TECHNISCHE UNIVERSITÄT WIEN

DIPLOMARBEIT

Lattice Walks in the Quarter Plane

ausgeführt am

Institut für
Diskrete Mathematik und Geometrie
TU Wien

durch
Andreas NESSMANN
Mat.-Nr. 01128984

betreut von
Univ.-Prof. Dipl.-Ing. Dr. techn. Michael Drmota

Wien, 9. August 2020

Kurzfassung

Eine Vielzahl von Problemen in der Kombinatorik lässt sich auf die Untersuchung von Gitterpfaden zurückführen. In der analytischen Kombinatorik wird solchen Pfaden eine erzeugende Funktion zugeordnet, deren Eigenschaften man mithilfe komplexer Analysis untersuchen kann, und die es dann wiederum ermöglichen, Rückschlüsse über die ursprüngliche kombinatorische Ausgangsfrage zu ziehen.
In dieser Arbeit werden Pfade mit kleinen Schritten in der Viertelebene betrachtet, und insbesondere die Frage, wann die resultierenden erzeugenden Funktionen holonomisch sind. Ausgehend von einer Publikation von M. Bousquet-Mélou und M. Mishna im Jahre 2010, in der die Autorinnen die Vermutung äußerten, dass die Holonomie gleichbedeutend mit der Endlichkeit einer gewissen Gruppe ist, die durch eine gewisse Menge erlaubter Schritte induziert wird, wurden in den letzten Jahren eine Vielzahl an Resultaten zu diesem Thema veröffentlicht, im Zuge derer die ursprüngliche Vermutung schlussendlich bewiesen wurde. Das Ziel dieser Arbeit ist einerseits, einen Überblick über den Beweis zu geben, andererseits aber auch eine Einführung sowohl über Gitterpfade an sich als auch über die überraschend vielseitige Theorie, die in Folge als Werkzeug dienen wird.
Im ersten Abschnitt der Arbeit wird eine kurze Einführung sowohl zu holonomischen Funktionen als auch zu Gitterpfaden im allgemeinen gegeben. Die Funktionalgleichung (die sog. kernel equation), die das Kernstück der späteren Untersuchungen ist, wird hergleitet, mit derer das Problem für Pfade in der ganzen, oder der Halbebene direkt gelöst werden kann. Im Folgenden wird die Gruppe eines Pfades eingeführt, und die Holonomie für endlichen Pfaden mit Hilfe von Orbitsummen gezeigt.
Im nächsten Abschnitt werden hauptsächlich die Pfade mit unendlicher Gruppe behandelt. Dazu werden verschiedene technische Hilfsmittel benötigt, insbesondere Theorie über die analytische Fortsetzung von Funktionen und die resultierenden Riemannflächen sowie ein wenig Theorie zu projektiven algebraischen Kurven, insbesondere elliptischen Kurven. Mithilfe dieser Werkzeuge werden im letzten Abschnitt zwei verschiedene Beweismöglichkeiten vorgestellt. Während die zugrundeliegende Idee, nämlich Polstellen gewisser Funktionen zu untersuchen, in beiden Fällen dieselbe ist, ist die Methodik überraschend verschieden: ein Beweis arbeitet direkt mit den Fortsetzungen der Gruppe und Funktionalgleichung auf der Riemannfläche und untersucht dort die Polstellen, der zweite ist hingegen weit algebraischer und verwendet Valuationen auf dem Funktionenkörper einer elliptischen Kurve, verbunden mit Galois-Theorie von Differenzenringen.

Abstract

Many combinatorial problems can be reduced to the study of lattice paths. In analytic combinatorics, one assigns to such paths a generating function, whose properties can then be studied using complex analysis. The results often allow conclusions about the initial combinatorial problem.

The topic of this work are paths with small steps in the quarter plane, in particular the question when the corresponding generating function is holonomic. In 2010, M. Bousquét-Melou and M. Mishna conjectured that holonomy of the generating function is equivalent to the finiteness of a certain group associated with the set of allowed steps. This leads to a multitude of publications, and eventually to the proof of the conjecture. The aim of this work is to give an overview of the proof, but also to provide an introduction to the surprisingly diverse theory which is used therein.

The first part of this work starts with a short introduction to holonomic functions and lattice paths in general. The kernel equation of the generating function is derived, which is in a sense the center piece of all that comes later. The kernel method for treating such equations is introduced, which immediately allows solving the question for walks in the plane and the half plane. Restricting ourselves to walks in the quarter plane, the group of a walk is introduced, and holonomy is proven for all paths with a finite group using orbit summation methods.

In the next part, the main topic of interest are those paths with infinite group. In order to effectively study those, some technical prerequisites are required, in particular a bit of theory about analytic continuation and Riemann surfaces, and basic information about Algebraic curves, in particular elliptic curves. Using these tools, two proofs for the non-holonomy in the infinite group case are outlined in the last section. While the underlying idea is always studying the poles of certain functions, the techniques applied are remarkably different. One proof works directly with meromorphic continuations of the kernel equation and the group on the Riemann surface, the other uses an algebraic approach, utilizing valuation theory on elliptic curves in connection with Galois theory of difference rings.

Acknowledgement

I would like to say a big thank you to my advisor, Prof. Dr. Michael Drmota, both for his always reliable and helpful advice and for pointing me towards analytic combinatorics and sparking my interest in the first place.
Without my family and their continuing and unconditional support and encouragement over many years I would never have come this far; and while supporting me for this thesis is by far not the only reason I wish to say a heartfelt thank you to all of them, it is as good an opportunity to do so as any.
Lastly, I would like to thank all my friends and relatives who have kept me sane during times of isolation.

Eidesstattliche Erklärung

Ich erkläre an Eides statt, dass ich die vorliegende Diplomarbeit selbstständig und ohne fremde Hilfe verfasst, andere als die angegebenen Quellen und Hilfsmittel nicht benutzt bzw. die wörtlich oder sinngemäß entnommenen Stellen als solche kenntlich gemacht habe.

