TECHNISCHE UNIVERSITÄT WIEN
Vienna University of Technology

D I P L O M A R B E I T

# Lattice Path Combinatorics

Ausgeführt am Institut für
Diskrete Mathematik und Geometrie
der Technischen Universität Wien

unter der Anleitung von Univ.Prof. Dipl.-Ing. Dr. Michael Drmota

durch

Michael Wallner
Siccardsburggasse 40/120
1100 Wien

Datum

Unterschrift

D I P L O M A  T H E S I S

# Lattice Path Combinatorics

*Author:*
Michael Wallner

*Supervisor:*
Univ.-Prof. Dipl.-Ing. Dr. Michael Drmota

**Abstract**

This thesis focuses on three big topics of lattice path theory: Directed lattice paths with focus on applications of the kernel method on the Euclidean lattice, walks confined to the quarter plane with focus on the model of small steps also on the Euclidean lattice and self-avoiding walks where the derivation of the exact value of the connective constant on the hexagonal lattice is presented. The nature of the generating functions (GFs) lies in the center of interest, namely the question concerning its rational, algebraic or holonomic (D-finite) character. The used definitions and the derived theory is put under a unified framework with the goal of giving a coherent and thorough but still deep and applied introduction to the theory of lattice paths.

Directed lattice paths possess a well understood structure, as their GF is always algebraic. This result is generalized to walks confined to the half-plane and it is shown how the kernel method can be used to derive similar results from the different view point of linear recurrence relations. The next natural generalization is the restriction to the quarter plane, where the nature of GFs gets much more complicated. For the class of walks with small steps a connection between the GF and the group of the walk is shown and a general result is derived. Fifty years ago the conjecture has been raised that the value of the connective constant on the hexagonal lattice equals $\sqrt{2+\sqrt{2}}$. The problem has been solved only recently and the given solution is an attractive example of the efficiency of interdisciplinary exchange (here combinatorics and complex analysis).

**Zusammenfassung**

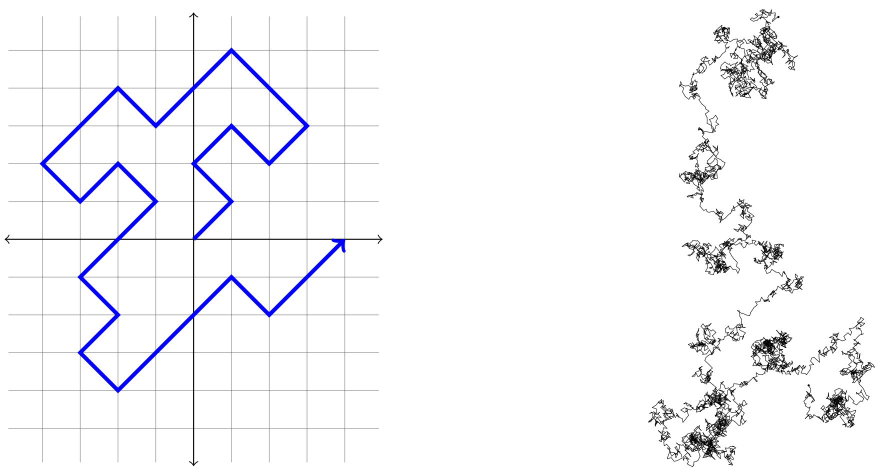
Diese Diplomarbeit befasst sich mit drei großen Gebieten der Gitterpunktpfadtheorie: Gerichteten Gitterpunktpfaden mit Fokus auf Anwendungen der „Kernel Method“ auf dem Euklidischen Gitter, Pfaden begrenzt auf den ersten Quadranten mit Fokus auf Modelle mit „kleinen Schritten“ ebenfalls auf dem Euklidischen Gitter und selbstvermeidenden Pfaden wobei die Herleitung des exakten Wertes der Gitterkonstante („connective constant“) auf dem hexagonalen Gitter präsentiert wird. Im Zentrum des Interesses liegt die Natur der Erzeugenden Funktionen (EF), im Konkreten wird die Frage behandelt, ob diese rational, algebraisch oder holonomisch sind. Die eingeführten Definitionen und die abgeleitete Theorie werden in einheitlicher Weise dargestellt, um eine verständliche und vollständige, aber dennoch tiefgehende und angewandte Einführung in obige Theorie zu geben.

Gerichtete Gitterpunktpfade stellen eine gut verstandene Klasse dar, deren EF stets algebraisch sind. Dieses Resultat wird auf Pfade verallgemeinert, welche nur auf der oberen Halbebene agieren und es wird gezeigt, wie die „Kernel Method“ verwendet werden kann, um ähnliche Resultate aus einem anderen, aber verwandten Blickwinkel der linearen Rekursionen abzuleiten. Die nächste natürliche Verallgemeinerung stellt die Einschränkung auf den ersten Quadranten dar, was in einer vielfältigeren Theorie der betreffenden EF resultiert. Für die Klasse von Pfaden mit „kleinen Schritten“ wird eine Verbindung zwischen EF und der Gruppe des Pfades gezeigt, wodurch ein allgemeines Ergebnis abgeleitet werden kann. Vor 50 Jahren wurde die Hypothese aufgestellt, dass die Gitterkonstante des hexagonalen Gitters gleich $\sqrt{2+\sqrt{2}}$ ist. Die erst vor kurzem präsentierte Lösung stellt ein ansprechendes Beispiel für den Erfolg von interdisziplinärem Austausch dar (hier Kombinatorik und Komplexe Analysis).

# Preface

The enumeration of lattice paths is a classical topic in combinatorics which is still a very active field of research. Its (and my) fascination is founded in the fact, that despite the easily understood construction of lattice paths, most of their properties remain unproven or even unknown. Figure 1 gives an intuition of this statement: In the small scale, lattice paths appear like mathematical doodles, but when looking at them a few steps further away, they show a completely different pattern. A fractal structure is visible, which gives a glimpse of the difficulties encountered in lattice path combinatorics. This justifies the richness of their applications, as they encode many combinatorial objects like trees, maps, permutations, lattice polygons, Young tableaux, queues, etc. [7].

The aim of this diploma thesis is to give a complete introduction to lattice path combinatorics by combining theory and practice, as the origins of this field lie in applied sciences like chemistry, physics and computer science. A unified framework is derived in order to present all methods and ideas as easily accessible as possible. All necessary derivations are made explicit and connections to other parts in literature are added.



(a) A Random Walk on a Euclidean Lattice

(b) A Random Walk with 5000 steps

Figure 1: Examples of two Random Walks in the Euclidean Plane

In Chapter 1 we present the basics of lattice path enumeration and answer the foremost question, about what kind of object a lattice path is at all. Moreover, a short recap of all needed concepts from discrete mathematics like formal power series is given and all necessary results from complex analysis in order to understand the subsequent chapters are stated.

Chapter 2 introduces the most important object for the following discussion: Generating Functions (GFs). Correspondingly, the symbolic method from analytic combinatorics is presented, which will deal as the standard tool to derive a functional equation on GFs. Finally, we give a characterization of the nature of GFs into rational, algebraic and holonomic functions, which proved to be very useful in this field. The derived theory is combined with classical examples of lattice path counting, like Dyck Paths or the Ballot Problem.

In Chapter 3 we investigate directed paths, which are walks with one fixed direction of increase. This theory includes walks confined to the half-space and we show the connections to the theory of linear recurrences. The most important tool in this context is the kernel method. We mainly follow the presentation of [3] in the first part and [8] in the second.

The next natural class of problems discussed in Chapter 4 are walks which are constrained to lie in the intersection of two rational half-spaces, where we choose the quarter plane (i.e. first quadrant). In particular the nature of the GFs for the class of walks with small steps is derived. This chapter is a nice example of how interdisciplinary work (algebra, analytic combinatorics and complex analysis) is able to deal with unsolved problems. The discussion is along the lines of [7].

Chapter 5 presents the up-to-date topic of self-avoiding walks. They are the object of choice to model polymers in chemistry. At the beginning we introduce some basic properties which lay the foundation for the recent proof of the value for the connective constant on the honeycomb lattice. We also draw some unmentioned connections between the used constants in the final remark after the proof. This exposition is mainly based on [10].

Michael Wallner

# Acknowledgements

First of all I want to thank my supervisor, Dr. Michael Drmota, for introducing the fascinating fields of discrete mathematics and number theory to me and his excellent support while writing this thesis.

Furthermore I want to thank my parents Margarete and Hans Wallner for their moral and financial support during my studies. Without their help all that would not have been possible.

Last but not least I am greatly indebted to my girlfriend, Birgit Ondra, for her constant support during all times of my studies and all her motivating words which encouraged me to explore the different fields of mathematics even further.

# Contents

# Chapter 1

# Preliminaries

## 1.1 What is a Lattice Path?

The central topic of investigation in this thesis are lattice paths. As the name suggests, they depend on a lattice, which can be described informally as a regular arrangement of points in the Euclidean space $\mathbb{R}^n$. Note, that they have many applications in physics, mathematics and computer science, like the solution of integer programming problems, cryptanalysis but they also appear in crystallography and sphere packing.

We start with a general and for our purpose suitable definition of the term *lattice*. Note, that there are various ways of how to define this term. A common and widely-used one is what we understand in the following under a periodic lattice (see below).

***Definition 1.1:*** A *lattice* $\Lambda = (V, E)$ is a mathematical model of a discrete space. It consists of two sets, a set $V \subset \mathbb{R}^n$ of vertices and a set $E \subset \mathbb{R}^n \times \mathbb{R}^n$ of edges, with no more than two edges between any two vertices. If two vectors are connected via an edge, we call them *nearest neighbors*.

A lattice is called

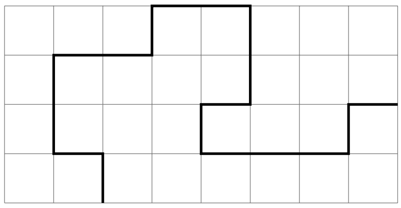
- *periodic*, if there exist vectors $v_1, \ldots, v_k$, such that the lattice is mapped to itself under any arbitrary translation $\sum_j \alpha_j v_j$ where $\alpha_j \in \mathbb{Z}$ for $j = 1, \ldots, k$. Vectors with this property are called *lattice vectors*.
- *Bravais lattice*, if any vector $r$ which is the difference between the position vectors of two lattice points is a valid lattice vector.

$\Diamond$

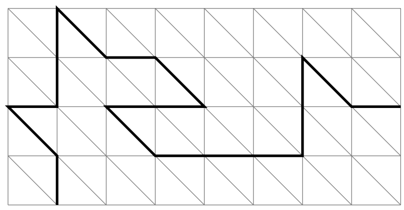
The importance of periodic lattices lies in the fact that they have a form of translation invariance. Thus, in this sense the Bravais lattice has the simplest possible form of translational invariance.

Some examples are shown in Figure 1.1. All of these lattices are periodic, but only the square lattice and the triangular lattice are of the Bravais type. This is due to the fact that there exist different types of nodes on these lattices. For example, in the case of the hexagonal lattice there are vertices with a horizontal edge on the left, and others with a horizontal edge on the right. Therefore there exist vectors which are the difference of two vertices, which do not map the lattice to itself.

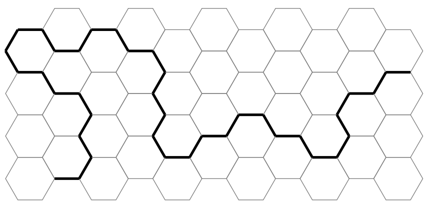
The expression “lattice” actually stems from physics. In mathematics and computer science lattices are also called *graphs* or *networks*.



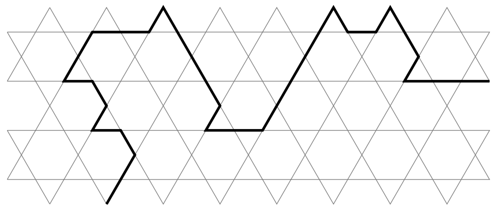
(a) Square Lattice



(b) Triangular Lattice



(c) Hexagonal Lattice



(d) Kagomé Lattice

Figure 1.1: Examples of Lattices

On a lattice we want to look at walks, that connect the vertices of the lattice. The basic component of a walk is a step, which essentially is nothing else than an edge.

***Definition 1.2:*** Let $\Lambda = (V, E)$. An $n$-step *lattice path* or *lattice walk* or *walk* from $s \in V$ to $x \in V$ is a sequence $\omega = (\omega_0, \omega_1, \dots, \omega_n)$ of elements in $V$, such that

1. $\omega_0 = s$, $\omega_n = x$,
2. $(\omega_i, \omega_{i+1}) \in E$.

The *length* $|\omega|$ of a lattice path is the number $n$ of steps (edges) in the sequence $\omega$. $\diamondsuit$

In most cases of this work we are going to work on the *Euclidean Lattice*, which we define to consist of the vertices $\mathbb{Z}^d$ and to be periodic. The edges are mostly defined through a so called step set. On this lattice an alternative definition via the step set can be used.

***Definition 1.3:*** A *step set* $\mathcal{S} \subset \mathbb{Z}^d$ is the fixed and finite set of possible steps. The primary examples which are considered are

(*nearest-neighbor model*) $\mathcal{S} = \{x \in \Lambda : \|x\|_1 = 1\}$,

(*spread-out model*) $\mathcal{S} = \{x \in \Lambda : 0 < \|x\|_\infty \leq L\}$,

where $L$ is a fixed integer. The elements of $\mathcal{S}$ are called *steps*. $\diamondsuit$

In Chapter 4 we are mainly going to work with a special kind of step set, so called small steps.

***Definition 1.4:*** If the step set $\mathcal{S}$ is a subset of $\{-1, 0, 1\}^2 \setminus \{(0, 0)\}$, then we say $\mathcal{S}$ is a set of *small steps*. $\Diamond$

In order to simplify notation, it is sometimes more convenient to use a more intuitive terminology by representing a step set by the corresponding points on a compass or by a small picture. In Figure 1.2 the full set of small steps is depicted. In this special case moving from $(1, 0)$ counterclockwise corresponds to **E**, **NE**, **N**, **NW**, **W**, **SW**, **S** and **SE**.

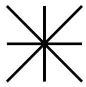


Figure 1.2: The full set of Small Steps

***Definition 1.5:*** An $n$-step *lattice path* or *lattice walk* or *walk* from $s \in \mathbb{Z}^d$ to $x \in \mathbb{Z}^d$ relative to $\mathcal{S}$ is a sequence $\omega = (\omega_0, \omega_1, \dots, \omega_n)$ of elements in $\mathbb{Z}^d$, such that

1. $\omega_0 = s$, $\omega_n = x$,
2. $\omega_{i+1} - \omega_i \in \mathcal{S}$.

The *length* $|\omega|$ of a lattice path is the number $n$ of steps in the sequence $\omega$. $\Diamond$

Comparing Definitions 1.2 and 1.5 we see, that in the second case $V = \mathbb{Z}^d$ and the set of possible edges $E$ is recursively defined over the set of allowed steps. The edge $(x, y) \in E$ exists if and only if $(y - x) \in \mathcal{S}$. The advantage of the second definition is its recursive character and its compact form, which is why we are going to choose this one for the remainder of this thesis. Note, that this definition can be adapted to apply for all lattices of Bravais type.

***Remark 1.6:*** In most cases we are going to consider the Euclidean Lattice. Here we will concretize Definition 1.5 to start from the origin $s = (0, 0)$, i.e. $\omega_0 = (0, 0)$. But this fact, will not represent a restriction on our discussion, as we are going to consider *homogeneous* lattices, in the sense that the number of $n$-step walks starting from $s$ is independent for all values of $n$. This is a general property of periodic lattices, which we will not proof here.

For more details on the basic properties of lattices we refer to [21]. $\blacksquare$

In the remainder of this work, we are going to work in the Euclidean plane only. Here we can also describe a lattice path by a polygonal line. An example is shown in Figure 1.3, where an unrestricted walk on the lattice $\mathbb{Z}^2$ and the set of small steps, from which it was constructed, is shown. Unrestricted in this context means, that there are no boundaries on the domain (lattice), that we allow self-intersections and that the walk ends at an arbitrary point.

Obviously, there is another equivalent representation of a walk with a fixed start point, by the sequence of performed steps. In particular, the walk in Figure 1.3b is given by the sequence

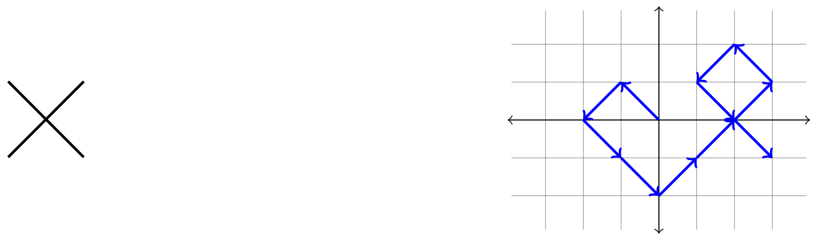
$$(\mathbf{NW}, \mathbf{SW}, \mathbf{SE}, \mathbf{SE}, \mathbf{NE}, \mathbf{NE}, \mathbf{NE}, \mathbf{NW}, \mathbf{SW}, \mathbf{SE}, \mathbf{SE})$$

or

$$(\nwarrow, \swarrow, \searrow, \searrow, \nearrow, \nearrow, \nearrow, \nwarrow, \swarrow, \searrow, \searrow).$$

The concept of steps is also useful for introducing weights on paths, which are needed for many applications.

***Definition 1.7:*** For a given step set $\mathcal{S} = \{s_1, \dots, s_k\}$ we define the respective *system of*



(a) $\mathcal{S} = \{\mathbf{NE}, \mathbf{SE}, \mathbf{NW}, \mathbf{SW}\}$ (b) Unrestricted Walk with Loops and 11 steps

Figure 1.3: Unrestricted Path with Loops in $\mathbb{Z}^2$

*weights* as $\Pi = \{w_1, \ldots, w_k\}$ with $w_j > 0$ the associated weight to step $s_j$ for $j = 1, \ldots, k$. The *weight of a path* is defined as the product of the weights of its individual steps. $\Diamond$

Some useful choices are:

- $w_j = 1$: Combinatorial paths in the standard sense;
- $w_j \in \mathbb{N}$: Paths with colored steps, i.e. $w_j = 2$ means that the associated step has two possible colors;
- $\sum_j w_j = 1$: Probabilistic model of paths, i.e. step $s_j$ is chosen with probability $w_j$.

## 1.2 Formal Power Series

Formal power series are the central object of investigation. For a ring $R$ we denote by $R[z]$ the ring of polynomials in $z$ with coefficients in $R$.

***Definition 1.8:*** Let $R$ be a ring with unity. The ring of *formal power series* $R[[z]]$ consists of all formal sums of the form

$$\sum_{n\geq 0} a_n z^n = a_0 + a_1 z + a_2 z^2 + \ldots,$$

with coefficients $a_n \in R$.

The *sum* of two formal power series $\sum_{n\geq 0} a_n z^n$, $\sum_{n\geq 0} b_n z^n$ is defined by

$$\sum_{n\geq 0} a_n z^n + \sum_{n\geq 0} b_n z^n = \sum_{n\geq 0} (a_n + b_n) z^n$$

and their *product* by

$$\sum_{n\geq 0} a_n z^n \cdot \sum_{n\geq 0} b_n z^n = \sum_{n\geq 0} \left( \sum_{k=0}^{n} a_k b_{n-k} \right) z^n.$$

$\Diamond$

***Definition 1.9:*** Let $A(z) = \sum_{n\geq 0} a_n z^n$ be a formal power series. We define the linear operator $[z^n]A(z)$ as

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

called the *coefficient extraction operator*. ◊

The coefficient extraction operator satisfies the following identity for all suitable $k$, i.e. all expressions have to be well-defined.

$$[z^{n-k}]A(z) = [z^n]z^k A(z). \tag{1.1}$$

***Definition 1.10:*** Let $R$ be a ring with unity and $A(z) = \sum_{n\geq 0} a_n z^n \in R[[z]]$. Then the *formal derivative* $A'(z)$ is given by

$$A'(z) = \sum_{n\geq 0}(n+1)a_{n+1}z^n.$$

The formal derivative fulfils all known rules from real analysis for derivatives, i.e. linearity, product-, quotient- and chain-rule, etc. ◊

We introduce a topology on the ring of formal power series. Via this we are able to consider limits.

***Definition 1.11:*** Let $R$ be a ring with unity, $A(z) = \sum_{n\geq 0} a_n z^n \in R[[z]]$. The *valuation* is a function $v : R[[z]] \to \mathbb{N} \cup \{\infty\}$ defined as

$$v(A(z)) = \begin{cases} \infty, & \text{if } A(z) \equiv 0, \\ \min\{n \mid a_n \neq 0\}, & \text{otherwise.} \end{cases}$$

Let $B(z) = \sum_{n\geq 0} b_n z^n$. The *distance* between two formal power series is defined as

$$d(A(z), B(z)) = 2^{-v(A(z)-B(z))}.$$

◊

Let $\varepsilon > 0$. If $d(A(z), B(z)) < \varepsilon$ then $v(A(z) - B(z)) > \log_2 \varepsilon$. This implies, that $[z^k]A(z) = [z^k]B(z)$ for all $k \leq \log_2 \varepsilon$. In other words, a small value $\varepsilon$ means, that the first "few" coefficients of the two formal power series are equal, and they may only differ in terms of high order.

***Theorem 1.12:*** The metric space $\langle R[[z]], d\rangle$ employing the formal topology is complete.

A sketch of the proof is given in [13, pp. 731]. More details on formal power series can be found in [16, 44].

In the end, we want to recall some important power series expansions:

$$\frac{1}{1-x} = \sum_{n\geq 0} x^n, \qquad e^x = \sum_{n\geq 0} \frac{1}{n!}x^n,$$

$$(1+x)^\alpha = \sum_{n\geq 0} \binom{\alpha}{n} x^n, \qquad \log(1+x) = \sum_{n\geq 0} \frac{(-1)^{n+1}}{n!} x^n,$$

where $\binom{\alpha}{n} = \alpha(\alpha-1)\cdots(\alpha-n+1)/n!$.

## 1.3 Asymptotic Notation

These definitions draw from [13, Chapter A.2], where more examples can be found.

Let $S$ be a set and $s_0 \in S$. We assume a notion of neighborhood to exist in $S$, e.g. $S = \mathbb{C}$ and $s_0 = 0$. Two functions $f, g : S \setminus \{s_0\} \to \mathbb{R}(\mathbb{C})$ are given.

- *$\mathcal{O}$-notation:* Denote

$$f(s) \underset{s \to s_0}{=} \mathcal{O}(g(s))$$

  if the ratio $f(s)/g(s)$ stays bound as $s \to s_0$ in $S$. In other words, there exists a neighborhood $V$ of $s_0$ and a constant $C > 0$, such that

$$|f(s)| \leq C|g(s)| \qquad s \in V,\ s \neq s_0.$$

  This is also known, as "Big-Oh-notation".

- *$\sim$-notation:* Denote

$$f(s) \underset{s \to s_0}{\sim} \mathcal{O}(g(s))$$

  if the ratio $f(s)/g(s)$ tends to 1 as $s \to s_0$ in $S$. One also says *$f$ and $g$ are asymptotically equivalent* (as $s$ tends to $s_0$).

- *o-notation:* Denote

$$f(s) \underset{s \to s_0}{=} o(g(s))$$

  if the ratio $f(s)/g(s)$ tends to 0 as $s \to s_0$ in $S$. In other words, for any $\varepsilon > 0$, there exists a neighborhood $V$ of $s_0$, such that

$$|f(s)| \leq \varepsilon|g(s)| \qquad s \in V,\ s \neq s_0.$$

  This is also known, as "little-Oh-notation".

## 1.4 Complex Analysis

We assume basic understanding of complex analysis, however we want to cite some important theorems which are going to be applied. The definitions of analytic, holomorphic and meromorphic functions as well as the basics of the analysis of singularities are left to more focused texts. The following theorems are taken from [27].

Let $U_r(w)$ be the ball in $\mathbb{C}$ around $w$ with radius $r$ and with respect to $\|\cdot\|_2 = |\cdot|$. We denote the path $\gamma : [0, 2\pi] \to \mathbb{C}$ with $\gamma(t) = w + r\exp(it)$ as $\vec{\partial} U_\rho(w)$.

***Theorem 1.13 (Cauchy's Integralformula):*** Let $f : D \to Y$ be holomorphic. For fixed $w \in D$ it holds that

$$f(w) = \frac{1}{2\pi i} \int_\gamma \frac{f(\zeta)}{\zeta - w}\, d\zeta,$$

for every closed, continuous, piecewise continuous differentiable path $\gamma : [0, 2\pi] \to D \setminus \{w\}$, which in $D \setminus \{w\}$ is homotopic to $\vec{\partial} U_\rho(w)$, where $\rho > 0$ such that $K_\rho(w) \subseteq D$.

The above statement can be generalized to derivatives, as every holomorphic functions is infinitely differentiable: Under the same conditions as in the last theorem it holds that

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

For a holomorphic $f : D \setminus \{w\} \to \mathbb{C}$, with the Laurent series $f(z) = \sum_{n=-\infty}^{\infty} a_n (z - w)^n$ the coefficient $a_{-1} =: \operatorname{Res}(f, w)$ is called *residue* of $f$ at $w$.

***Theorem 1.14 (Residue Theorem):*** Let $D \subseteq \mathbb{C}$ be open, $w_1, \dots, w_n \in D$ and $f : D \setminus \{w_1, \dots, w_n\} \to \mathbb{C}$ holomorphic. Let $\gamma : [0, 2\pi] \to D \setminus \{w_1, \dots, w_n\}$ be a closed, continuous, piecewise continuous differentiable path which is null-homotopic in $D$, i.e. $n(\gamma, z) = 0$ for all $z \in \mathbb{C} \setminus D$, then

$$\frac{1}{2\pi i} \int_\gamma f(\zeta), d\zeta = \sum_{j=1}^{n} \operatorname{Res}(f, w_j) n(\gamma, w_j).$$

The following lemma gives a way to calculate the residue for a special type of functions.

***Lemma 1.15:*** Let $D \subseteq \mathbb{C}$ be open, $w \in D$ and $f = \frac{h}{g}$ for two on $D$ holomorphic functions $g$ and $h$. Furthermore, let $h(w) \neq 0$ and assume $w$ is a simple root of $g$ (i.e. multiplicity 1). Then

$$\operatorname{Res}\left(\frac{h}{g}, w\right) = \frac{h(w)}{g'(w)}.$$

# Chapter 2

# Analytic Combinatorics

*"Combinatorics, the branch of mathematics concerned with the theory of enumeration, or combinations and permutations, in order to solve problems about the possibility of constructing arrangements of objects which satisfy specified conditions."*[1]

The focus of this thesis with regards to the preceding definition lies on the enumeration of objects, which are mostly described by recursions and boundary conditions, namely lattice paths. A standard tool in this context are generating functions which were introduced as formal power series whose coefficients give the sizes of a sought family of objects with respect to a parameter encoded in the exponent. A very colorful description from Wilf[2] [46] says

*"A generating function is a clothesline on which we hang up a sequence of numbers for display."*[3]

It describes quite vividly the idea of generating functions. This tool has led to many new insights in the field of combinatorics, by introducing new possible solution strategies. Their importance can be seen in the vast amount of available literature, like the books from Stanley[4] [43, 44] which, among other things, introduce a classification of generating functions, which has proved to be useful and applicable for lattice path combinatorics.

Furthermore they served as a link for interdisciplinary applications of techniques from different branches of mathematics. One very important field, which found entrance to combinatorics, is complex analysis. It revolutionized the field and founded the new branch of Analytic Combinatorics. The fathers of this development are Flajolet[5] and Sedgewick[6] in [13]. They interpret the formerly only algebraically investigated formal power series as complex analytic functions on their radii of convergence. This allows the extraction of the asymptotic behavior and much more.

[1] CollinsDictionary.com, `http://www.collinsdictionary.com/dictionary/english/combinatorics`, accessed 12/08/2013.

[2] Herbert Wilf, 13.6.1931-7.1.2012

[3] Wilf, generatingfunctionology, p. 1

[4] Richard Peter Stanley, 23.6.1944-

[5] Philippe Flajolet, 1.12.1948-22.3.2011

[6] Robert Sedgewick, 20.12.1946-

The structure of the subsequent chapter was inspired by [26, Chapter 4] and gives an introduction to symbolic methods, using [13, 43, 44, 46].

## 2.1 Combinatorial Classes and Ordinary Generating Functions

Following [13, pp. 16] we give a short introduction to the symbolic method. In particular, we emphasize on the topics important for lattice path combinatorics.

***Definition 2.1:*** A *combinatorial class*, or simply a *class*, is a finite or denumerable set on which a size function is defined, satisfying the following conditions:

1. the size of an element is a non-negative integer;
2. the number of elements of any given size is finite.

$\Diamond$

If $\mathcal{A}$ is a class, the size of an element $\alpha \in \mathcal{A}$ is denoted by $|\alpha|$, or $|\alpha|_{\mathcal{A}}$ in the few cases where the underlying class is not clear from the context. Using this size function, we decompose $\mathcal{A}$ into disjoint subclasses $\mathcal{A}_n$, which contain all elements of $\mathcal{A}$ of size $n$ and we denote the cardinality of these subsets by $a_n = \operatorname{card}(\mathcal{A}_n)$.

In accordance with this definition we define the class $\mathcal{W} = \mathcal{W}_{\mathcal{S},\Lambda}$ to be the set of all walks on a lattice $\Lambda$ with respect to the step set $\mathcal{S} = \mathcal{S}_\Lambda$. Here, $|\omega|$ is the length of a walk $\omega \in \mathcal{W}$.

***Definition 2.2:*** The *counting sequence* of a combinatorial class $\mathcal{A}$ is defined as the sequence of integers $(a_n)_{n\geq 0}$. $\Diamond$

***Definition 2.3:*** Two combinatorial classes $\mathcal{A}$ and $\mathcal{B}$ are said to be *(combinatorial) isomorphic*, which is written $\mathcal{A} \cong \mathcal{B}$, if and only if their counting sequences are identical. This condition is equivalent to the existence of a bijection from $\mathcal{A}$ to $\mathcal{B}$ that preserves size. One also says $\mathcal{A}$ and $\mathcal{B}$ are *bijectively equivalent*. $\Diamond$

Note, that this bijection, despite it needs to exist, is not always easy to be found nor does it have to behave in a nice and natural manner. The enumerative information of a class is stored in the formal power series $A(z)$.

***Definition 2.4:*** The *ordinary generating function* (OGF) of a sequence $(a_n)_{n\geq 0}$ is the formal power series

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

The OGF of a combinatorial class $\mathcal{A}$ is the generating function for the counting sequence $a_n = \operatorname{card}(\mathcal{A}_n),\ n \geq 0$. Equivalently, the combinatorial form

$$A(z) = \sum_{\alpha \in \mathcal{A}} z^{|\alpha|}$$

is employed. We say the variable $z$ *marks* the size in the generating function. $\Diamond$

Note, that there are two special classes:

| Class | Nr. of elements | Weights | OGF |
|---|---|---|---|
| Empty class $\mathcal{E}$ | 1 | 0 | $E(z) = 1$ |
| Atomic class $\mathcal{Z}$ | 1 | 1 | $Z(z) = z$ |

Here is a brief summary of the introduced naming convention:

| Class | Subclasses of elements of size $n$ | Cardinality of subclasses | OGF |
|---|---|---|---|
| $\mathcal{A}$ | $\mathcal{A}_n$ | $a_n$ | $A(z)$ |

Generating functions are elements of the ring of formal power series $\mathbb{C}[[z]]$, thus they can be manipulated algebraically. Two basic operations are the sum and the Cauchy product, which we want to introduce now.

Firstly, let $\mathcal{A}$ and $\mathcal{B}$ be two disjoint classes. Their union $\mathcal{C} = \mathcal{A} \,\dot{\cup}\, \mathcal{B}$ represents a new class with size defined consistently as

$$|\gamma|_{\mathcal{C}} = \begin{cases} |\gamma|_{\mathcal{A}}, & \text{if } \eta \in \mathcal{A}, \\ |\gamma|_{\mathcal{B}}, & \text{if } \eta \in \mathcal{B}. \end{cases}$$

This translates naturally into $c_n = a_n + b_n$, which concluded the intuition for the generating function of $\mathcal{C}$:

$$C(z) = A(z) + B(z) = \sum_{n \geq 0} (a_n + b_n) z^n.$$

Secondly, their Cartesian product $\mathcal{C} = \mathcal{A} \times \mathcal{B} = \{\gamma = (\alpha, \beta) \mid \alpha \in \mathcal{A}, \beta \in \mathcal{B}\}$ represents a new class with size defined consistently as

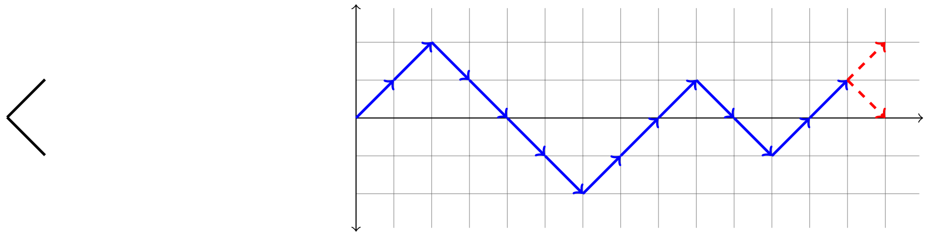
$$|\gamma|_{\mathcal{C}} = |\alpha|_{\mathcal{A}} + |\beta|_{\mathcal{B}}.$$

In this case we have to consider all possibilities in the manner of a Cauchy product, hence $c_n = \sum_{k=0}^{n} a_k b_{n-k}$, and we conclude as anticipated

$$C(z) = A(z) \cdot B(z) = \sum_{n \geq 0} \left( \sum_{k=0}^{n} a_k b_{n-k} \right) z^n.$$

The true power resulting from the *symbolic method*, is best understood by examples. Let's consider two cases, which apply the above definitions and operations.

***Example 2.5 (Unrestricted Paths):*** Consider the class $\mathcal{W}$ of unrestricted lattice paths employing the step set $\mathcal{S} = \{\mathbf{NE}, \mathbf{SE}\}$ and illustrated in Figure 2.1a. There are many ways to describe the construction of lattice paths. The most natural way is a step-by-step construction, from which one can deduce a recursive definition for the number of sought lattice



(a) $\mathcal{S} = \{\mathbf{NE}, \mathbf{SE}\}$ (b) Two possible extensions of an unrestricted path with a **NE**- or **SE**-step

Figure 2.1: Unrestricted Path

paths. Alternatively, one can deduce a construction for the combinatorial classes, which we want to demonstrate here.

