| \frac{4^s}{1+9 \cdot 4^s} \frac{n^s}{s(s+1)} \right| dt \\ &= \frac{(4n)^{-R}}{R(R-1)} \int_{-R}^R \frac{1}{|1+9 \cdot 4^{-R} 4^{it}|} dt \\ &\leq \frac{1}{(4n)^R R(R-1) (1-9 \cdot 4^{-R})} \int_{-R}^R dt \\ &= \frac{2R}{n^R R(R-1) (4^R - 9)} = \frac{2}{n^R (R-1) (4^R - 9)} \end{aligned}$$

$\Gamma_3$  vanishes like  $\Gamma_2$ .

- $\Gamma_4$ ;  $s = \sigma - iR$ ;  $ds = d\sigma$

This is a variation of the estimate for  $\Gamma_2$ .

We are ready to evaluate  $a_n$ .

$$\begin{aligned}
a_n &= \frac{\sqrt{3}}{4}n + \frac{n}{2\pi i} \int_{1-i\infty}^{1+i\infty} A(s)n^s \frac{ds}{s(s+1)} \\
&= \frac{\sqrt{3}}{4}n - \frac{\sqrt{3}}{4}10 \frac{n}{2\pi i} \left( 2\pi i \left( \frac{1}{10} - \frac{1}{13} \frac{1}{n} + \frac{1}{9 \log 4} \sum_{k \in \mathbb{Z}} \frac{n^{\rho+\chi_k}}{(\rho+\chi_k)(\rho+\chi_k+1)} \right) \right) \\
&= \frac{\sqrt{3}}{4}n - \frac{\sqrt{3}}{4}n + \frac{\sqrt{3}}{4} \frac{10}{13} - \frac{\sqrt{3}}{4} \frac{10}{9 \log 4} n \sum_{k \in \mathbb{Z}} \frac{n^{\rho+\chi_k}}{(\rho+\chi_k)(\rho+\chi_k+1)} \\
&= \frac{\sqrt{3}}{4} \frac{10}{13} - \frac{\sqrt{3}}{4} \frac{10}{9 \log 4} n \sum_{k \in \mathbb{Z}} \frac{n^{\rho+\chi_k}}{(\rho+\chi_k)(\rho+\chi_k+1)}
\end{aligned}$$

This formula prompts two questions. How can it be verified? What does it mean? We will treat these in order.

### 3.4 Verification

Recall that

$$a_{4^r} = \frac{\sqrt{3}}{4} \left( \frac{10}{13} + \frac{3}{13} \left( -\frac{4}{9} \right)^r \right),$$

where  $r \geq 1$ . Clearly the exact formula must agree with these values at  $n = 4^r$ . We have

$$\begin{aligned}
a_{4^r} &= \frac{\sqrt{3}}{4} \frac{10}{13} - \frac{\sqrt{3}}{4} \frac{10}{9 \log 4} 4^r \sum_{k \in \mathbb{Z}} \frac{4^{r(\rho+\chi_k)}}{(\rho+\chi_k)(\rho+\chi_k+1)} \\
&= \frac{\sqrt{3}}{4} \frac{10}{13} - \frac{\sqrt{3}}{4} \frac{10}{9 \log 4} \left( -\frac{4}{9} \right)^r \sum_{k \in \mathbb{Z}} \frac{1}{(\rho+\chi_k)(\rho+\chi_k+1)}.
\end{aligned}$$

We require the value of

$$\sum_{k \in \mathbb{Z}} \frac{1}{(\rho+\chi_k)(\rho+\chi_k+1)}.$$

We sum the series by the Cauchy Residue Theorem. The appropriate function is

$$C(s) = \frac{1}{1+9 \cdot 4^s} \frac{1}{s(s+1)}.$$

We use the contour  $\Gamma'_1 \cup \Gamma_2 \cup \Gamma_3 \cup \Gamma_4$ , where  $\Gamma'_1 = \{R+it \mid |t| \leq R\}$ . We must show that  $\lim_{R \rightarrow \infty} \int_{\Gamma'_1, \Gamma_2, \Gamma_3, \Gamma_4} C(s) ds = 0$ .

- $\Gamma'_1$ ;  $s = R + it$ ;  $ds = idt$ ;

$$\begin{aligned} \left| \int_{\Gamma'_1} C(s) ds \right| &\leq \int_{-R}^R \frac{1}{|1 + 9 \cdot 4^{R+it}|} \frac{1}{|s(s+1)|} dt \\ &\leq \frac{1}{R(R+1)} \frac{2R}{9 \cdot 4^R - 1} = \frac{2}{(R+1)(9 \cdot 4^R - 1)} \end{aligned}$$

- $\Gamma_2$ ;  $s = \sigma + iR$ ;  $ds = d\sigma$ ;

$$\begin{aligned} \left| \int_{\Gamma_2} C(s) ds \right| &\leq \int_{-R}^R \frac{1}{|1 + 9 \cdot 4^\sigma e^{2j\pi i}|} \frac{1}{|s(s+1)|} d\sigma \\ &\leq \frac{1}{1 + 9 \cdot 4^{-R}} \frac{2R}{R^2} = \frac{1}{1 + 9 \cdot 4^{-R}} \frac{2}{R} \end{aligned}$$

- $\Gamma_3$ ;  $s = -R + it$ ;  $ds = idt$ ;

$$\begin{aligned} \left| \int_{\Gamma_3} C(s) ds \right| &\leq \int_{-R}^R \frac{1}{|1 + 9 \cdot 4^{-R+it}|} \frac{1}{|s(s+1)|} dt \\ &\leq \frac{1}{R(R-1)} \frac{2R}{1 - 9 \cdot 4^{-R}} = \frac{2}{(R-1)(1 - 9 \cdot 4^{-R})} \end{aligned}$$

- $\Gamma_4$ ;  $s = \sigma - iR$ ;  $ds = d\sigma$ ;

This case is similar to  $\Gamma_2$ .

This shows that  $\lim_{R \rightarrow \infty} \int_{\Gamma'_1, \Gamma_{2,3,4}} C(s) ds = 0$  and hence  $\sum_{s \in \text{Sing}(C(s))} \text{Res}(C(s); s = \zeta) = 0$ . We compute the residues of  $C(s)$ .

$$\begin{aligned} \text{Res}(C(s); s = 0) &= \lim_{s \rightarrow 0} \frac{1}{1 + 9 \cdot 4^s} \frac{1}{s+1} = \frac{1}{10} \\ \text{Res}(C(s); s = -1) &= \lim_{s \rightarrow -1} \frac{1}{1 + 9 \cdot 4^s} \frac{1}{s} = \frac{4}{4 + 9} \frac{1}{-1} = -\frac{4}{13} \\ \text{Res}(C(s); s = \rho + \chi_k) &= \lim_{s \rightarrow \rho + \chi_k} \frac{s - (\rho + \chi_k)}{1 + 9 \cdot 4^s} \frac{1}{s(s+1)} \\ &= \frac{1}{(\rho + \chi_k)(\rho + \chi_k + 1)} \lim_{s \rightarrow \rho + \chi_k} \frac{1}{9 \log 4 \cdot 4^s} \\ &= \frac{1}{(\rho + \chi_k)(\rho + \chi_k + 1)} \frac{1}{-9 \log 4 \cdot \frac{1}{9}} = -\frac{1}{\log 4} \frac{1}{(\rho + \chi_k)(\rho + \chi_k + 1)} \end{aligned}$$

This yields

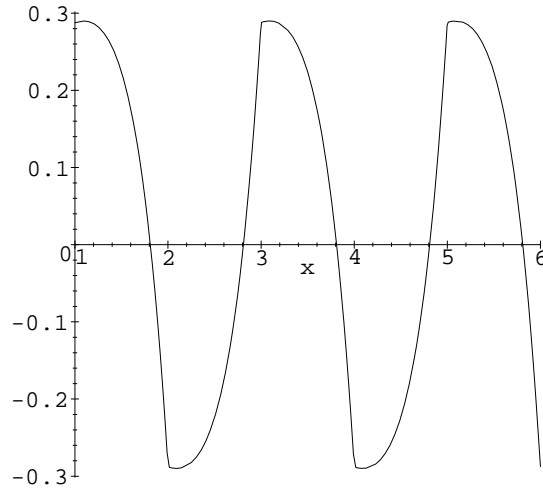
$$\sum_{k \in \mathbb{Z}} \frac{1}{(\rho + \chi_k)(\rho + \chi_k + 1)} = \log 4 \left( \frac{1}{10} - \frac{4}{13} \right) = -\log 4 \frac{27}{130}.$$

We can now verify the exact formula for  $n = 4^r$ .

$$\begin{aligned}
 a_{4^r} &= \frac{\sqrt{3}}{4} \frac{10}{13} - \frac{\sqrt{3}}{4} \frac{10}{9 \log 4} \left(-\frac{4}{9}\right)^r \sum_{k \in \mathbb{Z}} \frac{1}{(\rho + \chi_k)(\rho + \chi_k + 1)} \\
 &= \frac{\sqrt{3}}{4} \left( \frac{10}{13} + \frac{10}{9 \log 4} \log 4 \frac{27}{130} \left(-\frac{4}{9}\right)^r \right) \\
 &= \frac{\sqrt{3}}{4} \left( \frac{10}{13} + \frac{3}{13} \left(-\frac{4}{9}\right)^r \right)
 \end{aligned}$$

The exact and the restricted formula agree as expected.

### 3.5 Interpretation



plot of  $F(u)$  on a logarithmic scale for  $n \in [1, 4^6]$  and  $|k| < 256$

Note that

$$\begin{aligned}
 n^{\rho + \log_4 9} \sum_{k \in \mathbb{Z}} \frac{n^{\chi_k}}{(\rho + \chi_k)(\rho + \chi_k + 1)} &= e^{\log n \frac{i\pi}{\log 4}} \sum_{k \in \mathbb{Z}} \frac{e^{\log n \frac{2\pi ik}{\log 4}}}{(\rho + \chi_k)(\rho + \chi_k + 1)} \\
 &= e^{i\pi \log_4 n} \sum_{k \in \mathbb{Z}} \frac{e^{2\pi ik \log_4 n}}{(\rho + \chi_k)(\rho + \chi_k + 1)}.
 \end{aligned}$$

With

$$F(u) = e^{i\pi u} \sum_{k \in \mathbb{Z}} \frac{e^{2\pi iku}}{(\rho + \chi_k)(\rho + \chi_k + 1)}$$

we have

$$\begin{aligned} a_n &= \frac{\sqrt{3}}{4} \frac{10}{13} - \frac{\sqrt{3}}{4} \frac{10}{9 \log 4} n^{1 - \log_4 9} F(\log_4 n) \\ &= \frac{\sqrt{3}}{4} \frac{10}{13} - \frac{\sqrt{3}}{4} \frac{10}{9 \log 4} \frac{1}{n^{\log_4 \frac{9}{4}}} F(\log_4 n). \end{aligned}$$

The plot of  $F(u)$  shows the behaviour of  $a_n$  quite clearly. A pass from  $4^{r-1} + 1$  to  $4^r$  corresponds to a single level; on the graph these levels are unit intervals  $[r - 1, r)$ . The parity of the level determines whether the area increases or decreases; hence the alternating increase and decrease in  $F(u)$ . The formula for  $a_n$  further indicates that the fluctuation around the mean has an absolute value of  $\mathcal{O}\left(1/n^{\log_4 \frac{9}{4}}\right)$ , i.e.

$$\left| a_n - \frac{\sqrt{3}}{4} \frac{10}{13} \right| \in \mathcal{O}\left(\frac{1}{n^{\log_4 \frac{9}{4}}}\right).$$

## Chapter 4

### Digital Sums

The main topic of this chapter is the average order of digital sums in various bases. The results of [FGK<sup>+</sup>94] show how to obtain a Fourier series expansion of the sum-of-digits function with constant or exponential weights. We begin with basic definitions; then we demonstrate the method of [FGK<sup>+</sup>94] by considering alternating digital sums in base  $q$ , a special case of a kind of base known as a *Cantor representation*. We extend this method to periodic weights by giving a new proof of a result concerning alternating digital sums from [KPT85]. Thereafter we return to the general problem, and show that some classes of Cantor digital sums can also be dealt with by this method. We use the Mellin-Perron formula to obtain an asymptotic result that is similar to a theorem in [KPT85].

#### 4.1 Definitions

**Definition 4.1.1** Let  $q \in \mathbb{Z}^+$  and  $q \geq 2$ . For  $n \in \mathbb{N}$ ,

$$(d_r d_{r-1} \dots d_1 d_0)_q$$

denotes the unique  $q$ -ary representation of  $n$ , i.e.

$$n = \sum_{j=0}^r d_j q^j$$

where  $0 \leq d_j < q$ .

**Example.** We have  $m = q^{v_q(m)} m'$  with  $q \nmid m'$ , hence  $(m)_q = (m')_q \underbrace{000 \dots 000}_{v_q(m) \text{ 0 digits}}$ . (Side-by-side placement of two base- $q$  expansions means concatenation, not product.)

Cantor representations of integers generalize the concept of base- $q$  representations.

**Definition 4.1.2** Let  $\{q(j)\}_{j \geq 0} \subseteq \mathbb{Z}^+$  be a sequence of positive integers such that  $q(0) = 1$  and  $q(j) > 1$  when  $j \geq 1$ . Let  $\kappa(j) = \prod_{k=0}^j q(k)$  for  $j \geq 0$ . For  $n \in \mathbb{N}$ ,

$$(d_r d_{r-1} \dots d_1 d_0)_\kappa$$

denotes the unique base- $\kappa$  or Cantor representation of  $n$  with respect to  $\kappa$ , i.e.

$$n = \sum_{j=0}^r d_j \kappa(j)$$

where  $0 \leq d_j < q(j)$ .

It should be pointed out that  $\{\kappa(j)\}$  fulfills the requirements of Definition 2.7.3.

**Lemma 4.1.1** *The number of trailing zeros in the base- $\kappa$  representation of  $n$  is given by  $v_\kappa(n)$ .*

To see this, note that  $\sum_{j=0}^r (q(j) - 1)\kappa(j) < \kappa(r + 1)$ . This holds for  $r = 0$  and with

$$\kappa(r + 1) + (q(r + 1) - 1)\kappa(r + 1) = \kappa(r + 1)q(r + 1) < \kappa(r + 2)$$

for all  $r \in \mathbb{Z}^+$ .

Now  $\kappa(l) \mid n$  implies  $n \equiv \sum_{j=0}^r d_j \kappa(j) \equiv \sum_{j=0}^{l-1} d_j \kappa(j) \equiv 0 \pmod{\kappa(l)}$ . Because  $0 \leq \sum_{j=0}^{l-1} d_j \kappa(j) < \kappa(l)$ , this requires  $d_0 = d_1 = \dots = d_{l-1} = 0$ .

**Example.** The (2, 3)-number system;  $\{q(j)\} = \{1, 2, 3, 2, 3, 2, 3, \dots\}$  and  $\{\kappa(j)\} = \{1, 2, 6, 12, 36, \dots\}$ .