Wien, am 9. August 2020

Andreas Nessmann

Contents

1 Basic Notions

1.1 Generating Functions, Holonomy

Suppose we have any sequence $A = (a_n)_{n=0}^{\infty}$ of complex numbers. A natural question arising would be how to describe this sequence. If, for instance, A was generated randomly, then - even though asymptotic or suchlike statements may very well be possible - by the very nature of its construction the only way to give a complete description of A would be to list all its entries. However, since many of the sequences we are interested in are not purely random but instead arise out of a natural context, for example when counting a certain type of objects, they often have a lot more internal structure which we can try to exploit. The concept of holonomy, which will be one of the main interests in this thesis, can be understood as arising through a classification of such sequences - if they have, in a certain sense, enough internal structure, they are called holonomic. The main goal will be to decide, and in particular, *how* to decide for some particular sequences whether they are holonomic or not.
We can start by noticing that we can assign to our sequence A a formal power series $A(z)$ via

$$A(z) := \sum_{n=0}^{\infty} a_n z^n. \tag{1}$$

This formal power series is called the *generating function* of A. Clearly, no information is lost here, since we can immediately regain the sequence A from $A(z)$ via the coefficient extraction operator

$$[z^k]A(z) = [z^k]\sum_{n=0}^{\infty} a_n z^n := a_k. \tag{2}$$

In particular, the formal power series over $\mathbb{C}$ (or, in fact, any set) are in bijection with the sequences in $\mathbb{C}$ (or any other set). Also note that the formal power series over $\mathbb{C}$ (or any ring) form a ring with coefficient-wise addition and the Cauchy product. We will denote this ring by $\mathbb{C}[[z]]$. Its quotient field is given by $\mathbb{C}((z))$, the space of formal Laurent series,where one allows for finitely many negative powers of z (see, for instance, [4, chapter 2]). Also note that the coefficient extraction is a linear operator. At this point, the name of a generating *function* is slightly misleading: because we do not have any information about the sequence A, and we thus have no knowledge about possible convergence, it does not necessarily make sense to consider $A(z)$ a mapping, or to evaluate at any point z. Suppose, however, we *do* have such additional knowledge, and somehow we know that

$$|a_n| \leq C^n \tag{3}$$

for some constant C. This isn't as hard a restriction as it may appear at first; in many application finding such an a-priori estimate is fairly simple. And in this case, complex analysis now tells us that we can, in fact, consider our until now purely formal power series an analytic function, as it is surely convergent on the open disk around 0 with some radius $R < \frac{1}{C}$. At this point, we suddenly have a whole new box of tools to play with; namely complex analysis. Applying these analytic methods to our very discrete objects is the main idea of analytic combinatorics.
To give an example of why this can be useful, consider, for instance, that $a_n = 1$ for all n. Admittedly, this is not a very exciting or hard to describe sequence, but neither is its generating

function: for $|z| < 1$, we have

$$A(z) = \sum_{n=0}^{\infty} 1 \cdot z^n = \frac{1}{1-z} \tag{4}$$

via a geometric series. So instead of our infinite sequence, we can now simply consider the rational function $\frac{1}{1-z}$. Note that, as the power series representation around 0 of any analytic function is unique, we still haven't lost any information about our original sequence. There is a wide range of theory about formal power series in general and how changes to the generating function correspond to changes to the sequence (see, for instance, [1], [2]), but all we need for now is stated in the following easy lemma.

Lemma 1.1.1. *Let $A = (a_n)_{n=0}^{\infty}, B = (b_n)_{n=0}^{\infty}$ be sequences with corresponding generating functions $A(z), B(z)$. Then:*

1. *the sequence $(a_n + b_n)_{n=0}^{\infty}$ corresponds to the generating function $A(z) + B(z)$,*
2. *the sequence $\left(\sum_{i+j=n} a_i b_j\right)_{n=0}^{\infty}$ corresponds to the generating function $A(z) \cdot B(z)$.*
3. *the sequence $(n \cdot a_n)_{n=0}^{\infty}$ corresponds to the generating function $z\frac{\partial}{\partial z}A(z)$,*

Proof.

1. By definition of the generating function.
2. By definitions of the Cauchy product and generating functions.
3. The formal derivative $\frac{\partial}{\partial z}A(z)$ is given by $\sum_{n=1}^{\infty} n a_n z^{n-1}$. Multiplying the equation by z yields the result.

□

Note in particular that this lemma holds true no matter for all formal series, no matter whether they are convergent or not, as in the former case the formal derivative coincides with the analytic one, and all computations in the lemma are of purely algebraic nature.

1.1.1 Algebraic Functions

We have by now assigned each sequence a formal power series and vice versa, but have not yet made any steps towards classifying them according to their simplicity, as was original goal. To do so, we could start by considering some fundamentally simple cases and work our way up from there. Intuitively, the easiest power series to work with are polynomials, that is, finite power series. These correspond to finite sequences (or, more precisely, to those sequences that end up being constant 0 from a certain index on). So, our favourite class of power series is simply $\mathbb{C}[z]$, the ring of complex polynomials in one variable. It turns out that, if we consider polynomials closer together the more of their beginning coefficients coincide and formalize this via a metric, then the formal power series are the completion of the polynomial ring - similar to $\mathbb{R}$ being the completion of $\mathbb{Q}$. One of the very basic distinction of real numbers is between algebraic and transcendental numbers, and it turns out that we can do quite the same here[1].

[1]This is by far not the only analogue between polynomials and integers - in fact, there is a very rich theory coming from treating polynomials (in particular those over finite fields) like one would integers, which is in many ways surprisingly a lot easier than the usual number theory. See for instance [3].

Definition 1.1.1. *A formal power series $P(z) \in \mathbb{C}[[z]]$ is called* ***algebraic*** *if there is an $n \in \mathbb{N}^+$ and there are polynomials $p_0(z), \ldots, p_n(z) \in \mathbb{C}[z]$, such that*

$$p_n(z)P(z)^n + p_{n-1}(z)P(z)^{n-1} + \cdots + p_0(z) = 0. \tag{5}$$

The minimal n for which such an equation exists is called the degree of $P(z)$.
A power series that is not algebraic is called ***transcendental****.*

Note that we could just as well assume our $p_0, \ldots, p_n$ to be fractions of polynomials, as we can always clear denominators. If we were to allow for these, then we could also assume that $p_n = 1$. Thus, if we denote by $K := \mathrm{Quot}\,(\mathbb{C}[z])$ the quotient field of $\mathbb{C}[z]$, that is, $K = \left\{\frac{p}{q} : p, q \in \mathbb{C}[z], q \neq 0\right\}$, the field of **rational functions** over $\mathbb{C}$, then this definition implies that a power series $P(z)$ is algebraic if and only if the dimension of the K-vector space $\langle P \rangle$ generated by $P(z)$ is finite. Note also, that in this case $\langle P(z) \rangle \cong K[x]/\langle \mu(x) \rangle$, where μ is the minimal polynomial of $P(z)$, via the K-vector space isomorphism $x \mapsto P(z)$. As a consequence, we immediately obtain

Lemma 1.1.2. *Let $P(z), Q(z) \in \mathbb{C}[[z]]$ be algebraic power series. Then:*

1. *$P(z) + Q(z)$ is algebraic,*
2. *$P(z) \cdot Q(z)$ is algebraic,*
3. *if $P(0) \neq 0$, then $\frac{1}{P(z)}$ is algebraic,*
4. *$P'(z)$ is algebraic.*

Proof.

1. Let $\mu(x), \nu(x)$ be the minimal polynomials over K of $P(z), Q(z)$ respectively. Then, $K[x,y]/\langle \mu(x), \nu(y) \rangle$ is a finite-dimensional vector space of dimension $\deg \mu \cdot \deg \nu$. As we can injectively embed $\langle P(z), Q(z) \rangle$ into $K[x,y]$ via $P(z) \mapsto x, Q(z) \mapsto y$, we obtain $\dim_K \langle P(z), Q(z) \rangle \leq \dim_K (K[x,y]/\langle \mu(x), \nu(y) \rangle) < \infty$, and we are done.