A member of the class $\mathcal{W}$ is either the empty path or a path of non-zero length $n$. In the latter case we can construct a path of length $n+1$ by extending the path by one step out of the step set $\mathcal{S}$ and the resulting path is also a member of $\mathcal{W}$. This informal description is visualized in Figure 2.1b and translates into

$$\mathcal{W} = \underbrace{\mathcal{E}}_{\text{empty walk}} \ \dot{\cup} \ \underbrace{\mathcal{W} \times \mathcal{Z}_{\mathbf{NE}}}_{\text{append } \mathbf{NE}\text{-step}} \ \dot{\cup} \ \underbrace{\mathcal{W} \times \mathcal{Z}_{\mathbf{SE}}}_{\text{append } \mathbf{SE}\text{-step}}$$

As we do not distinguish between **NE**- and **SE**-steps the class $\mathcal{Z}_{\mathbf{NE}} \cong \mathcal{Z}_{\mathbf{SE}} \cong \mathcal{Z}$. Hence, we are able to apply the symbolic method by translating this equation into an equation on the corresponding generating functions:

$$W(z) = 1 + zW(z) + zW(z) = 1 + 2zW(z). \tag{2.1}$$

This equation can be solved algebraically and we get the solution

$$W(z) = \frac{1}{1-2z}. \tag{2.2}$$

In this case we extract the coefficients easily and get that the number of $n$-step unrestricted lattice paths with respect to the step set $\mathcal{S}$ starting from the origin is

$$w_n = [z^n]W(z) = [z^n]\frac{1}{1-2z} = [z^n]\sum_{k\geq 0} 2^k z^k = 2^n.$$

Note, that in this case it was quite easy to solve the functional equation (2.1). But in most general cases we are not able to deduce such a simple form for the solution and all we get is a relation on the functional equation. The following chapters will demonstrate different techniques on how to deal with these cases and how to extract enough information out of this equation, in order to decide on certain properties of the solution (which are partly introduced in this chapter), but without explicitly solving it. ■

***Remark 2.6:*** From Algebra we know that solutions of algebraic equations are unique up to multiplicity of roots. Recalling the definition of combinatorial isomorphic classes this gives us an easy way of checking such isomorphisms. If the generating functions of two classes satisfy the same functional equation, then the coefficient sequences satisfy the same recursion. In order to prove isomorphism, all which is left, is to check the “start values”, this can also be achieved by comparison of the first “few” (depending on the order of the recursion/equation) terms of the sequence. A straightforward example of two classes whose generating functions fulfil the same functional equation are the empty class $\mathcal{E}$ and the atomic class $\mathcal{Z}$. Both OGF satisfy the equation $A(z)^2 = A(z)$, but they are not the same, as $E(z) = 1$ and $Z(z) = z$, respectively. ■

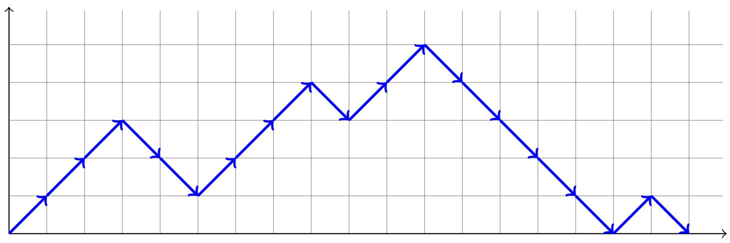


Figure 2.2: Dyck Path of length 18

***Example 2.7 (Dyck Paths, [13, pp. 319]):*** The probably most famous example of a class of lattice paths is the class of *Dyck Paths* $\mathcal{D}$. These are paths on the same step set $\mathcal{S} = \{\mathbf{NE}, \mathbf{SE}\}$ as before, but implying the restrictions, that they start at the origin, never leave the first quadrant and end on the $x$-axis. An example is shown in Figure 2.2.

As before, we are able to construct a functional equation for the OGF $D(z)$ of Dyck Paths using the introduced operations: The technique we will apply is known as *First passage decomposition*. Basically it decomposes an arbitrary path $\omega \in \mathcal{D}$ into two, possibly empty paths also belonging to $\mathcal{D}$.

A member of the class $\mathcal{D}$ is either the empty path or a path of non-zero length. If it is of non-zero length, after the initial point of contact at the origin, there will be another point of contact with the $x$-axis. Denote the first such second point as $x_0$. Now consider the path from the origin to $x_0$ without the initial **NE**- and the final **SE**-step. This, possible empty sub-path is also a legitimate Dyck path that belongs to $\mathcal{D}$. (Recall that the empty path is also a member of $\mathcal{D}$.) After the “first passage”, which ends at $x_0$, there will be another path starting at $x_0$ and ending on the $x$-axis. This path could be empty as well, but it is, as before, again a Dyck Path. The described procedure is depicted in Figure 2.3.

This informal description translates into

$$\mathcal{D} = \underbrace{\mathcal{E}}_{\text{empty walk}} \ \dot{\cup} \ \underbrace{\mathcal{Z}_{\mathbf{NE}} \times \mathcal{D} \times \mathcal{Z}_{\mathbf{SE}}}_{\text{first passage}} \times \mathcal{D}.$$

The symbolic method gives with the same reasoning as before

$$D(z) = 1 + z \cdot D(z) \cdot z \cdot D(z) = 1 + z^2 (D(z))^2. \tag{2.3}$$

Here we obtained a quadratic functional equation, which has the two possible solutions

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

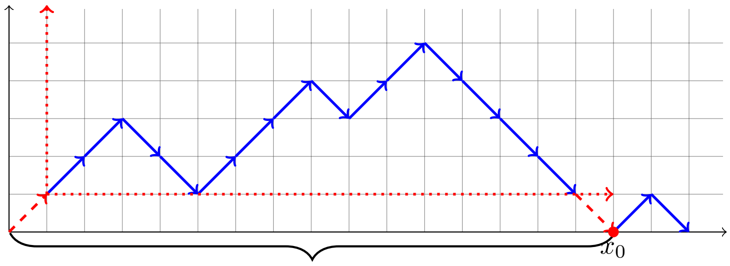



Figure 2.3: First Passage Decomposition of Dyck Path

Taking a closer look at $D_+(z)$, we see, that it possesses a singularity at 0, which corresponds to the constant term of the formal power series, and ought to be 1. Hence, we can dismiss this branch and arrive at the final solution

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

After using Newton's expansion theorem for general exponents and some elementary manipulations of binomial coefficients we get

$$d_n = [z^n]D(z) = \frac{1}{n+1}\binom{2n}{n} = C_n,$$

the $n$-th *Catalan number* (*EIS* **A000108**[7]), as the number of $n$-step Dyck-Paths. ∎

In the last two examples we have seen, that the sought-after OGFs may be the solutions of algebraic equations, compare (2.1) and (2.3). But in the case of our first example, the OGF is even a rational function, see (2.2). Naturally the question for a general classification of all possible generating functions arises. Stanley introduces in [43, Chapter 6] a hierarchy shown in (2.4), which answers this question and is presented in the subsequent section.

## 2.2 Classification of Ordinary Generating Functions

Throughout this whole chapter let $K$ be a field with characteristic $\operatorname{char} K = 0$, and $F$ be an arbitrary formal power series with coefficients in $K$, hence an element from the ring $K[[z]]$. The goal of this section is to introduce the three concepts of *rational*, *algebraic* and *D-finite* or *holonomic* functions. As seen before are algebraic functions a natural generalization of rational functions, analogously are $D$-finite functions a natural generalization of algebraic functions.

[7]Catalan numbers; `http://oeis.org/A000108`, accessed 26/08/2013.

Thus we get the hierarchy

$$\begin{array}{c} D\text{-finite/holonomic} \\ \uparrow \\ \text{algebraic} \\ \uparrow \\ \text{rational} \end{array} \tag{2.4}$$

Stanley remarks, that this hierarchy is by far not exhaustive, as various classes could be added, but these three seem the most useful for enumerative combinatorics.

***Definition 2.8:*** A formal power series $F \in K[[z]]$ is *rational* if there exist polynomials $P(z), Q(z) \in K[z]$, with $Q(z) \neq 0$, such that

$$F = \frac{P(z)}{Q(z)}.$$

$\Diamond$

As mentioned before we have already seen a rational OGF in (2.2). Note, that rationality corresponds to a linear recurrence relation, which follows immediately from rearranging the above definition to $F(z)Q(z) = P(z)$ in the language of OGFs. The concept of algebraic functions is a natural generalization to higher degrees.

***Definition 2.9:*** A formal power series $F \in K[[z]]$ is *algebraic* if there exist polynomials $P_0(z), P_1(z), \ldots, P_d(z) \in K[z]$, not all 0, such that

$$P_d(z)F^d + P_{d-1}(z)F^{d-1} + \ldots + P_1(z)F + P_0(z) = 0.$$

The smallest positive integer $d$ for which this equation holds is called the *degree* of $F$. $\Diamond$

***Example 2.10:*** As seen in Example 2.7 the OGF $D(z) = \frac{1-\sqrt{1-4z}}{2z}$ of Dyck paths satisfies

$$z^2 D(z)^2 - D(z) + 1 = 0.$$

Thus, $D$ is algebraic and of degree 2. $\blacksquare$

But there exists a larger class of functions, which encloses all algebraic functions: the $D$-finite (short for *differentiably finite*) or holonomic functions.

***Definition 2.11:*** A formal power series $F \in K[[z]]$ is *D-finite* or *holonomic*, if there exist polynomials $P_0(z), P_1(z), \ldots, P_d(z) \in K[z]$, with $P_d(z) \neq 0$, such that

$$P_d(z)F^{(d)} + P_{d-1}(z)F^{(d-1)} + \ldots + P_1(z)F' + P_0(z)F = 0, \tag{2.5}$$

where $F^{(j)} = d^j F/dz^j$ and $d \in \mathbb{N}$ is the order of the differential equation. $\Diamond$

***Remark 2.12:*** The historical source of holonomic functions is found in the theory of linear recursions. A sequence $(f_n)_{n\in\mathbb{N}}$ of complex numbers is *holonomic* or *P-recursive* (short for *polynomially recursive*) if it satisfies a homogeneous linear recurrence relation of finite degree with polynomial coefficients, i.e.

$$p_d(n)f_{n+d} + p_{d-1}(n)f_{n+d-1} + \cdots + p_0(n)f_n = 0, \qquad n \geq 0,$$

for some polynomials $p_i(x) \in \mathbb{C}[x]$. Let $F(z) = \sum_{n\geq 0} f_n z^n$ be the formal power series formed by the sequence $(f_n)_{n\in\mathbb{N}}$. As anticipated by the naming convention, a sequence is holonomic if and only if its generating function is holonomic, see [43, Proposition 6.4.3]. ■

***Proposition 2.13 [43, Proposition 6.4.1]:*** Let $U \in K[[z]]$. The following three conditions are equivalent:

(i) $U$ is holonomic.

(ii) There exist polynomials $Q_0(z), \dots, Q_m(z), Q(z) \in K[z]$, with $Q_m(z) \neq 0$, such that

$$Q_m(z)U^{(m)} + Q_{m-1}(z)U^{(m-1)} + \ldots + Q_1(z)U' + Q_0(z)U = Q(z). \tag{2.6}$$

(iii) The vector space over $K(z)$ spanned by $U$ and all its derivatives $U', U'', \dots$ is finite-dimensional, i.e.

$$\dim_{K(z)} \left[K(z)U + K(z)U' + K(z)U'' + \ldots\right] < \infty.$$

*Proof:* (i) ⇒ (ii): Trivial.

(ii) ⇒ (iii): Suppose (2.6) holds and $t$ is the degree of $Q(z)$. After differentiating (2.6) $t+1$ times we get an equation in the form of (2.5), with $d = m+t+1$ and $P_d(z) = Q_m(z) \neq 0$. Solving for $U^{(d)}$ yields

$$U^{(d)} = h_0(z)U + h_1(z)U' + \ldots + h_{d-1}(z)U^{(d-1)}, \tag{2.7}$$

with polynomials $h_0(z), \dots, h_{d-1} \in K[z] \subset K(z)$. Differentiating this expression with respect to $z$ we get

$$\begin{aligned} U^{(d+1)} &= \tilde{h}_0(z)U + \tilde{h}_1(z)U' + \ldots + \tilde{h}_{d-1}(z)U^{(d-1)} + \tilde{h}_d(z)U^{(d)} \\ &\in K(z)U + K(z)U' + \ldots + K(z)U^{(d-1)}, \end{aligned}$$

with polynomials $\tilde{h}_0(z), \dots, \tilde{h}_d(z) \in K[z]$ and the last member relation holds due to (2.7). By induction it holds that,

$$U^{(d+k)} \in K(z)U + K(z)U' + \ldots + K(z)U^{(d-1)},$$

for all $k \geq 0$.

(iii) ⇒ (i): Suppose

$$\dim_{K(z)} \left[K(z)U + K(z)U' + K(z)U'' + \ldots\right] = d.$$

Thus $u, u', \dots, u^{(d)}$ are linearly dependent over $K(z)$. This dependence relation, after clearing the denominators so that the coefficients are polynomials in $K[z]$, results in an equation of the form (2.5). □

***Example 2.14:*** The following functions are holonomic:

1. $U = \frac{z-2}{3z+4}$, as $(z-2)(3z+4)U' - 10U = 0$.
2. $U = e^z$, as $U' = U$ and $U = \log(z)$, as $zU' = 1$ or $zU'' + U' = 0$.
3. $U = z^m e^{az}$, as $U' = (mz + a)U$.
4. $U = \cos(z)$, as $U'' = -U$. The same holds obviously for $\sin(z)$.
5. $U = \sum_{n \geq 0} n! z^n$, since $(zU)' = \sum_{n \geq 0}(n+1)! z^n$ This implies that $z(zU)' + 1 = U$ or reordered in the form of (2.6): $z^2 U' + (z-1)U = -1$.

■

We end this section with the proof of the missing link between holonomic and algebraic functions.

***Theorem 2.15 [43, Proposition 6.4.6]:*** Let $U \in K[[z]]$ be algebraic of degree $d$, then $U$ is holonomic.

*Proof:* If $U(z)$ is algebraic, there is some polynomial $0 \neq P(z, y) \in K(z, y)$ of minimal degree such that $P(z, U) = 0$. It holds, that

$$0 = \frac{d}{dz} P(z, U) = \left.\frac{\partial P(z, y)}{\partial z}\right|_{y=U} + U' \left.\frac{\partial P(z, y)}{\partial y}\right|_{y=U}$$

Since $P(z, y)$ is of minimal degree, and therefore irreducible over $K(x)$, it follows, that $\partial P(z, y)/\partial y$ is non-zero (remember $\operatorname{char} K = 0$) and a polynomial in $y$ of smaller degree than $P$, so $\left.\frac{\partial P(z,y)}{\partial y}\right|_{y=U} \neq 0$. Hence, we get

$$U' = -\frac{\left.\frac{\partial P(z,y)}{\partial z}\right|_{y=U}}{\left.\frac{\partial P(z,y)}{\partial y}\right|_{y=U}} \in K(z, U).$$

In other words, $U'$ is a rational function in $z$ and $U$. By induction we get that $U^{(k)} \in K(z, U)$ for all $k \geq 0$. But due to the fact, that $U$ is algebraic, we get $\dim_{K(z)} K(z, U) = d$ and so it follows that $U, U', \ldots, U^{(d)}$ are linearly dependent over $K(z)$. This yields an equation of the form (2.5), which proves that $U$ is holonomic. □

***Example 2.16 [43, Ex. 6.1]:*** Not all holonomic functions are algebraic. Consider $U(z) = e^z$: If it would be algebraic of degree $d$ it would satisfy an equation of the form

$$P_d(z)e^{dz} + P_{d-1}(z)e^{(d-1)z} + \ldots + P_1(z)e^z + P_0(z) = 0,$$

where $P_0(z), \ldots, P_d(z) \in \mathbb{C}[z]$ and $P_d(z)$ is of minimal degree. Differentiating this equation and subtracting the initial one multiplied by $d$, gives

$$P_d' e^{dz} + \left(P_{d-1}' - P_{d-1}\right) e^{(d-1)z} + \ldots + \left(P_1' - (d-1)P_1\right) e^z + \left(P_0' - dP_0\right) = 0,$$

which either has degree less than $d$, and contradicts the fact that $U(z)$ is algebraic of degree $d$, or the degree is the same, which contradicts the choice of $P_d(z)$ to be of minimal degree. ■

The class of holonomic function enjoys rich closure properties. Note, that the following theorem mentions only the operations we are going to encounter in this thesis. For more details see [13, Theorem B.2].

***Theorem 2.17:*** The class of univariate holonomic functions is closed under the following operations: sum $(+)$, product $(\times)$, differentiation $(\partial_z)$, indefinite integration $(\int^z)$ and algebraic substitution ($z \mapsto y(z)$ for some algebraic function $y(z)$).

*Proof:* The proof is omitted here. A sketch of a proof can be found in [13, Theorem B.2], full details are discussed in [43, Chapter 6]. □

The discussion so far only considered univariate or ordinary generating functions, i.e. functions in one variable. In order to encode more information, it is sometimes necessary to introduce more than one variable. This fact has already been used in the proof of Theorem 2.15. The necessary theory is presented in the next section.

## 2.3 Multivariate Generating Functions

So far we have considered only univariate formal power series, but this concept can be easily generalized to multivariate formal power series. In the same manner OGFs generalize to *multivariate generating functions* (MGFs). As Flajolet and Sedgewick put it [13, Chapter III], the main advantage of several variables is the possibility to keep track of a collection of parameters defined over combinatorial objects. But their big applicability results from a straightforward generalization of the symbolic method, which proved so powerful in the ongoing discussion. The main message is, that we can use the symbolic method not just to count combinatorial objects but also to *quantify* their properties.

In the case of lattice path combinatorics we will need the notion of a trivariate generating function, with two parameters keeping track of the end-point in the first quadrant and one parameter encoding the length of a lattice path. This translates into

$$Q(x, y; z) = \sum_{i,j,n \geq 0} q(i, j; n) x^i y^j z^n.$$

Note, that it can also be interpreted as a formal power series in $z$ with coefficients in $\mathbb{Q}[x, y]$, where for all $n$, almost all coefficients $q(i, j; n)$ are zero. This interpretation somehow closes the circle and links MGFs with OGFs.

Another generalization is the usage of formal Laurent series instead of formal power series. All definitions and observations stay the same and can be straightforwardly adapted to this new case. As a short-hand we define

$$\bar{x} := \frac{1}{x}.$$

This notion allows us to encode paths of length $n$ ending anywhere in the Euclidean plane:

$$\hat{Q}(x, y; z) = \sum_{\substack{i,j \in \mathbb{Z} \\ n \geq 0}} \hat{q}(i, j; n) x^i y^j z^n.$$

In the following we want to discuss the changes as part of a hands-on example on the simplest MGF, a *bivariate generating function* (BGF). A rigorous introduction to MGF can be found in [13, Chapter III].

A BGF is a formal power series (formal Laurent series) in two variables. Hence, there are two possible parameters which we could keep track of. One suitable definition for lattice paths, is the use of one variable as the length of the path, and the second one as the final height of the path, i.e. the stopping $y$-coordinate.

$$F(y; z) = \sum_{\substack{i \in \mathbb{Z} \\ n \geq 0}} f(i; n) y^i z^n,$$

where the coefficients $[z^n]F(y; z)$ are in $\mathbb{Q}[y, \bar{y}]$, and for all $n$ almost all $f(i; n)$ are zero.

***Definition 2.18:*** The *positive part* of $F(y; z)$ in $y$ is the following series, which has coefficients in $\mathbb{Q}[y]$ as power series in $z$:

$$[y^{>}]F(y; z) := \sum_{\substack{i > 0 \\ n \geq 0}} f(i; n) y^i z^n.$$

Similarly we define the *negative, non-negative* and *non-positive* parts of $F(y; z)$ in $y$, which we denote respectively by $[y^{<}]F(y; z)$, $[y^{\geq}]F(y; z)$ and $[y^{\leq}]F(y; z)$. The operator $[y^{<}]F(y; z)$ is also called the *projection onto the pole part* of $F(y; z)$, i.e. the partial sum of $F(y; z)$ where all terms contain a negative index of $y$. $\Diamond$

***Example 2.19:*** We will continue the analysis started in Example 2.5 of unrestricted paths $\mathcal{W}$ starting from the origin and using the step set $\mathcal{S} = \{\mathbf{NE}, \mathbf{SE}\}$. We derived the following relation on the combinatorial classes

$$\mathcal{W} = \mathcal{E} \,\dot{\cup}\, \mathcal{W} \times \mathcal{Z}_{\mathbf{NE}} \,\dot{\cup}\, \mathcal{W} \times \mathcal{Z}_{\mathbf{SE}}.$$

The difference now, is that we have to distinguish between **NE**- and **SE**-steps. We kind of abuse the notation now, because a **NE**-step increases the height by one and hence corresponds to the generating function $y$, but a **SE**-step decreases the height by one and hence corresponds to $\bar{y} = \frac{1}{y}$. Additionally, both steps increase the length by 1. Note, that we will work in the ring of formal Laurent series $\mathbb{Z}[[y, \bar{y}]]$. Let's define the bivariate generating function associated with $\mathcal{W}$ as

$$W_2(y; z) = \sum_{\substack{i \in \mathbb{Z} \\ n \geq 0}} w(i; n) y^i z^n.$$

This gives

$$W_2(y; z) = 1 + yzW_2(y; z) + \frac{z}{y} W_2(y; z).$$

Solving this equation for $W_2$ results in

$$W_2(y;z) = \frac{1}{1 - z\left(y + \frac{1}{y}\right)}.$$

Next we will perform a coefficient extraction in order to get $w(j;m)$, the number of walks of length $m$ stopping at height $j$. Firstly, we start by fixing $n$ by $m$:

$$[z^m]W_2(y;z) = \left(y + \frac{1}{y}\right)^m.$$

This is a Laurent polynomial in $y$. Secondly, we apply the shift identity of the coefficient extraction (1.1) to get

$$\begin{aligned} w(j;m) &= [y^j]\left(y + \frac{1}{y}\right)^m = [y^{m+j}]\left(y^2+1\right)^m \\ &= \begin{cases} 0, & \text{for } m+j \equiv 1 \mod (2) \text{ or } |m| > j, \\ \binom{m}{\frac{m+j}{2}}, & \text{for } m+j \equiv 0 \mod (2). \end{cases} \end{aligned}$$

Note, that the BGF can be easily transformed into the OGF we found in Example 2.5, by substituting $y = 1$. This action sums over all possible heights at fixed length $n$:

$$\begin{aligned} W_2(1;z) &= \frac{1}{1-2z} = W(z) \\ \sum_{i\in\mathbb{Z}} w(i;m) &= \sum_{i=-m,-m+2,\dots m} \binom{m}{\frac{m+i}{2}} = \sum_{j=0}^{m} \binom{m}{j} = 2^m \end{aligned}$$

In general, we have to be careful here. We are only dealing with formal power series, which is the reason why insertion of special values for variables is in general not well-defined. So, we have to ensure that all operations are legitimate, e.g.: there are no singularities and all sums are finite, etc. ∎

The classification of multivariate formal power series can be directly generalized from the univariate case.

***Definition 2.20:*** Let $K$ be field of characteristic $\operatorname{char} K = 0$ and $F \in K[[z_1, \dots, z_n]]$ be a multivariate formal power series. We call $K$

- *rational*, if there exist polynomials $P, Q \in K[z_1, \dots, z_n]$ such that

$$QF = P,$$

- *algebraic*, if there exist polynomials $P_0, P_1, \dots, P_d \in K[z_1, \dots, z_n]$, not all 0, such that

$$P_d F^d + P_{d-1} F^{d-1} + \ldots + P_1 F + P_0 = 0.$$

  The smallest positive integer $d$ for which this equation holds is called the *degree* of $F$.

- *D-finite* or *holonomic*, if there exist polynomials $P_{\ell,i} \in K[z_1, \ldots, z_n]$, $i = 0, \ldots, n$, $\ell = 0, \ldots, d_i$ with $P_{d_i}(z) \neq 0$, $i = 0, \ldots, n$, such that

$$\sum_{\ell=0}^{d_i} P_{\ell,i} \frac{\partial^\ell F}{\partial z_i^\ell} = 0 \quad \text{for all } i = 1, \ldots, n, \tag{2.8}$$

where $d_i \in \mathbb{N}$ is the order of the partial differential equation in $z_i$.

$\diamondsuit$

The class of MGF is also closed under various operations. Note, that as in the univariate case the following theorem mentions only the operations we are going to encounter in this thesis. For more details see [13, Theorem B.3].

***Theorem 2.21:*** The class of multivariate holonomic functions is closed under the following operations: sum $(+)$, product $(\times)$, differentiation $(\partial)$, indefinite integration $(\int)$, algebraic substitution and specialization (setting some variable to a constant).

*Proof:* For the proof we refer to the paper from Lipshitz [34]. $\square$

In the proof of Theorem 4.14 we will need the following result:

***Proposition 2.22:*** If $F(x, y; z)$ is a rational power series in $z$, with coefficients in $\mathbb{C}(x)[y, \bar{y}]$, then $[y^>]F(x, y; z)$ is algebraic over $\mathbb{C}(x, y, z)$. If the latter series has coefficients in $\mathbb{C}[x, \bar{x}, y]$, its positive part in $x$, the series $[x^>][y^>]F(x, y; z)$, is a trivariate holonomic series.

*Proof:* The first statement is a simple adaption of [15, Theorem 6.1]. The series $F(x, y; z)$ is interpreted as element of $(\mathbb{C}(x)[y])[[\bar{y}, z]]$, where $\bar{y} \mathrel{\widehat{=}} x$ and $z \mathrel{\widehat{=}} y$ and instead of extracting the elements of the diagonal, the sub-series $\sum_{i \geq 0} \left([\bar{y}^0 z^i]F(x, y; z)\right) z^i$ is chosen. The proof idea is complete analogous, the key tool is to expand $F(x, y; z)$ in partial fractions of $y$ (resp. $z$).

The second statement follows from the fact, that the diagonal of a holonomic series is holonomic [46, Theorem 6.3.3]. $\square$

## 2.4 Why is it important to be holonomic?

After this short exposition on holonomic functions, one might ask why they are relevant to combinatorics, as they are one central topic of investigation in this work. The subsequent short summary of possible answers is derived from [31, pp. 12]. One unanswered question, is the size of this special class. Are at the end nearly all functions anyway holonomic? Flajolet, Gerhold and Salvy [12] conjecture the following

> *"... a rough heuristic in this range of problem is the following: Almost anything is non-holonomic unless it is holonomic by design."*

But they invalidate their self-called naive statement in the next sentence, as there are many known combinatorial structures of holonomic character which arise naturally from different fields. Some examples are the enumerations of $k$-regular graphs or the Apéry sequence

$$A_n = \sum_{k=0}^{n} \binom{n}{k}^2 \binom{n+k}{k}^2$$

for which a proof was needed that it satisfies the recurrence

$$(n+2)^3 B_{n+2} - (34n^3 + 153n^2 + 231n + 117) B_{n+1} + (n+1)^3 B_n = 0, \qquad n \geq 0,$$

with $B_0 = 1$ and $B_1 = 5$. This was the last missing link in the proof of the irrationality of $\zeta(3)$ [45]. A possible proof uses the closure properties of holonomic functions and associated algorithms, see [13, p. 752].

The intuition is, that holonomic functions possess a "nice" structure. This can be seen for example in the asymptotic growth rate of a holonomic sequence, that is typically of the form

$$a(n) \sim C e^{P(n^{1/r})} n^{\mu n} n^{\theta} \log(n)^{\beta},$$

where $P$ is a polynomial, $\beta, r, \mu \in \mathbb{N}$ and $C, \theta \in \mathbb{C}$, see [28, Theorem 2] and the original work [47]. This form is important because many applications in lattice path enumeration are interested in the asymptotic behavior.

Another important property of the class of holonomic functions, is the fact that the class is reasonable small but still large enough, so that it contains enough quantities that arise in applications. A small class is a class where strong assumptions are imposed on its elements. An example for a small class is the set of all polynomial functions. The advantage of small classes is that they admit a finite representation, e.g. in form of their coefficients. The biggest disadvantage of small classes is, that they mostly do not cover the important cases which arise in applications. But large classes, which are more interesting, normally do not admit a finite representation of its elements. Think of the set of all functions which allow a power series expansion. Now holonomic functions have proved to be a good compromise between these two extremes. For more details see Kauers compact introduction to the field of symbolic computation with regards to holonomic functions in [28].

But why do we look for such a class? One answer can be found in the field of symbolic computing. If a class of functions possesses a finite description it is most likely that efficient algorithms exist to manipulate its elements. And in the case of holonomic functions, such algorithms exist indeed. Examples for such algorithms are presented in [39]. These algorithms are also used in computer-aided proofs which provide completely new possibilities to tackle so far unsolved problems. The proof that the GF of Gessel[8] Walks is algebraic in [6] is a nice example, where the technique of guessing is applied to find a recurrence relation for the number of such walks of given length. First a finite number of terms of this sequence is computed numerically, then the algorithm tries to guess the recurrence relation based on these few values and lastly, one has to prove that this recurrence holds for all elements of the sequence. An introduction to this technique can also be found in [28, Chapter 4].

Last but not least we want to mention the closure properties of holonomic functions again , compare Theorems 2.17 and 2.21. They are the source for many algorithms on holonomic functions and justify their choice as a large but useful class of combinatorial objects.

[8] Ira Martin Gessel, 9.4.1951-

# Chapter 3

# Directed Lattice Paths

As an introduction to lattice path theory, we are going to consider *directed paths*. These are paths with a fixed direction of increase which we choose to be the positive horizontal axis. This is described by the allowed steps: if $(i, j) \in \mathcal{S}$ then $i > 0$. One first important observation, is that the geometric realization of the path always lives in the right half-plane $\mathbb{Z}_+ \times \mathbb{Z}$. But it essentially means that directed paths are one-dimensional objects.

The following chapter mainly focuses on the expositions of Banderier[1] and Flajolet given in [3]. But it also draws from [8] in terms of applications of the *kernel method* which will be introduced in this chapter.

***Definition 3.1:*** Along these restrictions, we introduce the following classes (see Table 3.1):

- A *bridge* is a path whose end-point $\omega_n$ lies on the $x$-axis;
- A *meander* is a path that lies in the quarter plane $\mathbb{Z}_+^2$;
- An *excursion* is a path that is at the same time a meander and a bridge, i.e. it connects the origin with a point lying on the $x$-axis and involves no point with negative $y$-coordinate.

Additionally, we call a family of paths or steps to be *simple* if each allowed step in $\mathcal{S}$ is of the form $(1, b)$ with $b \in \mathbb{Z}$. In this case, we denote $\mathcal{S} = \{b_1, \dots, b_k\}$. $\diamondsuit$

In the remainder of this section, if not specified differently, we will always consider simple paths.

***Definition 3.2:*** Let $\mathcal{S} = \{b_1, \dots, b_k\}$ be a simple set of steps, with $\Pi = \{w_1, \dots, w_k\}$ the corresponding system of weights. The *characteristic polynomial* of $\mathcal{S}$ is the Laurent polynomial $S(u)$, defined as

$$S(u) := \sum_{i=1}^{k} w_i u^{b_i}.$$

Define $c := -\min_i b_i$ and $d := \max_i b_i$ as the two extreme vertical amplitudes. We assume throughout this chapter $c, d > 0$.

[1] Cyril Banderier, 19.5.1975-

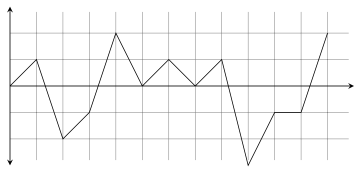
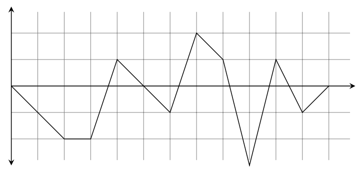
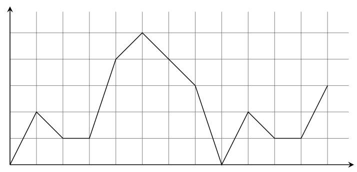
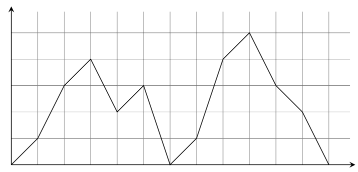
| | ending anywhere | ending at 0 |
|---|---|---|
| unconstrained<br>(on $\mathbb{Z}$) | <br>walk/path ($\mathcal{W}$)<br>$W(z) = \frac{1}{1-zS(1)}$ | <br>bridge ($\mathcal{B}$)<br>$B(z) = z \sum_{i=1}^{c} \frac{u_i'(z)}{u_i(z)}$ |
| constrained<br>(on $\mathbb{Z}_+$) | <br>meander ($\mathcal{M}$)<br>$M(z) = \frac{1}{1-zS(1)} \prod_{i=1}^{c} (1 - u_i(z))$ | <br>excursion ($\mathcal{E}$)<br>$E(z) = \frac{(-1)^{c-1}}{p_{-c}z} \prod_{i=1}^{c} u_i(z)$ |

Table 3.1: The four types of paths: walks, bridges, meanders and excursion and the corresponding GFs [3, Fig. 1].

The *characteristic curve* of the lattice paths determined by $\mathcal{S}$ is the plane algebraic curve defined by the equation

$$1 - zS(u) = 0, \qquad \text{or equivalently} \qquad K(u, z) := u^c - z\,(u^c S(u)) = 0. \tag{3.1}$$

The quantity $K(u, z)$ is the *kernel* of the lattice paths and the equation is also referred to as *kernel equation.* $\Diamond$

Remark, that the left equation in (3.1) is a rational function in $u$, while the second form is called its *entire version*, i.e. it contains no negative powers.

A useful property of (converging) power series/polynomials with positive coefficients is, that $|S(u)| \leq S(|u|)$, which follows from a straightforward application of the triangle inequality: $|\sum_{i=1}^{k} w_i u^{b_i}| \leq \sum_{i=1}^{k} w_i |u|^{b_i}$. Another important property of such power series is that they are monotonically increasing in the argument for positive values.

It increases readability to rewrite

$$S(u) = \sum_{i=-c}^{d} p_i u^i,$$

where not needed powers are canceled by zero coefficients.