$(n)_{10}$	$(n)_\kappa$	$v_\kappa(n)$	$(n)_{10}$	$(n)_\kappa$	$v_\kappa(n)$	$(n)_{10}$	$(n)_\kappa$	$v_\kappa(n)$	$(n)_{10}$	$(n)_\kappa$	$v_\kappa(n)$
0	0	-	4	20	1	8	110	1	12	1000	3
1	1	0	5	21	0	9	111	0	13	1001	0
2	10	1	6	100	2	10	120	1	14	1010	1
3	11	0	7	101	0	11	121	0	15	1011	0

**Definition 4.1.3** *A weight function  $w(j)$  is a function  $w : \mathbb{N} \mapsto \mathbb{C}$ . The weighted base- $\kappa$  digital sum  $v(n)$  of  $n = \sum_{j=0}^r d_j \kappa(j)$  is*

$$v(n) = \sum_{j=0}^r w(j) d_j.$$

An alternating digital sum uses the weight function  $w(j) = (-1)^j$ .

**Example.** Let  $m = q^r - 1$ ,  $r \in \mathbb{Z}^+$ . Hence  $(m)_q = \underbrace{(q-1)(q-1) \dots (q-1)}_{r \text{ digits}}$ . The corresponding alternating digital sum is

$$v(m) = \begin{cases} 0 & \text{if } r \text{ is even} \\ q-1 & \text{if } r \text{ is odd.} \end{cases}$$

## 4.2 Alternating digital sums

In the remainder of this section  $v(n)$  will always refer to an alternating digital sum, i.e. with weight function  $w(j) = (-1)^j$ . We have the following theorem.

**Theorem 4.2.1** [KPT85] *The average order*

$$\frac{1}{n} \sum_{k=1}^{n-1} v(k)$$

of the alternating digital sum  $v(n)$  is given by  $F(\log_q n)$  where  $F(u)$  is a Fourier series

$$F(u) = f_0 + \sum_{k \in \mathbb{Z}} f_k e^{(2k+1)\pi i u}$$

with coefficients

$$f_0 = \frac{q-1}{4} \text{ and } f_k = \frac{q+1}{(2k+1)\pi i} \zeta\left(\frac{(2k+1)\pi i}{\log q}\right) \left(1 + \frac{(2k+1)\pi i}{\log q}\right)^{-1}.$$

The proof in [KPT85] uses elementary methods and builds on an earlier result by Delange. The remainder of this section will present a self-contained proof that uses the Mellin-Perron formula for  $m = 1$ . The method is that of [FGK<sup>+</sup>94].

### 4.2.1 Application of the Mellin-Perron formula

We wish to evaluate  $\sum_{k=1}^{n-1} v(k)$ . Consider  $\nabla v(n) = v(n) - v(n-1)$ , i.e. the change in  $v(n)$  from  $n-1$  to  $n$ . The following diagram shows how to evaluate  $\nabla v(n)$ . We assume that  $(n)_q$  contains a prefix of unspecified length, which ends in the digit  $d+1$ , followed by  $v_q(n)$  zeros.

$$\begin{array}{rcc} w(j) = & \dots & \mp 1 \\ (n-1)_q = & \dots & d \\ (n)_q = & \dots & (d+1) \end{array} \left| \begin{array}{ccccc} \pm 1 & \mp 1 & \dots & -1 & 1 \\ (q-1) & (q-1) & \dots & (q-1) & (q-1) \\ 0 & 0 & \dots & 0 & 0 \end{array} \right.$$

There are  $v_q(n)$  columns to the right of the vertical line. Note that  $d < q-1$  by definition of  $v_q(n)$ . Using the diagram, we have

$$\nabla v(n) = -d(-1)^{v_q(n)} + (d+1)(-1)^{v_q(n)} - \begin{cases} 0 & \text{if } v_q(n) \text{ is even} \\ q-1 & \text{if } v_q(n) \text{ is odd} \end{cases}$$

or

$$\nabla v(n) = (-1)^{v_q(n)} - (q-1) (v_q(n) \bmod 2).$$

Note that  $v(n) = v(n) - v(0) = \sum_{l=1}^n \nabla v(l)$  and hence  $\sum_{k=1}^{n-1} v(k) = \sum_{k=1}^{n-1} \sum_{l=1}^k \nabla v(l)$ . We apply Lemma 2.9.3, i.e. the Mellin-Perron formula with  $m = 1$  and  $b_l = \nabla v(l)$  to obtain

$$\frac{1}{n} \sum_{k=1}^{n-1} \sum_{l=1}^k \nabla v(l) = \frac{1}{n} \frac{n}{2\pi i} \int_{c-i\infty}^{c+i\infty} V(s) n^s \frac{ds}{s(s+1)} = \frac{1}{2\pi i} \int_{c-i\infty}^{c+i\infty} V(s) n^s \frac{ds}{s(s+1)}$$

where  $V(s) = \sum \frac{\nabla v(n)}{n^s}$  is the Dirichlet generating function of  $\nabla v(n)$ .

The next step is to determine  $V(s)$  and hence  $c$ . Evidently

$$V(s) = \sum \frac{(-1)^{v_q(n)} - (q-1) (v_q(n) \bmod 2)}{n^s}.$$

These terms were evaluated earlier, when we discussed the analytic version of the fundamental theorem of arithmetic (Theorem 2.7.2). We have

$$\sum \frac{(-1)^{v_q(n)}}{n^s} = \zeta(s) \left(1 - 2 \frac{1}{q^s + 1}\right) \text{ and } \sum \frac{v_q(n) \bmod 2}{n^s} = \zeta(s) \frac{1}{q^s + 1}$$

with  $\sigma > 1$  and hence

$$V(s) = \zeta(s) \left(1 - 2 \frac{1}{q^s + 1}\right) - (q-1) \zeta(s) \frac{1}{q^s + 1} = \zeta(s) \left(1 - \frac{q+1}{q^s + 1}\right),$$

also with  $\sigma > 1$ . Any  $c > 1$  will suffice.

#### 4.2.2 Evaluating the integral

We evaluate this integral by means of the shifting lemma (Lemma 2.8.1), taking

$$\Phi(s) = \left(1 - \frac{q+1}{q^s + 1}\right) n^s = \frac{q^s - q}{q^s + 1} n^s \text{ and } T_j = \frac{2j\pi}{\log q}$$

with  $j > 0$ . Along the vertical segments situated at  $\alpha$  and  $c$  (recall that the contour is rectangular, with  $c > 1$  being the right vertical boundary, and  $\alpha \in (-1, 0)$  the left one)

$$|\Phi(s)| = \left| \frac{q^s - q}{q^s + 1} \right| n^\alpha \leq \frac{q^\alpha + q}{1 - q^\alpha} n^\alpha = M_\alpha \text{ and } |\Phi(s)| = \left| \frac{q^s - q}{q^s + 1} \right| n^c \leq \frac{q^c + q}{q^c - 1} = M_c$$

respectively. Along the two horizontal segments  $\langle \alpha, c \rangle \cap \{s \mid s = \sigma \pm iT_j\}$  we have

$$|\Phi(s)| = \left| \frac{q^\sigma - q}{q^\sigma + 1} \right| n^\sigma \leq \frac{q^c - q}{q^\alpha + 1} n^c = M_T.$$

Therefore the constant  $M$  required by the lemma is  $M = \max\{M_\alpha, M_c, M_T\}$ .

The poles of  $\Phi(s)\frac{\zeta(s)}{s(s+1)}$  in  $\langle \alpha, c \rangle$  are at  $s = 0$  and  $s = \rho + \chi_k$  where  $\rho = \frac{i\pi}{\log q}$  and  $\chi_k = \frac{2\pi ik}{\log q}$ ,  $k \in \mathbb{Z}$ . (There is no pole at  $s = 1$  because the pole of  $\zeta(s)$  at  $s = 1$  is simple and  $\Phi(1) = 0$ ; the zero of  $\Phi(s)$  cancels the pole, as in  $(1/(s-1) + \dots)(\phi_1(s-1) + \dots) = \phi_1 + \dots$ ) Hence the lemma gives

$$\begin{aligned} \frac{1}{2\pi i} \int_{c-i\infty}^{c+i\infty} \Phi(s) \frac{\zeta(s)}{s(s+1)} ds &= \frac{1}{2\pi i} \int_{\alpha-i\infty}^{\alpha+i\infty} \Phi(s) \frac{\zeta(s)}{s(s+1)} ds \\ &+ \text{Res} \left( \Phi(s) \frac{\zeta(s)}{s(s+1)}; s=0 \right) + \sum_{k=0} \text{Res} \left( \Phi(s) \frac{\zeta(s)}{s(s+1)}; s=\rho + \chi_k \right). \end{aligned}$$

Note that

$$\frac{1}{2\pi i} \int_{\alpha-i\infty}^{\alpha+i\infty} \Phi(s) \frac{\zeta(s)}{s(s+1)} ds = \frac{1}{2\pi i} \int_{\alpha-i\infty}^{\alpha+i\infty} \zeta(s) \frac{n^s}{s(s+1)} ds - (q+1) \frac{1}{2\pi i} \int_{\alpha-i\infty}^{\alpha+i\infty} \frac{1}{q^s+1} \zeta(s) \frac{n^s}{s(s+1)} ds.$$

Note also that the expansion  $\frac{1}{1+q^s} = 1 - q^s + q^{2s} - q^{3s} + \dots$  is convergent since  $|q^s| = q^\alpha \in (-1, 0)$ . Corollary 2.10.1 applies. We conclude that  $\frac{1}{2\pi i} \int_{\alpha-i\infty}^{\alpha+i\infty} \Phi(s) \frac{\zeta(s)}{s(s+1)} ds = 0$ ; both the constant term and the terms of the series expansion vanish.

It remains to compute the residues. Let  $\log_q n = u$ , and  $W(s) = \Phi(s) \frac{\zeta(s)}{s(s+1)}$ .

$$\begin{aligned} \text{Res}(W(s); s=0) &= \lim_{s \rightarrow 0} \Phi(s) \frac{\zeta(s)}{s+1} = \left(1 - \frac{q+1}{2}\right) \zeta(0) = \frac{q+1}{4} - \frac{1}{2} = \frac{q-1}{4} \\ \text{Res}(W(s); s=\rho + \chi_k) &= \lim_{s \rightarrow \rho + \chi_k} (s - \rho - \chi_k) \left(1 - \frac{q+1}{q^s+1}\right) n^s \frac{\zeta(s)}{s+1} \\ &= n^{\rho + \chi_k} \frac{\zeta(\rho + \chi_k)}{(\rho + \chi_k)(\rho + \chi_k + 1)} \lim_{s \rightarrow \rho + \chi_k} \left(s - \rho - \chi_k - (q+1) \frac{s - \rho - \chi_k}{q^s + 1}\right) \\ &= e^{u(2k+1)\pi i} \zeta\left(\frac{(2k+1)\pi i}{\log q}\right) \frac{(q+1) \log q}{(2k+1)\pi i} \left(1 + \frac{(2k+1)\pi i}{\log q}\right)^{-1} \frac{-1}{-\log q} \\ &= e^{u(2k+1)\pi i} \frac{q+1}{(2k+1)\pi i} \zeta\left(\frac{(2k+1)\pi i}{\log q}\right) \left(1 + \frac{(2k+1)\pi i}{\log q}\right)^{-1} \end{aligned}$$

This concludes the proof of Theorem 4.2.1.  $\blacksquare$

### 4.3 Periodic weights in general

The proof of Theorem 4.2.1 serves to illustrate the general method of treating digital sums with periodic weights. We will sketch the case

$$w(j) = \begin{cases} 2 & \text{if } j \equiv 0(3) \\ 5 & \text{if } j \equiv 1(3) \\ 11 & \text{if } j \equiv 2(3). \end{cases}$$

We consider the function  $\nabla v(n)$ , which is computed in a manner analogous to the case of an alternating digital sum ( $v(n)$  uses the new  $w(j)$ ). We have

$$\nabla v(n) = -dw(v_q(n)) + (d+1)w(v_q(n)) - 18(q-1) \left\lfloor \frac{v_q(n)}{3} \right\rfloor - \begin{cases} 0 & \text{if } v_q(n) \equiv 0(3) \\ 2(q-1) & \text{if } v_q(n) \equiv 1(3) \\ 7(q-1) & \text{if } v_q(n) \equiv 2(3). \end{cases}$$

With  $\lfloor n/3 \rfloor = n/3 - (n \bmod 3)/3$  this simplifies to

$$\nabla v(n) = -6(q-1)v_q(n) + \begin{cases} 2 & \text{if } v_q(n) \equiv 0(3) \\ 5 + 4(q-1) & \text{if } v_q(n) \equiv 1(3) \\ 11 + 5(q-1) & \text{if } v_q(n) \equiv 2(3). \end{cases}$$

This is the general form of  $\nabla v(n)$ , i.e.  $\nabla v(n)$  is a multiple of  $v_q(n)$  and a term linear in  $q$ , plus a second term linear in  $q$ , one for each of the residues of  $v_q(n)$  modulo the period length. We recall that

$$\sum \frac{v_q(n)}{n^s} = \frac{\zeta(s)}{q^s - 1} \quad \text{and} \quad \sum_{v_q(n) \equiv r(m)} \frac{1}{n^s} = \zeta(s) q^{(m-r-1)s} \prod_{v=1}^{m-1} (q^s - \omega_m^v)^{-1}.$$

The Dirichlet generating function  $V(s)$  of  $\nabla v(n)$  is a linear combination of these two kinds of terms, with poles corresponding to  $q^s - \omega_m^v = 0$  or

$$s = \frac{2\pi iv/m + 2\pi ik}{\log q}$$

for  $0 \leq v < m$  and  $k \in \mathbb{Z}$ . (The generating function of  $v_q(n)$  contributes  $v = 0$  and the residues  $r$  modulo  $m$  the rest. Note that there may be some cancellation of poles, such as that of  $s = 1$  in the  $v_q(n)$  term.)

Continuing the example, we have

$$\begin{aligned} V(s) &= -6(q-1) \frac{\zeta(s)}{q^s - 1} + \zeta(s) (2q^{2s} + 5q^s + 4q^s(q-1) + 11 + 5(q-1)) \frac{1}{(q^s - \omega_3)(q^s - \omega_3^2)} \\ &= \zeta(s) \left( \frac{2q^{2s} + 4q^{s+1} + q^s + 5q + 6}{(q^s - \omega_3)(q^s - \omega_3^2)} - 6 \frac{1}{(q^s - 1)/(q-1)} \right) \end{aligned}$$

At this point it is a matter of routine computation to obtain the Fourier series expansion of the average order of  $v(n)$ . Two questions must be considered.

- Does the shifting lemma apply to  $V(s)n^s/(s(s+1))$ , i.e. how do we choose  $M$  and  $T_j$  for  $\Phi(s) = V(s)n^s/\zeta(s)$ ?