2. Same as 1.

3. If $P(0) \neq 0$, $P(z)$ has an inverse $P^{-1}(z)$ in $\mathbb{C}[[z]]$. Multiplying the algebraicity equation for $P(z)$ by $\left(P^{-1}(z)\right)^n$ yields an algebraicity equation for $P^{-1}(z)$.

4. We show that $\langle P(z) \rangle$ is closed under differentiation, since it is of finite dimension there must then be a non-trivial linear relation of powers of $P'(z)$.
First, we note that if $P(z) \in K$, meaning $P(z)$ is a rational function, then so is its derivative, so the statement is obvious. We can thus assume that $P(z) \notin K$. Let again $\mu \in K[x]$ be the minimal polynomial of $P(z)$ over K. As $P(z) \notin K$, we have $\deg \mu \geq 2$. As $\mathrm{char}\, K = 0$, we have $\gcd(\mu, \mu') = 1$. Equivalently, there are $v, w \in K[x] : v\mu + w\mu' = 1$. In particular, this means $w\mu' = 1$ in $K[x]/\langle \mu(x) \rangle \cong \langle P(z) \rangle$.
We can now take the algebraicity equation of $P(z)$, namely, $\mu(P(z)) = 0$, and derive it with respect to z. Note that the coefficients of μ are rational polynomials in z, as μ is a polynomial in $\mathbb{C}(z)[x]$. We therefore obtain

$$\frac{\partial}{\partial z}\mu\left(P(z)\right) = \mu_z\left(P(z)\right) + P'(z)\mu'\left(P(z)\right), \tag{6}$$

where μ_z denotes the derivative with respect to z if we consider μ as a polynomial $\mu(z,x)$. Multiplying by the inverse w of μ', this yields

$$P'(z) = -\mu_z\left(P(z)\right) w\left(P(z)\right), \tag{7}$$

and as the right hand side lies in $\langle P(z)\rangle$, so must $P'(z)$. By simple computation, the derivative of each polynomial in $P(z)$ must therefore also lie in $\langle P(z)\rangle$, thus the vector space is closed under differentiation and the statement is shown.

□

Example.

1. The function $f(z) = \sqrt[n]{p(z)}$ is algebraic for any polynomial $p(z)$, as it is analytic and $f(z)^n - p(z) = 0$. By repeated application of the above lemma, so is the function $g(z) = \frac{\sqrt[3]{z^{17}-9}}{\sqrt{z-\sqrt{z^2+1}}}$.
2. The function $\exp(z)$ is *not* algebraic. Intuitively, this is because the highest power of $\exp(z)$ will always determine the asymptotic behavior in any potential algebraicity equation, and it cannot be sufficiently weighted down by polynomial functions.

While the class of algebraic functions has a lot of nice properties and applicable algorithms (see, for instance, [4, chapter 6]), it is a bit more restrictive than we would like. In particular, there are quite a few functions which also have nice properties but are not algebraic - for instance, as noted above, $\exp(z)$. We will therefore proceed to define another, more general class of functions.

1.1.2 Holonomic Functions

Definition 1.1.2. *A power series $P(z) \in \mathbb{C}[[z]]$ is called* ***holonomic*** *(or* ***D-finite****) if there is an $n \in \mathbb{N}^+$ and there are polynomials $p_0(z), \ldots, p_n(z) \in \mathbb{C}[z]$, such that*

$$p_n(z)\frac{\partial^n}{\partial z^n}P(z) + p_{n-1}(z)\frac{\partial^{n-1}}{\partial z^{n-1}}P(z) + \cdots + p_0(z)P(z) = 0. \tag{8}$$

A power series that is not holonomic will be called ***non-holonomic*** *or* ***D-infinite****.*

Essentially, all we did was change the condition for an algebraic function by switching out powers of $P(z)$ for derivatives. Maybe a bit surprisingly, this is in fact a stronger condition, as we already have all the work done to show

Lemma 1.1.3. *If a power series $P(z)$ is algebraic, then it is also holonomic.*

Proof.
By the proof of the previous lemma, we know that for algebraic $P(z)$, $\langle P(z)\rangle$ is closed under differentiation, and its dimension is finite. Let n be this dimension. Then, $P(z), \frac{\partial}{\partial z}P(z), \ldots, \frac{\partial^n}{\partial z^n}P(z)$ are contained in $\langle P(z)\rangle$, but cannot be linearly independent. Thus, there is a non-trivial linear relation between them and $P(z)$ is holonomic. □

Note that, while the construction of algebraic power series was done similarly to that of algebraic numbers, the concept of holonomic power series does not immediately admit such an analogue, due to the lack of a suitable derivative on $\mathbb{R}$. We could thus consider a correspondence of ring extensions[2] as follows:

$\mathbb{Q}$	$\mathbb{C}[z]$
Algebraic numbers	Algebraic power series
-	Holonomic power series
$\mathbb{R}$	$\mathbb{C}[[z]]$

So we have a class of power series whose analogue would have to be strictly "between" algebraic and real numbers. While there is no number field equivalent, notice that the construction of holonomic power series still proceeds in a manner very similar to that of algebraic power series. It is, in a way, a rather natural extension of that definition, and therefore maybe intuitively already worth looking at. But there is an even stronger reason why holonomic power series can be of interest. Recall that at the beginning we were trying to describe an infinite sequence, which we translated into a formal power series. Now let us stop for a moment and consider what holonomicity of a power series means for the sequence of coefficients. We know that

$$\frac{\partial^k}{\partial z^k}P(z) = \sum_{m=0}^{\infty}(m+k)\cdot\ldots\cdot(m+1)a_{m+k}z^n = \sum_{m=0}^{\infty}m^{\overline{k}}a_{m+k}z^m, \tag{9}$$

where $m^{\overline{k}} = \frac{(m+k)!}{m!}$. Let $d \in \mathbb{N}$ such that $\deg p_i(z) \le d$ for $i = 0, \ldots, n$, and $p_i(z) = b_0^i + b_1^i z + \cdots + b_d^i z^d$. Then by comparing coefficients we obtain

$$\sum_{i=0}^{n}\sum_{j=0}^{d} b_j^i (m-j)^{\overline{i}} a_{m+i-j} = 0. \tag{10}$$

Note that rearranging this for the a_k gives a linear equation in $a_{m+k}, \ldots, a_{m-d}$ with polynomial coefficients depending only on m, for $m \ge d$.
Now, suppose we have such an equation, that is, $a_{m+d}p_d(m) + \cdots + a_m p_0(m) = 0$ for $m \ge d$ and polynomials $p_i \in \mathbb{C}[x]$. We would like to regain a holonomy equation for the corresponding power series $A(z)$. This is not difficult, but requires some care about the approach. To illustrate the strategy, suppose we are given the simple equation

$$a_{m+1} + a_m = 0. \tag{11}$$

What we would like to do is just multiply the equation by z^m or z^{m+1}, and sum over all m, arriving at

$$\sum_{m\ge 0} a_{m+1}z^m + \sum_{m\ge 0} a_m z^m = 0. \tag{12}$$

The right summand is just $A(z)$, and the left one we would love to just apply the shift operator to - but we are missing the starting term to do so. Concretely, we arrive at

$$\frac{1}{z}\sum_{m\ge 0} a_{m+1}z^{m+1} + A(z) = \frac{1}{z}\left(A(z) - a_0\right) + A(z) = 0, \tag{13}$$

which would be an algebraic equation, but not a holonomy equation[3] for $A(z)$. So, we want to get rid of the starting terms. Fortunately, there is a operator we can use in order to get rid of

[2] It is possible to instead start out with the rational functions $\mathbb{C}(z)$ and then extend them to Laurent series, making the right column one of field extensions as well.