Examining equation (3.1) near $z = 0$ with respect to asymptotic considerations shows that the kernel equation can only be satisfied if one of the two relations

$$p_d z u^d \underset{z\to 0}{\sim} 1 \qquad \text{or} \qquad p_{-c} z u^{-c} \underset{z\to 0}{\sim} 1 \tag{3.2}$$

is satisfied. This is because $u$ can be interpreted as function of $z$ by the implicit definition of the kernel equation. Then it follows again by the kernel equation, that $u$ must be unbound or go to zero as $z$ tends to 0. If it would be bounded and not 0, the limit would lead to the contradiction $0 = 1$ in the kernel equation. Hence, $u \sim z^\alpha$ for $z \to 0$. This leads to the above result.

The entire version of the kernel equation is of degree $c + d$ in $u$ and it is known, that it has $c + d$ roots. These are the *branches* of a single algebraic curve, given by the kernel equation, which is then called the *characteristic curve*. As suggested by (3.2), one expects in the complex domain and for $z$ near 0, $c$ "small branches" that we write as $u_1, \dots, u_c$ and $d$ "large branches" denoted as $v_1 \cong u_{c+1}, \dots, v_d \cong u_{c+d}$, satisfying

$$u_j(z) \sim e^{2\pi i(j-1)/c}\,(p_{-c})^{1/c}\,z^{1/c}, \qquad v_\ell(z) \sim e^{2\pi i(1-\ell)/d}\,(p_{-d})^{-1/d}\,z^{-1/d}. \tag{3.3}$$

Written in formulas, this means for $z$ in a small enough neighborhood of 0, that

$$u^c - z(u^c S(u)) = -p_d z \prod_{i=1}^{c}(u - u_i(z)) \prod_{j=1}^{d}(u - v_j(z)). \tag{3.4}$$

In order to ensure uniqueness, we employ the standard approach and restrict ourself to the complex plane slit along the negative real axis. That allows us to talk about the individual branches in the sequel. More details about the theory of algebraic curves can be found in [1, 36].

The graph of branches is obtained by interchanging the axes in the graph of $1/S(u)$, with $u_1$ appearing as the real positive branch near the origin, see Figure 3.1[2].

## 3.1 Walks and Bridges

The first cases we are going to consider, are the unconstrained walks and bridges. These are the easiest models, but the following classification theorem shows already nicely, how the above theory of algebraic curves is applied.

In the following proof we will need the following definition

***Definition 3.3:*** A function $f : D \to \mathbb{R}$, $D \subseteq \mathbb{R}$ is called *unimodal*, if there exists a value $x_m \in D$, such that it is monotonically increasing for $x \le x_m$ and monotonically decreasing for $x \ge x_m$. $\lozenge$

By the definition it is clear, that the maximum is $f(x_m)$.

***Theorem 3.4 [3, Theorem 1]:*** The bivariate generating function of paths ($z$ marking size and $u$ marking final altitude) relative to a simple step set $\mathcal{S}$ with characteristic polynomial

[2] All plots created in Maple 12.0.

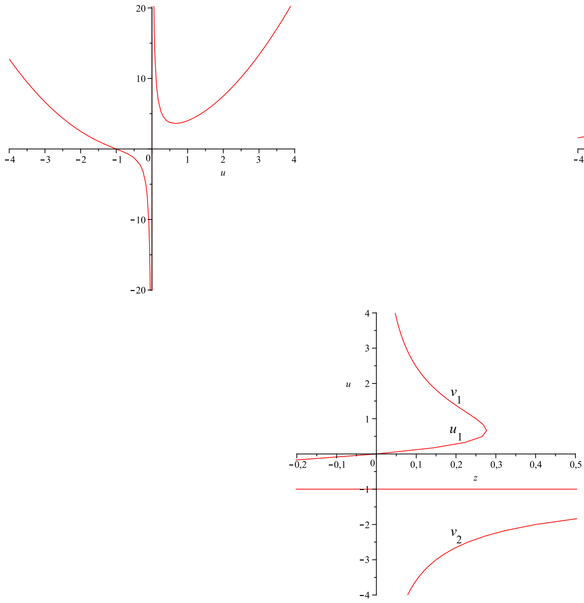


Figure 3.1: Graphs associated with the step set $\mathcal{S} = \{-1, 0, 1, 2\}$, with characteristic polynomial $S(u) = u^{-1} + 1 + u + u^2$. Top: the graphs of $S(u)$ and $1/S(u)$ for real $u$. Bottom: the three branches of the characteristic polynomial of the characteristic curve: a small one of order $z$ and two large ones of order $\pm z^{-1/2}$.

$S(u)$ is a rational function. It is given by

$$W(u; z) = \frac{1}{1 - zS(u)}.$$

The generating function of bridges is an algebraic function given by

$$B(z) = z \sum_{j=1}^{c} \frac{u_j'(z)}{u_j(z)} = z \frac{\mathrm{d}}{\mathrm{d}z} \log(u_1(z), \ldots, u_c(z)), \tag{3.5}$$

where $u_1, \ldots, u_c$ are all small branches of the characteristic curve (3.1). Generally the GF $W_k(z)$ of paths terminating at altitude $k$ is, for $-\infty < k < c$,

$$W_k(z) = z \sum_{j=1}^{c} \frac{u_j'(z)}{u_j(z)^{k+1}} = -\frac{z}{k} \frac{\mathrm{d}}{\mathrm{d}z} \left( \sum_{j=1}^{c} u_j(z)^{-k} \right), \tag{3.6}$$

and for $-d < k < \infty$,

$$W_k(z) = -z \sum_{j=1}^{d} \frac{v_j'(z)}{v_j(z)^{k+1}} = -\frac{z}{k} \frac{\mathrm{d}}{\mathrm{d}z} \left( \sum_{j=1}^{d} v_j(z)^{-k} \right), \tag{3.7}$$

where $v_1, \ldots, v_d$ are the large branches. For $W_0(z)$, the second form is to be taken in the limit sense $k \to 0$.

*Proof:* We start with a decomposition argument for walks: Fix $n \in \mathbb{N}$ and let $w_n(u) = [z^n]W(u;z)$ be the Laurent polynomial that describes the possible altitudes and the number of ways to reach them in $n$ steps. We have

- $w_0(u) = 1$, as we can never return to the origin, after leaving it;
- $w_1(u) = S(u)$, as length 1 corresponds to one step from the used step set $\mathcal{S}$;
- $w_{n+1}(u) = w_n(u)S(u)$, as a walk of length $n+1$ is constructed from a walk of length $n$ by appending an additional step from $\mathcal{S}$.

Hence, we obtain $w_n(u) = S(u)^n$ and therefore

$$W(u;z) = \sum_{n\geq 0} w_n(u)z^n = \sum_{n\geq 0} S(u)^n z^n = \frac{1}{1-zS(u)}.$$

We know from comparison with the geometric series, that this sum converges for $|z| < 1/S(|u|)$ where we have used, that $|S(u)| \leq S(|u|)$. Thus, it represents an analytic function in two variables, and beyond that it is entire in $z$ and of the Laurent type in $u$, because it involves arbitrary negative powers of $u$.

Next we are going to consider bridges. For positive $u$, the radius of convergence of $W(u;z)$ viewed as a function of $z$ is exactly $1/S(u)$. Due to the fact, that every bridge is also a special unconstrained walk (i.e. $[u^0]W(u;z) = B(z)$), we get that the number of bridges of given length $n$ is dominated by the number of walks, i.e. $B_n \leq w_n(1) = S(1)^n$. This implies, that the radius of convergence of $B(z)$ as a function of $z$ is at least $1/S(1)$.

We claim, that $1/S(u)$ is continuous and unimodal for $u \in (0,+\infty)$. The continuity is clear because $0 \notin (0,+\infty)$. It is unimodal, because $P$ is a convex function ($S''(u) = \sum_{i=-c}^{d} i(i-1)p_i u^{i-2} > 0$) that satisfies $1/S(0) = 1/S(\infty) = 0$.

Let $|z| < r$, with $r := \frac{1}{2}\frac{1}{S(1)}$. By the previous result, there exists an interval $(\alpha,\beta)$ such that for $\alpha \leq u \leq \beta$ we have $1/S(u) > r$. (The maximal possible interval would be $(1/S(\cdot))^{-1}(r,+\infty)$.) Reconsidering the properties of $W(u;z)$ we get from this consideration, that $W(u;z)$ is analytic in the open product domain

$$\mathcal{V} := \{z : |z| < r\} \times \{u : \alpha < |u| < \beta\}.$$

Thus, by Cauchy's Integralformula, Theorem 1.13, applied to the function $W(u;z)$ (viewed as a function of $u$) integrating over the closed circle $|z| = \frac{\alpha+\beta}{2}$ which lies completely in the domain $\mathcal{V}$, we get

$$B(z) = [u^0]W(u;z) = \frac{1}{2\pi i}\int_{|u|=(\alpha+\beta)/2} W(u;z)\frac{\mathrm{d}u}{u}.$$

By (3.3) we can choose $z$ small enough, so that all the large branches that escape to infinity lie outside of $|u| \leq (\alpha+\beta)/2$ and all the small branches are distinct. Then only the small branches remain inside and since they have only simple poles we are able to calculate the above integral with the Residue Theorem 1.14. The residues are

$$\operatorname{Res}\left(\frac{W(u;z)}{u}, u = u_j\right) = \operatorname{Res}\left(\frac{1}{u(1-zS(u))}, u = u_j\right) = -\frac{1}{zu_jS'(u_j)},$$

where we have used Lemma 1.15 with $h(u) = 1/u$ and $g(u) = 1 - zS(u)$. This value can be simplified, since differentiation of the characteristic curve yields

$$\begin{aligned} & -S(u) - zS'(u)u' = 0 \\ \Leftrightarrow \qquad & S'(u) = -\frac{S(u)}{zu'} \overset{(S(u)=\frac{1}{z})}{=} -\frac{1}{z^2u'} \end{aligned}$$

The integration contour can be shrunk to 0, which is legitimate since $W(u;z)$ remains $\mathcal{O}(1)$. Hence it can be chosen small enough so that only small branches contribute and the Residue Theorem gives

$$B(z) = \sum_{j=1}^{c} -\frac{1}{zu_jS'(u_j)} = z\sum_{i=1}^{c} \frac{u_j'(z)}{u_j(z)} \tag{3.8}$$

The same procedure is applicable to

$$W_k(z) = [u^k]W(u;z) = \frac{1}{2\pi i}\int_{|u|=(\alpha+\beta)/2} W(u;z)\frac{\mathrm{d}u}{u^{k+1}}.$$

the integration contour can be shrunk to zero, provided the integrand remains bounded as $u \to 0$. As the integrand is of order $u^{c-k-1}$ this requires $k \leq (c-1)$. Thus we get (3.6) by residue calculation involving small branches. In the same manner as above, the formula is valid in a small neighborhood of the origin. Therefore the identities are a posteriori valid as identities between formal power series.

In the case that $k > -d$ the residue calculation has to be adapted, by extending the contour to a large circle at $\infty$. By doing so, the large branches contribute and this shows (3.7).

It can be easily shown, that algebraic functions are closed under sums, products and multiplicative inverses (compare Theorem 2.17). Therefore, we see by (3.8) that $B(z)$ is algebraic, as all small branches $u_j(z)$ are algebraic. The same argument also shows, that $W_k(z)$ is algebraic. □

***Remark 3.5:*** The algebraic (holonomic) character of $B(z)$ also follows from the fact that $B(z) \equiv W_0(z)$ is equivalently given as the diagonal of a bivariate rational function

$$B(z) = \sum_{n\geq 0}\left([z^nu^{cn}]\frac{1}{1-zu^cS(u)}\right)z^n,$$

which follows immediately from Proposition 2.22. ■

After this short (and superficial) excursion into the field of algebraic curves we turn back to our lattice path problems. More details and a list of references on the theory of algebraic curves are given in [3]. First we show how the introduced theory is applied.

***Example 3.6 (Dyck Prefixes):*** The step set $\mathcal{S} = \{\mathbf{NE}, \mathbf{SE}\} = \{+1, -1\}$ corresponds to the walks of *Dyck prefixes*. The characteristic polynomial is $S(u) = u^{-1} + u$, and hence the characteristic curve reads

$$1 - z\left(\frac{1}{u} + u\right) = 0.$$

We see immediately from the step set, that $c = 1$ and $d = 1$. Therefore, the kernel equation is of degree 2:

$$u - z(1 + u^2) = 0.$$

There exists one small branch and one large branch. In this case, they can be easily computed, by solving the equation of degree 2:

$$\begin{aligned} u_1(z) &= \frac{1 - \sqrt{1 - 4z^2}}{2z} \underset{z \to 0}{\sim} z \\ v_1(z) &= \frac{1 + \sqrt{1 - 4z^2}}{2z} \underset{z \to 0}{\sim} \frac{1}{z} \end{aligned}$$

We see very good how the theory of algebraic curves predicts the solution in this example, compare (3.3). We used the fact, that $\sqrt{1 - 4z^2} = \sum_{n \geq 0} \binom{1/2}{n} (-4)^n z^{2n}$ in a small neighborhood of 0.

But what we really want, is top apply Theorem 3.4. This gives the GF for bridges in this case as

$$B(z) = z \frac{u_1'(z)}{u_1(z)} = \frac{1}{\sqrt{1 - 4z^2}} = 1 + 2z^2 + 6z^4 + 70z^8 + 252z^{10} + \dots$$

The coefficients are known as *EIS* **A000984**[3]

$$[z^n]B(z) = \binom{2n}{n} = [t^n](1 + t^2)^n \tag{3.9}$$

and called *central binomial numbers.* They are closely related to the Catalan numbers. ■

## 3.2 Meanders and Excursions

In this section we restrict the paths to the quarter plane $\mathbb{Z}_+^2$. After this introduction, we pursue this approach much further in Chapter 4, where we consider a special class of walks which are not necessary directed. Walks that stay in the first quadrant are called meanders and such whose final altitude is 0 are called excursions.

Let $f(k; n)$ be the number of meanders of size (i.e. length) $n$ that end at altitude $k$ and use step set $\mathcal{S}$. The corresponding BGF is

$$F(u; z) := \sum_{n,k \geq 0} f(k; n) u^k z^n,$$

which is now an entire series in both $z$ and $u$. With similar arguments as in the analysis of $W(u; z)$ in the previous chapter, one can show that $F(u; z)$ is bivariate analytic for $|u| \leq 1$ and $|z| \leq 1/S(1)$.

[3] Central binomial coefficients; `http://oeis.org/A000984`, accessed 26/08/2013.

We also need the polynomials $F_k(z)$ that describe the possible ways to reach altitude $k$. They are defined by

$$F(u;z) = \sum_{k\geq 0} F_k(z)u^k.$$

Analogously to the last section, we construct the walks recursively: A meander is either the empty one, or it has non-zero length $n+1$. In the second case it is constructed from a meander of length $n$ by appending a possible step from $\mathcal{S}$. But as we are restricted to the first quadrant, we are not allowed to construct a walk, that crosses the $x$-axis. Hence, we must not add a $y$-negative step to a walk of length $n$ that ends on the $x$-axis. This procedure translates directly into the language of generating functions as

$$F(u;z) = \underbrace{1}_{\text{empty path}} + \underbrace{zS(u)F(u;z)}_{\text{append step}} - \underbrace{z[u^<](S(u)F(u;z))}_{\text{paths leaving } \mathbb{Z}_+^2}, \tag{3.10}$$

where $[u^<]$ is the negative part in $u$ from Definition 2.18. This relation is the *fundamental functional equation* defining meanders. Rearranging this relation, reveals the kernel equation (3.1). Note that $S(u)$ involves only a finite number of negative powers (maximal $c$), so that

$$F(u;z)(1-zS(u)) = 1 - z\sum_{k=0}^{c-1} r_k(u)F_k(z), \tag{3.11}$$

for some Laurent polynomials $r_k(u)$ that are immediately computable via (3.10):

$$r_k(u) := [u^<](S(u)u^k) = \sum_{j=-c}^{-k-1} p_j u^{j+k}.$$

***Theorem 3.7 [3, Theorem 2]:*** For a simple set of steps, the BGF of meanders (with $z$ marking size and $u$ marking final altitude) relative to a simple set of paths $\mathcal{S}$ is algebraic. It is given in terms of small and large branches of the characteristic curve of $\mathcal{S}$ by

$$F(u;z) = \frac{\prod_{j=1}^{c}(u-u_j(z))}{u^c(1-zS(u))} = -\frac{1}{p_d z}\prod_{\ell=1}^{d}\frac{1}{u-v_\ell(z)}.$$

In particular the GF of excursions, $E(z) = F(0;z)$, satisfies

$$E(z) = \frac{(-1)^{c-1}}{p_{-c}z}\prod_{j=1}^{c} u_j(z) = -\frac{(-1)^{d-1}}{p_d z}\prod_{\ell=1}^{d}\frac{1}{v_\ell(z)}. \tag{3.12}$$

*Proof:* The main difficulty lies in the fact, that the fundamental equation (3.11) is massively undetermined. It involves the $c$ unknown functions $F_0(z),\ldots,F_{c-1}(z)$ and the also unknown bivariate function $F(u;z)$. The guiding idea is a method known as the *kernel method*. In a nutshell, we try to bind $z$ and $u$ in such a way, that the kernel $1-zS(u)$ and therefore the left-hand side vanishes.

As a first step we remove the negative coefficients of (3.11) by multiplying it with $u^c$. We obtain the entire kernel $K(u,z) = u^c - zu^cS(u)$ known from (3.1). From the discussion of the kernel equation $K(u,z) = 0$ we know, that there exist $c$ small branches $u_1(z),\dots,u_c(z)$ which satisfy this equation. By the theory of algebraic curves we are able to take $|z| < 1/S(1)$ and restrict $z$ to a small neighborhood of the origin in such a way that:

1. all small branches are distinct;
2. all the small branches satisfy $|u_j(z)| < 1$.

This justifies the substitution analytically, and provides us with a system of $c$ equations in the unknowns $F_0,\dots,F_{c-1}$:

$$u_1^c - z\sum_{k=0}^{c-1} u_1^c r_k(u_1)F_k = 0$$

$$\vdots$$

$$u_c^c - z\sum_{k=0}^{c-1} u_c^c r_k(u_c)F_k = 0$$

This linear system in $(F_k)_{k=0}^{c-1}$ is a variant of a Vandermonde matrix. Therefore its determinant is non-zero, as all small branches are distinct and by that it follows that this system is non-singular. Thus, each of the $F_k$ is an algebraic function expressible rationally in terms of the algebraic branches $u_j$.

Next we need an observation of Bousquet-Mélou [8]. Let

$$N(u;z) := u^c - z\sum_{k=0}^{c-1} u^c r_k(u)F_k, \tag{3.13}$$

and observe that (3.13) is a polynomial in $u$ with its roots at the small branches $u_j$. As its leading monomial is $u^c$ it factorizes to

$$N(u;z) = \prod_{j=1}^{c}(u - u_j(z)). \tag{3.14}$$

Now consider the constant term, it is at the same time

- $(-1)^c u_1 \cdots u_c$ (consider above factorization) and
- $-zp_cF_0$, which follows from (3.13) and the fact that only $r_0$ start from $u^{-c}$.

Hence, we get the GF for excursions $E(z) = F(0;z) = F_0(z)$.

The final result for meanders follows from the entire version of (3.11) and from the factorization (3.14)

$$F(u;z) = \frac{N(u;z)}{u^c(1 - zS(u))} = \frac{\prod_{j=1}^{c}(u - u_j(z))}{u^c(1 - zS(u))}.$$

The second identities for $F(u;z)$ and $E(z)$ stated in the theorem follow immediately by taking (3.4) into account. □

It is very easy to deduce the GFs for all paths and meanders from the last two theorems.

***Corollary 3.8 [3, Corollary 1]:*** The GFs of all paths ($W$) and all meanders ($M$) are

$$\begin{aligned} W(z) &= W(1;z) = \frac{1}{1 - zS(1)}, \\ M(z) &= F(1;z) = \frac{1}{1 - zS(1)} \prod_{j=1}^{c} (1 - u_j(z)) = -\frac{1}{p_d z} \prod_{\ell=1}^{d} \frac{1}{1 - v_\ell(z)}. \end{aligned} \tag{3.15}$$

A more interesting and non-trivial connection between bridges and excursions, can be easily deduced from their GFs by comparing (3.5) and (3.12). This is a nice example, of how a GF is able to point out new properties of a problem, as in this case it links two related but not directly connected problems.

***Corollary 3.9 [3, Corollary 2]:*** For the GFs of bridges ($B$) and excursions ($E$) holds

$$\begin{aligned} B(z) &= 1 + z \frac{\mathrm{d}}{\mathrm{d}z} (\log E(z)) = 1 + z \frac{E'(z)}{E(z)} \\ E(z) &= \exp\left( \int_0^z \frac{B(t) - 1}{t} \mathrm{d}t \right). \end{aligned}$$

***Example 3.10 (Dyck Paths and the Ballot Problem):*** Continuing Example 3.6, we ask for the number of paths with the step set $\mathcal{S} = \{+1, -1\}$ that end on the $x$-axis but never leave the first quadrant, i.e. never go below the $x$-axis. In the language of lattice paths, we want to determine the number of excursions for this given step set. We may directly apply Theorem 3.7 as we have already computed the small and large branch above and get for the GF of excursions

$$E(z) = \frac{1 - \sqrt{1 - 4z^2}}{2z^2} = \sum_{n \geq 0} \frac{1}{n+1} \binom{2n}{n} z^{2n} = \sum_{n \geq 0} C_n z^{2n},$$

where the coefficients $C_n$ are the Catalan numbers.

The *ballot problem* asks for the probability in a two candidate election between $A$ and $B$ that eventually ends in a tie, while $A$ is dominating $B$ throughout the poll. This problem can be modeled as a lattice path starting from the origin, with the steps **NE** representing a vote for candidate $A$ and **SE** being a vote for candidate $B$. The fact that it ends in a tie, translates into a walk that ends on the $x$-axis, and the condition of $A$ dominating $B$ is modeled by the restriction, that the walk must not leave the first quadrant. Hence, we are dealing with a Dyck Path.

The total number of possible walks from $(0,0)$ to $(2n, 0)$ is $\binom{2n}{n}$, which are the number of bridges with respect to this step set, compare (3.9). The asked probability is

$$\mathbb{P}(\text{tie, } A \text{ dominates } B \text{ throughout}) = \begin{cases} \frac{1}{n+1}, & \text{2n votes,} \\ 0, & \text{2n+1 votes.} \end{cases}$$

■

***Example 3.11:*** The step set from Figure 3.1 is $\mathcal{S} = \{-1, 0, 1, 2\}$. We know already from the figure, that there will be one small branch of order 1 and two large branches of order $-1/2$. The entire version of the characteristic equation is

$$u - z\left(1 + u + u^2 + u^3\right) = 0.$$

The one small branch is given by

$$u_1(z) = z + z^2 + 2z^3 + 5z^4 + 13z^5 + 36z^6 + 104z^7 + 309z^8 + \dots,$$

and the two large branches are conjugate

$$\begin{aligned} v_1(z) &= z^{1/2} - \tfrac{1}{2} - \tfrac{3}{8}z^{1/2} - \tfrac{1}{2}z - \tfrac{41}{128}z^{3/2} - \tfrac{1}{2}z^2 - \tfrac{763}{1024}z^{3/2} - z^3 + \dots, \\ v_2(z) &= -z^{1/2} - \tfrac{1}{2} + \tfrac{3}{8}z^{1/2} - \tfrac{1}{2}z + \tfrac{41}{128}z^{3/2} - \tfrac{1}{2}z^2 + \tfrac{763}{1024}z^{3/2} - z^3 + \dots. \end{aligned}$$

The first few terms of the GF for excursions are easily computed by (3.12)

$$E(z) = \frac{u_1(z)}{z} = 1 + z + 2z^2 + 5z^3 + 13z^4 + 36z^5 + 104z^6 + 309z^7 + \dots,$$

and similarly for meanders by (3.15)

$$M(z) = \frac{1 - u_1(z)}{1 - 4z} = 1 + 3z + 11z^2 + 42z^3 + 163z^4 + 639z^5 + \dots.$$

Obviously the second representations for $E(z)$ and $M(z)$ in terms of the large branches lead to the same result, but are much more complicated to calculate in this case. ■

## 3.3 Walks confined to the Half-Plane

The derived theory for directed lattice paths can be applied to classify and enumerate more complicated problems. The first generalization is the consideration of *real* 2-dimensional walks, which means that walks are not directed anymore but can vary in both coordinates. In other words, the walks are allowed to go back and forth. Additionally we introduce the restriction, that the walks are confined to the upper half-plane $\mathbb{Z} \times \mathbb{Z}_+$, shortly called half-plane in the sequel. We are going to construct a bijection between all these walks and directed meanders. The ideas of this approach are taken from [26, Chapter 6].

In detail, we are going to show, that the generating functions for walks confined to the half-plane are always algebraic. It holds, that they can be derived automatically using the kernel method, see Section 3.4 for details.

***Definition 3.12:*** A walk $\omega = (\omega_0, \omega_1, \dots, \omega_n)$ of length $n$ constructed from the step set $\mathcal{S} \subset \mathbb{Z}^2$ is called *confined to the half-plane* if all its points $\omega_k = (i, j)$ satisfy $j \geq 0$ for all $i = 1, \dots, n$ (see Figure 3.2). The associated class is called $\mathcal{H}$ and the number of walks of length $n$ is denoted by $h(n)$. ◊

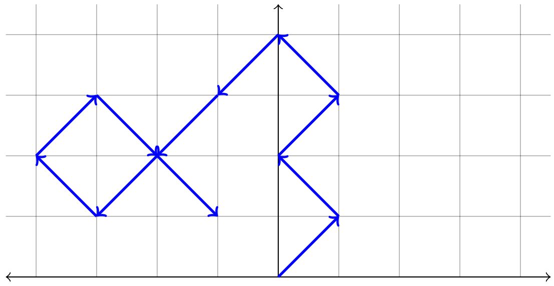


Figure 3.2: Example of Undirected Half-Plane Walk on $\mathcal{S} = \{\mathbf{NE}, \mathbf{NW}, \mathbf{SW}, \mathbf{SE}\}$

In the same way as in the case of directed walks, we are able to construct a recursion on the number $h(n)$ or equivalently on the associated trivariate generating function

$$H(x, y; z) = \sum_{n,j \geq 0; i \in \mathbb{Z}} h(i, j; n) x^i y^j z^n,$$

where $h(x, y; z)$ is the number of walks confined to the half-plane of length $n$ ending at $(i, j)$. Additionally we need a generalization of Definition 3.2:

***Definition 3.13:*** Let $\mathcal{S} = \{(a_1, b_1), \ldots, (a_k, b_k)\}$ be a set of steps and $\Pi = \{w_1, \ldots, w_k\}$ be the corresponding system of weights. The *characteristic polynomial* of $\mathcal{S}$ is the Laurent polynomial $T(x, y)$, defined as

$$T(x, y) := \sum_{i=1}^{k} w_i x^{a_i} y^{b_i}.$$

$\diamondsuit$

Now we are able to construct the functional equation on $H$: A walk is either empty or it is constructed from a walk of length one less, by appending a step from $\mathcal{S}$. But we have to be careful not to leave the half-plane.

$$H(x, y; z) = 1 + zT(x, y)H(x, y; z) - z[y^<](T(x, y)H(x, y; z)) \tag{3.16}$$

In essence, this is the same construction rule, which led to the functional equation on meanders in (3.10) and therefore the functional equations are quite similar. This is also the main observation, which will lead to the wanted bijection.

***Definition 3.14:*** The *horizontal projection* of a step set $\mathcal{S}$ with characteristic polynomial $T(x, y)$ is given by the (weighted) directed step set $\mathcal{S}_h$ with characteristic polynomial $S_h(y) = T(1, y)$. $\diamondsuit$

In other words we "forget" the $x$-coordinate of every step by setting $x = 1$. By doing so, we may get more steps in the same direction, which explains the possible weights. These weights can be understood like different colors for the same step in order to make them distinguishable but we obtain a *directed* step set.

***Example 3.15:*** Figure 3.4 shows how the horizontal projection transforms a walk confined to the half-plane into a directed walk, in particular a meander. Due to the coloring the transformation is reversible, and in particular it is the wanted bijection (see Proposition 3.16). The step set $\mathcal{S} = \{\mathbf{NE}, \mathbf{NW}, \mathbf{SW}, \mathbf{SE}\}$ is transformed into the step set $\mathcal{S}_h = \{\mathbf{NE}, \mathbf{SE}\}$ with weights $\Pi = (2, 2)$ or two colors for each step, represented by the dashed red and straight blue line (see Figure 3.3).

Note, that the transformation can also be interpreted as a flip of the $x$-negative steps along a vertical line. This is also a possible way to construct the associated meander, by iterative "flipping" of $x$-negative steps into $x$-positive ones starting from the origin. ∎

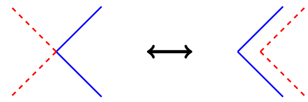


Figure 3.3: Horizontal Projection of $\mathcal{S} = \{\mathbf{NE}, \mathbf{NW}, \mathbf{SW}, \mathbf{SE}\}$

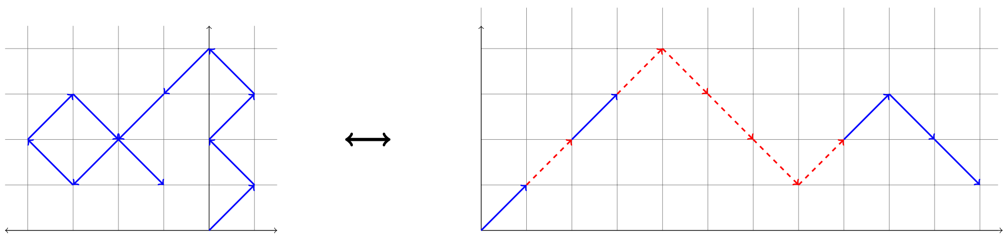


Figure 3.4: Horizontal Projection of Half-Plane Walk into its associated Meander

***Proposition 3.16:*** The class $\mathcal{H}$ of undirected half-plane walks on $\mathcal{S}$ is bijectively equivalent to the class $\mathcal{M}$ of directed meanders on $\mathcal{S}_h$. In particular is the length generating function of undirected half-plane walks always algebraic.

*Proof:* The functional equation for the GF of undirected half-plane walks $H(x, y, z)$ was derived in (3.16). The horizontal projections transforms it into

$$H(1, y; z) = 1 + zT(1, y)H(1, y; z) - z[y^<](T(1, y)H(1, y; z)).$$

Define the projection of the characteristic polynomial to be $S(y) := T(1, y)$, then we get

$$H(1, y; z) = 1 + zS(y)H(1, y; z) - z[y^<](S(y)H(1, y; z)).$$

But at the same time this equation is satisfied by $F_{\mathcal{S}_h}(y; z)$ which is the GF of meanders on the step set $\mathcal{S}_h$. Thus, $H(1, y; z)$ and $F_{\mathcal{S}_h}(y)$ satisfy the same functional equation and because of that their coefficients satisfy the same recursion. In addition the first few terms agree which implies that the counting sequences are the same, compare Definition 2.3.

In other words

$$H(1, y; z) \equiv F_{\mathcal{S}_h}(y; z).$$

The algebraic character of the length GF follows now directly from Theorem 3.7. □

***Remark 3.17:*** The above proof suppresses the $x$-coordinate by setting it to 1. This operation of horizontal projection is well suited to solve this problem, as only the $y$-coordinate of every step is significant in order to impose the half-space restriction. ■

***Example 3.18:*** Let's investigate the step set $\mathcal{S} = \{\mathbf{NE}, \mathbf{NW}, \mathbf{SW}, \mathbf{SE}\}$ of Example 3.15 further. In particular we want to derive the number of such walks of length $n$, i.e. $h(n)$. The characteristic polynomial of $\mathcal{S}$ is

$$T(x,y) = xy + \frac{x}{y} + \frac{y}{x} + \frac{1}{xy},$$

hence, the associated functional equation (3.16) for $H(x,y;z)$ is given by

$$H(x,y;z) = 1 + z\left(xy + \frac{x}{y} + \frac{y}{x} + \frac{1}{xy}\right)H(x,y;z) - z\left(\frac{x}{y} + \frac{1}{xy}\right)H(x,0;z).$$

The horizontal projection introduces weights on the step set and yields

$$H(1,y;z) = 1 + z\left(2y + \frac{2}{y}\right)H(1,y;z) - z\left(\frac{2}{y}\right)H(1,0;z).$$

By rearranging this equation we obtain the kernel equation (3.1) and we get

$$\underbrace{\left(1 - z\left(2y + \frac{2}{y}\right)\right)}_{=K(y,z)/y} H(1,y;z) = 1 - \frac{2z}{y}H(1,0;z).$$

Now we can apply the full theory of the previous sections. Note, that the kernel $K(y,z)$ is chosen to be entire, which is why the value in the equation above is divided by $y$. Solving $K(y,z) = 0$ gives the two branches

$$u_1(z) = \frac{1 - \sqrt{1 - 16z^2}}{4z} \qquad\qquad v_1(z) = \frac{1 + \sqrt{1 - 16z^2}}{4z}.$$

Corollary 3.8 tells us immediately the wanted GF for the number of walks confined to the half-plane with length $n$

$$H(1,1;z) = \frac{2}{1 - 4z + \sqrt{1 - 16z^2}} = 1 + 2z + 8z^2 + 24z^3 + 96z^4 + \ldots.$$

For the power series expansion Newton's expansion theorem was used. ■

## 3.4 Kernel Method and Linear Recurrences

We first give a short introduction on the origins of the kernel method. According to Banderier and Flajolet [3, pp. 55] the "kernel method" has been a part of the folklore of combinatorialists for some time. First references deal with a functional equation of the form

$$K(u,z)F(u;z) = A(u;z) + B(u;z)G(z) \tag{3.17}$$

with $F$, $G$ being the unknown functions and when there is only one small branch $u_1$, such that $K(u_1(z), z) = 0$. In this case a simple substitution solves the problem and $G(z) = -A(u_1(z); z)/B(u_1(z); z)$. One clear source of this is the exercise section of Knuth's book [29], published in 1968: The detailed solution to Exercise 2.2.1.1-4 presents a "new method for solving the ballot problem", for which the characteristic equation is quadratic. Another application can be found in Exercise 2.2.1.11 of the very same book.

But the topic is still alive in the mathematical community. The technique proved especially valuable in lattice path enumeration problems, as we have already seen in the previous sections. One very important contribution from the end of the last century comes from Bousquet-Mélou and Petkovšek whose paper offers deep insights into the usefulness of the technique if applied to multi-dimensional walks and recurrences [8]. We are going to discuss selected parts, which are especially useful for our goal of lattice path enumeration.

