Preface

The world is continuous, but the mind is discrete.

David Mumford

We seek to bridge some critical gaps between various fields of mathematics by studying the interplay between the continuous volume and the discrete volume of polytopes. Examples of polytopes in three dimensions include crystals, boxes, tetrahedra, and any convex object whose faces are all flat. It is amusing to see how many problems in combinatorics, number theory, and many other mathematical areas can be recast in the language of polytopes that exist in some Euclidean space. Conversely, the versatile structure of polytopes gives us number-theoretic and combinatorial information that flows naturally from their geometry.

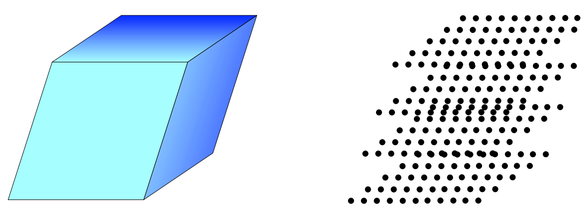

Fig. 0.1. Continuous and discrete volume.

The *discrete volume* of a body $\mathcal{P}$ can be described intuitively as the number of grid points that lie inside $\mathcal{P}$, given a fixed grid in Euclidean space. The *continuous volume* of $\mathcal{P}$ has the usual intuitive meaning of volume that we attach to everyday objects we see in the real world.

Indeed, the difference between the two realizations of volume can be thought of in physical terms as follows. On the one hand, the quantum-level grid imposed by the molecular structure of reality gives us a discrete notion of space and hence discrete volume. On the other hand, the Newtonian notion of continuous space gives us the continuous volume. We see things continuously at the Newtonian level, but in practice we often compute things discretely at the quantum level. Mathematically, the grid we impose in space—corresponding to the grid formed by the atoms that make up an object—helps us compute the usual continuous volume in very surprising and charming ways, as we shall discover.

In order to see the continuous/discrete interplay come to life among the three fields of combinatorics, number theory, and geometry, we begin our focus with the simple-to-state *coin-exchange problem* of Frobenius. The beauty of this concrete problem is that it is easy to grasp, it provides a useful computational tool, and yet it has most of the ingredients of the deeper theories that are developed here.

In the first chapter, we give detailed formulas that arise naturally from the Frobenius coin-exchange problem in order to demonstrate the interconnections between the three fields mentioned above. The coin-exchange problem provides a scaffold for identifying the connections between these fields. In the ensuing chapters we shed this scaffolding and focus on the interconnections themselves:

(1) Enumeration of integer points in polyhedra—combinatorics,
(2) Dedekind sums and finite Fourier series—number theory,
(3) Polygons and polytopes—geometry.

We place a strong emphasis on computational techniques, and on computing volumes by counting integer points using various old and new ideas. Thus, the formulas we get should not only be pretty (which they are!) but should also allow us to efficiently compute volumes by using some nice functions. In the very rare instances of mathematical exposition when we have a formulation that is both "easy to write" and "quickly computable," we have found a mathematical nugget. We have endeavored to fill this book with such mathematical nuggets.

Much of the material in this book is developed by the reader in the more than 200 exercises. Most chapters contain warm-up exercises that do not depend on the material in the chapter and can be assigned before the chapter is read. Some exercises are central, in the sense that current or later material depends on them. Those exercises are marked with ♣, and we give detailed hints for them at the end of the book. Most chapters also contain lists of open research problems.

It turns out that even a fifth grader can write an interesting paper on integer-point enumeration [144], while the subject lends itself to deep investigations that attract the current efforts of leading researchers. Thus, it is an area of mathematics that attracts our innocent childhood questions as well as our refined insight and deeper curiosity. The level of study is highly appropriate for a junior/senior undergraduate course in mathematics. In fact, this book is ideally suited to be used for a *capstone course*. Because the three topics outlined above lend themselves to more sophisticated exploration, our book has also been used effectively for an introductory graduate course.