[3] In this case, this would suffice to show the existence for a holonomy equation, as we have already established that every algebraic power series is also holonomic. But in other cases such occurrences might mix with derivatives, and then it would not be immediately obvious what happens.

starting terms - differentiation. Recall the correspondences

$$A(z) \leftrightarrow (a_n)_{n=0}^{\infty} \tag{14}$$
$$\frac{\partial}{\partial z}A(z) \leftrightarrow ((n+1)a_{n+1})_{n=0}^{\infty} \tag{15}$$
$$z\frac{\partial}{\partial z}A(z) \leftrightarrow (na_n)_{n=1}^{\infty}. \tag{16}$$

Using these correspondences, we can multiply our recursion by $(m+1)$ and then take the sum, arriving at

$$0 = \sum_{m\geq 0}(m+1)a_{m+1}z^m + \sum_{m\geq 0}(m+1)a_m z^m \tag{17}$$
$$= \sum_{m\geq 0}(m+1)a_{m+1}z^m + \sum_{m\geq 0} m a_m z^m + \sum_{m\geq 0} a_m z^m = \frac{\partial}{\partial z}A(z) + z\frac{\partial}{\partial z}A(z) + A(z) \tag{18}$$
$$= (1+z)\frac{\partial}{\partial z}A(z) + A(z), \tag{19}$$

which is what we want. This approach immediately generalizes for linear recursions with polynomial coefficients as follows:
Assume we have the recursion $a_{m+d}p_d(m+d) + \cdots + a_m p_0(m) = 0$. We can then:

1. Multiply the equation by $m^{\overline{d}}$, to get rid of the starting terms,

2. Multiply by z^m and take the sum, we will then have an equation of the form

$$\sum_{m\geq 0} q_d(m)m^{\overline{d}}a_{m+d}z^m + \sum_{m\geq 0} q_{d-1}(m)^{\overline{d-1}}a_{m+d-1}z^m + \cdots + \sum_{m\geq 0} q_0(m)a_m z^m = 0, \tag{20}$$

where the $q_i(m)$ are polynomials in m.

3. Rewrite $q_i(m)$ as a polynomial in $m+i$, that is, $q_i(m) = r_i(m+i)$, leading to an equation

$$\sum_{m\geq 0} r_d(m+d)m^{\overline{d}}a_{m+d}z^m + \sum_{m\geq 0} r_{d-1}(m+d-1)^{\overline{d-1}}a_{m+d-1}z^m + \cdots + \sum_{m\geq 0} r_0(m)a_m z^m = 0. \tag{21}$$

4. By the above correspondences, we now have the holonomy equation

$$r_d\left(z\frac{\partial}{\partial z}\right)\left(\frac{\partial^d}{\partial z^d}A(z)\right) + r_{d-1}\left(z\frac{\partial}{\partial z}\right)\left(\frac{\partial^{d-1}}{\partial z^{d-1}}\right)A(z) + \cdots + r_0\left(z\frac{\partial}{\partial z}\right)A(z) = 0, \tag{22}$$

where $r_d\left(z\frac{\partial}{\partial z}\right)$ denotes the operator $\sum_{i=0}^{\deg r_d} r_{d,i}\left(z\frac{\partial}{\partial z}\right)^i$, with $r_d(x) = \sum_{i=0}^{\deg r_d} r_{d,i}x^i$.

We have therefore justified the following definition:

Definition 1.1.3. *A sequence $(a_n)_{n=0}^{\infty}$ is called **holonomic** (or **P-recursive**) if there are a $d \in \mathbb{N}^+$ and polynomials $p_0(z), \ldots, p_n(z) \in \mathbb{C}[z]$, such that the recursion*

$$p_d(n+d)a_{n+d} + p_{d-1}(n+d-1)a_{n+d-1} + \cdots + p_0(n)a_n = 0 \tag{23}$$

holds for all $n \geq 0$.

As we just showed, a formal power series is holonomic if and only if the coefficients form a holonomic sequence. Furthermore, we have an explicit way of computing the holonomy equation or recursion from one another. This enables us to find either a polynomial recursion for the series expansion of some known holonomic function, or to find an ODE defining the generating function of some given holonomic sequence. It also becomes quite clear why holonomy is an interesting property: a holonomic sequence can be described by a *finite amount of information* in a fairly easy way, namely by the starting values and the recursion. To know the entire sequence we then only need to compute polynomials. There are, in fact, even better algorithms for computing holonomic sequences than just continuously reiterating the recursion, which make use of partial fraction decompositions, see [4, chapter 7]. Also, if a function is known (or suspected) to be holonomic, there is a variety of techniques to guess the corresponding recursion, and then inductively prove it is the right one, see [8].
Since we will be interested in finding out whether or not functions are holonomic, we should once again consider closure properties.

Lemma 1.1.4. *Let $P(z), Q(z) \in \mathbb{C}[[z]]$ be holonomic power series. Then:*

1. *$P(z) + Q(z)$ is holonomic,*
2. *$P(z) \cdot Q(z)$ is holonomic,*
3. *$\frac{\partial}{\partial z} P(z)$ is holonomic,*
4. *$\int P(z) \, \mathrm{d}\, z$ is holonomic,*
5. *if $Q(z)$ is algebraic, $Q(0) = 0$, then $P(Q(z))$ is holonomic.*

Proof.
In the following, denote by $\langle P(z) \rangle$ the vector space over $K = \mathbb{C}(z)$ spanned by $\{P(z), P'(z), \dots\}$. Note that $P(z)$ is holonomic precisely if $\langle P(z) \rangle$ is finite dimensional[4].

1. We have $\langle P(z) + Q(z) \rangle \subseteq \langle P(z) \rangle \oplus \langle Q(z) \rangle$, and thus $\dim \langle P(z) + Q(z) \rangle \leq \dim \langle P(z) \rangle + \dim \langle Q(z) \rangle < \infty$.