Finally it should be noted, that according to Banderier and Flajolet probabilists had known a lot since the early 1950s regarding related questions. The technique seems to share strong parallels with the so-called Wiener-Hopf approach. But for more details and references to related literature we refer to [3, p. 56].

In the remainder of this section we will discuss the results from [8]. We have seen in the previous results, that the enumeration of every lattice path problem starts with a recursive construction of the concerned walks. From this construction arises naturally a linear recurrence relation with constant coefficients, from which the associated generating functions are constructed. Hence we see, that the nature of the GF depends completely on the nature of the underlying recurrence relation. This is the reason why we focus our studies on these objects now.

First some words on the used notation:

- We write $\boldsymbol{u} = (u_1, u_2, \ldots, u_d)$ for $d$-tuples of numbers or indeterminates like: $\mathbf{0} = (0, 0, \ldots, 0)$, $\mathbf{1} = (1, 1, \ldots, 1)$.
- Vector-valued inequalities are defined in the usual way as $\boldsymbol{u} \geq \boldsymbol{v}$ when $u_i \geq v_i$ for all $1 \leq i \leq d$ and $\boldsymbol{u} < \boldsymbol{v}$ when $u_i < v_i$ for all $1 \leq i \leq d$.
- The monomial $x_1^{u_1} \cdots x_d^{u_d}$ is denoted as $\boldsymbol{x}^{\boldsymbol{u}}$.
- For $\boldsymbol{u}, \boldsymbol{v} \in \mathbb{Z}^d$ the scalar product $u_1 v_1 + \ldots + u_d v_d$ is denoted as $\boldsymbol{u} \cdot \boldsymbol{v}$.
- The convex hull of a set $H \subseteq \mathbb{R}^d$ is denoted as $\operatorname{conv} H$.

The general form of a *linear recurrence relation* with constant coefficients is

$$a_{\boldsymbol{n}} = c_{\boldsymbol{h}_1} a_{\boldsymbol{n}+\boldsymbol{h}_1} + c_{\boldsymbol{h}_2} a_{\boldsymbol{n}+\boldsymbol{h}_2} + \ldots + c_{\boldsymbol{h}_k} a_{\boldsymbol{n}+\boldsymbol{h}_k} \qquad \text{for } \boldsymbol{n} \geq \boldsymbol{s}.$$

## Existence and Uniqueness

In the beginning we ask ourselves, what conditions are necessary in order to guarantee the existence and uniqueness of a solution. These conditions can be derived in a more general context.

***Definition 3.19:*** Let $A$ be a non-empty set. A $d$-dimensional *recurrence equation* is of the form

$$a_{\boldsymbol{n}} = \Phi(a_{\boldsymbol{n}+\boldsymbol{h}_1}, a_{\boldsymbol{n}+\boldsymbol{h}_2}, \ldots, a_{\boldsymbol{n}+\boldsymbol{h}_k}) \qquad \text{for } \boldsymbol{n} \geq \boldsymbol{s}, \tag{3.18}$$

where $a : \mathbb{N}^d \to A$ is the unknown $d$-dimensional sequence of elements of $A$, $\Phi : A^k \to A$ is a given function, $H = \{\boldsymbol{h}_1, \boldsymbol{h}_2, \ldots, \boldsymbol{h}_k\} \subseteq \mathbb{Z}^d$ is the set of *shifts*, and $\boldsymbol{s} \in \mathbb{N}^d$ is the *starting point* satisfying $\boldsymbol{s} + H \subseteq \mathbb{N}^d$. A given function $\varphi$ specifies the *initial conditions*:

$$a_{\boldsymbol{n}} = \varphi(\boldsymbol{n}) \qquad \text{for } \boldsymbol{n} \geq \boldsymbol{0},\ \boldsymbol{n} \not\geq \boldsymbol{s}. \tag{3.19}$$

$\Diamond$

The idea of the proof is to characterize the set $H$ and find conditions for which there is an ordering of $\mathbb{N}^d$ of order type[4] $\omega$ such that the points $\boldsymbol{n} + \boldsymbol{h}_1, \boldsymbol{n} + \boldsymbol{h}_2, \ldots, \boldsymbol{n} + \boldsymbol{h}_k$ precede $\boldsymbol{n}$ in this ordering. The conclusion will be, that then there exists a unique solution of (3.18)-(3.19) and for any $\boldsymbol{n} \in \mathbb{N}^d$ it is possible to compute the value of $a_{\boldsymbol{n}}$ directly from these equations in a finite number of steps.

***Theorem 3.20 [8, Theorem 5]:*** Let $H \subseteq \mathbb{Z}^d$ be a non-empty set which satisfies

$$\{\boldsymbol{x} \in \mathbb{R}^d : \boldsymbol{x} \geq \boldsymbol{0}\} \cap \operatorname{conv} H = \emptyset. \tag{3.20}$$

Then there exists a unique $d$-dimensional sequence $a : \mathbb{N}^d \to A$ which satisfies (3.18)-(3.19).

We omit the proof here, because it brings no new insights on our applications.

***Remark 3.21:*** In [8] it is shown that condition (3.20) is equivalent to the fact that there exists $\boldsymbol{v} \in \mathbb{N}^d$, $\boldsymbol{v} > 0$, such that $\boldsymbol{v} \cdot \boldsymbol{h} < 0$ for all $\boldsymbol{h} \in H$. Furthermore it is also equivalent to the existence of an ordering that depends on $H$ that is well founded or can be extended to an ordering of $\mathbb{N}^d$ of order type $\omega$ (compare [8, Theorem 3]). We will not need these characterizations. ■

**<u>The nature of the solution</u>**

Due to the fact, that lattice path enumeration problems involve only linear recurrences with constant coefficients we turn back to these problems. Hence we study the recurrence relation

$$a_{\boldsymbol{n}} = \begin{cases} \sum_{\boldsymbol{h} \in H} c_{\boldsymbol{h}} a_{\boldsymbol{n}+\boldsymbol{h}}, & \text{for } \boldsymbol{n} \geq \boldsymbol{s}, \\ \varphi(\boldsymbol{n}), & \text{for } \boldsymbol{n} \geq 0 \text{ and } \boldsymbol{n} \not\geq \boldsymbol{s}, \end{cases} \tag{3.21}$$

where $(c_{\boldsymbol{h}})_{\boldsymbol{h} \in H}$ are given non-zero constants from $A$. Let $A$ be a field of characteristic zero and $H$ be a finite non-empty set from $\mathbb{Z}^d$ satisfying (3.20). We assume $\boldsymbol{s} \in \mathbb{N}^d$ and $\boldsymbol{s} + H \subseteq \mathbb{N}^d$.

Let $a$ be the unique solution of the above recurrence. The corresponding GF is given by

$$F_{\boldsymbol{s}}(\boldsymbol{x}) = \sum_{\boldsymbol{n} \geq \boldsymbol{s}} a_{\boldsymbol{n}} \boldsymbol{x}^{\boldsymbol{n}-\boldsymbol{s}}.$$

As a first step we transform the recurrence relation into a functional equation satisfied by the GF $F_{\boldsymbol{s}}(\boldsymbol{x})$. We proceed the standard way: Multiply (3.21) with $\boldsymbol{x}^{\boldsymbol{n}-\boldsymbol{s}}$ and sum over all $\boldsymbol{n} \geq \boldsymbol{s}$:

$$\begin{aligned} F_{\boldsymbol{s}}(\boldsymbol{x}) &= \sum_{\boldsymbol{h} \in H} c_{\boldsymbol{h}} \sum_{\boldsymbol{n} \geq \boldsymbol{s}} a_{\boldsymbol{n}+\boldsymbol{h}} \boldsymbol{x}^{\boldsymbol{n}-\boldsymbol{s}} = \sum_{\boldsymbol{h} \in H} c_{\boldsymbol{h}} \boldsymbol{x}^{-\boldsymbol{h}} \sum_{\boldsymbol{n} \geq \boldsymbol{s}+\boldsymbol{h}} a_{\boldsymbol{n}} \boldsymbol{x}^{\boldsymbol{n}-\boldsymbol{s}} \\ &= \sum_{\boldsymbol{h} \in H} c_{\boldsymbol{h}} \boldsymbol{x}^{-\boldsymbol{h}} \left( F_{\boldsymbol{s}}(\boldsymbol{x}) + P_{\boldsymbol{h}}(\boldsymbol{x}) - M_{\boldsymbol{h}}(\boldsymbol{x}) \right) \end{aligned} \tag{3.22}$$

[4] The ordinal number $\omega$ is the least infinite one and identified with the cardinal number $\aleph_0$.

where

$$P_{\boldsymbol{h}}(\boldsymbol{x}) = \sum_{\substack{\boldsymbol{n} \not\geq \boldsymbol{s} \\ \boldsymbol{n} \geq \boldsymbol{s}+\boldsymbol{h}}} a_{\boldsymbol{n}} \boldsymbol{x}^{\boldsymbol{n}-\boldsymbol{s}} = \sum_{\substack{\boldsymbol{n} \not\geq \boldsymbol{s} \\ \boldsymbol{n} \geq \boldsymbol{s}+\boldsymbol{h}}} \varphi(\boldsymbol{n}) \boldsymbol{x}^{\boldsymbol{n}-\boldsymbol{s}} \quad \text{and} \tag{3.23}$$

$$M_{\boldsymbol{h}}(\boldsymbol{x}) = \sum_{\substack{\boldsymbol{n} \geq \boldsymbol{s} \\ \boldsymbol{n} \not\geq \boldsymbol{s}+\boldsymbol{h}}} a_{\boldsymbol{n}} \boldsymbol{x}^{\boldsymbol{n}-\boldsymbol{s}}. \tag{3.24}$$

With these definitions we can express the simple relation between $F_{\boldsymbol{s}}(\boldsymbol{x})$ and the full GF $F(\boldsymbol{x})$ explicitly:

$$F(\boldsymbol{x}) = \sum_{\boldsymbol{n} \geq \boldsymbol{0}} a_{\boldsymbol{n}} \boldsymbol{x}^{\boldsymbol{n}} = \boldsymbol{x}^{\boldsymbol{s}} \left( \sum_{\boldsymbol{n} \geq \boldsymbol{0}} a_{\boldsymbol{n}} \boldsymbol{x}^{\boldsymbol{n}-\boldsymbol{s}} + \sum_{\substack{\boldsymbol{n} \not\geq \boldsymbol{s} \\ \boldsymbol{n} \geq \boldsymbol{0}}} a_{\boldsymbol{n}} \boldsymbol{x}^{\boldsymbol{n}-\boldsymbol{s}} \right) = \boldsymbol{x}^{\boldsymbol{s}} \left( F_{\boldsymbol{s}}(\boldsymbol{x}) + P_{-\boldsymbol{s}}(\boldsymbol{x}) \right). \tag{3.25}$$

Next we rewrite equation (3.22) with the structure of (3.17) in mind into

$$\left( 1 - \sum_{\boldsymbol{h} \in H} c_{\boldsymbol{h}} \boldsymbol{x}^{-\boldsymbol{h}} \right) F_{\boldsymbol{s}}(\boldsymbol{x}) = \sum_{\boldsymbol{h} \in H} c_{\boldsymbol{h}} \boldsymbol{x}^{-\boldsymbol{h}} \left( P_{\boldsymbol{h}}(\boldsymbol{x}) - M_{\boldsymbol{h}}(\boldsymbol{x}) \right). \tag{3.26}$$

To clear denominators on the left-hand side we introduce the notion of an *apex* which, as we will see shortly, is strongly related to the nature of the GF.

***Definition 3.22:*** Let $H \subset \mathbb{Z}^d$ be a finite set. The *apex* of $H$ is the point $\boldsymbol{p} = (p_1, p_2, \ldots, p_d) \in \mathbb{N}^d$, such that

$$p_i := \max\{h_i : \boldsymbol{h} \in H \cup \{\boldsymbol{0}\}\} \qquad i = 1, 2, \ldots, d.$$

$\Diamond$

Multiplying (3.26) by $\boldsymbol{x}^{\boldsymbol{p}}$ yields[5]

$$K(\boldsymbol{x}) F_{\boldsymbol{s}}(\boldsymbol{x}) = A(\boldsymbol{x}) - G(\boldsymbol{x}) \tag{3.27}$$

where

$$\begin{aligned} K(\boldsymbol{x}) &= \boldsymbol{x}^{\boldsymbol{p}} - \sum_{\boldsymbol{h} \in H} c_{\boldsymbol{h}} \boldsymbol{x}^{\boldsymbol{p}-\boldsymbol{h}}, \\ A(\boldsymbol{x}) &= \sum_{\boldsymbol{h} \in H} c_{\boldsymbol{h}} \boldsymbol{x}^{\boldsymbol{p}-\boldsymbol{h}} P_{\boldsymbol{h}}(\boldsymbol{x}), \\ G(\boldsymbol{x}) &= \sum_{\boldsymbol{h} \in H} c_{\boldsymbol{h}} \boldsymbol{x}^{\boldsymbol{p}-\boldsymbol{h}} M_{\boldsymbol{h}}(\boldsymbol{x}). \end{aligned}$$

The definition of the apex provides that $K(\boldsymbol{x})$ is a polynomial in $\boldsymbol{x}$ which is called the *kernel* of the recursion (compare Lemma 4.12). Note that the coefficients of $K(\boldsymbol{x})$ and $A(\boldsymbol{x})$ are given directly by the coefficients of the recurrence relation and by the initial conditions, respectively.

[5] In the original paper [8] a different naming convention is used: $K(\boldsymbol{x})$ is called $Q(\boldsymbol{x})$, $A(\boldsymbol{x})$ is called $K(\boldsymbol{x})$ and $G(\boldsymbol{x})$ is called $U(\boldsymbol{x})$. We want to stick with the intuitive convention of calling the kernel $K(\boldsymbol{x})$ here.

The coefficients of $G(\boldsymbol{x})$ can be computed from (3.21) but are not given explicitly. Therefore we call $A(\boldsymbol{x})$ the *known initial function* and $G(\boldsymbol{x})$ the *unknown initial function.*

If we now look at the functional equation (3.27) again (which is equivalent to our initial problem (3.21)) it seems like we have rather made the problem more complicated because there are now two unknown functions $F_{\boldsymbol{s}}(\boldsymbol{x})$ and $G(\boldsymbol{x})$. But we have also gained something: The initial conditions are now implicitly included and we have only one equation to deal with. Moreover, there is a strong connection between the two unknown functions, given over the kernel. The overall idea is, that if we are able to find $G(\boldsymbol{x})$ explicitly then the GF of the unique solution to (3.21) is given by

$$F_{\boldsymbol{s}}(\boldsymbol{x}) = \frac{A(\boldsymbol{x}) - G(\boldsymbol{x})}{K(\boldsymbol{x})}. \tag{3.28}$$

***Definition 3.23:*** Let $F(x_1, \ldots, x_d) = \sum_{\boldsymbol{n} \geq \boldsymbol{0}} a_{n_1,\ldots,n_d} \boldsymbol{x}^{\boldsymbol{n}}$ be a formal power series in $d$ variables. A *section* of $F$ is any sub-series of $F$ obtained by fixing some of the indices of $\boldsymbol{n} = (n_1, \ldots, n_d)$. ◊

Sections are important in this context, as all sections of rational (resp. algebraic, holonomic) series are also rational (resp. algebraic, holonomic). For details in the holonomic case we refer to [34, Proposition 2.5]. We start the analysis of (3.27) with some simple observations.

***Proposition 3.24 [8, Proposition 11]:*** Let $F_{\boldsymbol{s}}(\boldsymbol{x})$ be the GF of the unique solution of (3.21). Then the series $F_{\boldsymbol{s}}(\boldsymbol{x})$ is rational (resp. algebraic, holonomic) if and only if both its unknown initial function $A(\boldsymbol{x})$ and $G(\boldsymbol{x})$ are rational (resp. algebraic, holonomic).

*Proof:* Remember that rational, algebraic and holonomic power series are closed under the sum and the product, respectively (compare Theorem 2.21). Hence, if $A$ and $G$ are rational (resp. algebraic, holonomic) then by (3.28) $F_{\boldsymbol{s}}$ is also rational (resp. algebraic, holonomic).

Conversely, observe that for $\boldsymbol{h} \in H$, the series $M_{\boldsymbol{h}}$ from (3.24) is a finite linear combination of sections of $F_{\boldsymbol{s}}$. Consequently the same holds for $G$. Hence, if $F_{\boldsymbol{s}}$ is rational (resp. algebraic, holonomic), then so is $G(\boldsymbol{x})$. Finally (3.27) implies that the same holds for $A(\boldsymbol{x})$. □

***Remark 3.25:*** Above argumentation for $M_{\boldsymbol{h}}$ cannot be used for $P_{\boldsymbol{h}}$, as by (3.23) it is a linear combination of sections of the full GF $F(\boldsymbol{x}) = \sum_{\boldsymbol{n} \geq \boldsymbol{0}} a_{\boldsymbol{n}} \boldsymbol{x}^{\boldsymbol{n}}$, but not of sections of $F_{\boldsymbol{s}}(\boldsymbol{x})$. ■

The last proposition tells us that the nature of $F_{\boldsymbol{s}}$ is completely determined by the nature of $G$ because the nature of $A$ is known beforehand as it depends solely on the initial values given by the function $\varphi$ and is therefore explicitly known.

***Theorem 3.26 [8, Theorem 12]:*** Assume the apex $\boldsymbol{p}$ of $H$ is $\boldsymbol{0}$. Then the GF $F_{\boldsymbol{s}}(\boldsymbol{x})$ of the unique solution of (3.21) is rational if and only if the known initial function $A(\boldsymbol{x})$ itself is rational.

*Proof:* The assumption guarantees that for each $\boldsymbol{h} \in H$ we have $\boldsymbol{h} \leq \boldsymbol{0}$ and therefore $\boldsymbol{s}+\boldsymbol{h} \leq \boldsymbol{s}$ and $M_{\boldsymbol{h}}(\boldsymbol{x}) = \boldsymbol{0}$, compare (3.24). This implies $G(\boldsymbol{x}) = 0$ and so (3.28) simplifies to

$$F_{\boldsymbol{s}}(\boldsymbol{x}) = \frac{A(\boldsymbol{x})}{K(\boldsymbol{x})}.$$

We know that $K(\boldsymbol{x})$ is a polynomial in $\boldsymbol{x}$ and it follows that $F_{\boldsymbol{s}}(\boldsymbol{x})$ is rational if and only if $A(\boldsymbol{x})$ is. □

***Example 3.27 (Binomial Coefficients):*** Let's shade some light on the theory which was covered so far by considering an easy example on the well-known binomial coefficients. Therefore, let $n$ and $k$ be integers such that $0 \leq k \leq q$. We want to determine in how many ways we can choose a subset of $k$ objects from the set $\{1, 2, \ldots, n\}$? For now, we pretend that we don't know the answer.

Let $a_{n,k}$ be the answer to this question. First we need to derive a recurrence relation: Consider the collection of all possible subsets of $k$ of these $n$ objects. Next we fix the element $n$. If a subset of this collection contains the element $n$ it contains $k-1$ elements out of the set $\{1, 2, \ldots, n-1\}$. If it does not contain the element $n$ it consists out of $k$ elements from $\{1, 2, \ldots, n-1\}$. Therefore we obtain the recursion

$$a_{n,k} = a_{n-1,k-1} + a_{n-1,k}, \qquad \text{for } n, k \geq 1,$$

with the initial conditions $\varphi(m, 0) = a_{m,0} = 1$ for all $m \geq 0$ and $\varphi(0, \ell) = a_{0,\ell} = 0$ for all $\ell > 0$.

The set of shifts is given by $H = \{(-1, -1), (-1, 0)\}$ and the starting point by $\boldsymbol{s} = (1, 1)$. Hence, condition (3.20) is satisfied and there exists a unique solution to this problem.

The apex $\boldsymbol{p}$ is $(0, 0)$ and due to Theorem 3.26 the solution $F_{\boldsymbol{s}} = F_{(1,1)}$ is rational if and only if $A$ is rational. We want to check this condition.

Computing $P_{\boldsymbol{h}}$ for $\boldsymbol{h} \in H$ via (3.23) gives

$$P_{(-1,-1)}(x, y) = \sum_{\substack{(n,k) \ngeq (1,1) \\ (n,k) \geq (0,0)}} \varphi(n, k) x^{n-1} y^{k-1} = \sum_{\substack{n \geq 0 \\ k=0}} x^{n-1} y^{k-1} = \frac{1}{xy} \frac{1}{1-x},$$

$$P_{(-1,0)}(x, y) = \sum_{\substack{(n,k) \ngeq (1,1) \\ (n,k) \geq (0,1)}} \varphi(n, k) x^{n-1} y^{k-1} \equiv 0.$$

Therefore we get for the known initial function $A(x, y)$

$$A(x, y) = 1 \cdot xy P_{(-1,-1)}(x, y) + 1 \cdot x P_{(-1,0)}(x, y) = \frac{1}{1-x}.$$

Thus, Theorem 3.26 implies that $F_{\boldsymbol{s}}$ is rational. The kernel is given by

$$K(x, y) = 1 - x - xy$$

and due to that $F_{\boldsymbol{s}}$ is by (3.28) equal to

$$F_{\boldsymbol{s}}(x, y) = F_{(1,1)}(x, y) = \frac{1}{1-x} \frac{1}{1-x-xy}.$$

To check this answer, it is easy to compute the complete GF $F(x, y)$ by

$$F(x, y) = \sum_{n,k \geq 0} \binom{n}{k} x^n y^k = \sum_{n \geq 0} \left( \sum_{k=0}^{n} \binom{n}{k} y^k \right) x^n = \sum_{n \geq 0} (1+y)^n x^n = \frac{1}{1-x-xy}.$$

An easy computation shows, that our solution is correct. See also (3.25) for a connection between $F_{\boldsymbol{s}}$ and $F$. ∎

Obviously not all GFs are rational. The next theorem generalizes the last result and states a sufficient condition for algebraic GFs.

***Theorem 3.28 [8, Theorem 13]:*** Take $A = \mathbb{C}$ and assume that the apex $\boldsymbol{p}$ of $H$ has at most one positive coordinate. Then the GF $F_{\boldsymbol{s}}(\boldsymbol{x})$ of the unique solution of (3.21) is algebraic if and only if the known initial function $A(\boldsymbol{x})$ itself is algebraic.

*Proof:* Due to Proposition 3.24 it holds that if $F_{\boldsymbol{s}}$ is algebraic then so is $A$.

If $\boldsymbol{p} = \boldsymbol{0}$, then the proof is similar to that of Theorem 3.26. Assume now that exactly one coordinate of $\boldsymbol{p}$ is positive. Without loss of generality we assume that $p_1 = \ldots = p_{d-1} = 0$ and $p_d > 0$. The idea of the following proof is to investigate this special coordinate $x_d$. First we note

$$\begin{aligned} G(\boldsymbol{x}) &= \sum_{\boldsymbol{h}\in H} c_{\boldsymbol{h}} \boldsymbol{x}^{\boldsymbol{p}-\boldsymbol{h}} \sum_{\substack{\boldsymbol{n}\geq \boldsymbol{s} \\ \boldsymbol{n}\ngeq \boldsymbol{s}+\boldsymbol{h}}} a_{\boldsymbol{n}} \boldsymbol{x}^{\boldsymbol{n}-\boldsymbol{s}} \\ &= \sum_{\substack{\boldsymbol{h}\in H \\ h_d>0}} c_{\boldsymbol{h}} \boldsymbol{x}^{\boldsymbol{p}-\boldsymbol{h}} \sum_{n_d=s_d}^{s_d+h_d-1} \sum_{(n_1,\ldots,n_{d-1})\geq(s_1,\ldots,s_{d-1})} a_{\boldsymbol{n}} \boldsymbol{x}^{\boldsymbol{n}-\boldsymbol{s}}. \end{aligned}$$

Observe that all exponents of $x_d$ are positive, which is the reason that $G(\boldsymbol{x})$ is a polynomial in $x_d$ of degree at most $p_d - 1$. Rearranging (3.27) yields another representation of $G$:

$$G(\boldsymbol{x}) = A(\boldsymbol{x}) - K(\boldsymbol{x}) F_{\boldsymbol{s}}(\boldsymbol{x}), \tag{3.29}$$

with the kernel

$$K(\boldsymbol{x}) = x_d^{p_d} - \sum_{\boldsymbol{h}\in H} c_{\boldsymbol{h}} \boldsymbol{x}^{\boldsymbol{p}-\boldsymbol{h}},$$

which is a polynomial in $x_d$ of degree at most $p_d$. Next we want to apply what is commonly known as the "kernel method": We are going to prove that $K(\boldsymbol{x})$ regarded as polynomial in $x_d$ admits (at least) $p_d$ roots $\xi_i(x_1, \ldots, x_{d-1})$, counted with multiplicities, such that

$$\xi_i(0, \ldots, 0) = 0.$$

This condition is necessary, as we want to substitute $x_d$ with $\xi_i$ in (3.29). But this equation, interpreted as equation between converging power series, is only valid in some neighborhood of the origin. (Remark, it can be shown that $F_{\boldsymbol{s}}$ is analytic in a neighborhood of the origin, if there exist constants $m > 0$ and $\boldsymbol{u} \in \mathbb{R}^d$ such that $|\varphi(\boldsymbol{n})| \leq m^{\boldsymbol{u}\cdot\boldsymbol{n}}$ for all $\boldsymbol{n} \ngeq \boldsymbol{s}$, compare [8, Theorem 7].)

The existence of the $\xi_i$ follows directly from the observation that

$$K(0, \ldots, 0, x_d) = x_d^{p_d} - \sum_{\substack{\boldsymbol{h}\in H \\ h_1=\ldots=h_{d-1}=0}} c_{\boldsymbol{h}} x_d^{p_d-h_d}.$$

Because of the overall assumption (3.20) we have $\{\boldsymbol{x} \in \mathbb{R}^d : \boldsymbol{x} \geq 0\} \cap H = \emptyset$ and therefore $h_d < 0$ for all $\boldsymbol{h} \in H$ with $h_1 = \ldots = h_{d-1} = 0$ if such exist. This implies $p_d - h_d > p_d$ for all such $\boldsymbol{h}$ and therefore is $x_d = 0$ a root of $K(0, \ldots, 0, x_d)$ of multiplicity $p_d$. If no such $\boldsymbol{h} \in H$ exists, this is also true as in this case the sum vanishes.

Now we can replace $x_d$ by $\xi_i(x_1, \ldots, x_{d-1})$ in (3.29) and obtain, if $\xi_i$ is a root of multiplicity $m$, that

$$G(\xi_i) = A(\xi_i), \qquad \frac{\partial}{\partial x_d} G(\xi_i) = \frac{\partial}{\partial x_d} A(\xi_i), \qquad \ldots, \qquad \frac{\partial^{m-1}}{\partial x_d^{m-1}} G(\xi_i) = \frac{\partial^{m-1}}{\partial x_d^{m-1}} A(\xi_i),$$

where $G(\xi_i)$ is short for $G(x_1, \ldots, x_{d-1}, \xi_i(x_1, \ldots, x_{d-1}))$ and $A(\xi_i)$ analogously. Thus, the $p_d$ roots of $K$ provide a total of $p_d$ equations for the polynomial $G$ of degree at most $p_d - 1$. So we are able to reconstruct $G$ by means of the Hermite interpolation formula, which simplifies if $G$ has no multiple roots to the special case of the well-known Lagrange interpolation formula.

Because the $\xi_i$ are algebraic functions of $x_1, \ldots, x_{d-1}$ (they are the roots of the polynomial $K$ in $x_d$) this shows that $G(\boldsymbol{x})$ is algebraic provided $A(\boldsymbol{x})$ is. By Proposition 3.24 the same holds for $F_{\boldsymbol{s}}(\boldsymbol{x})$. □

***Remark 3.29:*** The above proof not only shows existence but is also constructive in the sense, that it provides an algorithm to compute the GF of the solution. However, it involves the explicit knowledge of the roots of the kernel equation, and as this equation could have quite high degree, it need not to be an efficient or even applicable algorithm.

In the case that the known initial function $A(\boldsymbol{x})$ itself is a polynomial (what happens quite often in enumerative combinatorics), the polynomial $A(\boldsymbol{x}) - G(\boldsymbol{x})$ has at least $p_d$ roots (namely $\xi_1, \ldots, \xi_{p_d}$, compare (3.27)). Under the conditions of Theorem 3.28 the polynomial $G(\boldsymbol{x})$ has degree at most $p_d - 1$ in $x_d$ and therefore $A$ has degree at least $p_d$. If $A$ has exactly degree $p_d$ then

$$A(\boldsymbol{x}) - G(\boldsymbol{x}) = \mathrm{lc}(A) \prod_{i=1}^{p_d} (x_d - \xi_i(x_1, \ldots, x_{d-1})),$$

where $\mathrm{lc}(A)$ denotes the leading coefficient of $K$ with respect to $x_d$. The degree of $K$ in $x_d$ is $p_d + r$, where $r = \max\{-h_d : \boldsymbol{h} \in H \cup \{\boldsymbol{0}\}\}$. The kernel $K$ has the $p_d$ roots $\xi_1, \ldots, \xi_{p_d}$ due to (3.27), and denote the remaining $r$ roots as $\mu_1, \ldots, \mu_r$. Hence, we get

$$K(\boldsymbol{x}) = \mathrm{lc}(K) \prod_{i=1}^{p_d} (x_d - \xi_i) \prod_{j=1}^{r} (x_d - \mu_j).$$

Then the formula (3.28) simplifies to

$$F_{\boldsymbol{s}}(\boldsymbol{x}) = \frac{A(\boldsymbol{x}) - G(\boldsymbol{x})}{K(\boldsymbol{x})} = \frac{\mathrm{lc}(A)}{\mathrm{lc}(K)} \prod_{j=1}^{r} \frac{1}{x_d - \mu_j}.$$

We will illustrate the usefulness of the derived formula in the next example. ■

***Example 3.30 (Generalized Dyck Paths, [8, Example 3]):*** The problem of Generalized Dyck Paths deals with a step set of the form $\mathcal{S} = \{(r_1, s_1), \ldots, (r_k, s_k)\}$ where $r_i, s_i \in \mathbb{Z}$

and $r_i > 0$ (directed problem). Let $b_{n,m}$ denote the number of paths going from $(0,0)$ to $(m,n)$ using only steps from $\mathcal{S}$ and staying within the first quadrant. In this problem we are mainly interested in the numbers $b_{m,0} =: b_m$. The recurrence relation is given by

$$b_{m,n} = \begin{cases} \sum_{i:(r_i,s_i)\leq(m,n)} b_{m-r_i,n-s_i}, & \text{if } m,n \geq 0 \text{ and } (m,n) \neq (0,0), \\ 1, & \text{if } m = n = 0. \end{cases}$$

This form does not fit the one we have been working with so far, compare (3.21), but can be easily brought into the desired form. Let $r := \max_{1\leq i\leq k} r_i$ and $s := \max_{1\leq i\leq k} s_i$. Interpret the family $(b_{m,n})_{m,n\in\mathbb{N}}$ as array $b$ and attach $r$ columns of zeros to the left of $b$ and $s$ rows of zeros below. Call the resulting array $a$ and number its rows and columns starting with 0. Then

$$b_{m,n} = a_{m+r,n+s}, \qquad \text{for } m,n \geq 0.$$

Finally only a technical step is missing: Let $(\rho,\sigma)$ be any step in $\mathcal{S}$ which is maximal with respect to the partial order $\leq$ (e.g. the lexicographically largest step). Set $a_{r-\rho,s-\sigma} := 1$. Then we attain the desired form (3.21)

$$a_{m,n} = \begin{cases} a_{m-r_1,n-s_1} + \ldots + a_{m-r_k,n-s_k}, & \text{if } m \geq r \text{ and } n \geq s, \\ \delta_{(m,n),(r-\rho,s-\sigma)}, & \text{if } m < r \text{ and } n < s. \end{cases}$$

The set of shifts is given by $H = \{(-r_1,-s_1),\ldots,(-r_k,-s_k)\}$, the starting point $\boldsymbol{s} = (r,s)$ and $c_{\boldsymbol{h}} = 1$ for all $\boldsymbol{h} \in H$. From the fact that $r_i > 0$ for all $i = 1,\ldots,k$ it follows that $\{\boldsymbol{x} \in \mathbb{R}^d : \boldsymbol{x} \geq 0\} \cap \operatorname{conv} H = \emptyset$ and by Theorem 3.20 there exists a unique solution.

The apex $\boldsymbol{p} = (0, \max\{0,t\})$ with $t := -\min\{s_1,\ldots,s_k\}$. We distinguish two cases:

(1) If $s_i \geq 0$ for all $i = 1,\ldots,k$, then the apex is $(0,0)$. Then we are in the case of Theorem 3.26 and the GF of $F_{\boldsymbol{s}}$ will be rational. Let's make this explicit:

The kernel is $K(x,y) = 1 - x^{r_1}y^{x_1} - \ldots - x^{r_k}y^{s_k}$ and the known initial function is $A(x,y) = 1$. Hence, we obtain for the GF $B(x,y)$ enumerating $b_{m,n}$

$$\begin{aligned} B(x,y) &= \sum_{m,n\geq 0} b_{m,n}x^m y^n = \sum_{m\geq r; n\geq s} a_{m,n}x^{m-r}y^{n-s} = F_{\boldsymbol{s}}(x,y) \\ &= \frac{1}{1 - x^{r_1}y^{s_1} - \ldots - x^{r_k}y^{s_k}}. \end{aligned}$$

The GF $B_0(x)$ for paths ending on the horizontal axis reads

$$B_0(x) = \sum_{m\geq 0} b_m x^m = G(x,0) = \frac{1}{1 - \sum_{1\leq i\leq k; s_i=0} x^{r_i}}.$$

(2) If there exists $i = 1,\ldots,k$ such that $s_i < 0$, then the apex of $H$ is $(0,t)$ with $t > 0$ and we are in the case of Theorem 3.28 and the corresponding GF is algebraic.