Note that all the poles of  $\Phi(s)$  are staggered along the imaginary axis. Furthermore, there is only a finite number of singularities in the interval  $2\pi i/\log q[k, k+1]$ . Hence we can choose  $T_j$  such that  $\Phi(s)$  is analytic on a closed rectangular band that includes  $\delta R_j$ . Because  $\Phi(s)$  is analytic there,  $|\Phi(s)|$  is bounded. We need to verify that this bound is independent of  $j$ . But  $\Phi(s)$  contains only terms in  $n^s$  and  $q^s$ , with  $|n^s| = n^\sigma$  and  $|q^s| = q^\sigma$ . This observation and the analyticity of  $\Phi(s)$  yield the claim.

- Does  $\int_{\alpha-i\infty}^{\alpha+i\infty} V(s)n^s/(s(s+1))ds$  vanish? (We need this in order to ensure that there are no terms other than the Fourier series; compare the proof of Theorem 5.1.2, where the corresponding term does not vanish.)

This is a question of expanding  $(q^s - \omega_m^v)^{-1}$  with  $0 \leq v < m$ . (We use the partial fraction decomposition of  $\prod_{v=1}^{m-1} (q^s - \omega_m^v)^{-1}$ .) Note that

$$\frac{1}{q^s - \omega_m^v} = -\frac{1}{\omega_m^v} \frac{1}{1 - q^s/\omega_m^v} = -\frac{1}{\omega_m^v} \left( 1 + \frac{q^s}{\omega_m^v} + \left( \frac{q^s}{\omega_m^v} \right)^2 + \dots \right)$$

converges since  $|q^s/\omega_m^v| = q^\alpha$  and  $\alpha \in (-1, 0)$ . Hence we may apply Corollary 2.10.1 to  $(qn)^{ks}$  ( $qn \in \mathbb{N}$  and  $1/\omega_m^{kv}$  is a constant factor with respect to  $s$ ). The integral vanishes as claimed.

The above observations lead to the following statement. *Suppose  $w(j)$  is a periodic weight function and  $v(n)$  the associated digital sum. Then the average order of  $v(n)$  can be expanded into a sum of Fourier series with terms corresponding to  $s = \frac{2\pi iv/m + 2\pi ik}{\log q}$ .*

#### 4.4 Intermezzo: digital sum paradigms

The preceding discussion should suffice to demonstrate that the problem of computing the average order of a general digital sum by Mellin-transform methods requires that the two following conditions hold.

- There must exist a closed form of  $\nabla v(n)$  in terms of a polynomial of the “number of trailing zeros”-function in  $\mathbb{Z}$  or  $\mathbb{Z}_q$ , or in exponentials of this function.

Recall that the “trailing zeros”-function is given by  $v_\kappa$  (Lemma 4.1.1.) It follows that the types of  $\nabla v(n)$  generated by periodic, constant or exponential weights all fit this condition.

- The corresponding Dirichlet generating function  $V(s)$  must have a closed form and  $V(s)n^s/\zeta(s)$  must satisfy the requirements of the shifting lemma.

The Lemmata 2.7.1 and 2.7.2 show that the behavior of  $\sum_{j=1} \kappa(j)^{-s}$  and  $\sum_{k=0} \kappa(mk+r)^{-s}$ ,  $m \geq 2$ ,  $0 \leq r < m$  determines that of  $V(s)$ . E.g. if  $\sum_{j=1} \kappa(j)^{-s}$  does not represent a meromorphic function in  $\langle \alpha, c \rangle$ ,  $V(s)$  fails the shifting lemma.

The above criteria constitute an informal quick test for the computability of a Fourier expansion by the Mellin-Perron formula for a given a digital sum problem. The significant part of the test is the investigation of the properties (read: analyticity and location of poles) of  $\sum_{j=1} \kappa(j)^{-s}$  and  $\sum_{k=0} \kappa(mk+r)^{-s}$ . We will present two additional examples in the remainder of this chapter. The first of these exhibits a  $V(s)$  that is well-behaved; the second shows how  $V(s)$  may fail the second condition.

#### 4.5 Digital sums relative to $\kappa$ when $\kappa(j+1)/\kappa(j) = q(j+1)$ is periodic

We treat the case

$$\{q(j)\} = \{1, 2, \dots, a+1, 2, \dots, a+1, \dots\}$$

and

$$\{\kappa(j)\} = \{1, 2!, \dots, (a+1)!, 2(a+1)! \dots (a+1)!^2 \dots\},$$

where  $a > 1$ . We select this case because it is one of a series of  $\kappa$  that have the factorial number system as their limit, see section 4.6.

**Step 1.** What is  $\nabla v(n)$ ? Every complete sequence of  $a$  zeros corresponds to digits  $a, a-1, \dots, 1$  lost from  $(n-1)_\kappa$ ; the remainder corresponds to digits  $r, r-1, \dots, 1$ , where  $r = v_\kappa(n) \bmod a$ . These digits are replaced by zeros; we gain a 1 in the first non-zero digit of  $(n-1)_\kappa$ . Hence

$$\nabla v(n) = 1 - \frac{1}{2}a(a+1) \left\lfloor \frac{v_\kappa(n)}{a} \right\rfloor - \frac{1}{2}(v_\kappa(n) \bmod a)(v_\kappa(n) \bmod a + 1).$$

**Step 2.** What is  $V(s)$ ? Using  $\lfloor v_\kappa(n)/a \rfloor = v_\kappa(n)/a - (v_\kappa(n) \bmod a)/a$ , the problem reduces to finding the Dirichlet generating functions of  $v_\kappa(n) \bmod a$ ,  $\frac{1}{2}(v_\kappa(n) \bmod a)(v_\kappa(n) \bmod a + 1)$  and  $v_\kappa(n)$ . We apply Lemma 2.7.2 to obtain the Dirichlet generating function of  $v_\kappa(n)$ . Note that  $\kappa(ak+r) = (a+1)^k(r+1)!$  and hence

$$\sum_{j=1} \frac{1}{\kappa(j)^s} = \sum_{r=1}^a \sum_{k=0} \frac{1}{\kappa(ak+r)^s} = \frac{(a+1)^s}{(a+1)^s - 1} \sum_{r=1}^a \frac{1}{(r+1)^s}.$$

This function contributes poles at  $(a+1)^s - 1 = 0$ . We use Lemma 2.7.1 to evaluate the remaining two types, i.e.

$$\sum \left\{ \begin{array}{c} v_\kappa(n) \bmod a \\ \frac{1}{2}(v_\kappa(n) \bmod a)(v_\kappa(n) \bmod a + 1) \end{array} \right\} \frac{1}{n^s} = \sum_{r=1}^{a-1} \left\{ \begin{array}{c} r \\ \frac{1}{2}r(r+1) \end{array} \right\} \sum_{v_\kappa(n) \equiv r(a)} \frac{1}{n^s}$$

and the relevant terms of the respective generating functions are

$$\frac{(a+1)!^s}{(a+1)!^s - 1} \sum_{r=1}^{a-1} r \left( \frac{1}{(r+1)!^s} - \frac{1}{(r+2)!^s} \right) \quad \text{and} \quad \frac{(a+1)!^s}{(a+1)!^s - 1} \sum_{r=1}^{a-1} \frac{1}{2} r(r+1) \left( \frac{1}{(r+1)!^s} - \frac{1}{(r+2)!^s} \right).$$

**Step 3.** Does the shifting lemma apply? We note that  $V(s)$  is the product of  $\zeta(s)$  and finite sum of terms, which are in turn the product of a meromorphic function with poles at  $s = 2\pi ik / \log(a+1)!$  and a finite sum of entire functions (the  $1/(r+1)!^s$  terms). Therefore the shifting lemma applies with  $T_j = \pi i(2j+1) / \log(a+1)!$ ; the term  $(a+1)!^s / ((a+1)!^s - 1)$  was evaluated in the base- $q$  problem and the norm of the terms from  $1/(r+1)!^s$  is bounded on the horizontal and constant on the vertical segments.

#### 4.6 Digital sums in the factorial number system

The factorial number system has  $q(j) = j+1$  and  $\kappa(j) = (j+1)!$ .

**Step 1.** What is  $\nabla v(n)$ ? We have  $v_\kappa(n)$  zeros, which correspond to the digits  $v_\kappa(n), \dots, 2, 1$ , hence  $\nabla v(n) = 1 - 1/2v_\kappa(n)(v_\kappa(n) + 1)$ .

**Step 2.** What is  $V(s)$ ? We use Lemma 2.7.2 to obtain

$$V(s) = \zeta(s) \left( 1 - \sum_{j=1}^{\infty} \frac{j}{\kappa(j)^s} \right) = \zeta(s) \left( 1 - \sum_{j=1}^{\infty} \frac{j}{(j+1)!^s} \right).$$

(Compare this with the corresponding function in the previous section. The contribution from the period length  $a$  vanishes and the finite sums in  $r/(r+1)!^s$  become an infinite series.)

**Step 3.** Does the shifting lemma apply? Recall that  $\sigma = 0$  is a natural boundary of  $\sum_{j=1}^{\infty} \frac{j}{(j+1)!^s}$  by Corollary 2.6.1. Hence we cannot continue this sum into the left half-plane. The shifting lemma does not apply.

We can however extract some information from the Mellin-Perron formula. Evidently the problem requires the evaluation of

$$\int_{c-i\infty}^{c+i\infty} \zeta(s) \left( 1 - \sum_{j=1}^{\infty} \frac{j}{(j+1)!^s} \right) \frac{n^s}{s(s+1)} ds.$$

With Corollary 2.10.1, this simplifies to

$$\frac{1}{2}(n-1) - \int_{c-i\infty}^{c+i\infty} \zeta(s) \left( \sum_{j=1}^{\infty} \frac{j}{(j+1)!^s} \right) \frac{n^s}{s(s+1)} ds.$$

(We could also have evaluated the first term directly, i.e. at Step 1.) Note that Theorem 2.10.1 applies to

$$j \int_{c-i\infty}^{c+i\infty} \zeta(s) \frac{1}{(j+1)!^s} \frac{n^s}{s(s+1)} ds$$

with  $r = (j+1)!$ ,  $a = 1$ . Hence

$$j \int_{c-i\infty}^{c+i\infty} \zeta(s) \frac{1}{(j+1)!^s} \frac{n^s}{s(s+1)} ds = j \sum_{k \in \mathbb{N}(n/(j+1)!-1)} \left( 1 - (k+1) \frac{(j+1)!}{n} \right).$$

Note that  $\mathbb{N}(n/(j+1)!-1) = \emptyset$  when  $(j+1)! \geq n$ . Therefore only a finite number of terms actually contribute to the integral in  $\sum_{j=1}^{\infty} j/(j+1)!^s$ , which justifies writing

$$\int_{c-i\infty}^{c+i\infty} \zeta(s) \left( \sum_{j=1}^{\infty} \frac{j}{(j+1)!^s} \right) \frac{n^s}{s(s+1)} ds = \sum_{j \geq 1}^{(j+1)! < n} j \sum_{k \in \mathbb{N}(n/(j+1)!-1)} \left( 1 - (k+1) \frac{(j+1)!}{n} \right).$$

By definition of  $v_\kappa(n)$ ,

$$\sum_{\mathbb{N}(n/(j+1)!-1)} = \begin{cases} \sum_0^{n/(j+1)!-2} & \text{if } j \leq v_\kappa(n) \\ \sum_0^{\lfloor n/(j+1)! \rfloor - 1} & \text{if } j > v_\kappa(n). \end{cases}$$

Hence the integral splits into two sums,

$$\begin{aligned} \sum_{j=1}^{v_\kappa(n)} j \left( \frac{n}{(j+1)!} - 1 \right) \left( 1 - \frac{(j+1)!}{2n} \frac{n}{(j+1)!} \right) &= -\frac{1}{4} v_\kappa(n)(v_\kappa(n)+1) + \frac{1}{2} n \sum_{j=1}^{v_\kappa(n)} \frac{j}{(j+1)!} \\ &= -\frac{1}{4} v_\kappa(n)(v_\kappa(n)+1) + \frac{1}{2} n \left( 1 - \frac{1}{(v_\kappa(n)+1)!} \right) \end{aligned}$$

and

$$\sum_{j=v_\kappa(n)+1}^{(j+1)! < n} j \left\lfloor \frac{n}{(j+1)!} \right\rfloor \left( 1 - \frac{(j+1)!}{2n} \left( \left\lfloor \frac{n}{(j+1)!} \right\rfloor + 1 \right) \right).$$

This formula has some utility as we shall see below. Nonetheless it must be pointed out that it can equally well be derived by an elementary counting argument (the reader is urged to supply this proof). There is no *qualitative* gain with respect to elementary methods because the function  $V(s)$  is not meromorphic and the terms of the series merely transcribe the problem definition.

We remark in passing that  $n = (r+1)!$  gives  $v_\kappa(n) = v_\kappa((r+1)!) = r$ , in which case the second term of the above sum drops out and  $1/n \sum_{m=1}^{n-1} v(m)$  is given by

$$\frac{1}{2}(n-1) - \left( -\frac{1}{4} v_\kappa(n)(v_\kappa(n)+1) + \frac{1}{2} n \left( 1 - \frac{1}{(v_\kappa(n)+1)!} \right) \right) = \frac{1}{4} r(r+1).$$

The average order of the sum of factorial digits of the first  $n = (r + 1)!$  non-negative integers is quadratic in the factorial inverse of  $n$ .

The work of this section suggests the following question. Considering the fact that

$$\frac{1}{n} \sum_{m=1}^{n-1} v(m) \sim \frac{1}{2} \log_q n \sim \frac{1}{2} \sum_{j=0}^{\kappa(j) < n} (q(j+1) - 1)$$

for  $q$ -ary digital sums,  $\kappa(j) = q^j$ ,  $q \geq 2$  and all  $n$ , and

$$\frac{1}{n} \sum_{m=1}^{n-1} v(m) = \frac{1}{2} \left( \frac{1}{2} r(r+1) \right) \sim \frac{1}{2} \sum_{j=0}^{\kappa(j) < n} (q(j+1) - 1)$$

for digital sums in the factorial number system,  $\kappa(j) = (j+1)!$ ,  $n = (r+1)!$ , what are the conditions such that

$$\frac{1}{2} \sum_{j=0}^{\kappa(j) < n} w(j) (q(j+1) - 1)$$

is the asymptotically dominant term of the general sum-of-digits function? This question will be answered in the next section.

We will need the following observation. Suppose  $q(j) \rightarrow \infty$  as  $j \rightarrow \infty$ . Then  $\kappa(j+1) < n$  implies

$$\frac{\kappa(j)}{n} \in o(1) \quad \text{as} \quad n \rightarrow \infty.$$

To see this, note that

$$\frac{\kappa(j)}{n} = \frac{\kappa(\kappa^{-1}(n) - 1)}{n} \prod_{k=j+1}^{\kappa^{-1}(n)-1} \frac{1}{q(k)} < \prod_{k=j+1}^{\kappa^{-1}(n)-1} \frac{1}{q(k)} \rightarrow 0 \quad \text{as} \quad n \rightarrow \infty.$$

#### 4.7 The general digital sum problem

We wish to examine the role of

$$\frac{1}{2} \sum_{j=0}^{\kappa(j) < n} w(j) (q(j+1) - 1)$$

in the behavior of

$$\frac{1}{n} \sum_{m=1}^{n-1} v(m)$$

for an arbitrary weight function and an arbitrary Cantor system  $\kappa$ . We require an expression of  $\nabla v(n)$ .