2. As, unlike in the algebraic case, our vector spaces are not multiplicatively closed, we need to consider the tensor product $\langle P(z) \rangle \otimes_K \langle Q(z) \rangle$. Clearly, the map defined by $\langle P(z) \rangle \times \langle Q(z) \rangle \to \mathbb{C}((x)) : (P^{(m)}(z), Q^{(n)}(z)) \mapsto P^{(m)}(z) Q^{(n)}(z)$ is bilinear, thus by the universal property it factors over the tensor product. The image of the factor must then contain all terms of the form $P^{(m)}(z) Q^{(n)}(z)$, and by the product rule thus contains all derivatives of $P(z)Q(z)$. As $\dim \langle P(z) \rangle \otimes_K \langle Q(z) \rangle = \dim \langle P(z) \rangle \cdot \dim \langle Q(z) \rangle < \infty$, we are done.

3. Clearly $\langle P'(z) \rangle \subseteq \langle P(z) \rangle$, the statement follows.

4. $\dim \left\langle \int P(z) \, \mathrm{d}\, z \right\rangle \leq \dim \langle P(z) \rangle + 1 < \infty$, as there is only one new generating element.

5. As $Q(z) = 0$, $P(Q(z))$ is again a formal power series. By iteration of chain and product rules, all derivatives of $P(Q(z))$ lie in the vector space V generated by the $\{Q(z), Q'(z), \dots, P(Q(z)), P'(Q(z)), \dots$ Hence it is sufficient to show that this vector space has finite dimension.
As $Q(z)$ is algebraic, we know by the proof of the closure properties for algebraic functions that the vector space $K[Q]$ is closed under differentiation, so it contains $\langle Q(z) \rangle$. Furthermore, we know it to be isomorphic to $K[x]/\mu_Q(x)$, and since the minimal polynomial is

[4] This is what inspires the name D-finite, namely *differentially* finite.

irreducible, $K[Q] = K(Q)$ is an extension field of K with $[K(Q) : K] < \infty$.
By holonomy of P, we know that $\langle P(y) \rangle$ is finite dimensional over $K[y]$, letting $y = Q(z)$ means that the dimension of the vector space generated by $\{P(Q(z)), P'(Q(z)), \dots\}$ over $K(Q(z))$ has finite dimension. Putting this together, we obtain
$\dim V \leq [K(Q) : K] \cdot \dim_{K(Q(z))} \langle P(Q(z)) \rangle < \infty$.

□

These closure properties are not complete; see for instance [4, chapter 7], or [5, 6.4].

Example.

1. The function $\exp(z)$ is not algebraic, but holonomic, as $\frac{\partial}{\partial z}\exp(z) - \exp(z) = 0$. Similarly, $\log(z)$ is holonomic.

2. By the previous lemma and the last example, the function $\exp(\sqrt[3]{x^4}) \cdot \log x$ is holonomic.

3. The sequence of harmonic numbers, $H_n := \sum_{i=1}^{n} \frac{1}{i}$, is holonomic. To see this, consider $H_{n+1} = H_n + \frac{1}{n}$, leading to

$$nH_{n+1} - nH_n = 1, \tag{24}$$
$$(n+1)H_{n+2} - (n+1)H_{n+1} = 1 \tag{25}$$
$$\Rightarrow (n+1)H_{n+2} - (2n+1)H_{n+1} + nH_n = 0. \tag{26}$$

4. The sequences $a_n = \log n$, $a_n = n^n$, $a_n = p_n$, where p_n is the n-th prime number, are not holonomic [6], [7].

While there is a rather obvious method for proving that a given power series is holonomic - that is, find a holonomy equation (often using computer algebra methods, as in [8]), and prove it holds -, it is not so obvious how one would go about disproving a power series is holonomic. Two important tools to do so are given below.

Theorem 1.1.1 (Structure Theorem, [7])**.** *Let $P(z)$ be a holonomic power series with holonomy equation* (8)*. Also assume that z_0 is a singularity of $P(z)$, and S a sector containing with vertex at z_0. Then there is a subsector $S' \subseteq S$ and a basis of solutions of* (8)*, such that, as $z \to z_0$, any solution $Y(z)$ in this basis has asymptotic expansion*

$$Y \sim \exp\left(p(z^{\frac{-1}{r}})\right) z^{\alpha} \sum_{j=0}^{\infty} q_j \left(\log(z)\right) z^{js}, \tag{27}$$

where p is a polynomial, $r \in \mathbb{N}, \alpha \in \mathbb{C}, 0 < s \in \mathbb{Q}$, and the q_j are polynomials of uniformly bounded degree, all depending on the chosen solution Y.

In particular, if we can show that the asymptotic growth of a function does not agree with the above, then this function cannot be holonomic. An outline of a proof is given in [7], using [9, Thm 19.1]. This strategy is used, for instance, in [6], [7]. Another property of holonomic functions, which will be of more interest to us, is the following:

Theorem 1.1.2 (Finiteness of singularities, [7])**.** *Let $P(z)$ be a holonomic power series. Then $P(z)$ has only finitely many singularities.*

Proof. By [10, Th 9.1], we have a (for set initial conditions) unique analytic solution for (8) in all of $\mathbb{C}$ except the zeros of $p_n(z)$. As $p_n(z)$ is a polynomial, it can only have finitely many roots, and by analytic continuation of a local solution we are done. □

1.1.3 Multivariate Case

The generating functions of interest to us in later sections will not be univariate, but will have multiple variables counting more than one parameter at a time. Such a generating function has general form

$$A(\overline{z}) = \sum_{n \in \mathbb{N}^d} a(n_1, \dots, n_d)\overline{z}^n, \tag{28}$$

where $\overline{z} = (z_1, \dots, z_d), n = (n_1, \dots, n_d), \overline{z}^n = z_1^{n_1} \cdot \dots \cdot z_d^{n_d}$.

Definition 1.1.4. *A multivariate power series $A(\overline{z})$ is called **holonomic** if the vector space over $\mathbb{C}(z_1, \dots, z_d)$ generated by all partial derivatives $\frac{\partial^k}{\partial \overline{z}^k}, k \in \mathbb{N}^d$, is finite.*

Lemma 1.1.5. *A multivariate power series $P(\overline{z})$ is holonomic precisely if for each $1 \leq i \leq d$ there are $n \in \mathbb{N}$ and polynomials $p_0, \dots, p_n \in \mathbb{C}[z_1, \dots, z_d]$, such that*

$$p_n(\overline{z})\frac{\partial^n}{\partial z_i^n}P(\overline{z}) + p_{n-1}(\overline{z})\frac{\partial^{n-1}}{\partial z_i^{n-1}}P(\overline{z}) + \dots + p_0(\overline{z})P(\overline{z}) = 0. \tag{29}$$

Put differently, a multivariate power series is holonomic precisely if it is holonomic in each variable, where all others are treated as parameters. Or, equivalently, when the power series is holonomic in each variable over the field of rational functions in all others.