Similarly as before we find

$$K(x,y) = y^t - x^{r_1}y^{s_1+t} - \ldots - x^{r_k}y^{s_k+t}, \qquad A(x,y) = y^t.$$

We are now in a special case discussed in Remark 3.29. The kernel $K$ has degree $t+s$ in $y$ with leading coefficient $-\sum_{i:s_i=s} x^{r_i}$. Assume $s > 0$ to avoid trivial cases, because we are confined to the first quadrant. According to the proof of Theorem 3.28 there exist $t$ roots $\xi_1(x), \dots, \xi_t(x)$ with $\xi_i(0) = 0$ satisfying $K(x, \xi_i(x)) = 0$. Let $\mu_1(x), \dots, \mu_s(x)$ be the $s$ remaining roots. Hence, we get

$$\begin{aligned} B(x,y) &= \sum_{m,n\geq 0} b_{m,n} x^m y^n = F_s(x,y) = \frac{1}{K(x,y)} \prod_{i=1}^{t} (y - \xi_i(x)) \\ &= -\frac{1}{\sum_{i:s_i=s} x^{r_i}} \prod_{j=1}^{s} \frac{1}{y - \mu_j(x)}, \\ B_0(x) &= \sum_{m\geq 0} b_m x^m = G(x,0) = \frac{(-1)^t}{K(x,0)} \prod_{i=1}^{t} \xi_i(x) = \frac{(-1)^{s+1}}{\sum_{i:s_i=s} x^{r_i}} \prod_{j=1}^{s} \frac{1}{\mu_j(x)}. \end{aligned} \tag{3.30}$$

As special cases, this example includes some well-known lattice path enumeration problems. Let us state some famous ones and the GF $B_0(x)$ of paths ending on the horizontal axis given by (3.30). Remark, that the set of steps $\mathcal{S}$ and the set of shifts $H$ are naturally connected via $H = -\mathcal{S} = \{(-r, -s) : (r, s) \in \mathcal{S}\}$. This is due to the fact, that lattice paths with the step set $\mathcal{S} = \{(r_1, s_1), \dots, (r_k, s_k)\}$ can be recursively constructed with the recurrence relation

$$a_{m,n} = a_{m-r_1,n-s_1} + \ldots + a_{m-r_k,n-s_k} \qquad \text{for all } (m,n) \geq (r,s).$$

- *Dyck Paths:* $\mathcal{S} = \{(1,-1),(1,1)\}$ with the kernel $K(x,y) = y - xy^2 - x$,
$$B_0(x) = \frac{1 - \sqrt{1-4x^2}}{2x^2}.$$
- *Motzkin Paths:* $\mathcal{S} = \{(1,-1),(1,1),(1,0)\}$ with the kernel $K(x,y) = y - xy^2 - x - xy$,
$$B_0(x) = \frac{1 - x - \sqrt{1-2x-3x^2}}{2x^2}.$$
- *Schröder Paths:* $\mathcal{S} = \{(1,-1),(1,1),(2,0)\}$ with the kernel $K(x,y) = y - xy^2 - x - x^2y$,
$$B_0(x) = \frac{1 - x^2 - \sqrt{1-6x^2+x^4}}{2x^2}.$$
- *Delannoy Paths:* $\mathcal{S} = \{(1,0),(0,1),(1,1)\}$ with the kernel $K(x,y) = 1 - x - xy - y$,
$$B(x,y) = \frac{x + xy + y}{1 - x - xy - y}, \qquad B_0(x) = \frac{1}{1-x}.$$
Note that the apex is $(0,0)$ and we get a rational GF.

■

At the end of this chapter we want to emphasize that not all recurrence relations need to have rational, algebraic or even holonomic solutions. There are also problems which have a non-holonomic and maybe even irrational solution. One example is the *Knight's Walk* studied in detail in [8]. Hereby we understand a walk that starts anywhere on the lines $x = 0, 1$ or $y = 0, 1$, takes only two kinds of steps $(-1, 2)$ and $(2, -1)$ and remains in the region $x \geq 2, y \geq 2$ once it leaves the starting point.

# Chapter 4

# Walks confined to the Quarter Plane

The properties of lattice paths are, as the name suggests, very dependent on the nature of the underlying lattice. These can differ in the chosen types (square lattice, hexagonal lattice, etc.) but also in their spacial expansion. By the latter we understand various types of restrictions, like the restriction to a half-plane or quarter plane of the lattice $\mathbb{Z}^2$. In the following chapter we want to study such walks confined to the positive quarter plane (or positive quadrant) $\mathbb{Z}_+^2$ also noted as $\mathbb{N}^2$.

The motivation in the choice of these walks lies in the vast number and richness of applications. The following list from [42], gives a glimpse of different fields, which are effected by this theory:

- *combinatorics:* Many combinatorial objects (e.g. maps, permutations, trees or Young tableaux) can be encoded in lattice walks, in particular by walks in the quarter plane, see [6, 7].

- *population biology:* The quarter plane is the natural space to parametrize any two-dimensional population, see [32].

- *probability theory:* One very recent topic are random walks in cones (quantum random walks, non-colliding random walks, etc.), see [21].

- *queuing theory:* Any two-dimensional queue can be modeled by random walks in the quarter plane.

- *complex analysis:* The inclusion $\mathbb{Z}_+^2 \subset \mathbb{C}$ turns out to be convenient for applying methods from complex analysis, see [11].

- *finance:* The dynamics of certain limit order books may be approximated by random processes in the quarter plane.

- etc.

## 4.1 Definitions

In addition to the definitions from the previous chapters we are going to need the following. First consider an adaption of Definition 3.13 for the characteristic polynomial, where all weights are set to 1.

***Definition 4.1:*** The *characteristic polynomial* of a step set $\mathcal{S} = \{(a_1, b_1), \ldots, (a_k, b_k)\}$ is the Laurent polynomial $S(x, y)$, which is the generating polynomial of the steps $\mathcal{S}$, defined by

$$S(x, y) = \sum_{i=1}^{k} x^{a_i} y^{b_i}.$$

Recall the notation $\bar{x} = \frac{1}{x}$, so that $\mathbb{C}[x, \bar{x}]$ is the ring of Laurent polynomials in $x$ with coefficients in $\mathbb{C}$. ◊

***Example 4.2:*** The characteristic polynomial for the full set of small steps $\{-1, 0, 1\}^2 \setminus \{(0, 0)\} = \{\mathbf{E}, \mathbf{NE}, \mathbf{N}, \mathbf{NW}, \mathbf{W}, \mathbf{SW}, \mathbf{S}, \mathbf{SE}\}$ is

$$\begin{aligned} S(x, y) &= x + xy + y + \frac{y}{x} + \frac{1}{x} + \frac{1}{xy} + \frac{1}{y} + \frac{x}{y} \\ &= x + xy + y + x\bar{y} + \bar{x} + \bar{x}\bar{y} + \bar{y} + x\bar{y}. \end{aligned}$$

■

When working with Laurent series, it will be useful to investigate the partial sum of all negative exponents in one variable. The extraction of this part from a GF was introduced in Definition 2.18 and will be a main tool in this chapter. In a similar fashion we distinguish among the steps with regards to their influence in the characteristic polynomial.

***Definition 4.3:*** A step $(i, j)$ is

- *$x$-positive* if $i > 0$,
- *$y$-positive* if $j > 0$,
- *$x$-negative* if $i < 0$,
- *$y$-negative* if $j < 0$.

◊

***Example 4.4:*** Continuing Example 4.2 and using the notation from Definition 2.18, we have

$$\begin{aligned} S(x, y) &= (\bar{x} + 1 + x)\,\bar{y} + \bar{x} + x + (x + 1 + \bar{x})\,y \qquad \in \mathbb{Z}(x)[y, \bar{y}], \\ [y^{<}]S(x, y) &= (\bar{x} + 1 + x)\,\bar{y}, \end{aligned}$$

as the Laurent polynomial of all $y$-negative steps.

The partition of all small steps in $x$/$y$-positive/negative steps is shown in Table 4.1. Obviously $x$-positive steps always need an east-, $y$-positive steps a north-, $x$-negative steps a west- and $y$-negative steps a south-content. ■

| Type | Steps | | |
|---|---|---|---|
| $x$-positive | **SE** | **E** | **NE** |
| $y$-positive | **NE** | **N** | **NW** |
| $x$-negative | **NW** | **W** | **SW** |
| $y$-negative | **SW** | **S** | **SE** |

Table 4.1: Partitioning of Small Steps in $x/y$-positive/negative Steps

## 4.2 Walks with Small Steps

This section is devoted to the analysis of walks confined to the quarter plane employing small steps, i.e. $\mathcal{S} = \{-1, 0, 1\}^2 \setminus \{(0, 0)\}$ (recall Definition 1.4). The investigation was started by Bousquet-Mélou and Mishna in [7] who laid the foundation for many follow-up papers on this topic [5, 6, 26, 30, 42]. It was the first to the author known approach to enumerate a "big" class of lattice paths restricted to the quarter plane.

A priori there are $2^8$ different problems of this type. But among these, some are trivial, some are equivalent to models confined to the half-plane and some are equivalent to others up to a $x/y$-symmetry. This analysis results in 79 inherently different problems.

Their core step introduces algebra or in particular Galois theory to the problem of lattice path enumeration. Each of the problems is associated with a group $G$ of birational transformations. This group is finite for 23 cases and infinite for the remaining 56 other cases.

Finally, they present a way for solving 22 of 23 problems, which are associated with a finite group and show that the corresponding generating functions are holonomic. In the mean time Bostan and Kauers provided a computer aided proof for the 23$^{\text{rd}}$ case in [6] and showed that its generating function is algebraic, hence holonomic.

For the 56 models with an infinite group the conjecture is raised from Bousquet-Mélou and Mishna that they have non-holonomic generating functions. This is proven in [30] by Kurkova and Raschel with analytic tools, like Riemann surfaces, universal covers, branches of meromorphic functions, etc.

In the remainder of this chapter, we want to present the derivation and analysis of Bousquet-Mélou and Mishna from [7], which shows, how a systematic approach is able to succeed on a so far unsolved problem. It is impressive to discover, that only the interdisciplinary and combined effort of different people and fields, like analytic combinatorics, complex analysis and algebra led to the successful solution of the problem.

### 4.2.1 Classification of Models with Small Steps

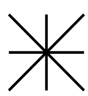


Figure 4.1: The full set of Small Steps

As we are investigating only small steps, there are only $2^8 = 256$ cases to study, all possible subsets of the full step set $\mathcal{S}$ shown in Figure 4.1. As a first simplification, we can discard trivial models, like $\mathcal{S} = \emptyset$ or $\mathcal{S} = \{\bar{x}\}$. Hence, we first consider some **easy algebraic cases**:

1. *No $x$-positive step*: If $\mathcal{S}$ contains no $x$-positive step, we can ignore its $x$-negative steps, as they would only lead us out of the quarter plane. Therefore, these walks degenerate to walks consisting of vertical steps on the vertical half-line. We know already that these are always algebraic and even rational if the degenerated set is $\mathcal{S} = \emptyset$, $\mathcal{S} = \{y\}$ or $\mathcal{S} = \{\bar{y}\}$. In fact we have already solved the algebraic case $\mathcal{S} = \{y, \bar{y}\}$ explicitly, as there is an obvious bijection to Dyck Prefixes confined to the first quadrant (i.e. meanders), compare Theorem 3.7 and Example 3.10 where the special case of Dyck Paths (i.e. excursions) is discussed.

   Another way to show that the GF is algebraic is to consider its recurrence relation. We define $a_{y,n}$ to be the number of steps ending at altitude $y$ after $n$ steps. Then

   $$\begin{aligned} a_{y,n} &= a_{y-1,n-1} + a_{y+1,n-1} && \text{for } y, n \geq 1 \\ a_{y,0} &= \delta_{y,0} \\ a_{0,n} &= [z^n]\frac{1-\sqrt{1-4z^2}}{z} = \frac{1}{n+1}C_n \end{aligned}$$

   The number $a_{0,n}$ represents all paths starting from $(0,0)$ and ending at $(0,0)$ which never go below the $x$-axis. Hence, as mentioned before, these are Dyck Paths and therefore we know its GF from Example 2.7. This gives the set of shift $H = \{(-1,-1),(1,-1)\}$, the starting point $\boldsymbol{s} = (1,1)$ and an algebraic known initial function $A(y,z)$. Additionally we see that the apex $\boldsymbol{p} = (1,0)$ has only one positive coordinate, which implies by Theorem 3.28 that its GF is algebraic.

2. *No $y$-positive step*: Symmetric arguments lead to the same conclusion, that this problem has an algebraic solution.

3. *No $x$-negative step*: The walks starting in the origin $(0,0)$ and employing this step set always stay in the right half-plane $x \geq 0$. If we now impose the restriction of staying in the quarter plane, this is equivalent to staying in the upper half-plane. The corresponding theory discussed in Section 3.3 implies that the GFs are always algebraic (compare Proposition 3.16).

4. *No $y$-negative step*: Again, by symmetry we obtain algebraic solutions.

This short analysis shows, that our step set has to contain $x$-positive, $x$-negative, $y$-positive and $y$-negative steps. In order to enumerate the solution we introduce the OGF $P(z) = \sum_{n=0}^{8} p_n z^n$ where $p_n$ is the number of models employing a step set of size $n$, which have not been covered yet. Observe, that there are in total 5 non-$x$-positive, non-$x$-negative, non-$y$-positive and non-$y$-negative steps each, shown in Figure 4.2. Thus, by an inclusion-exclusion argument we get the intermediate result

$$\begin{aligned} P_1(z) &= (1+z)^8 - 4(1+z)^5 + 2(1+z)^2 + 4(1+z)^3 - 4(1+z) + 1 \\ &= 2z^2 + 20z^3 + 50z^4 + 52z^5 + 28z^6 + 8z^7 + z^8. \end{aligned}$$

(a) no $x$-positive step (b) no $x$-negative step (c) no $y$-positive step (d) no $y$-negative step

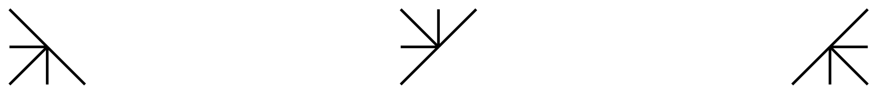
Figure 4.2: Small Steps which contain no $x/y$-positive/negative step, respectively

A short explanation of the above formula:

- $(1+z)^8$: all possible $2^8$ subsets
- $(1+z)^4$: counts sets with no $x$-positive step, ...
- $(1+z)^2$: counts sets with no $x$-positive nor $x$-negative step, ...
- $(1+z)^3$: counts sets with no $x$-positive nor $y$-positive step, ...
- $(1+z)$: counts sets with no $x$-positive nor $x$-negative nor $y$-positive step, ...
- 1: counts the empty set

Secondly, among the remaining 161 sets, there are some which do not contain any step with **both coordinates non-negative**: In this case the only quarter plane walk is the empty walk (see Figure 4.3a). These sets are subsets of $\{\bar{x}, \bar{y}, x\bar{y}, \bar{x}\bar{y}, \bar{x}y\}$. But due to the previous step the used step set must contain $x$-positive and $y$-positive steps, hence, $x\bar{y}$ and $\bar{x}y$ belong to $\mathcal{S}$. Therefore, we exclude $2^3$ step sets, and get

$$P_2(z) = P_1(z) - z^2(1+z)^3 = z^2 + 17z^3 + 47z^4 + 51z^5 + 28z^6 + 8z^7 + z^8.$$



(a) $x$-negative or $y$-negative (b) above first diagonal, $x$-cond. forces $y$-cond. (c) below first diagonal, $y$-cond. forces $x$-cond.

Figure 4.3: Step Sets with respect to Diagonals

Thirdly, there exist models in which **one of the quarter plane constraints implies the other**: In this case the model is equivalent to problems of walks confined to a half-space: their GF is always algebraic and can be derived using the kernel method [3].

Assume that all walks with steps in $\mathcal{S}$ that end at a non-negative abscissa automatically end at a non-negative ordinate (we say the $x$-condition forces/implies the $y$-condition). This can only be if $\bar{y}$ and $x\bar{y}$ are not part of $\mathcal{S}$. But as at this stage we need a $y$-negative step, $\bar{x}\bar{y}$ must be in $\mathcal{S}$. Under the prerequisite that the $x$-condition forces the $y$-condition, $x$ cannot be in $\mathcal{S}$, otherwise a walk could have non-negative final abscissa, but negative final ordinate. But now we exclude $x\bar{y}$ and $x$, which is why $xy$ must be in $\mathcal{S}$, because we need at least one $x$-positive step. This implies that $\mathcal{S} \setminus \{xy, \bar{x}\bar{y}\} \subseteq \{\bar{x}, y, \bar{x}y\}$. Observe, that these 3(5) steps are the ones lying (strictly) above the first diagonal, see Figure 4.3b.

Symmetric arguments lead to the fact, that if the $y$-condition implies the $x$ condition we must have: $\mathcal{S} \setminus \{xy, \bar{x}\bar{y}\} \subseteq \{x, \bar{y}, x\bar{y}\}$, which corresponds to the steps below the first diagonal, see Figure 4.3c. A final inclusion-exclusion arguments gives 138 remaining cases

$$P_3(z) = P_2(z) - 2z^2(1+z)^3 + z^2 = 11z^3 + 41z^4 + 49z^5 + 28z^6 + 8z^7 + z^8.$$

Lastly, we consider **symmetries**. When looking at the quarter plane embedded in the Euclidean plane, there is one obvious symmetry which jumps to one's eye: The $x/y$-symmetry or reflection along the first diagonal. This is the only one, which leaves the quarter plane fixed. (Note, that all 8 symmetries of the square act on the step sets, but none except this one maps the quarter plane to itself.) Thus, two step sets, which can be transformed into each other applying this symmetry lead to equivalent counting problems. From now on, we understand symmetry always with respect to the first diagonal.

We want to find all non-equivalent problems, so we eliminate these among the remaining 138 problems. We are going to achieve this by counting all symmetric models which are part of the previously counted ones. All others have a partner which is not the same and can be obtained through reflection, but possesses the same counting sequence. Hence, we successively repeat the last three steps:

1. There are 2 steps which are symmetric ($\nearrow$ and $\swarrow$) and 3 pairs of symmetric step subsets: $\{\uparrow, \rightarrow\}$, $\{\nwarrow, \searrow\}$ and $\{\leftarrow, \downarrow\}$. We discard all $\mathcal{S}$ which miss $x$- or $y$-positive, or $x$- or $y$-negative steps.

$$P_1^{\mathrm{sym}}(z) = (1+z)^2(1+z^2)^3 - 2(1+z)(1+z^2) + 1$$

2. No $\mathcal{S}$ is allowed, where all steps are negative in at least one coordinate. The ones which have to be excluded must obey $\{\bar{x}y, x\bar{y}\} \subseteq \mathcal{S}$. Hence, there are only two possible choices left: $\bar{x}\bar{y}$ and $\{\bar{x}, \bar{y}\}$.

$$P_2^{\mathrm{sym}}(z) = P_1^{\mathrm{sym}}(z) - z^2(1+z)(1+z^2)$$

3. One quarter plane constraint implies the other. Here, only $\mathcal{S} = \{xy, \bar{x}\bar{y}\}$ is to be excluded.

$$P_3^{\mathrm{sym}}(z) = P_2^{\mathrm{sym}}(z) - z^2 = 3z^3 + 5z^4 + 5z^5 + 4z^6 + 2z^7 + z^8$$

Therefore we get as the final result the generating polynomial

$$P(z) = \frac{1}{2}\left(P_3(z) + P_3^{\mathrm{sym}}(z)\right) = 7z^3 + 23z^4 + 27z^5 + 16z^6 + 5z^7 + z^8,$$

and a total of 79 non-equivalent models for step sets $\mathcal{S}$ with small steps.

In the next section we are going to introduce the core idea of the following proof, where we introduce algebra into the field of lattice path counting.

### 4.2.2 The Group of the Walk

Recall the characteristic polynomial $S(x,y) = \sum_{(i,j)\in\mathcal{S}} x^i y^j$. By fixing $x$ we can interpret it as Laurent polynomial in $y$ with coefficients in $\mathbb{Q}(x)$ and vice versa. This gives rise to the following representation:

$$\begin{aligned} S(x,y) &= A_{-1}(x)\bar{y} + A_0(x) + A_1(x)y \\ &= B_{-1}(y)\bar{x} + B_0(y) + B_1(y)x \end{aligned} \tag{4.1}$$

From now on we assume that all step sets $\mathcal{S}$ are containing $x$-positive, $x$-negative, $y$-positive and $y$-negative steps.

By this assumption we conclude that $A_1, B_1, A_{-1}$ and $B_{-1}$ are non-zero. This ensures, that the following transformations are well-defined

$$\Phi : (x,y) \mapsto \left(\bar{x}\frac{B_{-1}(y)}{B_1(y)}, y\right) \qquad \text{and} \qquad \Psi : (x,y) \mapsto \left(x, \bar{y}\frac{A_{-1}(x)}{A_1(x)}\right). \tag{4.2}$$

***Lemma 4.5:*** $S(x,y)$ is left unchanged by $\Phi$ and $\Psi$.

*Proof:* We calculate

$$\begin{aligned} S(\Phi(x,y)) &= B_{-1}(y)\overline{\bar{x}\frac{B_{-1}(y)}{B_1(y)}} + B_0(y) + B_1(y)\bar{x}\frac{B_{-1}(y)}{B_1(y)} \\ &= B_1(y)x + B_0(y) + B_{-1}(y)\bar{x} = S(x,y). \end{aligned}$$

Analogously we get $S(\Psi(x,y)) = S(x,y)$. □

***Lemma 4.6:*** $\Phi$ and $\Psi$ are involutions, and therefore birational transformations.

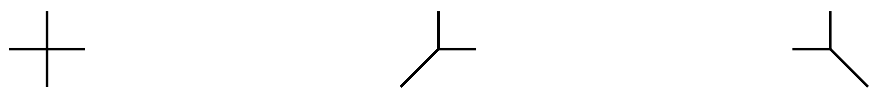
*Proof:* Recall, that an involution $I$ is characterized by $I \circ I = I$.

$$(\Phi \circ \Phi)(x,y) = \Phi\left(\bar{x}\frac{B_{-1}(y)}{B_1(y)}, y\right) = \left(\overline{\bar{x}\frac{B_{-1}(y)}{B_1(y)}}\frac{B_{-1}(y)}{B_1(y)}, y\right) = (x,y)$$

The result for $\Psi$ follows analogously. □

***Definition 4.7:*** The group $G(\mathcal{S})$ or short $G$ generated by $\Phi$ and $\Psi$ using composition is called *group of the walk* created by the step set $\mathcal{S}$. The *sign* of $g \in G$ is 1 ($-1$) if $g$ is the product of an even (odd) number of generators $\Phi$ and $\Psi$. ◊

This group is isomorphic to the dihedral group $D_n$ of order $2n$, with $n \in \mathbb{N} \cup \{\infty\}$. For the finite dimensional case this can be seen, by looking at the generator and relation definition of the dihedral group, i.e. $D_n \cong \langle a, b \mid a^2 = b^2 = (ab)^n = 1\rangle$ [22, Theorem I.6.13]. According to the previous lemma, for each $g \in G$, one has $S(g(x,y)) = S(x,y)$.



(a) $\mathcal{S} = \{\mathbf{N}, \mathbf{S}, \mathbf{E}, \mathbf{W}\}$ (b) $\mathcal{S} = \{\mathbf{E}, \mathbf{N}, \mathbf{SW}\}$ (c) $\mathcal{S} = \{\mathbf{N}, \mathbf{W}, \mathbf{SE}\}$

Figure 4.4: Step Sets for Example 4.8 on the Group of the Walk

***Example 4.8:***

1. Assume $\mathcal{S}$ is left unchanged by reflection across the horizontal line. This is equivalent to $S(x,y) = S(x,\bar{y})$, or that $A_1(x) = A_{-1}(x)$, or that $B_i(y) = B_i(\bar{y})$ for $i \in \{-1,0,1\}$. Set $C(y) = \frac{B_{-1}(y)}{B_1(y)}$ ($\neq 0$, due to the assumption at the beginning of this section), then the transformations are

$$\Phi : (x,y) \mapsto (C(y)\bar{x}, y) \qquad \text{and} \qquad \Psi : (x,y) \mapsto (x,\bar{y})\,,$$

and the orbit of $(x,y)$ under the action of $G$ evolves like

$$(x,y) \overset{\Phi}{\longleftrightarrow} (C(y)\bar{x}, y) \overset{\Psi}{\longleftrightarrow} (C(y)\bar{x}, \bar{y}) \overset{\Phi}{\longleftrightarrow} (x,\bar{y}) \overset{\Psi}{\longleftrightarrow} (x,y),$$

so that $G$ is of order 4.

Remark, that this group may not be the full group of transformations which leave $S(x,y)$ unchanged. For instance, apply the above to $\mathcal{S} = \{\mathbf{N},\mathbf{S},\mathbf{E},\mathbf{W}\}$ (compare Figure 4.4a), then the map $(x,y) \mapsto (y,x)$ leaves $S(x,y)$ unchanged, however the orbit of $(x,y)$ under $G$ is $\{(x,y),(\bar{x},y),(\bar{x},\bar{y}),(x,\bar{y})\}$ because $C(y) = 1$.

2. Let $\mathcal{S} = \{x, y, \bar{x}\bar{y}\}$, see Figure 4.4b. We get $A_1(x) = 1, A_{-1}(x) = \bar{x}, B_1(y) = 1, B_{-1}(y) = \bar{y}$. Then the transformations are

$$\Phi : (x,y) \mapsto (\bar{x}\bar{y}, y) \qquad \text{and} \qquad \Psi : (x,y) \mapsto (x,\bar{x}\bar{y})\,,$$

and they generate a group of order 6

$$(x,y) \overset{\Phi}{\longleftrightarrow} (\bar{x}\bar{y}, y) \overset{\Psi}{\longleftrightarrow} (\bar{x}\bar{y}, x) \overset{\Phi}{\longleftrightarrow} (y,x) \overset{\Psi}{\longleftrightarrow} (y,\bar{x}\bar{y}) \overset{\Phi}{\longleftrightarrow} (x,\bar{x}\bar{y}) \overset{\Psi}{\longleftrightarrow} (x,y).$$

3. Consider the case $\mathcal{S} = \{\bar{x}, y, x\bar{y}\}$ which results from the previous one by a rotation of 90 degrees, see Figure 4.4c. Here we get $A_1(x) = 1, A_{-1}(x) = x, B_1(y) = \bar{y}, B_{-1}(y) = 1$. The needed transformations are

$$\Phi : (x,y) \mapsto (\bar{x}y, y) \qquad \text{and} \qquad \Psi : (x,y) \mapsto (x,x\bar{y})\,,$$

and they also generate a group of order 6

$$(x,y) \overset{\Phi}{\longleftrightarrow} (\bar{x}y, y) \overset{\Psi}{\longleftrightarrow} (\bar{x}y, \bar{x}) \overset{\Phi}{\longleftrightarrow} (\bar{y},\bar{x}) \overset{\Psi}{\longleftrightarrow} (\bar{y},x\bar{y}) \overset{\Phi}{\longleftrightarrow} (x,x\bar{y}) \overset{\Psi}{\longleftrightarrow} (x,y).$$

■

The fact, that the last two examples both generated a group of order 6 is no coincidence, as the following lemma shows.

***Lemma 4.9 [7, Lemma 2]:*** Let $\mathcal{S}$ and $\tilde{\mathcal{S}}$ be two sets of steps differing by one of the 8 symmetries of the square. Then the groups $G(\mathcal{S})$ and $G(\tilde{\mathcal{S}})$ are isomorphic.

*Proof:* The group of symmetries of the square is the dihedral group $D_4$ which is generated by the two reflections $\Delta$ (across the first diagonal) and $V$ (across the vertical line). Therefore it is sufficient to prove the lemma for $\tilde{\mathcal{S}} = \Delta(\mathcal{S})$ and $\tilde{\mathcal{S}} = V(\mathcal{S})$. The transformations associated with $\mathcal{S}$ are denoted by $\Phi$ and $\Psi$, the ones associated with $\tilde{\mathcal{S}}$ are denoted by $\tilde{\Phi}$ and $\tilde{\Psi}$.

Assume $\tilde{\mathcal{S}} = \Delta(\mathcal{S})$. Then the values of $A_i(x)$ and $B_i(y)$ swap, i.e. $\tilde{A}_i(x) = B_i(x)$ and $\tilde{B}_i(y) = A_i(y)$. We reconstruct this swap with the involution $\delta(x,y) = (y,x)$.

$$\begin{aligned}\tilde{\Phi}(x,y) &= \left(\bar{x}\frac{\tilde{B}_{-1}(y)}{\tilde{B}_1(y)}, y\right) = \left(\bar{x}\frac{A_{-1}(y)}{A_1(y)}, y\right) = \delta\left(y, \bar{x}\frac{A_{-1}(y)}{A_1(y)}\right) = (\delta \circ \Psi \circ \delta)(x,y),\\ \tilde{\Psi} &= \delta \circ \Psi \circ \delta,\end{aligned}$$

where the second result follows similarly. Hence the groups $G(\mathcal{S})$ and $G(\tilde{\mathcal{S}})$ are conjugate by $\delta$ and therefore isomorphic.

Assume $\tilde{\mathcal{S}} = V(\mathcal{S})$. Then $\tilde{A}_i(x) = A_i(\bar{x})$ and $\tilde{B}_i(y) = B_{-i}(y)$. We reconstruct this symmetry by the involution $v(x,y) = (\bar{x}, y)$.

$$\begin{aligned}\tilde{\Phi}(x,y) &= \left(\bar{x}\frac{\tilde{B}_{-1}(y)}{\tilde{B}_1(y)}, y\right) = \left(\bar{x}\frac{B_1(y)}{B_{-1}(y)}, y\right) = v\left(x\frac{B_{-1}(y)}{B_1(y)}, y\right) = (v \circ \Phi \circ v)(x,y),\\ \tilde{\Psi}(x,y) &= \left(x, \bar{y}\frac{\tilde{A}_{-1}(x)}{\tilde{A}_1(x)}\right) = \left(x, \bar{y}\frac{A_{-1}(\bar{x})}{A_1(\bar{x})}\right) = v\left(\bar{x}, \bar{y}\frac{A_{-1}(\bar{x})}{A_1(\bar{x})}\right) = (v \circ \Psi \circ v)(x,y),\end{aligned}$$

Hence the groups $G(\mathcal{S})$ and $G(\tilde{\mathcal{S}})$ are conjugate by $v$. □

***Theorem 4.10 [7, Theorem 3]:*** Out of the 79 models under construction, exactly 23 are associated with a finite group:

- 16 have a vertical symmetry and thus a group of order 4,
- 4 have a group of order 6,
- 2 have a group of order 8.

*Proof:* From earlier considerations we know that $G(\mathcal{S})$ is a dihedral group of order $2n$ with $n \in \mathbb{N} \cup \{\infty\}$. As $\Phi$ and $\Psi$ are involutions, the group is of order $2n$ if and only if $\Theta := \Psi \circ \Phi$ is of order $n$. Now it's easy to check if a given group is of order $2n$. Compute all iterates of $\Theta$ and the first that is the identity gives you the half of the order.

By applying Lemma 4.9 to the first case of Example 4.8 we get, that the step sets with a vertical symmetry have order 4. It is straightforward to check the other assertions. All results are summarized in Tables 1-3 in [7]. The step sets of the walks we are talking about are shown in Figures 4.5, 4.6 and 4.7.

The difficult task, is to prove that the remaining models possess an infinite group. There are 2 different strategies depending on $\mathcal{S}$ necessary to achieve this goal. For sets shown in Figure 4.8 a **valuation** argument is used. These sets are special, as all elements $(i,j)$ satisfy $i+j \geq 0$. For the remaining cases a **fixed point argument** is used. It is interesting to note, that both strategies are necessary, i.e. the second one, does not solve the first cases.

**1. The valuation argument**

First, we introduce the general idea: We want to "project" the complicated case of a group in $(\mathbb{Z}[x,\bar{x},y,\bar{y}])^2$ to a simpler version in $\mathbb{Z}^2$, where we can perform simple calculations.

Let $z$ be an indeterminate and let $x$ and $y$ be Laurent series in $z$ with coefficients in $\mathbb{Q}$, and respective valuation $a$ and $b$. Furthermore, we assume that the *trailing coefficients* of these

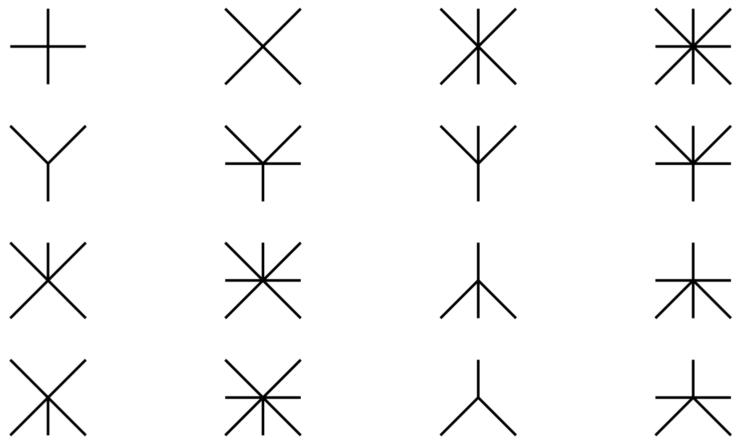


Figure 4.5: The 16 models whose group $G(\mathcal{S})$ is isomorphic to $D_2$. All of them possess a vertical symmetry and have a holonomic GF [7, Table 1].

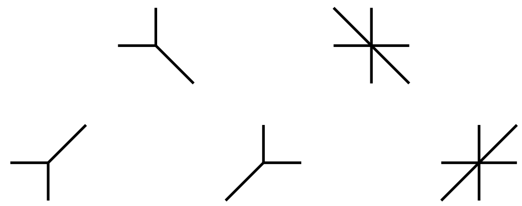


Figure 4.6: The 5 models whose group $G(\mathcal{S})$ is isomorphic to $D_3$. All have a holonomic GF, the last three even possess an algebraic GF [7, Table 2].

Figure 4.7: The 2 models whose group $G(\mathcal{S})$ is isomorphic to $D_4$. Both models have a holonomic GF, the second one even possesses an algebraic GF [7, Table 3].
The Second one is Gessel's Walk which was solved in [6].