To help the reader fully appreciate the scope of the connections between the continuous volume and the discrete volume, we begin the discourse in two dimensions, where we can easily draw pictures and quickly experiment. We gently introduce the functions we need in higher dimensions (Dedekind sums) by looking at the coin-exchange problem geometrically as the discrete volume of a generalized triangle, called a simplex.

The initial techniques are quite simple, essentially nothing more than expanding rational functions into partial fractions. Thus, the book is easily accessible to a student who has completed a standard college calculus and linear algebra curriculum. It would be useful to have a basic understanding of partial fraction expansions, infinite series, open and closed sets in $\mathbb{R}^d$, complex numbers (in particular, roots of unity), and modular arithmetic.

An important computational tool that is harnessed throughout the text is the *generating function* $f(x) = \sum_{m=0}^{\infty} a(m)\, x^m$, where the $a(m)$'s form any sequence of numbers that we are interested in analyzing. When the infinite sequence of numbers $a(m), m = 0, 1, 2, \dots$, is embedded into a single generating function $f(x)$, it is often true that for hitherto unforeseen reasons, we can rewrite the whole sum $f(x)$ in a surprisingly compact form. It is the rewriting of these generating functions that allows us to understand the combinatorics of the relevant sequence $a(m)$. For us, the sequence of numbers might be the number of ways to partition an integer into given coin denominations, or the number of points in an increasingly large body, and so on. Here we find yet another example of the interplay between the discrete and the continuous: we are given a *discrete* set of numbers $a(m)$, and we then carry out analysis on the generating function $f(x)$ in the *continuous* variable x.

What Is the Discrete Volume?

The physically intuitive description of the discrete volume given above rests on a sound mathematical footing as soon as we introduce the notion of a lattice. The grid is captured mathematically as the collection of all integer points in Euclidean space, namely $\mathbb{Z}^d = \{(x_1, \dots, x_d) : \text{ all } x_k \in \mathbb{Z}\}$. This discrete collection of equally spaced points is called a *lattice*. If we are given a geometric body $\mathcal{P}$, its discrete volume is simply defined as the number of lattice points inside $\mathcal{P}$, that is, the number of elements in the set $\mathbb{Z}^d \cap \mathcal{P}$.

Intuitively, if we shrink the lattice by a factor k and count the number of newly shrunken lattice points inside $\mathcal{P}$, we obtain a better approximation for the volume of $\mathcal{P}$, relative to the volume of a single cell of the shrunken lattice. It turns out that after the lattice is shrunk by an integer factor k, the number $\#\left(\mathcal{P} \cap \frac{1}{k}\mathbb{Z}^d\right)$ of shrunken lattice points inside an *integral polytope* $\mathcal{P}$

is magically a polynomial in k. This counting function $\#\left(\mathcal{P} \cap \frac{1}{k}\mathbb{Z}^d\right)$ is known as the *Ehrhart polynomial* of $\mathcal{P}$. If we kept shrinking the lattice by taking a limit, we would of course end up with the continuous volume that is given by the usual Riemannian integral definition of calculus:

$$\operatorname{vol} \mathcal{P} = \lim_{k \to \infty} \# \left(\mathcal{P} \cap \frac{1}{k} \mathbb{Z}^d \right) \frac{1}{k^d} \, .$$

However, pausing at fixed dilations of the lattice gives surprising flexibility for the computation of the volume of $\mathcal{P}$ and for the number of lattice points that are contained in $\mathcal{P}$.

Thus, when the body $\mathcal{P}$ is an integral polytope, the error terms that measure the discrepancy between the discrete volume and the usual continuous volume are quite nice; they are given by Ehrhart polynomials, and these enumeration polynomials are the content of Chapter 3.

The Fourier–Dedekind Sums Are the Building Blocks: Number Theory

Every polytope has a discrete volume that is expressible in terms of certain finite sums that are known as *Dedekind sums*. Before giving their definition, we first motivate these sums with some examples that illustrate their building-block behavior for lattice-point enumeration. To be concrete, consider for example a 1-dimensional polytope given by an interval $\mathcal{P} = [0, a]$, where a is any positive real number. It is clear that we need the greatest integer function $\lfloor x \rfloor$ to help us enumerate the lattice points in $\mathcal{P}$, and indeed the answer is $\lfloor a \rfloor + 1$.