Proof. If the series is holonomic, then equation (29) follows, because the iterated derivatives with respect to z_i must at some point be linearly dependent.
Suppose (29) holds, and let N be the maximum of all n occurring therein. Then in any derivative $\frac{\partial^m}{\partial \overline{z}^m}P(\overline{z})$ with $m_i > N$ for some i we can substitute according to (29), thus the maximum possible dimension of the vector space spanned by all partial derivatives is at most $N^d < \infty$. □

Lemma 1.1.6. *Let $P(\overline{z}), Q(\overline{z}) \in \mathbb{C}[[z_1, \dots, z_d]]$ be holonomic power series. Then:*

1. *$P(\overline{z}) + Q(\overline{z})$ is holonomic,*
2. *$P(\overline{z}) \cdot Q(\overline{z})$ is holonomic,*
3. *$\frac{\partial}{\partial z_i}P(\overline{z})$ is holonomic,*
4. *$\int P(\overline{z}) \, \mathrm{d}\, z_i$ is holonomic,*
5. *if $Q(\overline{z})$ is algebraic, $Q(0) = 0$, then $P(Q(\overline{z}))$ is holonomic,*
6. *for $t \in \mathbb{C}$, $1 \leq i \leq n$, $P(z_1, \dots, z_{i-1}, t, z_{i+1}, \dots, z_n)$ is holonomic.*

Proof.
With the same ideas as in the univariate case, just a bit more lengthy. See for instance [11]. □

So multivariate holonomic power series have essentially the same closure properties as the univariate ones, with the additional property that, if we substitute a number for any variable, the resulting power series with on variable less will still be holonomic. This allows one to effectively reduce the numbers of variables when trying to prove non-holonomy, as it suffices to find only *one* choice of parameters such that the remaining series is D-infinite to immediately gain the result for the whole multivariate series.

1.2 Lattice Walks, the Kernel Method

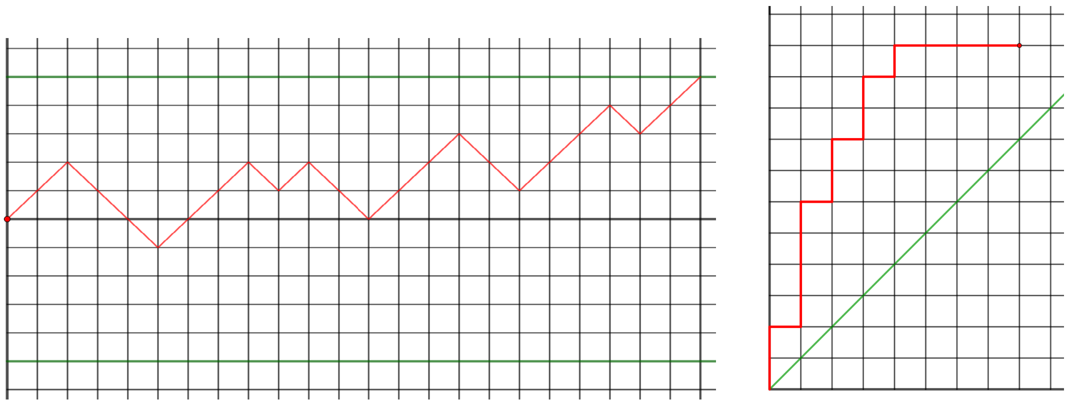
Lattice paths first appeared as a topic of interest in the late 19th century, as a method to illustrate and solve problems arising from probability theory (see [12] for a historical overview). In particular, they were closely linked to the so-called *Ballot problem* and the *Gambler's ruin*, so to give an idea of how such paths can arise in a natural context, these two problems shall be briefly illustrated here before a formal definition of lattice paths is introduced.
What would later become known as the *Gambler's ruin* started out with a letter from Blaise Pascal to Pierre de Fermat in 1654, the problem posed roughly translating to

> "Let two men play with three dice, the first player scoring a point whenever 11 is thrown, and the second whenever 14 is thrown. But instead of the points accumulating in the ordinary way, let a point be added to a player's score only if his opponent's score is nil, but otherwise let it be subtracted from his opponent's score. It is as if opposing points form pairs, and annihilate each other, so that the trailing player always has zero points. The winner is the first to reach twelve points; what are the relative chances of each player winning?"[13, p. 73]

Put a bit more concisely, the two players have winning probabilities as described, and the first player to lead by a margin of 12 wins is declared the victor. Both Pascal and Fermat were able to solve this problem, although one can merely speculate about their solution methods [13]. The problem generalizes quite naturally, however, by allowing for arbitrary winning probabilities and the winning margins.
The *Ballot problem*, as stated by Bertrand in 1887, translates to

> "We suppose that two candidates have been submitted to a vote in which the number of voters is μ. Candidate A obtains n votes and is elected; candidate B obtains $m = \mu - n$ votes. We ask for the probability that during the counting of votes, the number of votes for A is at all times greater than the number of votes for B." [14, p. 9]

Although named after him, the problem was not invented but merely rediscovered by Bertrand, and had already been solved before. The main reason why these two problems are so prominently linked to lattice paths is that there is a very natural geometric interpretation, which can be used to obtain a solution. In particular, there is an article of W. A. Whitworth from 1878 about the ballot problem, which is likely to be the first time a solution using lattice path enumeration was published [12, 15].
Whitworth assigned to each counting sequence of votes a path in $\mathbb{N} \times \mathbb{N}$ as follows: we start at $(0, 0)$. Whenever a vote goes to candidate A, we go one step upwards. If the vote goes to candidate B, we go one step to the right. After all votes are counted, we must end up at the point (m, n). Clearly, the condition that, at any given moment, A has more votes than B, is the same as our path never going below the diagonal $\{(x, x) : x \in \mathbb{N}\}$. This interpretation enables him to then use geometric arguments in order to arrive at a solution. Similarly, the gambler's ruin can be interpreted in terms of paths on $\mathbb{N} \times \mathbb{Z}$ by taking steps $(1, 1)$ if the first player wins, and $(-1, 1)$ if the second player wins. We are then interested in the number of paths hitting one of the lines $y = \pm m$ before the other, where m is the winning margin, with the number of paths weighted by the probability of the steps. These visualisations are illustrated in Figure 1 on page 11.
Therefore, the computation of probabilities can - in these two cases - reduced to counting paths fulfilling certain conditions. By today, a lot of theory and methods to deal with these kinds of enumeration problems have been developed, for an overview see eg [12, 16]. In parallel, a lot



(a) Visualisation of the Gambler's ruin.

(b) Visualisation of the Ballot problem.

Figure 1

of applications has emerged - lattice paths have applications in physics, biology and material sciences [17], polymer chemistry [18], and are commonly used in finance, to model stock markets [19]. In many of these applications, one of the great values of random walks lies in being an analogon for Brownian motion in a discrete setting [20]. In addition, lattice paths by themselves have proven to be interesting objects to study, with many connections to other areas of mathematics [12, 21].

The following only aims to give the most basic definitions and results; for an a bit more thorough introduction see for instance [16, 21].