Define  $s : \mathbb{N} \mapsto \mathbb{C}$  by

$$\begin{aligned} s(v) &= w(v) - \sum_{j=0}^{v-1} w(j) (q(j+1) - 1) \\ &= w(0) + w(v) - w(0) - \sum_{j=1}^v w(j-1) (q(j) - 1) = w(0) + \sum_{j=1}^v (w(j) - w(j-1)q(j)). \end{aligned}$$

Evidently

$$\nabla v(n) = s(v_\kappa(n)).$$

By Lemma 2.7.2, the Dirichlet generating function  $V(s)$  is

$$\zeta(s) \left( w(0) + \sum \frac{1}{\kappa(j)^s} (w(j) - w(j-1)q(j)) \right).$$

We proceed by the same method that was used to obtain a formula for the average order of digital sums in the factorial number system in the previous section. The  $w(0)$  term corresponds to

$$w(0) \frac{n-1}{2}$$

by Corollary 2.10.1. Theorem 2.10.1 is used to evaluate the series. We have

$$\int_{c-i\infty}^{c+i\infty} \zeta(s) \frac{1}{\kappa(j)^s} \frac{n^s}{s(s+1)} ds = \sum_{k \in \mathbb{N}(n/\kappa(j)-1)} \left( 1 - (k+1) \frac{\kappa(j)}{n} \right).$$

The sum is zero when  $\kappa(j) \geq n$ . There are two cases when  $\kappa(j) < n$ .

**Case 1.**  $\kappa(j) \mid n$

$$\sum_{k=0}^{n/\kappa(j)-2} \left( 1 - (k+1) \frac{\kappa(j)}{n} \right) = \left( \frac{n}{\kappa(j)} - 1 \right) \left( 1 - \frac{1}{2} \frac{\kappa(j)}{n} \frac{n}{\kappa(j)} \right) = \frac{1}{2} \left( \frac{n}{\kappa(j)} - 1 \right)$$

**Case 2.**  $\kappa(j) \nmid n$

$$\begin{aligned} & \sum_{k=0}^{\lfloor n/\kappa(j) \rfloor - 1} \left( 1 - (k+1) \frac{\kappa(j)}{n} \right) = \left\lfloor \frac{n}{\kappa(j)} \right\rfloor \left( 1 - \frac{1}{2} \frac{\kappa(j)}{n} \left( \left\lfloor \frac{n}{\kappa(j)} \right\rfloor + 1 \right) \right) \\ &= \left( \frac{n}{\kappa(j)} - \frac{n \bmod \kappa(j)}{\kappa(j)} \right) \left( 1 - \frac{1}{2} \frac{\kappa(j)}{n} \left( \frac{n}{\kappa(j)} - \frac{n \bmod \kappa(j)}{\kappa(j)} + 1 \right) \right) \\ &= \left( \frac{n}{\kappa(j)} - \frac{n \bmod \kappa(j)}{\kappa(j)} \right) \left( \frac{1}{2} - \frac{1}{2} \frac{\kappa(j)}{n} + \frac{1}{2} \frac{n \bmod \kappa(j)}{n} \right) \\ &= \frac{1}{2} \left( \frac{n}{\kappa(j)} - 1 \right) + \frac{1}{2} \frac{n \bmod \kappa(j)}{\kappa(j)} - \frac{1}{2} \left( \frac{n \bmod \kappa(j)}{\kappa(j)} - \frac{n \bmod \kappa(j)}{n} + \frac{(n \bmod \kappa(j))^2}{n\kappa(j)} \right) \\ &= \frac{1}{2} \left( \frac{n}{\kappa(j)} - 1 \right) + \frac{1}{2} \frac{n \bmod \kappa(j)}{n} \left( 1 - \frac{n \bmod \kappa(j)}{\kappa(j)} \right) \end{aligned}$$

Evaluating the contribution from the

$$\frac{1}{2} \left( \frac{n}{\kappa(j)} - 1 \right)$$

terms, we have

$$\begin{aligned}
& \sum_{j=1}^{\kappa(j)<n} \frac{1}{2} \left( \frac{n}{\kappa(j)} - 1 \right) (w(j) - w(j-1)q(j)) \\
= & \frac{1}{2}n \sum_{j=1}^{\kappa(j)<n} \left( \frac{w(j)}{\kappa(j)} - \frac{w(j-1)}{\kappa(j-1)} \right) - \frac{1}{2} \sum_{j=1}^{\kappa(j)<n} w(j) + \frac{1}{2} \sum_{j=1}^{\kappa(j)<n} w(j-1)q(j) \\
= & \frac{1}{2}n \left( \frac{w(\kappa^{-1}(n)-1)}{\kappa(\kappa^{-1}(n)-1)} - w(0) \right) + \frac{1}{2}w(0) - \frac{1}{2} \sum_{j=0}^{\kappa(j)<n} w(j) \\
& - \frac{1}{2}w(\kappa^{-1}(n)-1)q(\kappa^{-1}(n)) + \frac{1}{2} \sum_{j=0}^{\kappa(j)<n} w(j)q(j+1) \\
= & -w(0)\frac{n-1}{2} + \frac{1}{2} \sum_{j=0}^{\kappa(j)<n} w(j) (q(j+1) - 1) + \frac{1}{2}w(\kappa^{-1}(n)-1) \left( \frac{n}{\kappa(\kappa^{-1}(n)-1)} - q(\kappa^{-1}(n)) \right)
\end{aligned}$$

It remains to evaluate the contribution from the second term of the case 2 sum. With

$$\mu(n, j) = (n \bmod \kappa(j))(\kappa(j) - n \bmod \kappa(j)),$$

$$\mu(n, \kappa^{-1}(n)) = n(\kappa(\kappa^{-1}(n)) - n)$$

and

$$\frac{1}{2} \frac{n \bmod \kappa(j)}{n} \left( 1 - \frac{n \bmod \kappa(j)}{\kappa(j)} \right) = \frac{1}{2n\kappa(j)} \mu(n, j),$$

this becomes

$$\begin{aligned}
& \frac{1}{2n} \sum_{j=v_{\kappa}(n)+1}^{\kappa(j)<n} \mu(n, j) \left( \frac{w(j)}{\kappa(j)} - \frac{w(j-1)}{\kappa(j-1)} \right) \\
= & \frac{1}{2n} \left( \sum_{j=v_{\kappa}(n)}^{\kappa(j)<n} \frac{w(j)}{\kappa(j)} (\mu(n, j) - \mu(n, j+1)) - \mu(n, v_{\kappa}(n)) \frac{w(v_{\kappa}(n))}{\kappa(v_{\kappa}(n))} + \mu(n, \kappa^{-1}(n)) \frac{w(\kappa^{-1}(n)-1)}{\kappa(\kappa^{-1}(n)-1)} \right) \\
= & -\frac{1}{2n} \sum_{j=v_{\kappa}(n)}^{\kappa(j)<n} \frac{w(j)}{\kappa(j)} (\mu(n, j+1) - \mu(n, j)) + \frac{1}{2}w(\kappa^{-1}(n)-1) \left( q(\kappa^{-1}(n)) - \frac{n}{\kappa(\kappa^{-1}(n)-1)} \right)
\end{aligned}$$

We combine these results to obtain the following theorem.

**Theorem 4.7.1** *Let  $w(j)$  be an arbitrary weight function,  $\kappa$  any Cantor system, and define*

$$\mu(n, j) = (n \bmod \kappa(j))(\kappa(j) - n \bmod \kappa(j)).$$

*The sum-of-digits function for  $w$  and  $\kappa$  is given by*

$$\frac{1}{n} \sum_{m=1}^{n-1} v(m) = \frac{1}{2} \sum_{j=0}^{\kappa(j)<n} w(j) (q(j+1) - 1) - \frac{1}{2n} \sum_{j=v_{\kappa}(n)}^{\kappa(j)<n} \frac{w(j)}{\kappa(j)} (\mu(n, j+1) - \mu(n, j)).$$

We can use this theorem to answer the question posed in the previous section. In the following, we will assume that the  $w(j)$  are positive and that  $q(j) \rightarrow \infty$  as  $j \rightarrow \infty$ . The term

$$\frac{1}{2} \sum_{j=0}^{\kappa(j) < n} w(j) (q(j+1) - 1)$$

will dominate asymptotically if it dominates the second term. It is not difficult to see that

$$\mu(n, j) \leq \frac{1}{4} \kappa(j)^2$$

when  $\kappa(j) < n$ . This gives the following estimate for the  $-\mu(n, j)$  part of the second term.

$$\sum_{j=v_{\kappa(n)}}^{\kappa(j) < n} \frac{w(j)}{n\kappa(j)} \mu(n, j) \leq \sum_{j=v_{\kappa(n)}}^{\kappa(j) < n} w(j) q(j) \frac{\kappa(j-1)}{4n} \leq \sum_{j=v_{\kappa(n)}}^{\kappa(j) < n} w(j) (q(j+1) - 1) \frac{\kappa(j-1)}{4n}$$

Using our earlier observation and a term-by-term comparison we see that the first term dominates the  $-\mu(n, j)$  part asymptotically.

We split the  $\mu(n, j+1)$  part into  $\sum_{j=v_{\kappa(n)}}^{\kappa(j+1) < n}$  and  $j = \kappa^{-1}(n) - 1$ . For the first part we again have

$$\begin{aligned} \sum_{j=v_{\kappa(n)}}^{\kappa(j+1) < n} \frac{w(j)}{n\kappa(j)} \mu(n, j+1) &\leq \sum_{j=v_{\kappa(n)}}^{\kappa(j+1) < n} w(j) (q(j+1) - 1) \frac{\kappa(j)}{4n} \frac{q(j+1)^2}{q(j+1) - 1} \\ &\leq \sum_{j=v_{\kappa(n)}}^{\kappa(j+1) < n} w(j) (q(j+1) - 1) \frac{\kappa(j)}{4n} \left( q(j+1) + 1 + \frac{1}{q(j+1) - 1} \right). \end{aligned}$$

With

$$1 + \frac{1}{q(j+1) - 1} \leq 2$$

this is less than or equal to

$$\sum_{j=v_{\kappa(n)}}^{\kappa(j+1) < n} w(j) (q(j+1) - 1) \left( \frac{\kappa(j+1)}{4n} + \frac{\kappa(j)}{2n} \right).$$

This part is also dominated by the first term, except for the first half of the sum when  $j = \kappa^{-1}(n) - 2$ , which is

$$w(\kappa^{-1}(n) - 2) (q(\kappa^{-1}(n) - 1) - 1) \frac{\kappa(\kappa^{-1}(n) - 1)}{4n}.$$

Suppose we have

$$w(v-1)(q(v)-1) \in o\left(\frac{1}{q(v)} \sum_{j=0}^{v-1} w(j)(q(j+1)-1)\right)$$

for  $v$  sufficiently large. Then certainly

$$w(v-1)(q(v)-1)q(v)\frac{\kappa(v-1)}{4n} \in o(w(v-1)(q(v)-1)q(v)) \subset o\left(\sum_{j=0}^v w(j)(q(j+1)-1)\right)$$

for  $\kappa(v) < n$ . Taking  $v = \kappa^{-1}(n) - 1$ , we see that we have a sufficient condition for the first term to dominate. It remains to test  $j = \kappa^{-1}(n) - 1$ , in which case  $w(j)/(nk(j))\mu(n, j+1)$  becomes

$$\begin{aligned} \frac{w(\kappa^{-1}(n)-1)}{\kappa(\kappa^{-1}(n)-1)}(\kappa(\kappa^{-1}(n))-n) &< \frac{w(\kappa^{-1}(n)-1)}{\kappa(\kappa^{-1}(n)-1)}(\kappa(\kappa^{-1}(n))-\kappa(\kappa^{-1}(n)-1)) \\ &= w(\kappa^{-1}(n)-1)(q(\kappa^{-1}(n))-1) \end{aligned}$$

assuming  $\kappa(\kappa^{-1}(n)) \nmid n$  (if  $\kappa(\kappa^{-1}(n)) \mid n$  the term is zero and we are done). We need only take  $v = \kappa^{-1}(n)$  and point out that  $q(v) \rightarrow \infty$  as  $v \rightarrow \infty$ ; hence

$$w(v-1)(q(v)-1) \in o\left(\frac{1}{q(v)}\sum_{j=0}^{v-1} w(j)(q(j+1)-1)\right) \subset o\left(\sum_{j=0}^{v-1} w(j)(q(j+1)-1)\right)$$

and we have proved the following theorem.

**Theorem 4.7.2** *Let  $w$  be a weight function such that  $w(j) > 0$  and let  $q(j) \rightarrow \infty$  as  $j \rightarrow \infty$ . If*

$$w(v-1)(q(v)-1) \in o\left(\frac{1}{q(v)}\sum_{j=0}^{v-1} w(j)(q(j+1)-1)\right)$$

then

$$\frac{1}{n}\sum_{m=1}^{n-1} v(m) \sim \frac{1}{2}\sum_{j=0}^{\kappa(j)<n} w(j)(q(j+1)-1), \quad n \rightarrow \infty.$$

For example, the combination  $w(j) = (j+1)^{-\alpha}$ ,  $\alpha \leq -1$  and  $q(j) = j+1$  fulfills the conditions of the theorem.

#### 4.8 Notes

The survey [KPT85] defines the general digital sum problem for Cantor representations of integers and includes an extensive bibliography; the introduction to this chapter is modelled on [KPT85, p. 55-56]; I also consulted [KT84]. Theorem 4.2.1 can be found on [KPT85, p. 63-64]. Theorem 4.7.1 is similar to a result on [KPT85, p. 56], which is obtained with real-variable, Delange-type methods and contains a different form of the error term. A different proof of Theorem 4.7.2 is given on [KPT85, p. 58].

A more up-to-date introduction can be found on [FGK<sup>+</sup>94, p. 292-295]; the proof of Delange's theorem concerning binary digital sums is on [FGK<sup>+</sup>94, p. 297-299]; digital sums with exponential weights are treated on [FGK<sup>+</sup>94, p. 303-304].

## Chapter 5

### Counting sums of three squares

This chapter presents a new proof of a result due to Osbaldestin and Shiu concerning integers representable as sums of three squares. Their papers ([Shi88], [OS89]) use real-variable methods of the Delange type; we will use the Mellin-Perron formula for  $m = 1$ .

#### 5.1 Preliminaries

**Definition 5.1.1** Let  $Q$  be the set of integers  $n \in \mathbb{Z}^+$  representable as sums of three squares including 0.

**Lemma 5.1.1** If  $n = 4^l(8k + 7)$ , where  $l, k \in \mathbb{Z}^+$ , then  $n$  is not representable as a sum of three squares; i.e.  $n \in \bar{Q}$ .