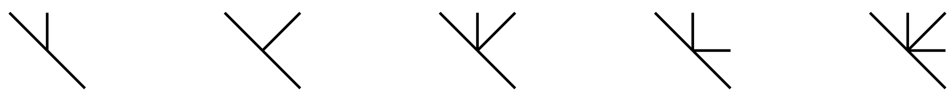


Figure 4.8: Five step sets with an infinite group, where $i + j \geq 0$ for all steps $(i, j)$

series, namely $[z^a]x$ and $[z^b]y$, are positive. Now define $(x', y) := \Phi(x, y)$. Then by (4.2) the trailing coefficient of $x'$ (and $y$) is positive, and the valuation of $x'$ (and $y$) only depends on $a$ and $b$:

$$\phi(a,b) := (\mathrm{val}(x'), \mathrm{val}(y)) = \begin{cases} (-a + b(v_{-1}^{(y)} - v_1^{(y)}), b), & \text{if } b \geq 0, \\ (-a + b(d_{-1}^{(y)} - d_1^{(y)}), b), & \text{if } b \leq 0, \end{cases}$$

where $v_i^{(y)}$ (respectively $d_i^{(y)}$) denotes the valuation (resp. degree) in $y$ of $B_i(y)$, for $i = \pm 1$. All this follows by the elementary construction of the inverse $\bar{x}(z)$ of the Laurent series $x(z)$ of valuation $a$: Let $x(z) = \sum_{n \geq a} x_n z^n$ and $\bar{x} = \sum_{n \geq \alpha} \bar{x}_n z^n$. A comparison of coefficients in

$$1 = x(z) \cdot \bar{x}(z) = \sum_{n \geq a+\alpha} \left( \sum_{k=0}^{n} x_{k+a} \bar{x}_{n-k+\alpha} \right) z^n,$$

gives that $\alpha = -a$ and that the trailing coefficient of $\bar{x}$ is $(x_a)^{-1}$ and therefore positive if $x_a$ is positive. All other parts follow analogously.

The two different cases are necessary, because a polynomial with degree $d$ has valuation $-d$ after evaluating at $1/z$, e.g. $p(z) = 1 + z^2$, $p(1/z) = 1/z^2 + 1$. Similarly, $(x, y') := \Psi(x, y)$ is defined, and the valuation of $x$ and $y'$ only depends on $a$ and $b$:

$$\psi(a,b) := (\mathrm{val}(x), \mathrm{val}(y')) = \begin{cases} (a, -b + a(v_{-1}^{(x)} - v_1^{(x)})), & \text{if } a \geq 0, \\ (a, -b + a(d_{-1}^{(x)} - d_1^{(x)})), & \text{if } a \leq 0, \end{cases}$$

where $v_i^{(x)}$ (respectively $d_i^{(x)}$) denotes the valuation (resp. degree) in $y$ of $A_i(x)$, for $i = \pm 1$.

This construction provides a simplification of the problem, as, in order to prove that $G$ is infinite, it suffices to prove that the group $G'$ generated by $\phi$ and $\psi$ is infinite. To prove the latter it is sufficient to exhibit $(a, b) \in \mathbb{Z}^2$, such that the orbit of $(a, b)$ under the action of $G'$ is infinite.

After these preparations we are ready to consider the particular case of the five sets shown in Figure 4.8. In all 5 cases $\bar{x}y$ and $x\bar{y}$ are part of $\mathcal{S}$ and they are also the only $x$-negative or $y$-negative steps respectively. Hence, $A_{-1}(x) = x$ and $B_{-1}(y) = y$ and $v_{-1}^{(x)} = d_{-1}^{(x)} = v_{-1}^{(y)} = d_{-1}^{(y)} = 1$. Additionally we know that $v_1^{(x)} = v_1^{(y)} = -1$, again because of the steps $\bar{x}y$ and $x\bar{y}$ which imply that $\bar{x}$ is a term in $A_1(x)$ and $\bar{y}$ is a term in $B_1(y)$. Applying this we get

$$\phi(a,b) = \begin{cases} (-a + 2b, b), & \text{if } b \geq 0, \\ \left(-a + b(1 - d_1^{(y)}), b\right), & \text{if } b \leq 0, \end{cases} \quad \psi(a,b) := \begin{cases} (a, 2a - b), & \text{if } a \geq 0, \\ \left(a, -b + a(1 - d_1^{(x)})\right), & \text{if } a \leq 0. \end{cases}$$

The only thing which is left, is to find a suitable pair for $(a, b) \in \mathbb{Z}^2$ that produces an infinite group. Consider $(\psi \circ \phi)(a, b) = (-a + 2b, -2a + 3b)$ with the constraints $b \geq 0$ and $2b \geq a$. Hence, e.g. $(a, b) = (1, 2)$ is a possible choice. It is easy to show by induction that

$$(\psi \circ \phi)^n(1,2) = (2n+1, 2n+2) \qquad \text{and} \qquad \phi(\psi \circ \phi)^n(1,2) = (2n+3, 2n+2).$$

As all these pairs have positive entries, we never need to know $d_1^{(y)}$ or $d_1^{(x)}$. This proves that the orbit of $(1, 2)$ under $G'$ is infinite, and so are the groups $G'$ and $G$.

**2. The fixed point argument**

There are 51 models remaining. Due to Lemma 4.9 this number can be reduced to 14 models to study. Note, that this is roughly a quarter and not an eighth, because we already considered $x/y$-symmetries. They are listed in [7, Table 5].

Let us introduce the general strategy again first: Assume $\Theta = \Psi \circ \Phi$ is well-defined in the neighborhood of $(a, b) \in \mathbb{C}^2$, and this point is fixed by $\Theta$, i.e. $\Theta(a, b) = (a, b)$. Note, that this implies that $a$ and $b$ are algebraic over $\mathbb{Q}$. Now, write $\Theta = (\Theta_1, \Theta_2)$ for the two coordinates of $\Theta$, i.e. $\Theta_i \in \mathbb{Q}(x, y)$, $i = 1, 2$. We look at the local expansion of $\Theta$ around $(a, b)$:

$$\Theta(a + u, b + v) = (a, b) + (u, v)J_{a,b} + \mathcal{O}(u^2) + \mathcal{O}(v^2) + \mathcal{O}(uv),$$

where $J_{a,b}$ is the Jacobian Matrix of $\Theta$ at $(a, b)$, i.e.

$$J_{a,b} = \begin{pmatrix} \frac{\partial \Theta_1}{\partial x}(a,b) & \frac{\partial \Theta_2}{\partial x}(a,b) \\ \frac{\partial \Theta_1}{\partial y}(a,b) & \frac{\partial \Theta_2}{\partial y}(a,b) \end{pmatrix}.$$

Iterating the above expansion yields, for $m \geq 1$:

$$\Theta^m(a + u, b + v) = (a, b) + (u, v)J^m_{a,b} + \mathcal{O}(u^2) + \mathcal{O}(v^2) + \mathcal{O}(uv).$$

Assume now, that $G(\mathcal{S})$ is finite of order $2n$, then $\Theta$ is of order $n$ which means that $\Theta^n(a + u, b + v) = (a, b) + (u, v)$. This can only be, if $J^n_{a,b}$ is the identity matrix. Let $\lambda$ be an eigenvalue of $J_{a,b}$ with eigenvector $v$, then we get

$$v = J^n_{a,b} v = \lambda^n v,$$

which implies, that all eigenvalues of $J_{a,b}$ are roots of unity.

This gives a general strategy for proving that a group $G(\mathcal{S})$ is infinite:

1. Find a fixed point $(a, b)$ of $\Theta$;
2. Compute its characteristic polynomial $\chi(X)$ of the Jacobian matrix $J_{a,b}$ ($\chi(X) \in \mathbb{Q}(a, b)[X]$);
3. Eliminate $a$ and $b$ from the equation $\chi(X) = 0$ to obtain a polynomial $\bar{\chi}(X) \in \mathbb{Q}[X]$ that vanishes at all eigenvalues of $J_{a,b}$;
4. If none of its factors is *cyclotomic*, we can conclude that $G(\mathcal{S})$ is infinite.

Remark, that as $a$ and $b$ are algebraic step is 3 always possible. Furthermore, step 4 is more effective than it looks like on the first view, as all cyclotomic polynomials of given degree are known. As short recap, remember its definition:

***Definition 4.11:*** The $n^{\text{th}}$ *cyclotomic polynomial* is given by

$$\Phi_n := \prod_{\substack{\omega \in \mathbb{C} \text{ primitive} \\ n^{\text{th}} \text{ root of unity}}} (X - \omega) = \prod_{\substack{1 \leq k < n \\ gcd(k,n)=1}} (X - e^{\frac{2\pi i k}{n}}) \in \mathbb{C}[X].$$

◊

From its definition it follows immediately, that the $\deg \Phi_n = \varphi(n)$, where $\varphi(n)$ is Euler's totient function.

Let us perform one case in detail. Let $\mathcal{S} = \{x, y, xy, \bar{x}\bar{y}\}$. Here we have

$$\Phi(x,y) = \left(\bar{x}\frac{\bar{y}}{1+y}, y\right) \qquad \text{and} \qquad \Psi(x,y) = \left(x, \bar{y}\frac{\bar{x}}{1+x}\right).$$

A straightforward calculation shows, that

$$\Theta(x,y) = (\Psi \circ \Phi)(x,y) = \left(\frac{1}{xy(1+y)}, \bar{y}\frac{(xy(1+y))^2}{1+xy(1+y)}\right).$$

Checking the fixed point condition, we get, that every pair $(a,a)$ such that $a^4 + a^3 = 1$ is fixed by $\Theta$. We skip the technical calculations of the Jacobian matrix, note that it might be easier by setting $C(x,y) = xy(1+y)$. At the end we get

$$J_{a,b} = \begin{pmatrix} -1 & 2-a^3 \\ & \\ a^3-2 & \frac{a(3a+2)}{(1+a)^2} \end{pmatrix}.$$

All we have used, was the fact, that $a^4+a^3 = 1$, or in an equivalent form $\frac{1}{a(a+1)} = a^2$, as $a \neq 0$. This equation can also be used to reduce the powers of $a$. As a characteristic polynomial we get

$$\chi(X) := \det(X\mathrm{Id} - J_{a,b}) = X^2 + (1 - a(2+3a)^7)X + a^6 - 4a^3 + 4 - a(2+3a)^7.$$

By eliminating $a$ from this expression, we obtain

$$\chi(X) = X^8 - 19X^7 - X^6 - 124X^5 + 3X^4 - 124X^3 - X^2 - 19X + 1.$$

This polynomial is irreducible and distinct from all cyclotomic polynomials of degree 8. This can be easily seen, by the fact, that cyclotomic polynomials have mostly small coefficients, in the sense, that the smallest order of a cyclotomic polynomial[1] containing a coefficient $\geq 2$ or $\leq -2$ is 105 (Note, that the smallest order for a coefficient $-124$ to occur is 40755.) Hence, none of its roots is a root of unity and that is why no power of $J_{a,b}$ is equal to the identity matrix. Thus, $G(\mathcal{S})$ is infinite. $\square$

### 4.2.3 Orbit Sums and a General Result

**A functional equation**

Let $\mathcal{S}$ be an arbitrary step set, and $\mathcal{Q}$ be the class of walks that start from the origin $(0,0)$, take their steps from $\mathcal{S}$ and always stay in the first quadrant. Let $q(i,j;n)$ be the number of such walks that have length $n$ and end at position $(i,j)$. The main question of concern is the trivariate generating function

$$Q(x,y;z) = \sum_{i,j,n\geq 0} q(i,j;n)x^i y^j z^n,$$

[1] Smallest order of cyclotomic polynomial containing $n$ or $-n$ as a coefficient; `http://oeis.org/A013594`, accessed 26/08/2013.

whose nature will be determined. As a short hand we denote $Q(x,z) \equiv Q(x,y;z)$.

***Lemma 4.12 [7, Lemma 4]:*** Following functional equaion characterizes $Q(x,y)$:

$$K(x,y)xyQ(x,y) = xy - zxA_{-1}(x)Q(x,0) - zyB_{-1}(y)Q(0,y) + z\epsilon Q(0,0), \tag{4.3}$$

where

$$K(x,y) = 1 - zS(x,y) = 1 - z \sum_{(i,j)\in\mathcal{S}} x^i y^j$$

is called the *kernel* of the equation. The polynomials $A_{-1}(x)$ and $B_{-1}(y)$ are the coefficients of $\bar{y}$ and $\bar{x}$ in $S(x,y)$, as described by (4.1), and $\epsilon$ is 1 if $(-1,-1)$ is one of the allowed steps, and 0 otherwise.

*Proof:* The structure of the functional equation is a direct consequence of a step-by-step construction of walks using only steps from $\mathcal{S}$: A walks is either the empty walk or it has non-zero length. A non-empty walk of length $n$ is obtained, by concatenating a step from $\mathcal{S}$ to a walk of length $n-1$. But we have to enforce the restriction, that the walks are confined to the quarter plane, hence we have to take special care at the borders. On the $x$-axis, any $y$-negative step leaves the quarter plane and such walks are removed again. Symmetrically for the $y$-axis. Denote the set of $y$-negative ($x$-negative) steps from $\mathcal{S}$ as $\mathcal{S}_y^-$ ($\mathcal{S}_x^+$) and the set of walks ending on the $x$-axis ($y$-axis) as $\mathcal{Q}_x^0$ ($\mathcal{Q}_y^0$). Finally, if $(-1,-1) \in \mathcal{S}$ we removed it twice with the previous operation, hence we have to add it again.

This procedure summarized in formulas on classes is given by

$$\mathcal{Q} = \mathcal{E} \ \dot\cup \ (\mathcal{Z}\times\mathcal{S}\times\mathcal{Q}) \setminus \left(\mathcal{Z}\times\mathcal{S}_y^-\times\mathcal{Q}_x^0\right) \setminus \left(\mathcal{Z}\times\mathcal{S}_x^+\times\mathcal{Q}_y^0\right) \ \dot\cup \ \begin{cases} \mathcal{Z}\times\mathcal{S}_{(-1,-1)}\times\mathcal{Q}_{0,0} & \text{if } (-1,-1)\in\mathcal{S}, \\ \emptyset & \text{otherwise,} \end{cases}$$

which translates by the symbolic method directly into the functional equation

$$Q(x,y) = 1 + zS(x,y)Q(x,y) - z\bar{y}A_{-1}(x)Q(x,0) - z\bar{x}B_{-1}(y)Q(0,y) + \epsilon z\bar{x}\bar{y}Q(0,0).$$

Multiplying this equation by $xy$, gives the equation of the lemma.

The fact, that it characterizes $Q(x,y)$ completely (as a power series in $z$) is justified by the fact, that the coefficient of $z^n$ in $Q(x,y)$ can be computed inductively using this equation. This reflects the recursive description given above. □

**Orbit sums**

We have seen in Lemma 4.5 that all transformation $g$ of the group $G$ leave the characteristic polynomial $S(x,y)$ unchanged. Thus, the kernel $K(x,y) = 1 - zS(x,y)$ is also left unchanged by them. Now we want to exploit this property: Write equation (4.3) as

$$K(x,y)xyQ(x,y) = xy - F(x) - G(y) + z\epsilon Q(0,0),$$

with $F(x) := zxA_{-1}(x)Q(x,0)$ and $G(y) := zyB_{-1}(y)Q(0,y)$. As a next step, replace $(x,y)$ by $(x',y) := \Phi(x,y)$:

$$K(x,y)x'yQ(x',y) = x'y - F(x') - G(y) + z\epsilon Q(0,0).$$

By taking the difference between these two equations and we get

$$K(x, y)\,(xyQ(x, y) - x'yQ(x', y)) = xy - x'y - F(x) + F(x').$$

Due to the fact, that $\Phi$ does not changed the $y$-value, the $G(y)$ term has disappeared. This process can be repeated, by adding to the last identity equation (4.3) evaluated at $(x', y') := \Psi(x', y) = (\Psi \circ \Phi)(x, y)$. This gives

$$K(x, y)\,(xyQ(x, y) - x'yQ(x', y) + x'y'Q(x', y')) = xy - x'y + x'y' - F(x) - G(y') + z\epsilon Q(0, 0).$$

This results in the disappearance of the term $F(x')$. If $G$ is finite of order $2n$, we can repeat this process by traversing the orbit of $(x, y)$ until we come back to $(\Psi \circ \Phi)^n(x, y) = (x, y)$. All known functions on the right-hand side eventually vanish (note, that $z\epsilon Q(0, 0)$ vanishes, as the order is even). For convenience we introduce the following definition

$$g(A(x, y)) := A(g(x, y)) \qquad \text{for } g \in G.$$

So we have proved the following proposition:

***Proposition 4.13 [7, Proposition 5]:*** Assume the group $G(\mathcal{S})$ is finite. Then

$$\sum_{g \in G} \operatorname{sign}(g) g(xyQ(x, y; z)) = \frac{1}{K(x, y; z)} \sum_{g \in G} \operatorname{sign}(g) g(xy).$$

The remarkable observation is, that the right-hand side is a rational function.

#### A general result

After the classification, the second main result is the proof that 22 of the 23 step sets, that generate a finite group are holonomic. As mentioned in the introduction, the holonomy of the 23$^{\text{rd}}$ walk, namely Gessel's walk, was shown in [6].

We show the general result, that applies to 19 of the 23 step sets. As the others have to be discussed in detail step-by-step, we refer to the original paper for the proofs. Before studying following Proposition recall Proposition 2.22.

***Theorem 4.14 [7, Proposition 8]:*** For the 23 models associated with a finite group, except from the four cases $\mathcal{S} = \{\bar{x}, \bar{y}, xy\}$, $\mathcal{S} = \{x, y, \bar{x}\bar{y}\}$, $\mathcal{S} = \{x, y, \bar{x}, \bar{y}, xy, \bar{x}\bar{y}\}$ and $\mathcal{S} = \{x, \bar{x}, xy, \bar{x}\bar{y}\}$, the following holds. The rational function

$$R(x, y; z) = \frac{1}{K(x, y; z)} \sum_{g \in G} \operatorname{sign}(g) g(xy)$$

is a power series in $z$ with coefficients in $\mathbb{Q}(x)[y, \bar{y}]$. Moreover, the positive part in $y$ of $R(x, y; z)$, denoted $R^+(x, y; z)$, is a power series in $z$ with coefficients in $\mathbb{Q}[x, \bar{x}, y]$. Extracting the positive part in $x$ of $R^+(x, y; z)$ gives $xyQ(x, y; z)$. In brief

$$xyQ(x, y; z) = [x^>][y^>]R(x, y; z). \tag{4.4}$$

In particular, $Q(x, y; z)$ is holonomic. The number of $n$-step walks ending at $(i, j)$ is

$$q(i, j; n) = [x^{i+1}y^{j+1}] \left( \sum_{g \in G} \operatorname{sign}(g) g(xy) \right) S(x, y)^n.$$

*Proof:* The theorem applies to all step sets of Table 4.5, the first two of Table 4.6 and the first one of Table 4.7.

We are going to distinguish two cases: First we cover the 16 models associated with a group of order 4 (compare Theorem 4.10), then we address the 3 remaining cases, namely $\mathcal{S} = \{\bar{x}, y, x\bar{y}\}$, $\mathcal{S} = \{x, \bar{x}, x\bar{y}, \bar{x}y\}$ and $\mathcal{S} = \{x, \bar{x}, y, \bar{y}, x\bar{y}, \bar{x}y\}$.

As seen in Theorem 4.10, all models of order 4 possess a vertical symmetry, i.e. $K(x, y) = K(\bar{x}, y)$. In a similar manner as in Example 4.8, the orbit for step sets with a vertical symmetry is obtained as

$$(x, y) \overset{\Phi}{\longleftrightarrow} (\bar{x}, y) \overset{\Psi}{\longleftrightarrow} (\bar{x}, C(x)\bar{y}) \overset{\Phi}{\longleftrightarrow} (x, C(x)\bar{y}) \overset{\Psi}{\longleftrightarrow} (x, y),$$

with $C(x) = \frac{A_{-1}(x)}{A_1(x)}$. This information is used to evaluate the orbit sum of Proposition 4.13:

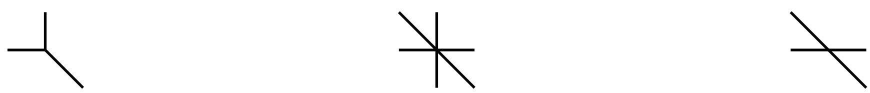
$$xyQ(x, y) - \bar{x}yQ(\bar{x}, y) + \bar{x}\bar{y}C(x)Q(\bar{x}, C(x)\bar{y}) - x\bar{y}C(x)Q(x, C(x)\bar{y}) = R(x, y).$$

Observe, that both sides of this identity are series in $z$ with coefficients in $\mathbb{Q}(x)[y, \bar{y}]$. For the positive part in $y$ only the first two terms of the left-hand side contribute, hence extracting it yields

$$xyQ(x, y) - \bar{x}yQ(\bar{x}, y) = R^+(x, y).$$

From the expression of the left-hand side, it is clear that $R^+(x, y)$ has coefficients in $Q[x, \bar{x}, y]$, because all terms are polynomials and no truly rational functions from $\mathbb{Q}(x)[y] \setminus \mathbb{Q}[x, \bar{x}, y]$ are involved. Extracting the positive part in $x$ shows (4.4), as the second term of the left-hand $\bar{x}yQ(\bar{x}, y)$ contains only powers of $\bar{x}$.

Secondly, we consider the cases $\mathcal{S} = \{\bar{x}, y, x\bar{y}\}$, $\mathcal{S} = \{x, \bar{x}, y, \bar{y}, x\bar{y}, \bar{x}y\}$ and $\mathcal{S} = \{x, \bar{x}, x\bar{y}, \bar{x}y\}$, shown in Figure 4.9.



(a) $\mathcal{S} = \{\bar{x}, y, x\bar{y}\}$ (b) $\mathcal{S} = \{x, \bar{x}, y, \bar{y}, x\bar{y}, \bar{x}y\}$ (c) $\mathcal{S} = \{x, \bar{x}, x\bar{y}, \bar{x}y\}$

Figure 4.9: Non-vertically symmetric step sets for which general result of Theorem 4.14 applies

Analogously, as in the first case we want to apply the result of orbit sums. Hence, we first need to consider the orbits of these step sets:

(a) $\mathcal{S} = \{\bar{x}, y, x\bar{y}\}$: same as (b).

(b) $\mathcal{S} = \{x, \bar{x}, y, \bar{y}, x\bar{y}, \bar{x}y\}$: $\Phi(x,y) = (\bar{x}y, y)$, $\Psi(x,y) = (x, x\bar{y})$

$$(x,y) \overset{\Phi}{\longleftrightarrow} (\bar{x}y, y) \overset{\Psi}{\longleftrightarrow} (\bar{x}y, \bar{x}) \overset{\Phi}{\longleftrightarrow} (\bar{y}, \bar{x}) \overset{\Psi}{\longleftrightarrow} (\bar{y}, x\bar{y}) \overset{\Phi}{\longleftrightarrow} (x, x\bar{y}) \overset{\Psi}{\longleftrightarrow} (x,y)$$

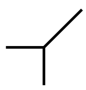
(c) $\mathcal{S} = \{x, \bar{x}, x\bar{y}, \bar{x}y\}$: $\Phi(x,y) = (\bar{x}y, y)$, $\Psi(x,y) = (x, x^2\bar{y})$

$$(x,y) \overset{\Phi}{\leftrightarrow} (\bar{x}y, y) \overset{\Psi}{\leftrightarrow} (\bar{x}y, \bar{x}^2y) \overset{\Phi}{\leftrightarrow} (\bar{x}, \bar{x}^2y) \overset{\Psi}{\leftrightarrow} (\bar{x}, \bar{y}) \overset{\Phi}{\leftrightarrow} (x\bar{y}, \bar{y}) \overset{\Psi}{\leftrightarrow} (x\bar{y}, x^2\bar{y}) \overset{\Phi}{\leftrightarrow} (x, x^2\bar{y}) \overset{\Psi}{\leftrightarrow} (x,y)$$

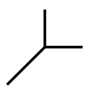
All elements in these orbits are of the form $(x^a y^b, x^c y^d)$ with $a, b, c, d \in \mathbb{Z}$. This special cases ensures that $R(x,y)$ is a series in $z$ with coefficients in $\mathbb{Q}[x, \bar{x}, y, \bar{y}]$ (compare Theorem 4.13). When extracting the positive part in $x$ and $y$ it is easily checked, that in each of the three cases only the term $xyQ(x,y)$ contributes on the left-hand side. Note, that all pairs except the first one contain either $\bar{x}$ or $\bar{y}$.

So in all cases Proposition 2.22 applies and it follows that $Q(x,y;z)$ is holonomic. The expression of $q(i,j;n)$ follows from a simple coefficient extraction. □

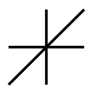
***Remark 4.15:*** The remaining 4 step sets, depicted in Figure 4.10, which also generate a finite group $G$ are associated with not only holonomic, but even algebraic generating functions. Bousquet-Mélou and Mishna show this for $\mathcal{S} = \{\bar{x}, \bar{y}, xy\}$, $\mathcal{S} = \{x, y, \bar{x}\bar{y}\}$ and $\mathcal{S} = \{x, y, \bar{x}, \bar{y}, xy, \bar{x}\bar{y}\}$. The proofs are given Propositions 13 – 15 in [7]. However, the last step set $\mathcal{S} = \{x, \bar{x}, xy, \bar{x}\bar{y}\}$ is solved by Bostan and Kauers in [6]. ■



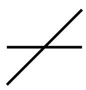
(a) $\mathcal{S} = \{\bar{x}, \bar{y}, xy\}$



(b) $\mathcal{S} = \{x, y, \bar{x}\bar{y}\}$



(c) $\mathcal{S} = \{x, y, \bar{x}, \bar{y}, xy, \bar{x}\bar{y}\}$



(d) $\mathcal{S} = \{x, \bar{x}, xy, \bar{x}\bar{y}\}$,
Gessel's Walk

Figure 4.10: The 4 exceptions to the general result which still allow a holonomic (even algebraic) generating function

# Chapter 5

# Self-Avoiding Walks

The interest in self-avoiding walks (SAWs) probably first arose in the field of chemistry in the 1950s, in order to model the behavior of polymer chains, whose physical volume prohibits multiple occupation of the same spatial point (i.e. excluded volume constraint). They were introduced by the famous chemist and Nobel laureate P. Flory[1] who derived the first non-rigorous results on SAWs [14]. But SAWs have also found interesting applications in different sciences, such as the physics of magnetic materials and the study of phase transitions [20]. Needless to say they are also of great interest as purely mathematical objects leading to rich mathematical theories and challenging questions [35].

Despite the basic definition of the problem and more than 70 years of mathematical investigation, rigorous results about properties of SAWs remain scarce. On the contrary there exists a huge amount of numerical information which supports many conjectures, and most of them are universally believed to be true, but they remain unproven. It is interesting to observe, that all exact and conjectural information we have, applies only to the model on the two-dimensional lattice [17]. The focus of this chapter lies on one special property of SAWs on a two-dimensional lattice. In particular we discuss the proof of the exact value of the *connective constant* on the hexagonal lattice in 2D which was found recently by H. Duminil-Copin[2] and S. Smirnov[3] [10].

## 5.1 Definitions

As in the previous part of this work we are going to work with the nearest-neighbor model which consists of the following step set

$$\mathcal{S} = \{x \in \mathbb{Z}^d : \|x\|_1 = 1\}. \tag{5.1}$$

***Definition 5.1:*** An $n$-step *self-avoiding walk* (SAW) from $0 \in \mathbb{Z}^d$ to any $x \in \mathbb{Z}^d$ relative to $\mathcal{S}$ is a sequence $\omega = (\omega_0, \omega_1, \dots, \omega_n)$ of elements in $\mathbb{Z}^d$, such that

[1] Paul John Flory, 19.6.1910-9.9.1985
[2] Hugo Duminil-Copin, 26.8.1985-
[3] Stanislav Konstantinowitsch Smirnov, 3.9.1970-

1. $\omega_0 = 0$, $\omega_n = x$,
2. $\omega_{i+1} - \omega_i \in \mathcal{S}$,
3. $\omega_i \neq \omega_j$ for all $i \neq j$ (*self-avoidance*).

Let $c_n$ denote the number of SAWs of length $n$. $\Diamond$

***Remark 5.2:*** The same reasoning as in Remark 1.6 leads to the conclusion, that Definition 5.1 could be generalized to SAWs starting at a specific point $s \in \mathbb{Z}^d$ by replacing the origin as first element of the sequence with $s$, i.e. $\omega_0 = s$. But this fact does not represent a restriction on our discussion as we are going to consider *homogeneous* lattices, in the sense that the number $c_n$ of $n$-step SAWs starting from $s$ is independent for all values of $n$.

Condition 2 simply means, that $|\omega_{i+1} - \omega_i| = 1$, for the Euclidean distance, as we are considering the nearest-neighbor model. $\blacksquare$

When studying SAWs there are three major questions which arise (nearly) naturally [17]:

1. How many SAWs are there of length $n$?
2. How "big" is the typical $n$-step SAW? How might we actually measure size?
3. What is the scaling limit of SAWs?

Concerning question 1 the asymptotic behavior of $c_n$ is of particular interest. From numerical experiments and intuitive reasoning by comparison to similar families of walks, the approximative number of $n$-step SAWs may be loosely stated as

$$c_n \approx \mu^n \tag{5.2}$$

where the growth rate $\mu$ is called the *connective constant* of the lattice [21, pp. 423].

The following lemma implies the existence of the connective constant.

***Lemma 5.3 (Fekete [40, p. 24/198]):*** Let $(a_n)_{n\in\mathbb{N}}$ be a sequence of non-negative real numbers satisfying the subadditivity inequality

$$a_{n+m} \leq a_n + a_m. \tag{5.3}$$

Then the sequence $\left(\frac{a_n}{n}\right)_{n\in\mathbb{N}}$ converges to its lower bound $\inf_{n\in\mathbb{N}} \frac{a_n}{n} < \infty$.

*Proof:* Observe from the non-negativity of $a_n$ that the sequence $a_n/n \in \mathbb{R}$ is bounded from below by zero. Hence, by the completeness property of the real numbers it possess a greatest lower bound or infimum $\alpha = \inf_{n\in\mathbb{N}} \frac{a_n}{n} \in \mathbb{R}$. This implies that for every $\varepsilon > 0$ exists an index $m \in \mathbb{N}$ such that

$$\frac{a_m}{m} < \alpha + \frac{\varepsilon}{2}.$$

Any number $n \in \mathbb{N}$ can be decomposed into $n = qm + r$ where $q, r \in \mathbb{N}$ and $0 \leq r \leq m-1$. We define $a_0 = 0$. Then we derive by the subadditivity

$$a_n = a_{qm+r} \leq \underbrace{a_m + a_m + \ldots a_m}_{q \text{ times}} + a_r = q a_m + a_r,$$

$$\frac{a_n}{n} \leq \frac{q a_m + a_r}{qm + r} = \frac{a_m}{m}\frac{qm}{qm+r} + \frac{a_r}{n} \leq \left(\alpha + \frac{\varepsilon}{2}\right)\frac{qm}{qm+r} + \frac{a_r}{n}.$$

By choosing $N$ large enough to provide $\frac{\max\{a_1,a_2,\ldots,a_{m-1}\}}{n} < \frac{\varepsilon}{2}$ for all $n \geq N$ we get

$$\alpha \leq \frac{a_n}{n} \leq \alpha + \frac{\varepsilon}{2} + \frac{\max\{a_1, a_2, \ldots, a_{m-1}\}}{n} \leq \alpha + \varepsilon, \qquad \forall n \geq N.$$

As $\varepsilon$ was arbitrary this proves the lemma. □

We will proof the next theorem in a more general context than we are going to need it. It formulates the necessary conditions for the connective constant to exist on a lattice which not necessarily corresponds with the Euclidean Lattice $\mathbb{Z}^d$ (compare Section 1.1).

***Theorem 5.4 (Hammersley-Morton, [21, p. 424]):*** Assume that a lattice has the following properties

(i) the lattice is homogenous in the sense of Remark 5.2,
(ii) for each positive integer $n$, at least one $n$-step SAW is possible (i.e. $c_n \geq 1$),
(iii) the number of edges leading out of any vertex is finite (the number leading into any vertex need not be finite).

Then

$$\lim_{n\to\infty} \frac{1}{n} \log c_n = \log \mu \tag{5.4}$$

exists, where $1 \leq \mu < \infty$. Moreover, for each value of $n$

$$\frac{1}{n} \log c_n \geq \log \mu. \tag{5.5}$$

*Proof:* First note that condition (iii) is clearly necessary, as $c_n$ has to be finite for every $n$ to be well-defined. An $(n+m)$-step SAW can be uniquely decomposed into an $n$-step SAW followed by a translation of an $m$-step SAW starting at the endpoint of the $n$-step walk. Hence, $c_n$ satisfies the following submultiplicative inequality

$$c_{n+m} \leq c_n c_m, \tag{5.6}$$

that is strict for sufficiently large values of $n$ and $m$. The reverse inequality is of course false. As $c_n \geq 1$ for all $n$ we may take logarithms to deduce a subadditive inequality in the manner of (5.3) for the sequence

$$a_n = \log c_n$$

of non-negative real numbers. The theorem now follows directly from Lemma 5.3. □

The result (5.4) finally makes precise what is meant by assertion (5.2):

$$c_n \sim f(n)\mu^n, \qquad \text{where } \log f(n) = o(n), \tag{5.7}$$

while the bound (5.5) may be used to derive upper bounds on the connective constant by careful enumeration of all possible SAWs of a finite number of steps.

Obviously the connective constant depends on the geometry of the lattice. For the nearest-neighbor model the connective constant only depends on the dimension $d$ of the lattice. Trivial bounds may be found by considering only walks that move in positive coordinate directions, and by counting walks that are restricted only to prevent immediate reversals of steps [4, p. 398]. Hereby we obtain

$$d^n \leq c_n \leq 2d(2d-1)^{n-1} \qquad \text{which implies} \qquad d \leq \mu \leq 2d-1 \tag{5.8}$$

as $\lim_{n\to\infty}(c_n)^{1/n} = \mu$.

For $d = 2$ the following rigorous bounds are known:

$$\mu \in [2.625622, 2.679193].$$

The lower bound is due to Jensen (2004) [24] via bridge enumeration, while the upper bound is due to Pönitz and Tittmann (2000) [41]. The estimate

$$\mu = 2.63815853031(3)$$

is given in [23] by Jensen, who achieved it with a parallel algorithm on a cluster consisting of 500 1GHz processors with a total peak speed over 1Tflop. The 3 in parentheses represents the spread (basically one standard deviation) among the approximants. Jensen remarks that this error bound should not be viewed as a measure of the true error as it cannot include possible systematic sources of error. On the square lattice it has been observed that $1/\mu$ is well approximated by the unique positive root of the polynomial $581x^4 + 7x^2 - 13$ [9, 25]. This "conjecture" has to be treated carefully, as it has been derived by the idea, that the connective constant of the honeycomb lattice $\mu_h = \sqrt{2+\sqrt{2}}$ (see Section 5.3) satisfies a quadratic equation in $1/\mu_h^2$. Therefore Conway, Enting and Guttmann tried to find an analogous equation with "small" integer coefficients for an approximation of the connective constant on the square lattice. This result remains a purely numerical observation, and later evidence has raised doubts about its validity [23, p. 11].