Definition 1.2.1. *A* ***lattice path*** *on a lattice $\Lambda \subset \mathbb{R}^d$ with edges E is a - not necessarily finite - sequence of points $(x_0, x_1, \dots), x_i \in \Lambda$. The differences $s_i := x_{i+1} - x_i, i > 0$ are called the* ***steps*** *of the path.*

By far the most common choice of Λ is a subset of an Euclidean lattice, that is $\Lambda \subseteq \mathbb{Z}^d$. This will also be the case in later sections. Since, according to the previous definition, arbitrary paths are nothing more than sequences on Λ and thus arguably only mildly interesting, usually one will impose some kind of additional restrictions to the paths in question. There is a variety of interesting choices here, with applications in many different combinatorial settings, see for instance [12, 16]. The most prevalent one, however, is restricting the steps of a path to a set $\mathcal{S} \subset \mathbb{R}^d$ of allowed steps.

Definition 1.2.2. *A* ***lattice path with steps*** *$\mathcal{S}$ on a lattice Λ is a lattice path on Λ which fulfils the additional condition that $s_i \in \mathcal{S}$ for $i \geq 0$.*

In most cases, $\mathcal{S}$ is chosen to be finite.
Now let us assume our given lattice Λ is Euclidean. We are interested in lattice path enumeration, so we will want to find some ways how to sort the - usually infinite - number of paths into classes of interest. One rather natural way to do so would be to take any finite path γ starting at a fixed point $O \in \Lambda$. We can then, since it is finite, assign to it first its length, and secondly its ending point $P = (p_1, \dots, p_d) \in \Lambda \subseteq \mathbb{Z}^d$. We can then denote by $a(P; k) \in \mathbb{N} \cup \{\infty\}$ the number of lattice paths of length k starting at O and ending at P. Therefore, provided that all $a(P; k)$

are finite, we can define a multivariate generating function for lattice paths on Λ with steps $\mathcal{S}$ starting at O:

$$F(\overline{x};t) := \sum_{P\in\Lambda, k\in\mathbb{N}} a(P;k)\overline{x}^P t^k, \tag{30}$$

where $\overline{x} := (x_1, \dots, x_d)$, $\overline{x}^P := x_1^{p_1} \cdot \dots \cdot x_d^{p_d}$.

Lemma 1.2.1. *Let $\mathcal{S}$ be finite. Then $a(P;k)$ is finite for all $P \in \Lambda, k \in \mathbb{N}$. Furthermore, the formal Laurent series $F(\overline{x};t)$ as defined above converges in each of its arguments on some open set $U \subset \mathbb{C}^{d+1}$. In particular, for each $n \in \mathbb{N}$, there is an open neighbourhood V of 0 in $\mathbb{C}$, such that $F(\overline{x};t)$ is absolutely convergent on $\left(U_n \setminus U_{\frac{1}{n}}\right)\times V = \left\{(x_1, \dots, x_d, t) \in \mathbb{C}^{d+1} : \frac{1}{n} < |x_i| < n, t \in V\right\}$.*

Proof.
As, by assumption, $S := |\mathcal{S}|$ is finite, there are at most S^k paths of length k. Therefore we have for all $k \in \mathbb{N}$

$$\sum_{P\in\Lambda} a(P;k) \leq S^k. \tag{31}$$

As all summands are positive, each $a(P;k)$ can be at most S^k and must thus be finite. Now consider the sum

$$\sum_{P\in\Lambda} \overline{x}^P \left(\sum_{k\in\mathbb{N}} a(P;k)t^k\right). \tag{32}$$

By the same argument as before, there can be at most S different points $P_1^1, P_2^1, \dots, P_S^1$, such that $a(P;1) \neq 0$. Inductively, there can be at most S^k points $P_1^k, \dots P_{S^k}^k$, such that $a(P;k) \neq 0$. If we let D be the longest distance any allowed step travels, ie $D := \max\{|s|, s \in \mathcal{S}\}$, then we also know that[5] $|P_i^k| \leq D^k$. We now partition our lattice Λ into sublattices $\Lambda_0, \Lambda_1, \dots$, such that Λ_i contains precisely those points P for which i is the first index such that $a(P;i) \neq 0$. Note that this does not necessarily cover the entire lattice, but all points which contribute to the sum. If $\frac{1}{n} < |x_i| < n$, then we have $|\overline{x}^P| < n^{|P|}$ with $P = |p_1| + \dots + |p_n|$. Clearly, for $P \in \Lambda_i$ we have $|P| \leq dD^i$. We then have, if we choose t sufficiently small such that $\lambda := S \cdot |t| < 1$ (remember that we have $a(P;k) \leq S^k$),

$$\sum_{m=0}^{\infty} \sum_{P\in\Lambda_m} |\overline{x}^P| \left(\sum_{k\geq 0} a(P;k)|t|^k\right) \leq \sum_{m=0}^{\infty} S^m n^{|P|} \left(\sum_{k\geq m} a(P;k)|t|^k\right) \tag{33}$$

$$\leq \sum_{m=0}^{\infty} S^m (n^{dD})^m \lambda^m \frac{1}{1-\lambda} < \infty \tag{34}$$

which is a geometric series and thus converges for t small enough such that $\lambda S n^{dD} < 1$. As we have shown absolute convergence on an open set U as described, uniform convergence follows, and we can arrange our series to be a convergent Laurent series on U in each variable, hence we are done.

□

[5] Strictly speaking, we would need to account for the starting point P_0 here, but it does not matter for the series convergence - we could also always choose D large enough such that it additionally includes the distance from the starting point to the origin.

1.2.1 Unrestricted Plane Walks

In order to get an impression of what a generating function of lattice paths as previously defined can look like and how to compute it, let us first consider paths on the lattice $\Lambda = \mathbb{Z}^d$, with any finite step set $\mathcal{S}$. First of all, notice that we can encode any step $s = (s_1, \ldots, s_d) \in \mathcal{S}$ by a Laurent monomial in the variables $x_1, \ldots, x_d$, via a Laurent monomial $\overline{x}^s := x_1^{s_1} \ldots x_d^{s_d}$, in analogue to previous notation. Correspondingly, we can associate to the entire step set a Laurent polynomial via

$$\mathcal{S}\left(\overline{x}\right) := \sum_{x \in \mathcal{S}} \overline{x}^s. \tag{35}$$

This Laurent polynomial is called the **characteristic polynomial** of a step set $\mathcal{S}$. For an illustration and examples see table 1 on page 15. In the following, it will usually be clear when we are talking about $\mathcal{S}$ as the set of steps, and when we are talking about $\mathcal{S}\left(\overline{x}\right)$ as the associated Laurent polynomial, in all other cases it will be explicitly pointed out.

We claim:

Lemma 1.2.2. *The generating function of unrestricted plane walks starting in a point O with a finite set $\mathcal{S}$ of steps is given by*

$$F(\overline{x}; t) = \frac{\overline{x}^O}{1 - \mathcal{S}(\overline{x})}. \tag{36}$$

In particular, it is a rational function.