**Proof.** Note that 0, 1, 4 are the only quadratic residues modulo 8. Hence  $x_1^2 + x_2^2 + x_3^2 \not\equiv 7(8)$ . Suppose  $4^l(8k + 7)$  cannot be represented as a sum of three squares and  $4^{l+1}(8k + 7)$  can, i.e.  $4^{l+1}(8k + 7) = x_1^2 + x_2^2 + x_3^2$ . This implies  $4^l(8k + 7) = (x_1/2)^2 + (x_2/2)^2 + (x_3/2)^2$ , a contradiction. (Note that  $x_1^2 + x_2^2 + x_3^2 \equiv 0 \pmod{4}$  implies  $x_{1,2,3} \equiv 0 \pmod{2}$ .) ■

In fact all  $n$  not of the form  $4^l(8k + 7)$  are representable as a sum of three squares. The following theorem is due to Gauss.

**Theorem 5.1.1** A positive integer  $n$  is representable as the sum of three squares if and only if there do not exist  $k, l \in \mathbb{Z}^+$  such that  $n = 4^l(8k + 7)$ .

We let  $k(n)$  be the characteristic function of  $\bar{Q}$ , i.e.

$$k(n) = \begin{cases} 1 & \text{if } n \in \bar{Q}, \\ 0 & \text{if } n \in Q \end{cases}$$

or equivalently

$$k(n) = \begin{cases} 1 & \text{if } n = 4^l(8k + 7), \text{ where } l, k \in \mathbb{Z}^+ \\ 0 & \text{otherwise.} \end{cases}$$

Note that  $8k + 7 = 4(2k + 1) + 3$ . This shows that  $k(n)$  is the characteristic function of those integers whose base-four representation ends in a 1 or a 3, followed by a 3, followed by a possibly empty string of zeros. Let

$$Q(N) = \sum_{n \in Q, 0 < n \leq N} 1 = N - \sum_{0 < n \leq N} k(n).$$

Define  $\Delta(N)$  as follows:

$$Q(N) = \frac{5}{6}N + \Delta(N),$$

i.e.

$$\Delta(N) = \frac{1}{6}N - \sum_{0 < n \leq N} k(n)$$

and let  $\Delta(0) = 0$ . Osbaldestin and Shiu consider the average order of  $\Delta(N)$ , which is given by

$$\begin{aligned} \frac{1}{N} \sum_{0 \leq n < N} \Delta(n) &= \frac{1}{N} \sum_{n=1}^{N-1} \left( \frac{1}{6}n - \sum_{0 < l \leq n} k(l) \right) \\ &= \frac{1}{N} \frac{1}{6} \frac{1}{2} (N-1)N - \frac{1}{N} \sum_{n=1}^{N-1} \sum_{l=1}^n k(l) = \frac{1}{12}N - \frac{1}{12} - \frac{1}{N} \sum_{n=1}^{N-1} \sum_{l=1}^n k(l). \end{aligned}$$

They prove the following theorem.

**Theorem 5.1.2** [OS89] *There exists a periodic function  $F(u)$  with period 1 such that for  $N \geq 1$ ,*

$$\frac{1}{N} \sum_{0 \leq n < N} \Delta(n) = \frac{3}{8}L + F(L) + \frac{\delta(N)}{N}$$

where

$$L = \frac{\log N}{\log 4} \text{ and } \delta(N) = \begin{cases} \frac{1}{8} & N \text{ odd,} \\ 0 & N \text{ even.} \end{cases}$$

The function  $F(u)$  is a Fourier series

$$\sum_{k \in \mathbb{Z}} c_k e^{2\pi i k u}$$

with coefficients

$$c_0 = -\frac{31}{48} - \frac{3}{8 \log 4} - \frac{1}{\log 4} (\zeta'(0, 7/16) + \zeta'(0, 15/16))$$

and

$$c_k = -\frac{1}{2\pi i k} \left( 1 + \frac{2\pi i k}{\log 4} \right)^{-1} \left( \zeta \left( \frac{2\pi i k}{\log 4}, \frac{7}{16} \right) + \zeta \left( \frac{2\pi i k}{\log 4}, \frac{15}{16} \right) \right), k \neq 0.$$

We present a new proof of this theorem in the remainder of this chapter.

## 5.2 Application of the Mellin-Perron formula

The closed form of the Dirichlet generating function  $K(s) = \sum k(n)/n^s$  of  $k(n)$  is obtained as follows.

$$\begin{aligned} K(s) &= \sum \frac{k(n)}{n^s} = \sum \frac{k(4n)}{(4n)^s} + \sum_{n \geq 0} \frac{k(4n+3)}{(4n+3)^s} \\ &= \frac{1}{4^s} \sum \frac{k(n)}{n^s} + \sum_{n \geq 0} \frac{k(16n+(13)_4)}{(16n+(13)_4)^s} + \sum_{n \geq 0} \frac{k(16n+(33)_4)}{(16n+(33)_4)^s} \\ &= \frac{1}{4^s} K(s) + \frac{1}{16^s} \sum_{n \geq 0} \frac{1}{(n+7/16)^s} + \frac{1}{16^s} \sum_{n \geq 0} \frac{1}{(n+15/16)^s} \end{aligned}$$

We conclude that

$$K(s) = \frac{4^s}{4^s - 1} \frac{1}{16^s} (\zeta(s, 7/16) + \zeta(s, 15/16)) = \frac{1}{4^s - 1} \frac{1}{4^s} (\zeta(s, 7/16) + \zeta(s, 15/16)).$$

Let

$$L(s) = K(s) \frac{N^s}{s(s+1)}.$$

The Mellin-Perron formula (Lemma 2.9.3) tells us that

$$\frac{1}{N} \sum_{n=1}^{N-1} \sum_{l=1}^n k(l) = \frac{1}{N} \frac{N}{2\pi i} \int_{3/2-i\infty}^{3/2+i\infty} L(s) ds = \frac{1}{2\pi i} \int_{3/2-i\infty}^{3/2+i\infty} L(s) ds.$$

We evaluate this integral by means of the shifting lemma (Lemma 2.8.1), taking

$$\Phi(s) = \frac{1}{4^s - 1} \frac{N^s}{4^s}$$

and

$$T_j = \frac{(2j+1)\pi}{\log 4}.$$

Note that

$$|\Phi(s)| = \left(\frac{N}{4}\right)^\sigma \frac{1}{|4^s - 1|}.$$

Along the vertical segments situated at  $\alpha$  and  $c$

$$|\Phi(s)| < \left(\frac{N}{4}\right)^\alpha \frac{1}{1-4^\alpha} = M_\alpha \quad \text{and} \quad |\Phi(s)| < \left(\frac{N}{4}\right)^c \frac{1}{4^c - 1} = M_c$$

respectively. Along the horizontal segments at  $\pm iT_j$

$$|\Phi(s)| < \left(\frac{N}{4}\right)^c \frac{1}{|-4^\sigma - 1|} < \left(\frac{N}{4}\right)^c \frac{1}{1+4^\alpha} = M_T.$$

Therefore the bound independent of  $\sigma, t$  on  $\Phi(s)$  that we require in order to apply the lemma is  $M = \max\{M_c, M_\alpha, M_T\}$ . We have

$$\begin{aligned} \frac{1}{2\pi i} \int_{3/2-i\infty}^{3/2+i\infty} L(s) ds &= \frac{1}{2\pi i} \int_{-1/4-i\infty}^{-1/4+i\infty} L(s) ds \\ &+ \operatorname{Res}(L(s); s=1) + \operatorname{Res}(L(s); s=0) + \sum_{k \in \mathbb{Z} \setminus \{0\}} \operatorname{Res}\left(L(s); s = \frac{2\pi i k}{\log 4}\right). \end{aligned}$$

Let  $\zeta_0 = \zeta(0, 7/16) + \zeta(0, 15/16) = 1/2 - 7/16 + 1/2 - 15/16 = -3/8$  and let  $\zeta_1 = \zeta'(0, 7/16) + \zeta'(0, 15/16)$ . Finally, let  $\chi_k = 2\pi i k / \log 4$  when  $k \neq 0$ .

$$\begin{aligned} \operatorname{Res}(L(s); s=0) &= \lim_{s \rightarrow 0} \left( K(s) \frac{N^s}{s(s+1)} s^2 \right)' = \lim_{s \rightarrow 0} \left( sK(s) \frac{N^s}{s+1} \right)' \\ &= \lim_{s \rightarrow 0} sK(s) (\log N N^s (s+1)^{-1} - N^s (s+1)^{-2}) + \lim_{s \rightarrow 0} (sK(s))' \frac{N^s}{s+1} \\ &= (\log N - 1) \lim_{s \rightarrow 0} \frac{s}{4^s - 1} \frac{1}{4^s} (\zeta(s, 7/16) + \zeta(s, 15/16)) + \\ &\quad \lim_{s \rightarrow 0} \frac{s}{4^s - 1} \frac{1}{4^s} (\zeta'(s, 7/16) + \zeta'(s, 15/16)) - \\ &\quad \lim_{s \rightarrow 0} \frac{s}{4^s - 1} \frac{\log 4}{4^s} (\zeta(s, 7/16) + \zeta(s, 15/16)) + \\ &\quad \lim_{s \rightarrow 0} \frac{4^s - 1 - \log 4 s 4^s}{(4^s - 1)^2} \frac{1}{4^s} (\zeta(s, 7/16) + \zeta(s, 15/16)) \\ &= (\log n - 1) \zeta_0 \lim_{s \rightarrow 0} \frac{1}{\log 4 4^s} + \\ &\quad \zeta_1 \lim_{s \rightarrow 0} \frac{1}{\log 4 4^s} - \log 4 \zeta_0 \lim_{s \rightarrow 0} \frac{1}{\log 4 4^s} - \zeta_0 \lim_{s \rightarrow 0} \frac{\log 4 \log 4 s 4^s}{2(4^s - 1) \log 4 4^s} \\ &= (\log_4 N - 1/\log 4 - 1) \zeta_0 + \frac{1}{\log 4} \zeta_1 - \zeta_0 \lim_{s \rightarrow 0} \frac{\log 4 s}{2(4^s - 1)} \\ &= (\log_4 N - 1/\log 4 - 1) \zeta_0 + \frac{1}{\log 4} \zeta_1 - \zeta_0 \lim_{s \rightarrow 0} \frac{1}{2 4^s} \\ &= -\frac{3}{8} (\log_4 N - 1/\log 4 - 3/2) + \frac{1}{\log 4} \zeta_1 \\ \operatorname{Res}(L(s); s=1) &= \lim_{s \rightarrow 1} (s-1) K(s) \frac{N^s}{s(s+1)} = \frac{N}{2} \lim_{s \rightarrow 1} \frac{s}{4^s - 1} \frac{1}{4^s} (s-1) (\zeta(s, 7/16) + \zeta(s, 15/16)) \\ &= \frac{N}{2} \frac{1}{12} 2 = \frac{N}{12} \\ \operatorname{Res}\left(L(s); s = \frac{2\pi i k}{\log 4}\right) &= \lim_{s \rightarrow \chi_k} (s - \chi_k) K(s) \frac{N^s}{s(s+1)} \\ &= \frac{e^{2\pi i k \log_4 N}}{\chi_k (\chi_k + 1)} \lim_{s \rightarrow \chi_k} \frac{s - \chi_k}{4^s - 1} \frac{1}{4^s} ((\zeta(s, 7/16) + \zeta(s, 15/16))) \\ &= \frac{e^{2\pi i k \log_4 N}}{\chi_k (\chi_k + 1)} ((\zeta(\chi_k, 7/16) + \zeta(\chi_k, 15/16))) \lim_{s \rightarrow \chi_k} \frac{1}{\log 4 4^s} \\ &= \frac{1}{\log 4} \frac{e^{2\pi i k \log_4 N}}{\chi_k (\chi_k + 1)} ((\zeta(\chi_k, 7/16) + \zeta(\chi_k, 15/16))) \end{aligned}$$

We note that

$$\sum_{k \in \mathbb{Z} \setminus \{0\}} \operatorname{Res} \left( L(s); s = \frac{2\pi ik}{\log 4} \right) = -(F(\log_4 N) - c_0).$$

In order to evaluate

$$\frac{1}{2\pi i} \int_{-1/4-i\infty}^{-1/4+i\infty} L(s) ds = \frac{1}{2\pi i} \int_{-1/4-i\infty}^{-1/4+i\infty} \frac{1}{4^s - 1} \frac{1}{4^s} (\zeta(s, 7/16) + \zeta(s, 15/16)) \frac{N^s}{s(s+1)} ds$$

we note that we may use the expansion

$$\frac{1}{4^s - 1} \frac{1}{4^s} = - \sum_{k=0}^{\infty} \frac{4^{ks}}{4^s} = -\frac{1}{4^s} - \sum_{k=1}^{\infty} 4^{ks}$$

since  $\sigma = -1/4$ . By Corollary 2.10.1 all integrals of the form

$$\frac{1}{2\pi i} \int_{-1/4-i\infty}^{-1/4+i\infty} (\zeta(s, 7/16) + \zeta(s, 15/16)) \frac{(4^k n)^s}{s(s+1)} ds$$

are zero. (The integer  $4^k N$  takes the place of the integer  $n$ .) We apply Corollary 2.10.2 in the manner shown in the associated example and obtain

$$\frac{1}{2\pi i} \int_{-1/4-i\infty}^{-1/4+i\infty} L(s) ds = -\frac{1}{2\pi i} \int_{-1/4-i\infty}^{-1/4+i\infty} \frac{1}{4^s} (\zeta(s, 7/16) + \zeta(s, 15/16)) \frac{N^s}{s(s+1)} ds = -\frac{\delta(N)}{N}.$$

We are now able to compute the average order of  $\Delta(n)$  and complete the proof of the Osbaldestin-Shiu result.

$$\begin{aligned} \frac{1}{N} \sum_{0 \leq n < N} \Delta(n) &= \frac{1}{12} N - \frac{1}{12} - \frac{1}{N} \sum_{n=1}^{N-1} \sum_{l=1}^n k(l) = \frac{1}{12} N - \frac{1}{12} - \frac{1}{2\pi i} \int_{3/2-i\infty}^{3/2+i\infty} L(s) ds \\ &= \frac{1}{12} N - \frac{1}{12} \\ &\quad - \left( -\frac{\delta(N)}{N} - \frac{3}{8} (\log_4 N - 1/\log 4 - 3/2) + \frac{1}{\log 4} \zeta_1 + \frac{N}{12} - (F(\log_4 N) - c_0) \right) \\ &= \frac{3}{8} \log_4 N + F(\log_4 N) + \frac{\delta(N)}{N} - c_0 - \frac{1}{12} - \frac{27}{48} - \frac{3}{8 \log 4} - \frac{1}{\log 4} \zeta_1 \\ &= \frac{3}{8} \log_4 N + F(\log_4 N) + \frac{\delta(N)}{N} \end{aligned}$$

This concludes the proof.  $\blacksquare$

### 5.3 Notes

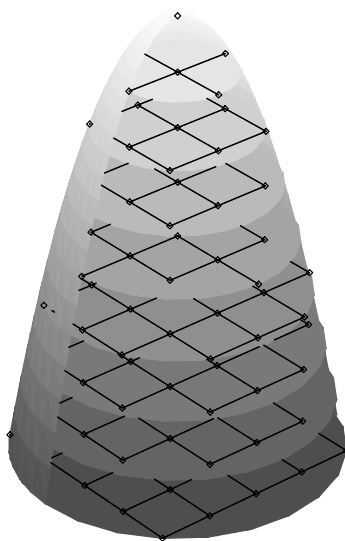
The introduction is modelled on [Kra81, p. 162-163]. The Fourier series expansion is developed on [OS89, p. 373-374].