Next, consider a 1-dimensional line segment that is sitting in the 2-dimensional plane. Let's pick our segment $\mathcal{P}$ so that it begins at the origin and ends at the lattice point (c, d). As becomes apparent after a moment's thought, the number of lattice points on this finite line segment involves an old friend, namely the greatest common divisor of c and d. The exact number of lattice points on the line segment is $\gcd(c, d) + 1$.

To unify both of these examples, consider a triangle $\mathcal{P}$ in the plane whose vertices have rational coordinates. It turns out that a certain finite sum is completely natural because it simultaneously extends both the greatest integer function and the greatest common divisor, although the latter is less obvious. An example of a Dedekind sum in two dimensions that arises naturally in the formula for the discrete volume of the rational triangle $\mathcal{P}$ is the following:

$$s(a, b) = \sum_{m=1}^{b-1} \left(\frac{m}{b} - \frac{1}{2} \right) \left(\frac{ma}{b} - \left\lfloor \frac{ma}{b} \right\rfloor - \frac{1}{2} \right) .$$

The definition makes use of the greatest integer function. Why do these sums also resemble the greatest common divisor? Luckily, the Dedekind sums satisfy a remarkable reciprocity law, quite similar to the Euclidean algorithm that computes the gcd. This reciprocity law allows the Dedekind sums to be computed in roughly $\log(b)$ steps rather than the b steps that are implied by the definition above. The reciprocity law for $s(a, b)$ lies at the heart of some amazing number theory that we treat in an elementary fashion, but that also comes from the deeper subject of modular forms and other modern tools.

We find ourselves in the fortunate position of viewing an important tip of an enormous mountain of ideas, submerged by the waters of geometry. As we delve more deeply into these waters, more and more hidden beauty unfolds for us, and the Dedekind sums are an indispensable tool that allows us to see further as the waters get deeper.

The Relevant Solids Are Polytopes: Geometry

The examples we have used, namely line segments and polygons in the plane, are special cases of polytopes in all dimensions. One way to define a polytope is to consider the *convex hull* of a finite collection of points in Euclidean space $\mathbb{R}^d$. That is, suppose someone gives us a set of points $\mathbf{v}_1, \dots, \mathbf{v}_n$ in $\mathbb{R}^d$. The polytope determined by the given points $\mathbf{v}_j$ is defined by all linear combinations $c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_n\mathbf{v}_n$, where the coefficients c_j are nonnegative real numbers that satisfy the relation $c_1 + c_2 + \cdots + c_n = 1$. This construction is called the *vertex description* of the polytope.

There is another equivalent definition, called the *hyperplane description* of the polytope. Namely, if someone hands us the linear inequalities that define a finite collection of half-spaces in $\mathbb{R}^d$, we can define the associated polytope as the simultaneous intersection of the half-spaces defined by the given inequalities.

There are some "obvious" facts about polytopes that are intuitively clear to most students but are, in fact, subtle and often nontrivial to prove from first principles. We will assume here one of these facts, namely that every polytope has both a vertex and a hyperplane description. This statement is intuitively clear, so that it can be assumed by novices without their ability to compute continuous and discrete volumes of polytopes being affected. For the reader who wants to peer into these subtleties, they are beautifully described in [89, 192]. Aside from assuming the "obvious" fact that any polytope has both a vertex and a hyperplane description, all theorems in the text are proved from first principles, with the exception of the last chapter, where we assume basic notions from complex analysis.

The text naturally flows into two parts, which we now explicate.

Part I

We have taken great care in making the content of the chapters flow seamlessly from one to the next, over the span of the first six chapters.

- Chapters 1 and 2 introduce some basic notions of generating functions, in the visually compelling context of discrete geometry, with an abundance of detailed motivating examples.