Table 5.1 states some results of upper and lower bounds for the connective constant $\mu$ on different lattices. Examples of the quoted lattices are shown in Figure 1.1.

| Lattice | Lower Bound | $\mu$ | Upper Bound |
|---|---|---|---|
| Square | 2.625622 | 2.63815853031(3) | 2.679193 |
| Triangular | 4.118935 | 4.150797226(26) | 4.25152 |
| Hexagonal | 1.841925 | 1.847759065... | 1.868832 |
| Kagomé | 2.548497 | 2.560576765(10) | 2.590301 |

Table 5.1: Jensen's collection of results (2004) for the connective constant $\mu$ from [24, p. 10]; The lower bounds are from Jensen; the upper bounds from Pönitz/Tittman, Alm, Alm/Parviainen and Guttman/Parviainen/Rechnitzer, respectively.

### 5.1.1 Critical Exponents

Models of statistical mechanics possess the characteristic feature, that there exist *critical exponents* at the the critical point which describe the asymptotic behavior on the large scale.

Numerical experiments lead to the conjecture, that these critical exponents are *universal* in the sense, that they depend only on the spatial dimension $d$ but not on specific details of the lattice in $\mathbb{R}^d$ [4, p. 399].

As seen in (5.7) the asymptotic behavior of the number of SAWs $c_n$ depends on a function $f(n)$. It is generally accepted in physics literature that for $n \to \infty$,

$$c_n \sim A\mu^n n^{\gamma-1}, \tag{5.9}$$

where $A \in \mathbb{R}$ is a lattice dependent constant and the exponent $\gamma$ depends only on the dimension of the lattice. From (5.5) we have that $c_n \geq \mu^n$, and so if the exponent $\gamma$ exists, it holds $\gamma \geq 1$. The predicted values of the critical exponent are:

$$\gamma = \begin{cases} 1 & d = 1, \\ \frac{43}{32} & d = 2, \\ 1.16\ldots & d = 3, \\ 1 & d \geq 4. \end{cases}$$

The case $d = 4$ is irregular, as it involves a logarithmic correction:

$$c_n \sim A\mu^n (\log n)^{1/4}.$$

The case $\gamma = 1$ of dimensions $d \geq 5$ reflects the intuition, that for large dimensions the defining restrictive behavior of a SAW does not influence its asymptotic number anymore. We can see that in the case of a simple random walk we have $c_n = |\mathcal{S}|^n$, i.e. $\mu = |\mathcal{S}|$ and $\gamma = 1$.

Note, that the critical exponent also has a probabilistic interpretation for SAWs. Drawing two $n$-step SAWs $\omega_1, \omega_2$ uniformly at random, it holds that

$$\mathbb{P}(\omega_1 \cap \omega_2 = \{0\}) = \frac{c_{2n}}{c_n^2} \sim C\frac{1}{n^{\gamma-1}}.$$

for a constant $C \in \mathbb{R}$. Hence, $\gamma$ is a measure of how likely it is for two SAWs to avoid each other.

Similar to previous conjectures, these too withstood all attempts of rigorous proofs so far and the best known bounds are more than 50 years old. In [18] Hammersley[4] and Welsh[5] proved that for all $d \geq 2$ and all $n \in \mathbb{N}$ there exists a constant $\kappa > 0$ depending on $d$ such that

$$\mu^n \leq c_n \leq \mu^n e^{\kappa\sqrt{n}}. \tag{5.10}$$

Although (5.10) is an improvement over $c_n \leq \mu^n e^{o(n)}$, which follows directly from (5.7), it is still much larger than the predicted growth rate $c_n \sim A\mu^n n^{\gamma-1}$.

One year later in 1963 Kesten proved that there exists a constant $A > 0$ such that

$$\left|\frac{c_{n+2}}{c_n} - \mu^2\right| \leq An^{-1/3}. \tag{5.11}$$

[4] John Michael Hammersley, 21.3.1920-2.5.2004

[5] Dominic James Anthony Welsh, 29.8.1938-

This inequality has an interesting application. Combining (5.11) with the trivial bounds $\mu \geq d$ and $c_{n+1} \leq (2d-1)c_n$, which were shown in (5.8), we obtain

$$c_n \leq \frac{4}{3}\mu^{-2}c_{n+2} \leq \frac{4}{3}\frac{2d-1}{d^2}c_{n+1} \leq c_{n+1}$$

for all $n$ large enough (see below). The last inequality holds because $d \geq 2$. The first inequality is a bit tricky. It follows from (5.11): As the right-hand side tends to zero for $n \to \infty$ we can choose it arbitrarily small, let's say we choose $0 < \varepsilon < \mu^2$. Then we deduce

$$\begin{aligned} \mu^2 - \frac{c_{n+2}}{c_n} &\leq \varepsilon && \Leftrightarrow \\ c_n \leq \frac{1}{\mu^2 - \varepsilon}c_{n+2} &= \frac{1}{1-\frac{\varepsilon}{\mu^2}}\mu^{-2}c_{n+2}. \end{aligned}$$

Hence, choosing $n$ as large as that $\varepsilon \leq \frac{\mu^2}{4}$ holds, we get the first inequality above. In this context it is interesting to note, that the proof of $c_n \leq c_{n+1}$ for *all* $n$ is a non-trivial result proved by O'Brien in 1990 [38]. Above argumentation is inspired by N. Madras, mentioned in the same paper.

## 5.2 Bridges and the Hammersley-Welsh Bound

In this section we introduce and study bridges, a specific class of SAWs and show that the number of bridges grows with the same exponential rate as the number of SAWs, namely $\mu^n$. In Section 5.3 an analogous result will play a key role in the derivation of the connective constant for the hexagonal lattice $\mathbb{H}$. The goal of this section is to prove the Hammersley-Welsh bound (5.10), which is founded on the idea of bridge decomposition. We are going to follow mainly the ideas of [4, Section 2.1].

***Definition 5.5:*** For an $n$-step SAW $\omega = (\omega_0, \omega_1, \dots, \omega_n)$, we denote by $\omega_i^{[1]}$ the first spacial coordinate of $\omega_i$ for $i = 0, 1, \dots, n$.

An $n$-step *bridge* is an $n$-step SAW $\omega$, such that

$$\omega_0^{[1]} < \omega_i^{[1]} \leq \omega_n^{[1]} \qquad \text{for } i = 1, 2, \dots, n.$$

Define $b_n$ as the number of $n$-step bridges with $\omega_0 = 0$ for $n > 1$ and $b_0 := 1$. ◊

We see directly from the definition, that the number of bridges is *supermultiplicative*, i.e.

$$b_{n+m} \geq b_n b_m, \tag{5.12}$$

in contrast to the number of SAWs $c_n$ which is submultiplicative by (5.6). Applying Lemma 5.3 to $-\log b_n$, we derive the existence of the *bridge growth constant*

$$\mu_{Bridge} = \lim_{n\to\infty} b_n^{1/n} = \sup_{n\geq 1} b_n^{1/n}. \tag{5.13}$$

Due to the definition, we get the trivial inequality $\mu_{Bridge} \leq \mu$ which implies directly that

$$b_n \leq \mu_{Bridge}^n \leq \mu^n. \tag{5.14}$$

Next we consider a generalization of bridges, by dropping the maximality property in the first component of the end-point $\omega_n$.

***Definition 5.6:*** An $n$-step *half-space walk* is an $n$-step SAW $\omega$ with

$$\omega_0^{[1]} < \omega_i^{[1]} \qquad\qquad \text{for } i = 1, 2, \dots, n.$$

Define $h_n$ as the number of $n$-step half-space walks with $\omega_0 = 0$ for $n > 1$ and $h_0 := 1$. $\Diamond$

***Definition 5.7:*** The *span* of an $n$-step SAW $\omega$ is defined as

$$\max_{0 \leq i \leq n} \omega_i^{[1]} - \min_{0 \leq i \leq n} \omega_i^{[1]}.$$

Define $b_{n,A}$ as the number of $n$-step bridges with span $A$. $\Diamond$

In two dimensions, the span can be interpreted as the "width" of a SAW.

In the study of bridges, the following asymptotic result on integer partitions from Hardy[6] and Ramanujan[7] will be useful.

***Theorem 5.8 [19, p. 260]:*** For an integer $N \geq 1$, let $P_D(N)$ denote the number of ways of writing $N = N_1 + N_2 + \dots N_k$ with $N_1 > N_2 > \dots > N_k \geq 1$ for any $k \geq 1$. Then

$$\log P_D(N) \sim \pi \sqrt{\frac{N}{3}} \tag{5.15}$$

for $N \to \infty$.

This structural property will appear in the following proof, while applying a bridge decomposition argument on any SAW $\omega$.

***Proposition 5.9:*** $h_n \leq P_D(n) b_n$ for all $n \geq 1$.

*Proof:* Let $\omega$ be any $n$-step SAW. Set $n_0 = 0$ and inductively define the following auxiliary variables

$$A_{i+1} = \max_{j > n_i} (-1)^i \left( \omega_j^{[1]} - \omega_{n_i}^{[1]} \right)$$

and

$$n_{i+1} = \max \left\{ j > n_i : (-1)^i \left( \omega_j^{[1]} - \omega_{n_i}^{[1]} \right) = A_{i+1} \right\}.$$

By construction, $n_1$ maximizes $\omega_j^{[1]}$, $n_2$ minimizes $\omega_j^{[1]}$ for $j > n_1$, $n_3$ maximizes $\omega_j^{[1]}$ for $j > n_3$, and so on in an alternating pattern. Additionally, these values are the last times these extrema are attained. The $A_i$ correspond to the largest distances in the first component while traversing through the walk starting from $n_i$ till the end. Compare Figure 5.1 for details.

As the walk $\omega$ is finite, this procedure stops at some step $K \geq 1$ where $n_K = n$. Observe, that the $A_i$ form a strictly monotonically decreasing sequence, as the $n_i$ are chosen to be maximal.

[6] Godfrey Harold Hardy, 7.2.1877-1.12.1947

[7] Srinivasa Ramanujan Aiyangar, 22.12.1887-26.4.1920

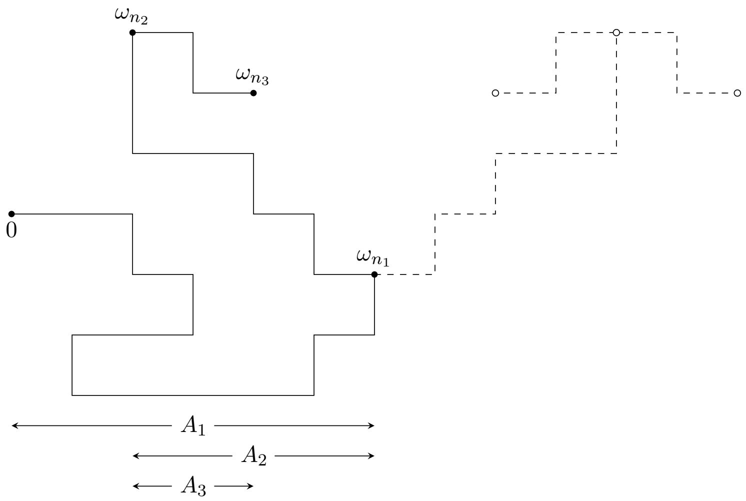



Figure 5.1: A half-space walk is decomposed into bridges, which are reflected to form a single bridge.

Furthermore note, that $K = 1$ if and only if $\omega$ is a bridge, as by definition $\omega_j^{[1]} \leq \omega_n^{[1]}$ for $j < n$. In that case $A_1$ is the span of $\omega$. Let $h_n[a_1, \ldots, a_k]$ denote the number of $n$-step half-space walks with $K = k$ and $A_i = a_i$ for $i = 1, \ldots, k$. It holds, that

$$h_n[a_1, a_2, a_3, \ldots, a_k] \leq h_n[a_1 + a_2, a_3, \ldots, a_k].$$

Consider a walk with the properties of $h_n[a_1, \ldots, a_k]$. By reflecting the walk $(\omega_j)_{j \geq n_1}$ across the hyper plain attained by fixing the first coordinate $x_1 = A_1$ we get a walk with $A_1 = a_1 + a_2$ (see Figure 5.1). By Repetition of this process we create a bridge of length $n$ and hence:

$$h_n[a_1, a_2, a_3, \ldots, a_k] \leq h_n[a_1 + \ldots + a_k] = b_{n, a_1 + \ldots + a_k}.$$

Thus, we get

$$\begin{aligned} h_n &= \sum_{k \geq 1} \sum_{a_1 > \ldots > a_k > 0} h_n[a_1, \ldots, a_k] \\ &\leq \sum_{k \geq 1} \sum_{a_1 > \ldots > a_k > 0} b_{n, a_1 + \ldots + a_k} \\ &= \sum_{k=1}^{n} P_D(k) b_{n,k}, \end{aligned}$$

due to the fact, that we are summing over all integer partitions of $n$ with distinct positive components. The number of partitions $P_D(k)$ is obviously monotonically increasing in $k$, hence we complete the proof by bounding $P_D(k) \leq P_D(n)$ and get

$$h_n \leq P_D(n) \sum_{k=1}^{n} b_{n,k} = P_D(n) b_n$$

as claimed. □

Finally, we need an elementary lemma about binomial equations.

***Lemma 5.10:*** Let $x, y \geq 0$. Then $\sqrt{x} + \sqrt{y} \leq \sqrt{2x + 2y}$.

*Proof:* As both sides are positive squaring the inequality gives

$$\begin{aligned} x + 2\sqrt{xy} + y &\leq 2x + 2y && \Leftrightarrow \\ 2\sqrt{xy} &\leq x + y && \Leftrightarrow \\ 0 &\leq (\sqrt{x} - \sqrt{y})^2. \end{aligned}$$

□

With the last three auxiliary results we are able to prove the Hammersley-Welsh bound [4,18].

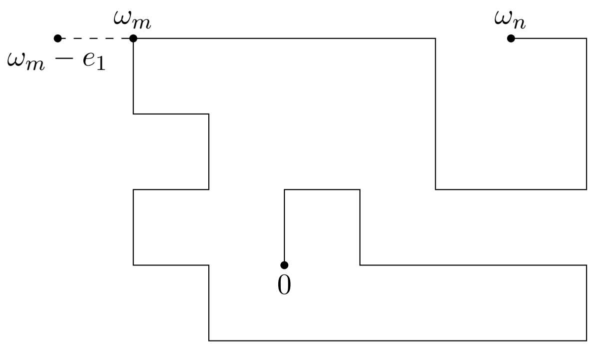



Figure 5.2: Decomposition of a SAW into two half-space walks.

***Theorem 5.11 (Hammersley-Welsh):*** Fix $B > \pi(\frac{2}{3})^{1/2}$. Then there is an $n_0 = n_0(B)$ independent of the dimension $d \geq 2$, such that

$$c_n \leq b_{n+1} e^{B\sqrt{n}} \leq \mu^{n+1} e^{B\sqrt{n}} \qquad \text{for } n \geq n_0.$$

*Proof:* First we show that any $n$-step SAW can be transformed into two half-space walks and we obtain

$$c_n \leq \sum_{m=0}^{n} h_{n-m} h_{m+1}. \tag{5.16}$$

This is achieved by using the decomposition sketched in Figure 5.2 as follows. Let $\omega$ be an $n$-step SAW and define

$$x_1 := \min_{0 \leq i \leq n} \omega_i^{[1]}, \qquad\qquad m := \max\left\{ i \;:\; \omega_i^{[1]} = x_1 \right\}.$$

Let $e_1$ be the unit vector in the first coordinate direction of $\mathbb{Z}^d$. Now we construct out of $\omega$ two half-space walks of length $n - m$ and $m + 1$.

1. By the choice of $m$ the walk $(\omega_j)_{j\geq m}$ is an $(n-m)$-step half-space walk starting at $\omega_m$. After translation it starts at the origin and is part of all half-space walks counted by $h_{n-m}$.
2. Again by the choice of $m$ the walk $(\omega_m - e_1, \omega_m, \omega_{m-1}, \dots, \omega_1, \omega_0)$ is an $(m+1)$-step half-space walk starting at $\omega_m - e_1$. With the same arguments as above it is counted by $h_{m+1}$.

Next, we apply Proposition 5.9 in (5.16) and use the supermultiplicity of $b_n$ (5.12) to get

$$\begin{aligned} c_n &\leq \sum_{m=0}^{n} P_D(n-m)P_D(m+1)b_{n-m}b_{m+1} \\ &\leq b_{n+1}\sum_{m=0}^{n} P_D(n-m)P_D(m+1). \end{aligned} \tag{5.17}$$

We want to apply the asymptotic result for integer partitions. By Theorem 5.8 we get for any $\varepsilon > 0$ a natural number $N_0 \geq 0$ such that

$$\log P_D(N) \leq (1+\varepsilon)\pi\sqrt{\frac{N}{3}} = \underbrace{(1+\varepsilon)\pi\sqrt{\frac{2}{3}}}_{=:B'(\varepsilon)}\sqrt{\frac{N}{2}} \qquad \text{for all } N > N_0.$$

As $\varepsilon > 0$ is arbitrary, we can choose $B' := B'(\varepsilon)$ small enough so that $B > B' > \pi\sqrt{\frac{2}{3}}$. Additionally, there exists a $K > 0$ such that

$$P_D(N) \leq K \exp\left(B'\sqrt{\frac{N}{2}}\right) \qquad \text{for all } N \geq 0.$$

For example set $K = \max\{P_D(1), \dots, P_D(N_0)\}$ in order to cover the first $N_0$ cases. Note, that this holds because $\exp(B'\sqrt{N/2}) \geq 1$ for all $N \geq 0$. Consequently we obtain

$$P_D(n-m)P_D(m+1) \leq K^2 \exp\left[B'\left(\sqrt{\frac{n-m}{2}} + \sqrt{\frac{m+1}{2}}\right)\right].$$

Applying Lemma 5.10 to (5.17) gives

$$c_n \leq (n+1)K^2 e^{B'\sqrt{n+1}} b_{n+1} = \exp\left[\underbrace{\left(\frac{\log\left((n+1)K^2\right)}{B\sqrt{n}} + \frac{B'}{B}\sqrt{1+\frac{1}{n}}\right)}_{\leq 1 \text{ for } n\geq n_0(B)} B\sqrt{n}\right] b_{n+1}.$$

Hence, by (5.14) the result follows:

$$c_n \leq e^{B\sqrt{n}} b_{n+1} \leq e^{B\sqrt{n}} \mu^{n+1} \qquad \text{for } n \geq n_0(B).$$

□

From the previous theorem we directly get:

***Corollary 5.12:*** For $n > n_0(B)$,

$$b_n \geq c_{n-1}e^{-B\sqrt{n-1}} \geq \mu^{n-1}e^{-B\sqrt{n-1}}. \tag{5.18}$$

In particular $b_n^{1/n} \to \mu$ and thus $\mu_{Bridge} = \mu$.

*Proof:* From (5.5) we know that $c_n \geq \mu^n$ for all $n \in \mathbb{N}$. Combining this with the result from Theorem 5.11 yields for $n > n_0(B)$

$$\mu^{n-1} \leq c_{n-1} \leq b_n e^{B\sqrt{n-1}} \leq \mu^n e^{B\sqrt{n-1}}.$$

Rearranging this inequality chain gives the asserted (5.18). Because of (5.13) taking the $n$-th root and letting $n \to \infty$ shows $\mu_{Bridge} = \mu$ . □

## 5.3 Connective Constant of the Honeycomb Lattice equals $\sqrt{2+\sqrt{2}}$

In the following we are mainly going to follow the notation and structure of Duminil-Copin and Smirnov [10] combined with definitions and remarks from [4, Section 3].

Throughout this section we consider SAWs on the hexagonal lattice $\mathbb{H}$ applying the nearest-neighbor model. Our primary goal of this section is the proof of the following theorem.

***Theorem 5.13:*** For the hexagonal lattice $\mathbb{H}$,

$$\mu = \sqrt{2+\sqrt{2}}.$$

***Remark 5.14:*** In 1982 based on Coulomb gas formalism the physicist B. Nienhuis [37] (see also [21, Section 7.6.5]) predicted the connective constant $\mu$ for the hexagonal lattice to be equal to $\sqrt{2+\sqrt{2}}$. The arguments in the following differ from the ones Nienhuis used, but they are similarly motivated by considerations of vertex operations in the $O(n)$-model. The interested reader is referred to [21, Section 7.6.1] for an introduction to the $O(n)$-model from statistical mechanics. ■

### 5.3.1 Notation

We are going to consider walks between *mid-edges* of $\mathbb{H}$, i.e. centers of edges of $\mathbb{H}$ (see Figure 5.3a). The set of all mid-edges will be called $H$. We define $\omega : a \to E$ for all walks $\omega$ starting at mid-edge $a \in H$ and ending at some mid-edge of $E \subset H$. If $E = \{b\}$ we also write $\omega : a \to b$. The *length* $\ell(\omega)$ of the walk $\omega$ is the number of vertices in $\mathbb{H}$ belonging to $\omega$.

We define the generating function

$$C(x) = \sum_{n \geq 0} c_n x^n \tag{5.19}$$

of SAWs. More often we will use the following representation

$$C(x) = \sum_{\omega : a \to H} x^{\ell(\omega)}. \tag{5.20}$$

As our lattice is homogeneous it does not depend on the choice of $a$ and is increasing in $x$, due to $\ell(\omega) \in \mathbb{N}$. To simplify the formulas below and in order to use notation which anticipates our conclusion we define the two important values

$$x_c := \frac{1}{\sqrt{2+\sqrt{2}}}, \tag{5.21}$$

$$j := e^{2\pi i/3}. \tag{5.22}$$

***Lemma 5.15:*** The value $\mu = \sqrt{2+\sqrt{2}}$ is equivalent to

1. $C(x) = +\infty$ for $x > x_c$ and
2. $C(x) < +\infty$ for $x < x_c$.

*Proof:* From calculus of power series we know, that $C(x)$ is associated with a radius of convergence $R \geq 0$. The power series diverges for $|x| > R$ and converges (absolutely) for $|x| < R$. This radius can be calculated from representation (5.19) of $C(x)$ using the coefficients $c_n$ in the following way

$$R = \frac{1}{\limsup\limits_{n\to\infty} c_n^{1/n}}.$$

We know from Hammersley and Morton (Theorem 5.4) that the limit of the sequence $(c_n^{1/n})_{n\in\mathbb{N}}$ exists and is equal to $\mu$. Hence,

$$R = \frac{1}{\mu}$$

proves the statement. □

Due to this reason the value $x_c$ is also called the *critical value* or *critical point* of the generating function $C(x)$.

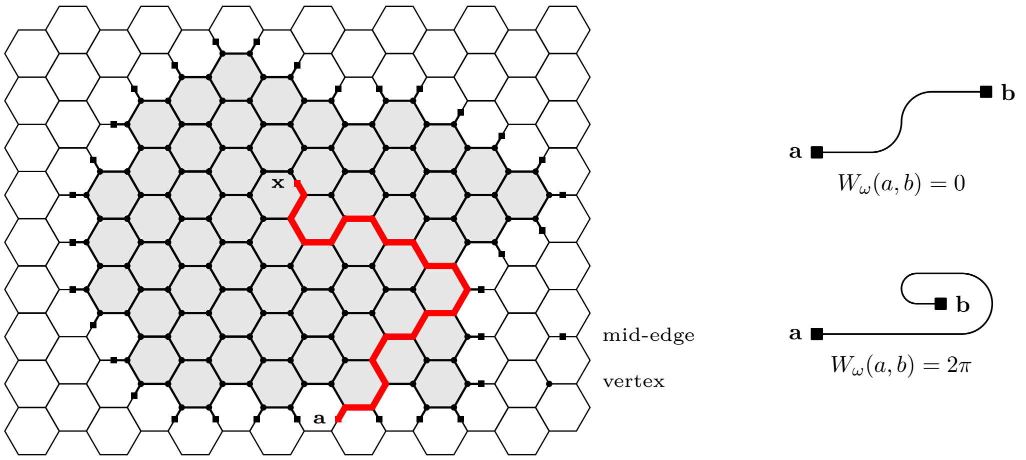
### 5.3.2 The Holomorphic Observable

The core idea of the proof is the use of a generalization of (5.20) which we call *holomorphic observable*[8]. This is achieved by introducing weights on individual walks which depend on its winding (see below).

***Definition 5.16:*** A *domain* $\Omega \subset H$ is the union of all mid-edges emanating from a given connected collection of vertices $V(\Omega)$ (see Figure 5.3a). A mid-edge $z$ belongs to $\Omega$ if at least one endpoint of its associated edge is in $V(\Omega)$.

The boundary $\partial\Omega$ consists of the mid-edges $z$ whose associated edge has exactly one endpoint in $V(\Omega)$. ◊

[8] In the original paper this function is called *parafermionic observable.*




(a) Hexagonal Domain $\Omega$ (b) Winding of SAW $\omega$

Figure 5.3: A hexagonal domain $\Omega$ in gray whose boundary mid-edges are pictured by small black squares. Vertices of $V(\Omega)$ correspond to circles and a sample walk from $a$ to $x$.

In the following we assume $\Omega$ to be simply-connected, i.e. having a connected complement in the Riemann sphere [27, Korollar 4.2.9].

***Definition 5.17:*** The *winding number* $W_\omega(a,b)$ of a SAW $\omega$ between mid-edges $a$ and $b$ (not necessarily the start and the end) is the total rotation in radians when $\omega$ is traversed from $a$ to $b$ (compare Figure 5.3b). $\diamond$

Now we are able to describe our main tool for the following analysis.

***Definition 5.18:*** Fix $a \in \partial\Omega$ and $\sigma \in \mathbb{R}$. The *holomorphic observable* for $z \in \Omega$ and $x \geq 0$ is defined by

$$F_x(z) = \sum_{\omega \subset \Omega : a \to z} e^{-i\sigma W_\omega(a,z)} x^{\ell(\omega)}. \tag{5.23}$$

$\diamond$

Note, that that the term $e^{-i\sigma W_{\omega(a,z)}}$ can be interpreted as a complex weight, which simplifies to a product of values $\lambda$ and $\bar{\lambda}$ per left or right turn of $\omega$, with

$$\lambda = \exp\left(-i\sigma\frac{\pi}{3}\right). \tag{5.24}$$

Here we used that a turn on the hexagonal lattice is exactly $\pi/3$ radians.

***Lemma 5.19:*** If $x = x_c$ and $\sigma = \frac{5}{8}$, then $F_z$ satisfies for every vertex $v \in V(\Omega)$

$$(p-v)F_{x_c}(p) + (q-v)F_{x_c}(q) + (r-v)F_{x_c}(r) = 0, \tag{5.25}$$

where $p, q, r$ are the mid-edges of the three edges adjacent to $v$ (see Figure 5.4a).


(a) Adjacent Mid-Edges (b) Edge Rotation

Figure 5.4: Hexagonal Lattice Details

*Proof:* Without loss of generality assume that $p, q$ and $r$ are oriented counter-clockwise around $v$. Next we interpret (5.25) as one sum over all walks $\omega$ starting at $a$ and ending at either $p, q$ or $r$, i.e. with $E = \{p, q, r\}$ it is equivalent to

$$\sum_{\omega: a \to E} c(\omega) = 0,$$

where $c(\omega)$ is one contribution in (5.25) over all possible SAWs $\omega$ ending at $p, q$ or $r$. For instance, if $\omega$ ends at mid-edge $p$, it will be

$$c(\omega) = (p - v)e^{-i\sigma W_\omega(a,p)} x^{\ell(\omega)}.$$

Next we want to partition the set of walks finishing in $E$ into pairs and triplets of walks in the following way (see Figure 5.5):

- Pairs: If a SAW $\omega_1$ visits all **three** mid-edges $p, q, r$, there exists a partition of the walk into a SAW plus (up to a half-edge) a self-avoiding loop from $v$ to $v$. One can associate to $\omega_1$ the walk $\omega_2$ passing through the same edges, but traversing the loop from $v$ to $v$ in the other direction. Like this, walks visiting all three mid-edges can be grouped in pairs.

- Triplets: If a walk $\omega_1$ visits only **one** mid-edge, it can be associated to two walks $\omega_2$ and $\omega_3$ that visit exactly **two** mid-edges by extending the walk one step further to the two possible choices. Now consider the reverse: A walk that visits exactly two mid-edges is naturally associated to a walk visiting only one mid-edge, by erasing the last step. Thus, walks visiting one or two mid-edges can be grouped in triplets.

As last step we will show that the sum of contributions for each pair or triplet vanishes which implies naturally that the total sum is zero.

Firstly, let $\omega_1$ and $\omega_2$ be two associated walks from the first case. Without loss of generality, we assume that $\omega_1$ ends at $q$, $\omega_2$ ends at $r$ and both pass through $p$. Let $a$ denote their common starting point. Due to their association they are equal up to the mid-edge $p$, and follow an almost complete loop in two opposite directions. We deduce

$$\ell(\omega_1) = \ell(\omega_2) \qquad \text{and} \qquad \begin{cases} W_{\omega_1}(a,q) = W_{\omega_1}(a,p) + W_{\omega_1}(p,q) = W_{\omega_1}(a,p) - \frac{4\pi}{3}, \\ W_{\omega_2}(a,r) = W_{\omega_2}(a,p) + W_{\omega_2}(p,r) = W_{\omega_1}(a,p) + \frac{4\pi}{3}. \end{cases}$$

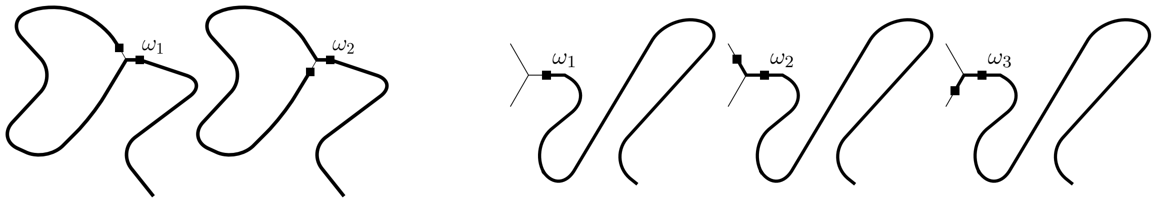



Figure 5.5: Left: Pair of walks visiting three mid-edges and grouped together. Right: Triplet of walks, one visiting one mid-edge, the two others visiting two mid-edges each. All matched together [10, p. 4].

For the evaluation of $W_{\omega_1}(p,q)$ we used that $a$ is on the boundary $\partial\Omega$ and $\Omega$ is simply-connected. These conditions guarantee that $\omega_1$ does not traverse in one or more loops around the starting point $a$. This would increase the value by $2\pi k$ for $k \in \mathbb{Z}$ counting the loops around $a$ in positive or negative direction.

Using notation (5.24) for $\lambda$ and $j = e^{2\pi i/3}$ we conclude

$$\begin{aligned} c(\omega_1) + c(\omega_2) &= (q-v)e^{-i\sigma W_{\omega_1}(a,q)} x_c^{\ell(\omega_1)} + (r-v)e^{-i\sigma W_{\omega_2}(a,r)} x_c^{\ell(\omega_2)} \\ &= e^{-i\sigma W_{\omega_1}(a,p)} x_c^{\ell(\omega_1)} \left((q-v)e^{i\sigma\frac{4\pi}{3}} + (r-v)e^{-i\sigma\frac{4\pi}{3}}\right) \\ &= (p-v)e^{-i\sigma W_{\omega_1}(a,p)} x_c^{\ell(\omega_1)} \left(j\bar{\lambda}^4 + \bar{j}\lambda^4\right). \end{aligned}$$

The last equality holds, as $(p-v)e^{2\pi i/3} = (q-v)$ and $(p-v)e^{-2\pi i/3} = (r-v)$ (compare Figure 5.4b).

Now we set $\sigma = \frac{5}{8}$ and obtain

$$j\bar{\lambda}^4 + \bar{j}\lambda^4 = e^{i\frac{3\pi}{2}} + e^{-i\frac{3\pi}{2}} = 2\cos\left(\frac{3\pi}{2}\right) = 0, \tag{5.26}$$

and hence $c(\omega_1) + c(\omega_2) = 0$.

Secondly, let $\omega_1, \omega_2, \omega_3$ be three walks matched like in the second case. Without loss of generality, assume $\omega_1$ ends at $p$ and that $\omega_2$ and $\omega_3$ extend to $q$ and $r$, respectively. With similar arguments as before we see

$$\ell(\omega_2) = \ell(\omega_3) = \ell(\omega_1) + 1 \quad \text{and} \quad \begin{cases} W_{\omega_2}(a,q) = W_{\omega_2}(a,p) + W_{\omega_2}(p,q) = W_{\omega_1}(a,p) - \frac{\pi}{3}, \\ W_{\omega_3}(a,r) = W_{\omega_3}(a,p) + W_{\omega_3}(p,r) = W_{\omega_1}(a,p) + \frac{\pi}{3}. \end{cases}$$

These values give

$$c(\omega_1) + c(\omega_2) + c(\omega_3) = (p-v)e^{-i\sigma W_{\omega_1}(a,p)} x_c^{\ell(\omega_1)} \left(1 + x_c j\bar{\lambda} + x_c\bar{j}\lambda\right). \tag{5.27}$$

Due to the choice of $\sigma = \frac{5}{8}$ we get $\lambda = \exp\left(-i\frac{5\pi}{24}\right)$. Hence, we get analogously $j\bar{\lambda} + \bar{j}\lambda = 2\cos(\frac{7\pi}{8}) = -2\cos(\frac{\pi}{8})$. Therefore the claim that (5.27) vanishes is equivalent to

$$2\cos\left(\frac{\pi}{8}\right) \cdot x_c = 1. \tag{5.28}$$

This equation holds, because $2\cos(\frac{\pi}{8}) = \sqrt{2+\sqrt{2}}$ which can be easily proved by using the identity $\cos\left(\frac{x}{2}\right) = \operatorname{sign}(\cos(\frac{x}{2}))\sqrt{\frac{1+\cos(x)}{2}}$. This is the only point in the entire proof of Theorem 5.13 where we explicitly need the choice $x = x_c = 1/\sqrt{2+\sqrt{2}}$.