## Chapter 6

### A paraboloid and the lattice points that it contains

This chapter treats a counting problem related to the number of lattice points inside a certain paraboloid  $P_n$ . The instance  $P_8$  is shown below. The paraboloid  $P_n$  is the surface

$$\{(x, y, z) \mid x^2 + y^2 = n - z, z \in [0, n]\} \quad \text{or} \quad \{(r \cos \theta, r \sin \theta, n - r^2) \mid 0 \leq \theta \leq 2\pi, 0 \leq r \leq \sqrt{n}\}.$$



The paraboloid  $P_8 = \{(x, y, z) \mid x^2 + y^2 = 8 - z, z \in [0, 8]\}$  and the lattice points inside/on  $P_8$ .

We are concerned with the number of lattice points, i.e. integer points *inside* and *on*  $P_n$ . The figure shows the lattice points at each integer height  $z = 0, 1, \dots, n$ . We will count these lattice points according to the following weight function:

$$w(a, b, c) = \begin{cases} 1 & \text{if } a^2 + b^2 < n - c \\ \frac{1}{2} & \text{if } a^2 + b^2 = n - c; \end{cases}$$

where  $a, b \in \mathbb{Z}$  and  $c \in \mathbb{N}$ ; i.e. a lattice point has weight 1 if it is located inside  $P_n$  and weight  $1/2$  if it lies on  $P_n$ . We define the the sum  $V_n$  as

$$V_n = \sum_{a^2+b^2 < n-c, a, b \in \mathbb{Z}, c \in \mathbb{N}} 1 + \sum_{a^2+b^2 = n-c, a, b \in \mathbb{Z}, c \in \mathbb{N}} \frac{1}{2}.$$

We will examine the average order of  $V_n$ . This is the measure

$$\frac{1}{N} \sum_{n=0}^{N-1} V_n.$$

### 6.1 What to look for

By definition the dominant term of  $V_n$  should be of the same order as the volume of  $P_n$ . The volume of  $P_n$  is easily seen to be

$$\frac{1}{2} \pi n^2,$$

by the following argument. We may say informally that the volume  $V$  of a paraboloid of height  $h$  is obtained by integrating a disk of radius  $\sqrt{z}$  from zero to  $h$  along the  $z$  axis, i.e.

$$V = \int_0^h \pi \sqrt{z}^2 dz = \frac{1}{2} \pi h^2.$$

Alternatively we can use the appropriate formula from basic calculus, i.e.

$$V = \int_0^{2\pi} \int_0^{\sqrt{h}} \int_{r^2}^h r dz dr d\theta = 2\pi \int_0^{\sqrt{h}} hr - r^3 dr = 2\pi \left( \frac{1}{2} h \sqrt{h}^2 - \frac{1}{4} \sqrt{h}^4 \right) = \frac{1}{2} \pi h^2.$$

It follows that the average order of  $V_n$  will be approximately

$$\frac{\pi}{2N} \sum_{n=0}^{N-1} n^2 = \frac{\pi}{2N} \frac{1}{6} (N-1)N(2N-1) = \frac{1}{12} \pi (2N^2 - 3N + 1) = \frac{1}{6} \pi N^2 - \frac{1}{4} \pi N + \frac{1}{12} \pi,$$

at least in the higher-order terms. We will see this estimate confirmed by the end of this chapter.

### 6.2 Preliminaries

We recall the following theorem by Gauss and Jacobi.

**Theorem 6.2.1** *Let  $r_2(n)$  be the number of integer solutions  $x, y \in \mathbb{Z}$  of  $x^2 + y^2 = n$ , i.e.*

$$r_2(n) = |\{(x, y) \mid x, y \in \mathbb{Z}, x^2 + y^2 = n\}|$$

and let

$$\chi(n) = \begin{cases} 0 & n \text{ even} \\ (-1)^{\frac{n-1}{2}} & n \text{ odd.} \end{cases}$$

Then

$$r_2(n) = 4 \left( \sum_{d|n, d \equiv 1(4)} 1 - \sum_{d|n, d \equiv 3(4)} 1 \right) = 4(1 * \chi(n)).$$

**Proof.** Recall that we outlined Jacobi's proof of this theorem in the introduction. The most straightforward proof, however, uses the field  $\mathbb{Q}[i]$  of Gaussian integers. Consider  $n \in \mathbb{N}$ , with the prime factorization

$$n = i^{3l}(1+i)^{2l} \prod_{k=1}^r \pi_k^{l_k} \overline{\pi}_k^{l_k} \prod_{k=1}^s \rho_k^{m_k}.$$

Suppose  $n = x^2 + y^2 = (x+iy)(x-iy)$ . Then  $x+iy \mid n$  and

$$x+iy = i^g(1+i)^{g'} \prod_{k=1}^r \pi_k^{g_k} \overline{\pi}_k^{g'_k} \prod_{k=1}^s \rho_k^{h_k}$$

where  $0 \leq g \leq 3$ ,  $0 \leq g' \leq 2l$ ,  $0 \leq g_k, g'_k \leq l_k$  and  $0 \leq h_k \leq m_k$ . This implies

$$x-iy = i^{-g} i^{3g'} (1+i)^{g'} \prod_{k=1}^r \overline{\pi}_k^{g'_k} \pi_k^{g_k} \prod_{k=1}^s \rho_k^{h_k},$$

where we have used  $\overline{i} = 1/i$  and  $\overline{1+i} = -i(1+i) = i^3(1+i)$ . If we compare the sums of the exponents in the factorizations of  $x+iy$  and  $x-iy$  to the exponents in the factorization of  $n$ , we find that  $g' = l$ ,  $g_k + g'_k = l_k$  and  $2h_k = m_k$ . Hence  $n = x^2 + y^2$  is solvable only iff all of the  $m_k$  are even. If they are, the solutions  $(x, y)$  are given by

$$x+iy = i^g(1+i)^l \prod_{k=1}^r \pi_k^{g_k} \overline{\pi}_k^{l_k - g_k} \prod_{k=1}^s \rho_k^{m_k/2}$$

where  $0 \leq g \leq 3$  and  $0 \leq g_k \leq l_k$ , i.e. there are  $4 \prod_{k=1}^r (1+l_k)$  solutions.

We have proved that

$$r_2(n) = \begin{cases} 0 & \text{if at least one of the } m_k \text{ is odd} \\ 4 \prod_{k=1}^r (1+l_k) & \text{if all of the } m_k \text{ are even.} \end{cases}$$

We must now show that this equals

$$4 \left( \sum_{d|n, d \equiv 1(4)} 1 - \sum_{d|n, d \equiv 3(4)} 1 \right).$$

Note that  $1^2 \equiv 1(4)$  and  $3^2 \equiv 1(4)$ . Suppose  $s = 0$ . All the  $m_k$  are zero and hence even. The number of divisors  $d \equiv 1(4)$  is  $\prod_{k=1}^n (1 + l_k)$  and we are done. Now consider  $s \geq 1$ . We have

$$\sum_{d|n, d \equiv 1(4)} 1 = \prod_{k=1}^n (1 + l_k) \sum_{\mu_k \in [0, m_k], \sum \mu_k \equiv 0(2)} 1 = a_{M,s} \prod_{k=1}^n (1 + l_k)$$

and

$$\sum_{d|n, d \equiv 3(4)} 1 = \prod_{k=1}^n (1 + l_k) \sum_{\mu_k \in [0, m_k], \sum \mu_k \equiv 1(2)} 1 = b_{M,s} \prod_{k=1}^n (1 + l_k).$$

Note that

$$a_{M,1} = \left\lfloor \frac{m_1}{2} \right\rfloor + 1$$

and

$$b_{M,1} = \left\lceil \frac{m_1}{2} \right\rceil.$$

Furthermore,

$$a_{M,s+1} = a_{M,s} \left( \left\lfloor \frac{m_{s+1}}{2} \right\rfloor + 1 \right) + b_{M,s} \left\lceil \frac{m_{s+1}}{2} \right\rceil$$

and

$$b_{M,s+1} = a_{M,s} \left\lceil \frac{m_{s+1}}{2} \right\rceil + b_{M,s} \left( \left\lfloor \frac{m_{s+1}}{2} \right\rfloor + 1 \right).$$

This yields

$$a_{M,1} - b_{M,1} = \left\lfloor \frac{m_1}{2} \right\rfloor + 1 - \left\lceil \frac{m_1}{2} \right\rceil$$

and

$$a_{M,s+1} - b_{M,s+1} = (a_{M,s} - b_{M,s}) \left( \left\lfloor \frac{m_{s+1}}{2} \right\rfloor + 1 - \left\lceil \frac{m_{s+1}}{2} \right\rceil \right).$$

We conclude that

$$a_{M,s} - b_{M,s} = \prod_{k=1}^s \left( \left\lfloor \frac{m_k}{2} \right\rfloor + 1 - \left\lceil \frac{m_k}{2} \right\rceil \right).$$

Note that

$$\left\lfloor \frac{m}{2} \right\rfloor + 1 - \left\lceil \frac{m}{2} \right\rceil = \begin{cases} 1 & \text{if } m \equiv 0(2) \\ 0 & \text{if } m \equiv 1(2). \end{cases}$$

Therefore the product is 1 iff all the  $m_k$  are even and 0 otherwise.  $\blacksquare$

We use this theorem to evaluate  $V_n$ . Evidently the contribution from a disk of radius 0 is  $1/2$ , and

$$1 + \sum_{l=1}^{k-1} r_2(l) + \frac{1}{2} r_2(k)$$

for a disk of radius  $\sqrt{k}$  where  $0 < k \leq n$ . The total for  $P_n$  is

$$\frac{1}{2} + \sum_{k=1}^n \left( 1 + \sum_{l=1}^{k-1} r_2(l) + \frac{1}{2} r_2(k) \right) = \frac{1}{2} + n + \sum_{k=1}^n \sum_{l=1}^{k-1} r_2(l) + \frac{1}{2} \sum_{k=1}^n r_2(k).$$

The average order of  $V_n$  is thus given by

$$\frac{1}{N} \sum_{n=0}^{N-1} \left( \frac{1}{2} + n + \sum_{k=1}^n \sum_{l=1}^{k-1} r_2(l) + \frac{1}{2} \sum_{k=1}^n r_2(k) \right) = \frac{1}{2} N + \frac{1}{N} \sum_{n=0}^{N-1} \sum_{k=1}^n \sum_{l=1}^{k-1} r_2(l) + \frac{1}{2N} \sum_{n=0}^{N-1} \sum_{k=1}^n r_2(k).$$

Now

$$2 \sum_{n=0}^{N-1} \sum_{k=1}^n \sum_{l=1}^{k-1} r_2(l) = 2 \sum_{n=0}^{N-1} \sum_{k=1}^n (n-k) r_2(k) = 2 \sum_{n=1}^{N-1} \frac{1}{2} (N-1-n)(N-n) r_2(n)$$

and

$$\sum_{n=0}^{N-1} \sum_{k=1}^n r_2(k) = \sum_{n=1}^{N-1} (N-n) r_2(n).$$

Hence

$$\begin{aligned} \frac{1}{N} \sum_{n=0}^{N-1} V_n &= \frac{1}{2} N + \frac{1}{2N} \left( 2 \sum_{n=0}^{N-1} \sum_{k=1}^n \sum_{l=1}^{k-1} r_2(l) + \sum_{n=0}^{N-1} \sum_{k=1}^n r_2(k) \right) \\ &= \frac{1}{2} N + \frac{1}{2N} \sum_{n=1}^{N-1} (N-n)^2 r_2(n) = \frac{1}{2} N + \frac{1}{2} N \sum_{n=1}^{N-1} \left( 1 - \frac{n}{N} \right)^2 r_2(n). \end{aligned}$$

The next step is to evaluate the sum.

### 6.3 Application of the Mellin-Perron formula with $m = 2$

The Mellin-Perron formula for  $m = 2$  (Theorem 2.9.2) states that

$$\frac{1}{2} \sum_{1 \leq n < N} r_2(n) \left( 1 - \frac{n}{N} \right)^2 = \frac{1}{2\pi i} \int_{c-i\infty}^{c+i\infty} \left( \sum_{n \geq 1} \frac{r_2(n)}{n^s} \right) N^s \frac{ds}{s(s+1)(s+2)}$$

where  $c$  lies in the half-plane of convergence of  $\sum r_2(n)/n^s$ .

By Theorem 2.7.2 the Dirichlet generating function  $R(s) = \sum r_2(n)/n^s$  of  $r_2(n) = 4(1 * \chi(n))$  is given by

$$\begin{aligned} \sum \frac{r_2(n)}{n^s} &= 4\zeta(s) \sum \frac{\chi(n)}{n^s} = 4\zeta(s) \left( \sum_{k=0} \frac{1}{(4k+1)^s} - \sum_{k=0} \frac{1}{(4k+3)^s} \right) \\ &= \frac{1}{4^{s-1}} \zeta(s) \left( \zeta \left( s, \frac{1}{4} \right) - \zeta \left( s, \frac{3}{4} \right) \right). \end{aligned}$$

We will evaluate the Mellin inversion integral by the Cauchy residue theorem; therefore we must study the poles of

$$V(s) = \frac{1}{4^{s-1}} \zeta(s) \left( \zeta \left( s, \frac{1}{4} \right) - \zeta \left( s, \frac{3}{4} \right) \right) \frac{N^s}{s(s+1)(s+2)}.$$

- The pole at  $s = 1$ .

All three of  $\zeta(s)$ ,  $\zeta(s, 1/4)$  and  $\zeta(s, 3/4)$  have a simple pole with residue 1 at  $s = 1$ . So does

$$\zeta(s) \left( \zeta \left( s, \frac{1}{4} \right) - \zeta \left( s, \frac{3}{4} \right) \right),$$

because the poles at  $s = 1$  of  $\zeta(s, 1/4)$  and  $\zeta(s, 3/4)$  cancel. Hence

$$\operatorname{Res}(V(s); s = 1) = \frac{1}{6} N^4 \sum_{k=0}^{\infty} \frac{(-1)^k}{2k+1}.$$

By a basic result from calculus

$$\tan^{-1}(x) = x - \frac{1}{3}x^2 + \frac{1}{5}x^3 - \frac{1}{7}x^5 + \dots$$

for  $|x| \leq 1$  and in particular,

$$\frac{\pi}{4} = \sum_{k=0}^{\infty} \frac{(-1)^k}{2k+1}.$$

*Ergo,*

$$\operatorname{Res}(V(s); s = 1) = \frac{1}{6} \pi N.$$

- The pole at  $s = 0$ .