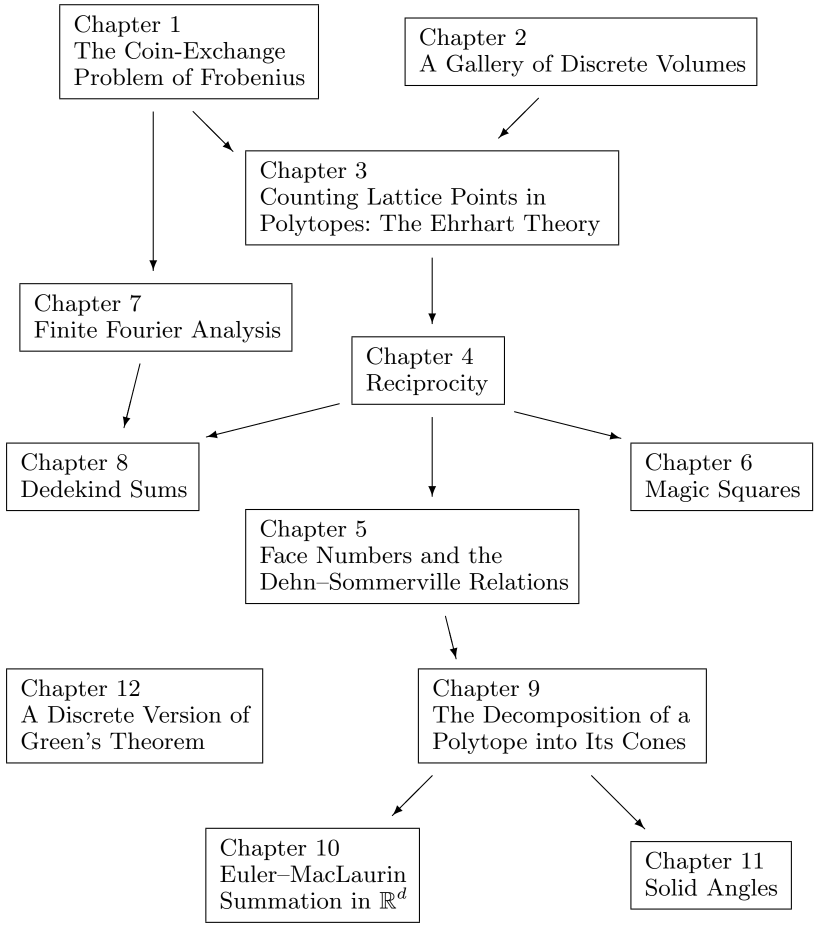

Fig. 0.2. The partially ordered set of chapter dependencies.

- Chapters 3, 4, and 5 develop the full Ehrhart theory of discrete volumes of rational polytopes.
- Chapter 6 is a "dessert" chapter, in that it enables us to use the theory developed to treat the enumeration of *magic squares*, an ancient topic that enjoys active current research.

Part II

We now begin anew.

- Having attained experience with numerous examples and results about integer polytopes, we are ready to learn about the *Dedekind sums* of Chapter 8, which form the atomic pieces of the discrete volume polynomials. On the other hand, to fully understand Dedekind sums, we need to understand *finite Fourier analysis*, which we therefore develop from first principles in Chapter 7, using only partial fractions.

- Chapter 9 answers a simple yet tricky question: how does the finite geometric series in one dimension extend to higher-dimensional polytopes? *Brion's theorem* give the elegant and decisive answer to this question.
- Chapter 10 extends the interplay between the continuous volume and the discrete volume of a polytope (already studied in detail in Part I) by introducing *Euler–Maclaurin summation* formulas in all dimensions. These formulas compare the continuous Fourier transform of a polytope to its discrete Fourier transform, yet the material is completely self-contained.
- Chapter 11 develops an exciting extension of Ehrhart theory that defines and studies the *solid angles* of a polytope; these are the natural extensions of 2-dimensional angles to higher dimensions.
- Finally, we end with another "dessert" chapter that uses complex analytic methods to find an integral formula for the discrepancy between the discrete and continuous areas enclosed by a closed curve in the plane.