Finally the claim of the lemma follows by summing over all pairs and triplets of walks. □

***Remark 5.20:*** As seen in the above proof, the coefficients of (5.25) are three cube roots of unity multiplied by $p - v$ (compare Figure 5.4b). So its left hand-side can be interpreted as a discrete $dz$-integral along an elementary contour on the dual lattice. The fact that the integral of the holomorphic observable along discrete contours vanishes justifies its name, as it suggests that it is *discrete holomorphic.* ■

### 5.3.3 Proof of Theorem 5.13

To reduce the complexity of the problem, we consider a vertical strip domain $S_T$ composed of $T$ strips of hexagons, and its finite version $S_{T,L}$ cut at heights $\pm L$ at angles $\pm\pi/3$, see Figure 5.6. In other words, position a hexagonal lattice $\mathbb{H}$ of meshsize[9] 1 in $\mathbb{C}$ so that there exists a horizontal edge e with the mid-edge $a$ being 0. Then[10]

$$V(S_T) = \{z \in V(\mathbb{H}) : 0 \le \operatorname{Re}(z) \le 3\frac{T+1}{2}\},$$
$$V(S_{T,L}) = \{z \in V(S_T) : |2\operatorname{Im}(z) - \operatorname{Re}(z)| \le \sqrt{3}(2L+1)\}.$$

The left and right boundaries of $S_T$ and $S_{T,L}$ are denoted by $\alpha$ and $\beta$, respectively. The top and bottom boundaries of $S_{T,L}$ are denoted by $\varepsilon$ and $\bar{\varepsilon}$, respectively. In accordance with these declarations we define the following positive partition functions

$$A_{T,L}(x) = \sum_{\omega\subset S_{T,L}:a\to\alpha\setminus\{a\}} x^{\ell(\omega)},$$
$$B_{T,L}(x) = \sum_{\omega\subset S_{T,L}:a\to\beta} x^{\ell(\omega)},$$
$$E_{T,L}(x) = \sum_{\omega\subset S_{T,L}:a\to\varepsilon\cup\bar{\varepsilon}} x^{\ell(\omega)}.$$

This decomposition allows us to deduce a global identity from equation (5.25) without complex but positive weights on the introduced strip domain.

***Lemma 5.21:*** For $x = x_c$, the following identity holds

$$1 = c_\alpha A_{T,L}(x_c) + B_{T,L}(x_c) + c_\varepsilon E_{T,L}(x_c), \tag{5.29}$$

with positive coefficients $c_\alpha = \cos\left(\frac{3\pi}{8}\right)$ and $c_\varepsilon = \cos\left(\frac{\pi}{4}\right)$.

[9]The diameter of one hexagon.

[10]In the original paper the upper bound in $V(S_T)$ is specified as $(3T+1)/2$, and $V(S_{T,L})$ is bounded by $|\sqrt{3}\operatorname{Im}(z) - \operatorname{Re}(z)| \le 3L$, but the used definition of meshsize is not stated.

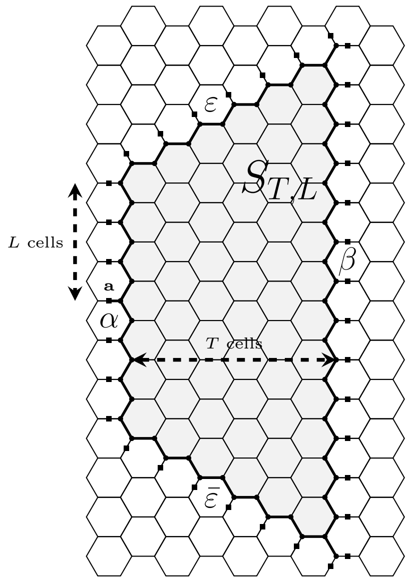



Figure 5.6: Domain $S_{T,L}$ and boundary intervals $\alpha, \beta, \varepsilon$ and $\bar{\varepsilon}$ [10, p. 5].

*Proof:* To improve readability we fix $x = x_c$ and drop it from notation, e.g. $F := F_x$. As a first step we sum (5.25) over all vertices in $V(S_{T,L})$. All contributions of interior mid-edges vanish, because both end-points are parts of this sum. Particularly, fix one interior mid-edge $p \in H$ with two end-points $v_1, v_2 \in V(S_{T,L})$. The total sum contains the two terms $(p - v_1)F(p)$ and $(p - v_2)F(p)$, whose sum vanishes as $(p - v_1) = -(p - v_2)$. Hence, we deduce

$$0 = -\sum_{z\in\alpha} F(z) + \sum_{z\in\beta} F(z) + j\sum_{z\in\varepsilon} F(z) + \bar{j}\sum_{z\in\bar{\varepsilon}} F(z). \tag{5.30}$$

Next we are going to consider these sums individually. The winding from any SAW from $a$ to the top part of $\alpha$ is $\pi$, while the winding to the bottom part is $-\pi$. Furthermore, we can use the symmetry of our domain with respect to the real axis, which implies $F(\bar{z}) = \overline{F(z)}$ and the fact that the only SAW from $a$ to $a$ has length 0. This gives for the first sum

$$\begin{aligned}\sum_{z\in\alpha} F(z) &= F(a) + \sum_{z\in\alpha\backslash\{a\}} F(z) = F(a) + \frac{1}{2}\sum_{z\in\alpha\backslash\{a\}} (F(z) + F(\bar{z})) = 1 + \frac{e^{-i\sigma\pi} + e^{i\sigma\pi}}{2} A_{T,L} \\ &= 1 + \cos(\sigma\pi)\, A_{T,L} = 1 - \cos\left(\frac{3\pi}{8}\right) A_{T,L} = 1 - c_\alpha A_{T,L}.\end{aligned}$$

Similarly we get, that the winding number from $a$ to any half-edge in $\beta$, $\varepsilon$ or $\bar{\varepsilon}$ is 0, $\frac{2\pi}{3}$ or $-\frac{2\pi}{3}$ and therefore, again using symmetry, for the other sums

$$\sum_{z\in\beta} F(z) = B_{T,L},$$

$$\begin{aligned} j\sum_{z\in\varepsilon} F(z) + \bar{j}\sum_{z\in\bar{\varepsilon}} F(z) &= \frac{j}{2} e^{-i\sigma\frac{2\pi}{3}} E_{T,L} + \frac{\bar{j}}{2} e^{i\sigma\frac{2\pi}{3}} E_{T,L} = \frac{e^{i\frac{2\pi}{3}(1-\sigma)} + e^{-i\frac{2\pi}{3}(1-\sigma)}}{2} E_{T,L} \\ &= \cos\left(\frac{2\pi}{3}(1-\sigma)\right) E_{T,L} = \cos\left(\frac{\pi}{4}\right) E_{T,L}(x) = c_\varepsilon E_{T,L}.\end{aligned}$$

The lemma follows now directly by inserting the last three formulas into (5.30). $\square$

Next, we are going to extend the finite strip domain $S_{T,L}$ to its unbounded version $S_T$. Therefore, observe that the sequences $\left(A_{T,L}(x)\right)_{L>0}$ and $\left(B_{T,L}(x)\right)_{L>0}$ are increasing in $L$ and are bounded for $x < x_c$, due to (5.29) and their monotonicity in $x$. Thus they have limits

$$A_T(x) = \lim_{L\to\infty} A_{T,L}(x) = \sum_{\omega\subset S_T: a\to\alpha\setminus\{a\}} x^{\ell(\omega)},$$
$$B_T(x) = \lim_{L\to\infty} B_{T,L}(x) = \sum_{\omega\subset S_T: a\to\beta} x^{\ell(\omega)}.$$

As the coefficients $c_\alpha$ and $c_\varepsilon$ in (5.30) are positive for $x = x_c$, this identity implies that $\left(E_{T,L}(x)\right)_{L>0}$ decreases and converges to a limit $E_T(x) = \lim_{L\to\infty} E_{T,L}(x)$. Hence, we obtain

$$1 = c_\alpha A_T(x_c) + B_T(x_c) + c_\varepsilon E_T(x_c). \tag{5.31}$$

For the final step we need one last definition which is an adaption of Definition 5.5.

***Definition 5.22:*** A *bridge* on $\mathbb{H}$ is a SAW, which never revisits the vertical line through its starting point and never visits a vertical line to the right of the vertical line through its endpoint. Furthermore it starts and ends at the mid-edge of a horizontal edge.

The *width* of a bridge is the horizontal distance between its start- and end-point. $\diamondsuit$

We denote the number of $n$-step bridges in $\mathbb{H}$, which start from 0 by $b_n$ and its generating function by $B(z) = \sum_{n\geq 0} b_n z^n$.

*Proof of Theorem 5.13:* First note, that similar arguments as in the proof of Hammersley-Welsh in Section 5.2 on the hexagonal lattice lead to the conclusion $\mu_{Bridge} = \mu$ also on $\mathbb{H}$. Hence, it is sufficient to show that

$$\mu_{Bridge} = \sqrt{2+\sqrt{2}}.$$

Observe, that the bridge generating function is given by $B(z) = \sum_{T=0}^{\infty} B_T(z)$.

First, assume $x < x_c$. By (5.31) we get $B_T(x) \leq 1$ as all constants (i.e. $c_\alpha$ and $c_\varepsilon$) are positive. Now note, that a bridge of width $T$ has at least length $T$, i.e. $B_T(x) = \sum_{n\geq T} b_n x^n$. We obtain

$$B_T(x) \leq \left(\frac{x}{x_c}\right)^T B_T(x_c) \leq \left(\frac{x}{x_c}\right)^T,$$

and hence $B(x)$ is finite for $x < x_c$ as we found a summable geometric series as a majorant.

It remains to show, that $Z(x_c) = \infty$ or $B(x_c) = \infty$ which implies that $\mu \geq \sqrt{2+\sqrt{2}}$ by Lemma 5.15. This is done, by considering two separate cases. First suppose that for some $T$, $E_T(x_c) > 0$. As observed above $E_{T,L}(x_c)$ decreases in $L$ and therefore

$$Z(x_c) \geq \sum_{L>0} E_{T,L}(x_c) \geq \sum_{L>0} E_T(x_c) = \infty.$$

On the contrary, assume $E_T(x_c) = 0$ for all $T$, which simplifies (5.31) to

$$1 = c_\alpha A_T(x_c) + B_T(x_c). \tag{5.32}$$

Let us consider the geometry of walks $\omega$ contributing to $A_{T+1}(x_c)$ but not to $A_T(x_c)$. These walks must visit some vertex adjacent to the right edge of $S_{T+1}$. Cutting the walk $\omega$ at the first such point (and adding half-edges to the two halves), we uniquely decompose it into two bridges of width $T+1$ in $S_{T+1}$, which together are one step longer than $\omega$. We conclude

$$A_{T+1}(x_c) - A_T(x_c) \leq x_c \left(B_{T+1}(x_c)\right)^2. \tag{5.33}$$

Combining (5.32) for two consecutive values of $T$ with (5.33) we get

$$\begin{aligned} 0 &= \left(c_\alpha A_{T+1}(x_c) + B_{T+1}(x_c)\right) - \left(c_\alpha A_T(x_c) + B_T(x_c)\right) \\ &\leq c_\alpha x_c \left(B_{T+1}(x_c)\right)^2 + B_{T+1}(x_c) - B_T(x_c), \end{aligned}$$

which is equivalent to

$$c_\alpha x_c \left(B_{T+1}(x_c)\right)^2 + B_{T+1}(x_c) \geq B_T(x_c).$$

We are going to use this relation to show by induction, that

$$B_T(x_c) \geq \min\{B_1(x_c), 1/(c_\alpha x_c)\}/T$$

for every $T \geq 1$.

For brevity define $\zeta := c_\alpha x_c > 0$, $\eta := \min\{B_1(x_c), 1/\zeta\} \geq 0$ and write $B_T = B_T(x_c)$. Hence, we show $B_T \geq \frac{\eta}{T}$ by induction. By the choice of $\eta$ the case $T = 1$ is trivial. Consider $T \to T+1$: From the induction hypothesis (IH) we get

$$cB_{T+1}^2 + B_{T+1} \geq B_T \overset{\text{(IH)}}{\geq} \frac{\eta}{T}.$$

Solving the quadratic equation $cB_{T+1}^2 + B_{T+1} - \frac{\eta}{T} = 0$ gives the only possible result $B_{T+1} = \frac{-1+\sqrt{1+4\zeta\eta/T}}{2\zeta}$ because $B_{T+1} \geq 0$. Now we use this value, to check the induction assumption:

$$\begin{aligned} B_{T+1} = \frac{-1+\sqrt{1+4\frac{\zeta\eta}{T}}}{2\zeta} &\overset{!}{\geq} \frac{\eta}{T+1} && \Leftrightarrow \\ 1 + 4\frac{\zeta\eta}{T} &\overset{!}{\geq} 1 + 4\frac{\zeta\eta}{T+1} + 4\left(\frac{\zeta\eta}{T+1}\right)^2 && \Leftrightarrow \\ \frac{1}{T} - \frac{1}{T+1} &\overset{!}{\geq} \frac{\zeta\eta}{(T+1)^2} && \Leftrightarrow \\ \frac{1}{T} &\geq \frac{\zeta\eta}{T+1} \end{aligned}$$

The last inequality holds, as $\zeta\eta = \min\{\zeta B_1, 1\} \leq 1$.

This implies

$$B(x_c) \geq \sum_{T>0} B_T(x_c) = \infty$$

and completes the proof. □

***Remark 5.23:*** The value of the connective constant was motivated in Remark 5.14 by empirical experiments performed by Nienhuis, but the value of $\sigma$ in the holomorphic observable (5.23) seems quite arbitrary or at least suitable for doing the purpose. Clearly these two values are interconnected, but how is not completely obvious. We want to give a motivation of these two values here, by showing how they arise naturally out of the usage of the holomorphic observable, which was the main tool in the entire process.

The power of the holomorphic observable is founded in its discrete holomorphy shown in Lemma 5.19. The main reasons for this property can be found in equations (5.26) and (5.28), in particular we needed

$$j\bar{\lambda}^4 + \bar{j}\lambda^4 = 0, \tag{a}$$

$$1 + x_c\left(j\bar{\lambda} + \bar{j}\lambda\right) = 0. \tag{b}$$

Equation (a) is equivalent to

$$\exp\left(\frac{2\pi i}{3}(1+2\sigma)\right) = -\exp\left(-\frac{2\pi i}{3}(1+2\sigma)\right) = \exp\left(\frac{\pi i}{3}(1-4\sigma)\right).$$

Considering that the complex exponential function is $2\pi i$-periodic we get

$$\begin{aligned} & \frac{2\pi i}{3}(1+2\sigma) = \frac{\pi i}{3}(1-4\sigma) + 2\pi i k \\ \Rightarrow \qquad & \sigma = \frac{6k-1}{8} \qquad \text{for } k \in \mathbb{Z}. \end{aligned}$$

Setting $k = 1$ gives the known value $\sigma = 5/8$.

Equation (b) can now be used to extract the critical value $x_c$. Observe $j\bar{\lambda} + \bar{j}\lambda = 2\,\mathrm{Re}(j\bar{\lambda}) = 2\cos((2+\sigma)\frac{\pi}{3})$. Inserting the above value for $\sigma$ we obtain

$$x_c\left(2\cos\left(\frac{6k+15}{24}\pi\right)\right) = -1.$$

Note, that by its definition as the reciprocal of the connective constant, $x_c$ must be positive as $\mu$ is positive. Thus, the above equation implies that $2\cos(\frac{6k+15}{24}\pi) < 0$. Up to periodicity the only values we get are $k \in \{0,1,2,3\}$ and as the cosine function is symmetric with respect to $\pi$, we only need to consider $k = 0$ and $k = 1$.

By using the identity $\cos\left(\frac{x}{2}\right) = \mathrm{sign}(\cos(\frac{x}{2}))\sqrt{\frac{1+\cos(x)}{2}}$, which was also needed in the end of the proof of Lemma 5.19 and by observing that the cosine is negative for these two values of $k$, we get:

$$x_c = -\left(2\cos\left(\frac{2k+5}{8}\pi\right)\right)^{-1} = \left(2\sqrt{\frac{1+\cos\left(\frac{2k+5}{4}\pi\right)}{2}}\right)^{-1} = \begin{cases} 1/\sqrt{2-\sqrt{2}}, & k=0, \\ 1/\sqrt{2+\sqrt{2}}, & k=1. \end{cases}$$

Up to now there are two possibilities for $\sigma$ and $x_c$ which would do the job:

| $k$ | $\sigma$ | $x_c$ |
|---|---|---|
| 0 | $-\frac{1}{8}$ | $1/\sqrt{2-\sqrt{2}}$ |
| 1 | $\frac{5}{8}$ | $1/\sqrt{2+\sqrt{2}}$ |

As a next step in the proof we restricted ourselves to the vertical strip domain $S_T$ and introduced partition functions $A_{T,L}(x), B_{T,L}(x)$ and $E_{T,L}(x)$ in Section 5.3.3. These functions were needed to translate the discrete holomorphy equation (5.25) in Lemma 5.21 to the vertical strip domain. As a result we got two coefficients $c_\alpha$ and $c_\varepsilon$, which were derived as

$$c_\alpha = -\cos(\sigma\pi) = \begin{cases} -\cos\left(\frac{\pi}{8}\right), & k=0, \\ \cos\left(\frac{3\pi}{8}\right), & k=1, \end{cases}$$

$$c_\varepsilon = \cos\left(\frac{2\pi}{3}(1-\sigma)\right) = \begin{cases} \cos\left(\frac{3\pi}{4}\right), & k=0, \\ \cos\left(\frac{\pi}{4}\right), & k=1. \end{cases}$$

These coefficients were needed for the derivations of the limits $A_T(x), B_T(x)$ and $E_T(x)$ of the partition functions as $L$ tends to infinity. In particular, the limits existed, because the partition functions and the coefficients $c_\alpha$ and $c_\varepsilon$ in (5.29) were positive. But if we consider the above coefficients for $k=0$, we see, that they are negative, which does not allow the used argumentation. Hence, the only possible values for $x_c$ and $\sigma$ (up to periodicity) are the used ones and we see that they arise naturally from the holomorphic observable:

| $k$ | $\sigma$ | $x_c$ | $c_\alpha$ | $c_\varepsilon$ |
|---|---|---|---|---|
| 0 | $-\frac{1}{8}$ | $1/\sqrt{2-\sqrt{2}}$ | $-\cos\left(\frac{\pi}{8}\right) = -\frac{\sqrt{2+\sqrt{2}}}{2} < 0$ | $\cos\left(\frac{3\pi}{4}\right) = -\frac{\sqrt{2}}{2} < 0$ |
| 1 | $\frac{5}{8}$ | $1/\sqrt{2+\sqrt{2}}$ | $\cos\left(\frac{3\pi}{8}\right) = \frac{\sqrt{2-\sqrt{2}}}{2} > 0$ | $\cos\left(\frac{\pi}{4}\right) = \frac{\sqrt{2}}{2} > 0$ |

■

## 5.4 Open Problems

The knowledge of the connective constant brings us one step closer to reveal the true nature of lattice paths. However, Nienhuis proposed more than just the value of this parameter. In [37] he also states the conjecture that the asymptotic behavior for the number of SAWs is

$$c_n \sim A n^{\gamma-1} \sqrt{2+\sqrt{2}}^{\,n}, \tag{5.34}$$

with $\gamma = 43/32$ and $A \in \mathbb{R}$ as $n \to \infty$. Moreover, he argued that the *mean square displacement* $\langle |\omega_n|^2 \rangle$ representing a measure for the distance a random walker covers on average (with respect to a certain number of experiments) over a certain period of time is equal to

$$\langle |\omega_n|^2 \rangle = \frac{1}{c_n} \sum_{\omega \ n-\text{step SAW}} |\omega_n|^2 = n^{2\nu + o(1)},$$

with $\nu = 3/4$. These two problems remain open questions, where the best known rigorous bounds on (5.34) have been introduced in Sections 5.1.1 and 5.2.

A possible way to solve these problems could be found in the theory of *Schramm-Loewner Evolutions*. In [33] Lawler, Schramm and Werner show that $\gamma$ and $\nu$ could be computed if the SAW would possess a *conformally invariant scaling limit*.

We want to give a minimal overview from [17] to introduce the mentioned objects, see [33] for more details.

What do we understand under the scaling limit? So far our lattices where fixed, in particular we had been dealing with a subset of $\mathbb{Z}^d$. There are many ways to change a lattice, but we are interested in keeping the essential properties which define a lattice, i.e. we do not want to change a square-lattice into a hexagonal lattice. Therefore, a possible way is to change the *mesh size* which is the minimal distance between any two adjacent points. We denote this size by $\delta$.

Consider a simply-connected domain $\Omega \neq \mathbb{C}$, with an underlying lattice $\Lambda$ with mesh size $\delta$ (need not to be hexagonal) contained in $\Omega$. The largest portion of $\Lambda$ that is contained in $\Omega$ is denoted as $\Omega_\delta$. Take two distinct points $a$, $b$ on the boundary of $\Omega$ and denote the vertices of $\Omega_\delta$ which are closest as $a_\delta$ and $b_\delta$, respectively (see Figure 5.7). Consider the set of SAWs $\omega(\Omega_\delta)$ on the finite domain $\Omega_\delta$ from $a_\delta$ to $b_\delta$ that remain inside $\Omega_\delta$. A probability measure $\mathbb{P}_{x,\delta}$ is defined on $\omega(\Omega_\delta)$ by assigning to $\omega$ a weight proportional to $x^{|\omega|}$. The reason for that is that the walks are of different lengths.

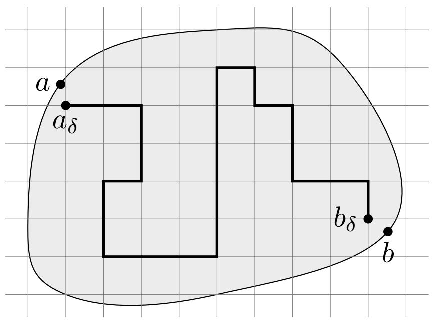



Figure 5.7: Discretization of Domain $\Omega$ with underlying Square Lattice

The idea is to let $\delta \to 0$. We expect that the nature of the walks will depend on the value of $x$. Let $\omega_\delta \in \omega(\Omega_\delta)$ and recall that $x_c$ denotes the critical point. We distinguish 3 cases for $\delta \to 0$, which all hold only in distribution of the before defined probability measure:

- For $x < x_c$ we have that $\omega_\delta$ tends to a straight line.
- For $x > x_c$ it is expected that $\omega_\delta$ becomes space-filling.
- For $x = x_c$ it is conjectured that $\omega_\delta$ becomes a random continuous curve, and is conformally invariant.

Under *conformally invariant*, we understand that the behavior of the curves do not change when they are transformed under a conformal mapping, i.e. $f : U \to \mathbb{C}$ is holomorphic and injective and therefore preserves angles.

More precisely, in [33] it is stated as follows:

***Conjecture 5.24:*** Let $\Omega$ be a simply-connected domain (not equal to $\mathbb{C}$) with two distinct points $a$, $b$ on its boundary. For $x = x_c$ the law of $\omega_\delta$ in $(\Omega_\delta, a_\delta, b_\delta)$ converges, when $\delta \to 0$ to the (chordal) Schramm-Loewner Evolution with parameter $\kappa = 8/3$ in $\Omega$ from $a$ to $b$.

A *Schramm-Loewner Evolution with parameter* $\kappa$ ($\mathrm{SLE}_\kappa$) can be described as follows. Let $\mathbb{C}^+$ be the upper half-plane of $\mathbb{C}$. Consider a path $\omega$ starting at the boundary and ending at an internal vertex. Then $\mathbb{C}^+ \setminus \omega$ is a slit upper half-plane (complement of the path) and therefore simply-connected. From the Riemann Mapping Theorem we know, that it can be conformally mapped to the upper half-plane. The mapping satisfies a differential equation, the *Löwner Equation* and can be alternatively described by a real-valued function. In detail, the Löwner Equation generates a set of conformal maps, driven by a continuous real-valued function. Schramm's idea was to use a Brownian Motion $B_t$ as the driving function.

In formula: Let $B_t$, $t \geq 0$ be a Standard Brownian Motion on $\mathbb{R}$ and let $\kappa$ be a real parameter. Then $\mathrm{SLE}_\kappa$ is the family of conformal maps $g_t$, $t \geq 0$ defined by the Löwner Equation

$$\frac{\partial}{\partial t} g_t(z) = \frac{2}{g_t(z) - \sqrt{\kappa} B_t}, \qquad g_0(z) = z.$$

This is actually called *chordal* $SLE_\kappa$ as it describes paths growing from the boundary and ending on the boundary.

Therefore, understanding SAWs boils down to understanding the theory of $\mathrm{SLE}_\kappa$. The parameter $\kappa$ plays a very important role. For $\kappa > 0$ the curve is always fractal, and becomes more fractal the higher values are chosen. The following summary is taken from [2].

- For $\kappa \leq 4$ the curve is almost surely simple, i.e. it does not touch itself and it does not intersect the real line ($\omega \cap \mathbb{R} = \{0\}$).
- For $4 < \kappa < 8$ the curve touches itself but does not cross itself. Furthermore it intersects parts of the real line ($\omega \cap \mathbb{R} \subsetneq \mathbb{R}$).
- For $\kappa \geq 8$ the curve is space-filling and hits every point on the real axis ($\omega \cap \mathbb{R} = \mathbb{R}$).

# List of Figures

# Bibliography


[1] Shreeram S. Abhyankar. *Algebraic geometry for scientists and engineers*, volume 35 of *Mathematical Surveys and Monographs*. American Mathematical Society, Providence, RI, 1990.

[2] Tom Alberts. An introduction to the schramm-loewner evolution. Talk, `http://www.math.caltech.edu/~alberts/talks/SLEintro.pdf`, 2008. Online; accessed 26/08/2013.

[3] Cyril Banderier and Philippe Flajolet. Basic analytic combinatorics of directed lattice paths. *Theoret. Comput. Sci.*, 281(1-2):37–80, 2002. Selected papers in honour of Maurice Nivat.

[4] Roland Bauerschmidt, Hugo Duminil-Copin, Jesse Goodman, and Gordon Slade. Lectures on self-avoiding walks. *Probability and Statistical Physics in Two and More Dimensions (D. Ellwood, CM Newman, V. Sidoravicius, and W. Werner, eds.), Clay Mathematics Institute Proceedings*, 15:395–476, 2012.

[5] Alin Bostan and Manuel Kauers. Automatic classification of restricted lattice walks. In *21st International Conference on Formal Power Series and Algebraic Combinatorics (FPSAC 2009)*, Discrete Math. Theor. Comput. Sci. Proc., AK, pages 201–215. Assoc. Discrete Math. Theor. Comput. Sci., Nancy, 2009.

[6] Alin Bostan and Manuel Kauers. The complete generating function for Gessel walks is algebraic. *Proc. Amer. Math. Soc.*, 138(9):3063–3078, 2010. With an appendix by Mark van Hoeij.

[7] Mireille Bousquet-Mélou and Marni Mishna. Walks with small steps in the quarter plane. In *Algorithmic probability and combinatorics*, volume 520 of *Contemp. Math.*, pages 1–39. Amer. Math. Soc., Providence, RI, 2010.

[8] Mireille Bousquet-Mélou and Marko Petkovšek. Linear recurrences with constant coefficients: the multivariate case. *Discrete Mathematics*, 225(1–3):51 – 75, 2000. FPSAC'98.

[9] Andrew R. Conway, Ian G. Enting, and Anthony J. Guttmann. Algebraic techniques for enumerating self-avoiding walks on the square lattice. *J. Phys. A*, 26(7):1519–1534, 1993.

[10] Hugo Duminil-Copin and Stanislav Smirnov. The connective constant of the honeycomb lattice equals $\sqrt{2+\sqrt{2}}$. *Ann. of Math. (2)*, 175(3):1653–1665, 2012.

[11] Guy Fayolle, Roudolf Iasnogorodski, and Vadim Malyshev. *Random walks in the quarter-plane*, volume 40 of *Applications of Mathematics (New York)*. Springer-Verlag, Berlin, 1999. Algebraic methods, boundary value problems and applications.

[12] Philippe Flajolet, Stefan Gerhold, and Bruno Salvy. On the non-holonomic character of logarithms, powers, and the $n$th prime function. *Electron. J. Combin.*, 11(2):Article 2, 16, 2004/06.

[13] Philippe Flajolet and Robert Sedgewick. *Analytic Combinatorics*. Cambridge University Press, New York, NY, USA, 1 edition, 2009.

[14] Paul J. Flory. *Principles of Polymer Chemistry*. Cornell University Press, London, 1953.

[15] Ira M. Gessel. A factorization for formal laurent series and lattice path enumeration. *Journal of Combinatorial Theory, Series A*, 28(3):321 – 337, 1980.

[16] Bernhard Gittenberger. Diskrete Methoden. Course Notes, Institut für Diskrete Mathematik und Geometrie, Technische Universität Wien. Version Summer term 2010. `http://www.dmg.tuwien.ac.at/gittenberger/dmeth.pdf`. Online; accessed 08/07/2013.

[17] Anthony J. Guttmann. Self-avoiding walks and polygons – an overview, December 2012.

[18] John M. Hammersley and Dominic J. A. Welsh. Further results on the rate of convergence to the connective constant of the hypercubical lattice. *Quart. J. Math. Oxford Ser. (2)*, 13:108–110, 1962.

[19] Godfrey H. Hardy and Srinivasa Ramanujan. Asymptotic formulæ for the distribution of integers of various types [Proc. London Math. Soc. (2) **16** (1917), 112–132]. In *Collected papers of Srinivasa Ramanujan*, pages 245–261. AMS Chelsea Publ., Providence, RI, 2000.

[20] Brian Hayes. How to avoid yourself. *American Scientist*, 86(4), July-August 1998. `http://www.americanscientist.org/issues/pub/how-to-avoid-yourself/`. Online; accessed 18/05/2013.

[21] Barry D. Hughes. *Random walks and random environments. Vol. 1.* Oxford Science Publications. The Clarendon Press Oxford University Press, New York, 1995. Random walks.

[22] Thomas W. Hungerford. *Algebra*, volume 73 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, 1980. Reprint of the 1974 original.

[23] Iwan Jensen. A parallel algorithm for the enumeration of self-avoiding polygons on the square lattice. *J. Phys. A*, 36(21):5731–5745, 2003.

[24] Iwan Jensen. Improved lower bounds on the connective constants for two-dimensional self-avoiding walks. *Journal of Physics A: Mathematical and General*, 37(48):11521, 2004.

[25] Iwan Jensen and Anthony J. Guttmann. Self-avoiding polygons on the square lattice. *J. Phys. A*, 32(26):4867–4876, 1999.

[26] Samuel Johnson. Analytic combinatorics of planar lattice paths. Master's thesis, Simon Fraser University, 2012.

[27] Michael Kaltenbäck and Harald Woracek. Komplexe Analysis für Technische Mathematik. Course Notes, Institut ASC, Technische Universität Wien. Version 25.6.2013. `http://asc.tuwien.ac.at/~funkana/kaltenbaeck/mykana/mykana.pdf`. Online; accessed 07/08/2013.

[28] Manuel Kauers. The holonomic toolkit. In *Computer Algebra in Quantum Field Theory: Integration, Summation and Special Functions, Texts and Monographs in Symbolic Computation*. Springer, 2013.

[29] Donald E. Knuth. *The art of computer programming. Vol. 1: Fundamental algorithms.* Second printing. Addison-Wesley Publishing Co., Reading, Mass.-London-Don Mills, Ont, 1969.

[30] Irina Kurkova and Kilian Raschel. On the functions counting walks with small steps in the quarter plane. *Publications mathématiques de l'IHÉS*, 116(1):69–114, 2012.

[31] David Fernand Laferrière. Classification of walks in wedges. Master's thesis, Simon Fraser University, 2007.

[32] Pauline Lafitte-Godillon, Kilian Raschel, and Viet Chi Tran. Extinction probabilities for a distylous plant population modeled by an inhomogeneous random walk on the positive quadrant. *SIAM J. Appl. Math.*, 73(2):700–722, 2013.

[33] Gregory F. Lawler, Oded Schramm, and Wendelin Werner. On the scaling limit of planar self-avoiding walk. In *Fractal geometry and applications: a jubilee of Benoît Mandelbrot, Part 2*, volume 72 of *Proc. Sympos. Pure Math.*, pages 339–364. Amer. Math. Soc., Providence, RI, 2004.

[34] Leonard Lipshitz. $D$-finite power series. *J. Algebra*, 122(2):353–373, 1989.

[35] Neal Madras and Gordon Slade. *The self-avoiding walk.* Modern Birkhäuser Classics. Birkhäuser/Springer, New York, 2013. Reprint of the 1993 original.

[36] Rick Miranda. *Algebraic curves and Riemann surfaces*, volume 5. Amer Mathematical Society, 1995.

[37] Bernard Nienhuis. Exact critical point and critical exponents of $\mathrm{O}(n)$ models in two dimensions. *Phys. Rev. Lett.*, 49(15):1062–1065, 1982.

[38] George L. O'Brien. Monotonicity of the number of self-avoiding walks. *J. Statist. Phys.*, 59(3-4):969–979, 1990.

[39] Marko Petkovšek, Herbert B. Wilf, and Doron Zeilberger. $A = B$. AK Peters, Ltd, Wellesley, MA, first edition edition, 1996.

[40] George Pólya and Gabor Szegő. *Problems and theorems in analysis. I.* Classics in Mathematics. Springer-Verlag, Berlin, 1998. Series, integral calculus, theory of functions, Translated from the German by Dorothee Aeppli, Reprint of the 1978 English translation.

[41] André Pönitz and Peter Tittmann. Improved upper bounds for self-avoiding walks in $\mathbf{Z}^d$. *Electron. J. Combin.*, 7:Research Paper 21, 10 pp. (electronic), 2000.

[42] Kilian Raschel. Walks in the quarter plane: analytic approach and applications. To be published in ESAIM:Proceedings, 2013, CNRS & Universite de Tours, `http://www.lmpt.univ-tours.fr/~raschel/MAS2012Raschel.pdf`. Online; accessed 03/08/2013.

[43] Richard P. Stanley. *Enumerative combinatorics. Vol. 2*, volume 62 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 1999. With a foreword by Gian-Carlo Rota and appendix 1 by Sergey Fomin.

[44] Richard P. Stanley. *Enumerative combinatorics. Volume 1*, volume 49 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, second edition, 2011.

[45] Alfred Van der Poorten and R Apéry. A proof that euler missed... *The Mathematical Intelligencer*, 1(4):195–203, 1979.

[46] Herbert S. Wilf. *generatingfunctionology*. Academic Press Inc., Boston, MA, second edition, 1994.

[47] Jet Wimp and Doron Zeilberger. Resurrecting the asymptotics of linear recurrences. *J. Math. Anal. Appl.*, 111(1):162–176, 1985.

# Index