This pole is simple and we have

$$\operatorname{Res}(V(s); s = 0) = 4\zeta(0) \left( \zeta \left( 0, \frac{1}{4} \right) - \zeta \left( 0, \frac{3}{4} \right) \right) \frac{1}{2} = 2 \left( -\frac{1}{2} \right) \left( \frac{1}{2} - \frac{1}{4} - \frac{1}{2} + \frac{3}{4} \right) = -\frac{1}{2}.$$

- The pole at  $s = -1$ .

Note that

$$\zeta \left( -1, \frac{1}{4} \right) - \zeta \left( -1, \frac{3}{4} \right) = -\frac{1}{2} \left( B_2 \left( \frac{1}{4} \right) - B_2 \left( \frac{3}{4} \right) \right) = 0$$

by Theorem 2.6.7. The zero of this term at  $s = -1$  cancels the pole.

- The pole at  $s = -2$ .

We have

$$\zeta(-2) = -\frac{1}{3} B_3(1) = 0,$$

again by Theorem 2.6.7. The zero of  $\zeta(s)$  at  $s = -2$  cancels the pole.

It follows that

$$\frac{1}{2\pi i} \int_{c-i\infty}^{c+i\infty} V(s) ds = \frac{1}{6} \pi N - \frac{1}{2} + \frac{1}{2\pi i} \int_{\alpha-i\infty}^{\alpha+i\infty} V(s) ds$$

for all  $\alpha \in (-1, 0)$ , if we can prove that we may shift the integral, i.e. that the contribution along the two horizontal segments vanishes in the limit. This is an immediate consequence of the generalized shifting lemma (Lemma 2.8.2). We apply the lemma with  $m = 2$ ,  $a_1 = 1$ ,  $a_2 = 1/4$  and  $a_3 = 3/4$ , respectively. The function  $\Phi(s)$  is given by  $4(N/4)^s$ . Because there are no poles on the imaginary axis, we may set  $T_j = j$ . It follows that  $M = 4(N/4)^c$  is the constant required by the lemma.

We need to estimate the remainder term

$$\frac{1}{2\pi i} \int_{\alpha-i\infty}^{\alpha+i\infty} V(s) ds = \frac{2}{\pi} \left(\frac{N}{4}\right)^\alpha \int_{-\infty}^{\infty} \left(\frac{N}{4}\right)^{it} \zeta\left(s, \frac{1}{4}\right) - \zeta\left(s, \frac{3}{4}\right) \frac{dt}{s(s+1)(s+2)}$$

where we have set  $s = \alpha + it$  and  $ds = idt$ . Note that

$$\left| \int_{-\infty}^{\infty} \left(\frac{N}{4}\right)^{it} \zeta\left(s, \frac{1}{4}\right) - \zeta\left(s, \frac{3}{4}\right) \frac{dt}{s(s+1)(s+2)} \right|$$

is bounded by

$$\left( \int_{-\infty}^{\alpha} + \int_{\alpha}^{-\infty} + \int_{-\infty}^{\infty} \right) |\zeta(s)| \left( \left| \zeta\left(s, \frac{1}{4}\right) \right| + \left| \zeta\left(s, \frac{3}{4}\right) \right| \right) \frac{dt}{|s| |s+1| |s+2|}.$$

The middle integral is the integral of an analytic function along a finite curve; hence it contributes a constant  $C_1(\alpha)$ . The sum of the outer two integrals may be estimated as follows.

$$2 \int_{-\alpha}^{\infty} |\zeta(s)| \left( \left| \zeta\left(s, \frac{1}{4}\right) \right| + \left| \zeta\left(s, \frac{3}{4}\right) \right| \right) \frac{dt}{t^3} \leq 4 \int_{-\alpha}^{\infty} \frac{(t^{1/2} \log t)^2}{t^3} dt$$

(We used Theorem 2.6.8 in the last step.) This integral converges and contributes  $C_2(\alpha)$  to the remainder term.

The next theorem summarizes the above results.

**Theorem 6.3.1** *For all  $\alpha \in (-1, 0)$ , the relation*

$$\left| \frac{1}{N} \sum_{n=0}^{N-1} V_n - \frac{1}{6} \pi N^2 \right| \in \mathcal{O}(N^{1+\alpha})$$

*describes the average order*

$$\frac{1}{N} \sum_{n=0}^{N-1} V_n$$

*of the number  $V_n$  of lattice points inside and on the paraboloid*

$$P_n = \{(x, y, z) \mid x^2 + y^2 = n - z, z \in [0, n]\}$$

*where interior points have weight 1 and points on the surface of  $P_n$  have weight 1/2. The constant in the  $N^{1+\alpha}$  term depends on  $\alpha$ ,  $C = C(\alpha)$ .*

This result matches our earlier estimate, which was based on the volume of  $P_n$ .

## 6.4 Significance of the result

It is worth pointing out how our result differs from those that can be obtained with more elementary methods. There are two elementary approaches that apply to this problem. We describe them below; then we compare their results to ours.

### 6.4.1 Volume plus error in the boundary surface

In our earlier historical review, we mentioned a means of quickly discovering first approximations to lattice point sums. This is the schema “exact volume plus an error term in the order of the boundary surface”. It applies to lattice sums where the weights are *uniform*, i.e. constant no matter where in or on the surface a lattice point is located. Our measure is not uniform, but the schema still applies. To see why, note that  $V_n$  becomes uniform if we add the lattice points on  $P_n$ , where each has weight  $1/2$ . The modified sum assigns unity weights to every point. However, the error with respect to  $V_n$  is clearly on the order of the surface of  $P_n$  (these are the extra points that were added to  $V_n$ ).

What does this schema say about  $\frac{1}{N} \sum_{n=0}^{N-1} V_n$ ? We already know the volume of  $P_n$ . The surface area is obtained by substituting  $z = f(x, y) = x^2 + y^2 = r^2$  in

$$\int \int_F \sqrt{1 + f_r^2 + \frac{1}{r^2} f_\theta^2} r dr d\theta.$$

(This formula for the surface area in polar coordinates is a result from basic calculus.) Let  $A_n$  be the surface area of  $P_n$ . We thus have

$$\begin{aligned} A_n &= \int_0^{2\pi} \int_0^{\sqrt{h}} \sqrt{1 + 4r^2} r dr d\theta \\ &= 2\pi \left[ \frac{2}{3} \frac{1}{8} (1 + 4r^2)^{\frac{3}{2}} \right]_0^{\sqrt{h}} = \frac{\pi}{6} \left( (1 + 4h)^{\frac{3}{2}} - 1 \right) \end{aligned}$$

With

$$\frac{1}{N} \sum_{n=0}^{N-1} \frac{\pi}{6} \left( (1 + 4n)^{\frac{3}{2}} - 1 \right) \in \mathcal{O} \left( N^{\frac{3}{2}} \right),$$

we find

$$\left| \frac{1}{N} \sum_{n=0}^{N-1} V_n - \frac{1}{6} \pi N^2 \right| \in \mathcal{O} \left( N^{\frac{3}{2}} \right).$$

### 6.4.2 Use of estimates for $r_2(n)$

The *circle problem* is the problem of estimating  $\theta = \inf \beta$  in

$$\left| \sum_{n \leq x} r_2(n) - \pi x \right| \in \mathcal{O}(x^\beta).$$

We discuss  $\beta$  in a moment.

Recall that

$$V_n = \frac{1}{2} + n + \sum_{k=1}^n \sum_{l=1}^{k-1} r_2(l) + \frac{1}{2} \sum_{k=1}^n r_2(k).$$

With the above definition of  $\beta$ , we have

$$\left| \sum_{k=1}^n \sum_{l=1}^{k-1} r_2(l) - \sum_{k=1}^n \pi(k-1) \right| = \left| \sum_{k=1}^n \sum_{l=1}^{k-1} r_2(l) - \frac{1}{2} \pi(n-1)n \right| \in \mathcal{O} \left( \sum_{k=1}^n (k-1)^\beta \right) = \mathcal{O}(n^{\beta+1}).$$

We also have

$$\left| \frac{1}{2} \sum_{k=1}^n r_2(k) - \frac{1}{2} \pi n \right| \in \mathcal{O}(n^\beta).$$

This yields

$$\left| V_n - \frac{1}{2} \pi n^2 \right| \in \mathcal{O}(n^{\beta+1}).$$

The term  $\frac{1}{2} \pi n^2$  is the volume of  $P_n$ ; whence

$$\left| \frac{1}{N} \sum_{n=0}^{N-1} V_n - \frac{1}{6} \pi N^2 \right| \in \mathcal{O} \left( \frac{1}{N} N^{\beta+2} \right) = \mathcal{O}(N^{\beta+1}).$$

However, a classic result by Landau (1915) and a modern one by Wen-Lin Yin (1962), taken together show that

$$\frac{1}{4} \leq \beta \leq \frac{12}{37}.$$

### 6.4.3 Comparison of the two elementary methods to ours

It is now readily apparent that neither of these two methods is as good as the Mellin-Perron approach. The first produces an error in  $\mathcal{O}(N^{\frac{3}{2}})$ , off by  $\sqrt{N}$  at least from our  $N^{1+\alpha}$ ,  $\forall \alpha \in (-1, 0)$ . The second comes closer; it gives  $\mathcal{O}(N^{1+\frac{12}{37}+\epsilon})$ , for any  $\epsilon \in \mathbb{R}^+$ , but we still have  $1 + \alpha < 1 + \frac{12}{37} + \epsilon$ , and we know that even the best possible improvement of the circle problem bound will only give  $1 + \frac{1}{4} > 1 + \alpha$ .

It is also worth pointing out that the two elementary methods only tell us about the order of the error term. We *know* what the error term is:

$$\frac{1}{N} \sum_{n=0}^{N-1} V_n = \frac{1}{6} \pi N^2 + \frac{N}{2\pi i} \int_{\alpha-i\infty}^{\alpha+i\infty} \frac{1}{4^{s-1}} \zeta(s) \left( \zeta \left( s, \frac{1}{4} \right) - \zeta \left( s, \frac{3}{4} \right) \right) \frac{N^s}{s(s+1)(s+2)} ds.$$

### 6.5 Notes

An elementary proof of Theorem 6.2.1 can be found on [Kra81, p. 73, 165-167]. This proof is based on the Dirichlet approximation theorem. A second proof is given on [HST91, p. 47, 59-62], where the result is derived from Minkowski's lattice point theorem. Another version of the first proof can be found on [Fri82, p. 8-17]; the proof that we present in this chapter is substantially based on [Fri82, p. 13-15] (this author's proof that the product formula for  $r_2(n)$  is equal to the divisor difference seems to be new). For an introduction to Gaussian integers and an elementary, detailed discussion of Theorem 6.2.1, consult [Sie64, p. 417-435]. The circle problem is discussed on [Fri82, p. 47-61] and [Kra81, p. 191-192].

## Bibliography

- [Apo86] Tom M. Apostol. *Introduction to analytic number theory*. Springer-Verlag, New York, 1986.
- [BMP55a] California Institute of Technology Bateman Manuscript Project. *Higher transcendental functions. Based, in part, on notes left by Harry Bateman, and compiled by the staff of the Bateman Manuscript Project. [Director: Arthur Erdlyi. Research associates: Wilhelm Magnus, Fritz Oberhettinger, and Francesco G. Tricomi]*, volume 3. McGraw-Hill, New York, 1953-1955.
- [BMP55b] California Institute of Technology Bateman Manuscript Project. *Higher transcendental functions. Based, in part, on notes left by Harry Bateman, and compiled by the staff of the Bateman Manuscript Project. [Director: Arthur Erdlyi. Research associates: Wilhelm Magnus, Fritz Oberhettinger, and Francesco G. Tricomi]*, volume 2. McGraw-Hill, New York, 1953-1955.
- [BMP55c] California Institute of Technology Bateman Manuscript Project. *Higher transcendental functions. Based, in part, on notes left by Harry Bateman, and compiled by the staff of the Bateman Manuscript Project. [Director: Arthur Erdlyi. Research associates: Wilhelm Magnus, Fritz Oberhettinger, and Francesco G. Tricomi]*, volume 1. McGraw-Hill, New York, 1953-1955.
- [Bus40] Bush. An asymptotic formula for the average sum of the digits of integers. *Amer. Math. Monthly*, 47:154–156, 1940.
- [CDM91] R. Casas, J. Diaz, and C. Martinez. Statistics on random trees. In *Annual International Colloquium on Automata, Languages and Programming*, pages 186–203, 1991.
- [Cha40] M. C. Chakrabarti. On the limit points of a function connected with the three-square problem. *Bull. Calcutta Math. Soc.*, 32:1–6, 1940.
- [Cla82] Colin W. Clark. *Elementary mathematical analysis*. Wadsworth Publishers of Canada, Belmont, Calif., 1982.
- [Del75] Hubert Delange. Sur la fonction sommatoire de la fonction somme des chiffres. *Enseign. Math.*, 21:31–47, 1975.
- [Det84] John W. Dettman. *Applied complex variables*. Dover Publications Inc., New York, 1984.
- [DG52] Drazin and Griffith. On the decimal representation of integers. *Proc. Cambridge Phil. Soc* (4), 48:555–565, 1952.
- [FG94] Philippe Flajolet and Mordecai Golin. Mellin transforms and asymptotics: The mergesort recurrence. *Acta Informatica*, 31:673–696, 1994.
- [FGD95] Philippe Flajolet, Xavier Gourdon, and Philippe Dumas. Mellin transforms and asymptotics: Harmonic sums. *Theoretical Computer Science*, 144(1–2):3–58, June 1995.

- [FGK<sup>+</sup>94] Flajolet, Grabner, Kirschenhofer, Prodinger, and Tichy. Mellin transforms and asymptotics: Digital sums. *Theoretical Computer Science*, 123:291–314, 1994.
- [Fri82] Francois Fricker. *Einfuehrung in die Gitterpunktlehre*. Number 73 in Lehrbuecher und Monographien aus dem Gebiete der exakten Wissenschaften. Mathematische Reihe. Birkhaeuser, Basel, 1982.
- [FS93] Philippe Flajolet and Robert Sedgewick. The average case analysis of algorithms: Counting and generating functions. Research Report 1888, Institut de Recherche en Informatique et en Automatique, 1993. 116 pages.
- [HST91] Edmund Hlawka, Johannes Schoissengeier, and Rudolf Taschner. *Geometric and analytic number theory*. Universitext. Springer-Verlag, Berlin, New York, 1991.
- [Kar92] Anatolii Alekseevich Karatsuba. *The Riemann zeta-function*. De Gruyter expositions in mathematics. Walter de Gruyter, New York, 1992.
- [KPT85] P. Kirschenhofer, H. Prodinger, and R. F. Tichy. Über die ziffernsumme natürlicher zahlen und verwandte probleme. In Edmund Hlawka, editor, *Zahlentheoretische Analysis*, number 1114 in Lecture Notes in Mathematics, pages 55–65. Springer Verlag, 1985. Wiener Seminarberichte 1980-82.
- [Kra81] E. Kraetzel. *Zahlentheorie*. VEB Deutscher Verlag der Wissenschaften, Berlin, 1981.
- [KT84] P. Kirschenhofer and R. F. Tichy. On the distribution of digits in the cantor representation of integers. *Journal of Number Theory*, 18:121–134, 1984.
- [Lan08] E. Landau. Ueber die einteilung der positiven ganzen zahlen in vier klassen nach der minderstzahl der zu ihrer additiven zusammensetzung erforderlichen quadrate. *Arch. Math. Phys.* (3), 13:303–312, 1908.
- [Lan62] Edmund Landau. *Ausgewählte Abhandlungen zur Gitterpunktlehre*. Edited by Arnold Walfisz. Deutscher Verlag der Wissenschaften, Berlin, 1962.
- [Lan93] Serge Lang. *Complex Analysis*. Graduate texts in mathematics. Springer-Verlag, New York, 3rd. ed. edition, 1993.
- [Man72] Szolim Mandelbrojt. *Dirichlet series. Principles and methods*. Dordrecht Reidel, Dordrecht, Netherlands, 1972.
- [Mar87] Jerrold E. Marsden. *Basic complex analysis*. W. H. Freeman, New York, 1987.
- [OS89] A.H. Osbaldestin and P. Shiu. A correlated digital sum problem associated with sums of three squares. *Bulletin of the London Mathematical Society*, 21, 1989.
- [Shi88] P. Shiu. Counting sums of three squares. *Bulletin of the London Mathematical Society*, 20, 1988.
- [Sie64] Waclaw Sierpinski. *Elementary theory of numbers*. Number 42 in Monografie matematyczne. Polska Akademia Nauk, Warszawa, 1964.