Because polytopes are both theoretically useful (in triangulated manifolds, for example) and practically essential (in computer graphics, for example) we use them to link results in number theory and combinatorics. There are many research papers being written on these interconnections, even as we speak, and it is impossible to capture them all here; however, we hope that these modest beginnings will give the reader who is unfamiliar with these fields a good sense of their beauty, inexorable connectedness, and utility. We have written a gentle invitation to what we consider a gorgeous world of counting and of links between the fields of combinatorics, number theory, and geometry for the general mathematical reader.

There are a number of excellent books that have a nontrivial intersection with ours and contain material that complements the topics discussed here. We heartily recommend the monographs of Barvinok [12] (on general convexity topics), Ehrhart [80] (the historic introduction to Ehrhart theory), Ewald [81] (on connections to algebraic geometry), Hibi [95] (on the interplay of algebraic combinatorics with polytopes), Miller–Sturmfels [131] (on computational commutative algebra), and Stanley [171] (on general enumerative problems in combinatorics).

Acknowledgments

We have had the good fortune of receiving help from many gracious people in the process of writing this book. First and foremost, we thank the students of the classes in which we could try out this material, at Binghamton University (SUNY), San Francisco State University, and Temple University. We are indebted to our MSRI/Banff 2005 graduate summer school students. We give special thanks to Kristin Camenga and Kevin Woods, who ran the problem sessions for this summer school, detected numerous typos, and provided us with many interesting suggestions for this book. We are grateful for the generous support for the summer school from the Mathematical Sciences Research Institute, the Pacific Institute of Mathematics, and the Banff International Research Station.

Many colleagues supported this endeavor, and we are particularly grateful to everyone notifying us of mistakes, typos, and good suggestions: Daniel Antonetti, Alexander Barvinok, Nathanael Berglund, Andrew Beyer, Tristram Bogart, Garry Bowlin, Benjamin Braun, Robin Chapman, Yitwah Cheung, Jessica Cuomo, Dimitros Dais, Jesus De Loera, David Desario, Mike Develin, Ricardo Diaz, Michael Dobbins, Jeff Doker, Han Minh Duong, Richard Ehrenborg, David Einstein, Joseph Gubeladze, Christian Haase, Mary Halloran, Friedrich Hirzebruch, Brian Hopkins, Serkan Hoşten, Benjamin Howard, Piotr Maciak, Evgeny Materov, Asia Matthews, Peter McMullen, Martín Mereb, Ezra Miller, Mel Nathanson, Julian Pfeifle, Peter Pleasants, Jorge Ramírez Alfonsín, Bruce Reznick, Adrian Riskin, Junro Sato, Melissa Simmons, Richard Stanley, Bernd Sturmfels, Thorsten Theobald, Read Vanderbilt, Andrew Van Herick, Sven Verdoolaege, Michèle Vergne, Julie Von Bergen, Neil Weickel, Carl Woll, Zhiqiang Xu, Jon Yaggie, Ruriko Yoshida, Thomas Zaslavsky, Günter Ziegler, and two anonymous referees. We will collect corrections, updates, etc., at the Internet website

`math.sfsu.edu/beck/ccd.html`.

We are indebted to the Springer editorial staff, first and foremost to Mark Spencer, for facilitating the publication process in an always friendly and supportive spirit. We thank David Kramer for the impeccable copyediting, Frank Ganz for sharing his LaTeX expertise, and Felix Portnoy for the seamless production process.

Matthias Beck would like to express his deepest gratitude to Tendai Chitewere, for her patience, support, and unconditional love. He thanks his family for always being there for him. Sinai Robins would like to thank Michal Robins, Shani Robins, and Gabriel Robins for their relentless support and understanding during the completion of this project. We both thank all the cafés we have inhabited over the past five years for enabling us to turn their coffee into theorems.

San Francisco
Philadelphia
August 2006

Matthias Beck
Sinai Robins

Contents

Part II Beyond the Basics