- [Tri95] Claude Tricot. *Curves and fractal dimension / Claude Tricot; with a foreword by Michel Mendès France*. Springer-Verlag, New York, 1995.
- [Tro68] Trollope. An explicit representation for binary digital sums. *Math. Mag.*, 41:21–25, 1968.
- [WW15] E.T Whittaker and G.N. Watson. *A course of modern analysis*. Cambridge: University Press, Cambridge, 2nd ed., completely rev. edition, 1915.

## Index

- absolute value, vii
- analytic continuation
  - Dirichlet series, of, 22
  - harmonic sums, and, 22
  - power series, by, 67
    - example, 68
  - references for, 89
- analytic function
  - complete, 67
  - Dirichlet series, defined by, 72
- analytic function element, 67
- analytic version of the fundamental theorem of arithmetic, 75
- analyticity
  - complex function  $f(z)$ , of, 65
- arithmetical function
  - definition of, 71
  - multiplicative, completely multiplicative, 75
- asymptotic expansion, ix, 22
  - harmonic numbers, of, 16
- asymptotic sequence, ix
  
- Bellman, 45
- Bernoulli polynomials, 74
- binary tree, 6, 8, 59
  - decomposition of, 9
  - decomposition of set of pairs of, 11
  - external node, 9
  - internal node, 9
  - intersection, 10
    - size function, 10
  - ordinary generating functions, and, 12
  - with  $n$  internal nodes, number of, 10
- binomial coefficient, ix
- boundary point, 64
- Bush, 45, 60
  
- Cantor base, 43
- Cantor representation, 101
- cartesian product, 8
- Catalan number, 10
- Cauchy criterion, 66
- Chakrabarti, 52, 60
  
- circle problem, 128
- coefficient
  - power series, ix
- combinatorial enumeration, 59
- complex analysis
  - basics
    - references for, 89
- continuity
  - complex function  $f(z)$ , of, 65
  - uniform, 65
- contour
  - rectangular, 19
  - simple, closed, and/or smooth, 64
- convergence
  - absolute
    - sequence of complex numbers, of, 66
    - pointwise, 66
    - sequence of complex numbers, of, 66
- curve
  - simple, closed, and/or smooth, 64
  
- DeBruijn, 6
- Delange, 46, 60
- Delange's method
  - applied to sum of three squares, 54
  - principal steps, 48
- $\nabla v(n)$ 
  - evaluation diagram for, 103
- derivative
  - complex function  $f(z)$ , of, 65
- digital sum problem
  - canonical, 42
- dimension
  - curves constructed by similarities, of, 41
  - fractal
    - curves, of, 60
- diophantine equation
  - lattice point, relation to, 58
- Dirichlet convolution, 5
- Dirichlet product, 74
- Dirichlet series, 1, 3, 5
  - abscissae of convergence of, 71
  - analytic continuation, and, 67

- analyticity of, 72
- definition of, 71
- Dirichlet product, and, 75
- Mellin-Perron formula, and, 26
- references for, 90
- discontinuous factor, 2, 60
  - finite sums, and, 26
- divide-and-conquer, 1
- divisor, vii
- domain, 64
- Drazin, 45, 60
- dyadic rationals, 36
- Eisenstein, 50
- elliptic function, 59
- $\epsilon$ -neighborhood, 64
- Euler, 6
- Euler totient function, vii
- expansion
  - infinity, about, 21
  - zero, about, 21
- factorial number system, 108
- fast decrease, 22
- Flajolet, 6, 49, 60
- folklore theorem of combinatorial enumeration
  - EGF, 5
  - OGF, 4
- Fourier series
  - fractal
    - Delange's analysis of, 48
    - sum of three squares, of, as open problem, 52
  - fractal fluctuation, and, 30
  - fractal ornament, and, 38
    - interpreted, 99
- fractal, 1
  - Fourier series
    - digital sums, and, 44
    - function-theoretic meaning, 35
- fractal ornament
  - approximation at  $n = 4^r$ , 93
  - area recurrence, 38
  - Fourier series, and, 38, 99
  - Mellin-Perron formula, and, 94
  - picture of, 91
  - problem statement, 92
  - result verified, 96
  - verbal description of, 38, 92
- function
  - meromorphic, 70
- fundamental strip, 17, 22
  - example computation, 18
- fundamental theorem of arithmetic, viii
  - for Gaussian integers, viii
- G. Bouligand
  - similarity exponent, of, 41
- Gauss, 50, 121
- Gaussian integers, 122
- generating function, 1, 3, 8
  - Dirichlet, 3, 5, 124
    - $v_q(n) \bmod 2$ , of, 78
    - $v_q(n)$ , of, 79
  - divide-and-conquer recurrences, and, 32
  - harmonic sum, and, 22
  - Mellin transform, and, 14
  - poles of, 14
  - exponential, 1, 3–5, 59
  - ordinary, 1, 3–5, 8, 59
    - binary tree, and, 12
  - probabilistic, 1
  - product of, 4
  - relation to complexity classes, 6
- Griffith, 45, 60
- half-plane, 14, 17
- Hardy, 8, 50, 58
- Harmonic numbers, 14, 59
- harmonic sum, 59
- Heaviside step function, 83
- historical review
  - format, of, 36
- $H_m(x)$ 
  - definition and Mellin transform of, 83
- Hurwitz zeta function, 73
- integer point, 6, 7
- integers, vii
  - Gaussian, vii
    - properties of, viii
- interior point, 64
- Jacobi, 6, 8, 50, 121
- $\kappa^{-1}(n)$

- definition of, 77
- $\kappa(n)$ 
  - definition of, 76
- Kirschenhofer, 60
- Knuth, 6
- Koch curve, 92
  - definition of, 41
- Koch curve, alternating
  - definition of, 41
- Landau, 52, 60, 128
- lattice point, 1, 6, 8
  - diophantine equation, relation to, 58
  - inside a circle, 7
  - volume-and-boundary heuristic, 59
- lattice point measure
  - $\zeta(s)$ , and, 57
  - intuitive motivation for, 57
  - not self-similar, 56
  - paraboloid, associated with, 55
  - visual interpretation of, 58
- Laurent expansion, 69
- limit
  - complex function  $f(z)$ , of, 65
  - left, right, 62
  - superior, inferior, 63
- limit point, 64
- linearity and rescaling, 22
  - Mellin transform, and, 84
- Mellin summation formula, generalized, 22
- Mellin transform, 2, 5, 59
  - analytic continuation, and, 67
  - convergence of, 17
  - definition of, 83
  - domain, 16
  - fundamental strip of, 83
  - harmonic numbers, and, 14
  - harmonic sum, of, 13, 22
  - image, 16
  - inverse, 13, 15, 18, 124
    - definition of, 85
- Mellin-Perron formula, 1, 2, 24
  - Cauchy residue theorem, and, 24
  - discontinuous factor, and, 24
  - Fourier series, and, 29
  - fractal ornament, and, 94
  - history, 24
  - Hurwitz  $\zeta$ -function integrals, and, 87
  - Karatsuba multiplication, and, 28
  - Landau's use of, 24
  - Mellin inversion, and, 24
  - proof by Mellin inversion, 86
  - references for, 90
  - singularities, and, 27
  - solution expansion, and, 24
  - statement of, 85
  - sums of three squares, and, 117
  - use when  $m = 1$ , explained, 86
- mergesort, 6
  - complexity
    - best, average and worst, 33
  - divide-and-conquer, as, 30
  - problem definition, 30
  - recurrence, 6, 8, 29, 30, 60
    - best-case of, and binary digital sums, 34
- meromorphic function
  - example of, 23
- Minkowski, 129
- Minkowski sausage, 40
- Mirsky, 45
- modulus, vii
- multiplicative self-similarity
  - definition of, 30
  - digital sums, and, 43
  - Dirichlet series, and, 35
  - Fourier series, implies, 33
  - sequence associated with  $r_3(n)$ , and, 50
  - substitution, generation by, 35
- natural boundary
  - Dirichlet series, of, associated with factorial number system, 108
- non-divisor, vii
- norm, vii
- numbers
  - complex, vii
  - natural, vii
  - real, vii
- Odlyzko, 6
- open strip, 83
- paraboloid

- equation of
  - rectangular coordinates, 120
  - spherical coordinates, 120
- interior lattice points, and
  - picture of, 120
- lattice point
  - Mellin-Perron formula, and, 124
- verbal description of, 55
- volume, 121
- $\phi(n)$ , vii
- plane curve
  - dimension of, 40
  - example computation, 41
- point set, 63
  - closed, 64
  - open, 64
- pole, 5, 14
  - definition of, 70
  - order of, 70
  - simple, 20, 56, 95, 125
    - definition of, 70
- power series, 3
- prime factorization
  - digital sums, and, 43
- principal part, 70
- probability distribution
  - uniform, 10
- Pythagoras, 58
- $q$ -ary representation
  - definition of, and example, 100
- $r_2(n)$ , 7, 56, 58
  - evaluation by factorization of Gaussian integers, 122
  - evaluation by theta function identity, 8
  - Gauss-Jacobi theorem on, 121
- $r_3(n)$ , 50
- Ramanujan, 8, 50, 58
- random channel networks, 1
- recurrence
  - divide-and-conquer
    - general solution of, 31
    - integral for, 32
    - translation of poles into Fourier series, 33
- region, 64
- register allocation strategy, 1
- relatively prime, vii
- residue, 14
  - compute it, how, 70
- Riemann, 6
- Riemann surface, 68
- Riemann zeta function, 73
  - references for, 90
- Riemann's functional equation for  $\zeta(s)$ , 73
- $r_k(n)$ , 58
- self-similar curves
  - definition of, 40
- sequence
  - ascending, of positive reals, 63
- sequence of analytic functions
  - analyticity of, 66
- set, ix
  - complement of, ix
- Shapiro, 45
- shifting lemma
  - application to digital sums, 103
  - application to lattice point integral, 126
  - generalized, 81
  - statement of, 80
- Shiu, 53
- similarity
  - planar
    - definition of, 39
- similarity exponent
  - G. Bouligand, of, 41
- simplicity criterion, 40
- $\text{Sing}(f(z))$ 
  - definition of, 70
- singularity, 69
  - essential, 70
  - removable, 70
  - transform function, of, 13
- size function, 3, 4, 8
- slow increase, 22
- smoothness
  - tangent, and, 64
- sorting networks, 1
- sum
  - binary digital, 42
  - digital, 1
    - alternating, 43
    - alternating, definition of, 102

- alternating, Fourier series for, 104
- alternating, fractal Fourier series and, 44
- alternating, represented by generator, 44
- asymptotically dominant term of, 110
- asymptotically dominant term, theorem on, 114
- basic approximations, 45
- Cantor base, and, 43
- Delange's method, 46
- dominant term, informal evaluation of, 45
- error term for sum of three squares, for, 53
- evaluation paradigm, 106
- factorial number system, in, 108
- Flajolet's contribution to evaluation of, 49
- general formula, 113
- general problem, 110
- $\kappa(j+1)/\kappa(j)$  periodic, 107
- period weights, with, 104
- prime factorization, and, 43
- references for, 114
- three phases of research into, 45
- weight function, and, 43
- finite
  - discontinuous factor, and, 26
- first  $n$  integers, of, ix
- first  $n$  squares, of, ix
- geometric progression, of, ix
- harmonic, 1, 5, 13, 21
  - amplitudes of, 13, 22
  - base function of, 13, 22
  - Dirichlet generating function, and, 22
  - evaluation paradigm, 16
  - expansion of by Mellin inversion, 19
  - frequencies of, 13, 22
  - harmonic numbers, of, 14
  - Mellin transform, and, 13
- no index variable, with, ix
- pyramidal, 58
- residues, of, 20
- three squares, of, 1
  - digital sum, as, 53
  - error term, 50
  - Fourier series for, 54
  - recurrence for, 52
  - references for, 119
- three squares, of, Chakrabarti's paper re., 52
- three squares, of, Landau's paper re., 52
- surface area in polar coordinates, 127
- theorem
  - binomial, ix
  - Cauchy residue, 14, 15, 19, 124
    - fractal ornament, applied to, 94
    - Mellin-Perron formula, and, 24
    - statement of, 71
  - Delange's, on digital sums, 46
  - Fabry-Pólya's, 72
  - fundamental, of arithmetic
    - analytic version, 75
  - Gauss's, on integers representable as sums of three squares, 115
  - Gauss-Jacobi, 50, 52, 61
  - Mandelbrojt's gap, 72
  - Mapping, 21, 22, 60
  - Minkowski's lattice point, 58, 61, 129
  - Osbaldestin and Shiu's, on  $\Delta(N)$ , 116
  - Whittaker-Watson's, on  $\zeta(s)$ 
    - application to paraboloid, 126
    - statement of, 74
- theta function, 7, 8, 59
- Tichy, 60
- \*, 74
- Tricot, 60
- Trollope, 45, 60
- (2,3)-number system, 101
- vertical strip, 16
- $v_\kappa(n)$ 
  - definition of, 101
- $v(n)$ 
  - digital sums, and, 43
- volume
  - $k$ -ball, of, 59
- $v_q(n)$ 
  - definition of, 77
- Weierstrass  $M$ -test, 66
- weight function
  - definition of, and example, 101
  - lattice points, for, 120
- winding number, 71
- Yin, Wen-Lin, 128
- $\zeta(s)$

- analytic continuation to all of  $\mathbb{C}$ , of, 73
  - definition of, 73
  - Euler product of, 76
  - lattice point measure, and, 57
  - pole at  $s = 1$  of, 73
  - Riemann's functional equation for, 73
  - special values at  $\mathbb{Z}^- \cup \{0\}$ , 74
  - special values at  $s = 0$ , 74
- $[z^n]$ , ix