Proof.
First note that for all $P \in \Lambda, k \geq 0$ we have

$$a(P; k+1) = \sum_{s \in \mathcal{S}} a(P - s; k), \tag{37}$$

because the last point has precisely one predecessor for each step. Now, we can multiply this by $\overline{x}^P$ and sum over all $P \in \Lambda$ ($a(P; k)$ is 0 for almost all points), arriving at

$$\sum_{P \in \Lambda} a(P; k+1)\overline{x}^P = \sum_{P \in \Lambda} \left(\sum_{s \in \mathcal{S}} a(P - s; k)\overline{x}^P \right) = \sum_{P \in \Lambda} \left(\sum_{s \in \mathcal{S}} \overline{x}^s a(P - s; k)\overline{x}^{P-s} \right). \tag{38}$$

We can rearrange the latter (finite) sum, leading to

$$\sum_{P \in \Lambda} a(P; k+1)\overline{x}^P = \sum_{s \in \mathcal{S}} \overline{x}^s \left(\sum_{P \in \Lambda} a(P - s; k)\overline{x}^{P-s} \right), \tag{39}$$

and as $\Lambda = \mathbb{Z}^d$ is translation invariant, this yields

$$\sum_{P \in \Lambda} a(P; k+1)\overline{x}^P = \sum_{s \in \mathcal{S}} \overline{x}^s \left(\sum_{P \in \Lambda} a(P; k)\overline{x}^P \right) = \mathcal{S}(\overline{x}) \cdot \left(\sum_{P \in \Lambda} a(P; k)\overline{x}^P \right). \tag{40}$$

Now, multiplying the latter equation by t^k and taking the sum, we arrive at

$$\sum_{P \in \Lambda, k \in \mathbb{N}^{>0}} a(P; k)t^k = \mathcal{S}(\overline{x}) \cdot \left(\sum_{P \in \Lambda, k \in \mathbb{N}} a(P; k)t^k \right) \quad \Leftrightarrow \tag{41}$$

$$F(\overline{x}; t) - \overline{x}^O = \mathcal{S}(\overline{x})F(\overline{x}, t), \tag{42}$$

because clearly $a(P;0)$ is 1 if P is the starting point O, and 0 otherwise. Thus, we obtain the desired statement. Clearly, by expanding the fraction by an appropriate term, we can make the denominator a polynomial, thus $F(\overline{x};t)$ is rational.

$\square$

1.2.2 Walks in the Half Plane

As a next preparatory step, we will now continue by considering walks in the half plane $\Lambda = \mathbb{Z}\times\mathbb{N}$. It will turn out that the computation of a generating function is nowhere near as straightforward as in the previous case. It is, however, a very good illustration of some of the obstacles which will arise when considering quarter plane walks later on, and the main ideal in bypassing them. In the following, we will restrict our step set $\mathcal{S}$, such that each step goes at most unit length in the second coordinate, so for each $s = (s_1, s_2) \in \mathcal{S}$, we have $|s_2| \leq 1$. This restriction is in fact not necessary for how the following pans out, in fact, we could even consider $\Lambda = \mathbb{Z}^d \times \mathbb{N}$ for some $d \in \mathbb{N}$ without changing the idea of the proof. The reasons for restricting oursleves to this special case are essentially twofold: firstly, while the solution method otherwise is the same, the computation would be somewhat more technical, which is of no use when only aiming to illustrate the main ideas. Secondly, the quarter plane walks we will study later all fulfil these restrictions, anyway. For a more general solution, see eg [22].
First, let us see why we need to adapt our approach from the proof of Lemma 1.2.2 does not work out. If we start by a recursive definition as in (37), this clearly holds only for points $P = (p_1, p_2)$ with $p_2 \geq 1$, because else we count impossible paths from the negative half plane as well. So, consequently, when taking the sums afterward, we would need to account for these paths - leading to

$$\sum_{P\in\Lambda} a(P;k+1)\overline{x}^P = \sum_{s\in\mathcal{S}} \overline{x}^s \left(\sum_{P\in\Lambda} a(P-s;k)\overline{x}^{P-s}\right) - \sum_{s\in\mathcal{S}^+} \left(\sum_{p_2=0} a(P-s;k)\overline{x}^{P-s}\right), \tag{43}$$

where $\mathcal{S}^+$ is the set of steps with positive second coordinate. This recursion is a lot less tangible than the previous one. Note, for instance, that $a(P-s;k)$ is not immediately defined, as $P-s$ might not even lie in Λ. This can be remedied by first considering all paths in $\mathbb{Z}\times\mathbb{Z}$, and then only counting those which do fulfil the conditions, and eventually this can probably be worked out to a solution. But there is a similar, more methodical approach.
Our aim is to not only consider one generating function $F((x,y);t)$, but rather to consider a set of generating functions $F_y(x;t)$, where $F_y(x;t)$ is a generating function in only two variables, fixing the height of the end point. Now, clearly, if we know all these functions, we have what we want; more precisely, we have

$$F\left((x,y);t\right) = \sum_{n\in\mathbb{N}} y^n F_n(x;t). \tag{44}$$

Correspondingly, the number of paths of length k ending at a given x-coordinate m, given the height n is fixed, shall be denoted by $a_n(m,k)$. It will turn out that this innocuous additional structure makes our life a lot easier, because the relations between the $F_n(x;t)$ are now fairly simple. Let $\mathcal{S}^+, \mathcal{S}^-, \mathcal{S}^0$ be the sets of steps with positive, negative, and 0 second coordinates respectively. We can assign to each its characteristic polynomial as before, ie $\mathcal{S}^+(x,y)$, but will also assign to each its characteristic polynomial in x only, that is, $\mathcal{S}^+(x) := \mathcal{S}^+(x,1)$. See table 1 for two examples.

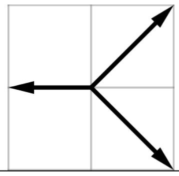
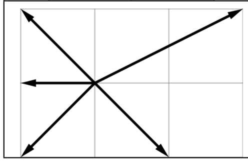
Steps	$\mathcal{S}(x,y)$	$\mathcal{S}^+(x,y)$	$\mathcal{S}^+(x)$
	$xy + xy^{-1} + x^{-1}$	xy	x
	$x^2y + xy^{-1} + x^{-1}y + x^{-1} + x^{-1}y^{-1}$	$x^2y + x^{-1}y$	$x^2 + x^{-1}$

Table 1: Two possible step sets with corresponding characteristic Laurent polynomials $\mathcal{S}(x,y)$, along with the respective $\mathcal{S}^+(x,y), \mathcal{S}^+(x)$.

For $n \geq 1$, we then have

$$a_n(m;k+1) = \sum_{s\in\mathcal{S}^+} a_{n-1}(m-s_1;k) + \sum_{s\in\mathcal{S}^0} a_n(m-s_1;k) + \sum_{s\in\mathcal{S}^-} a_{n+1}(m-s_1;k). \tag{45}$$

Again, we can multiply by $x^k t^{k+1}$, sum over $m \in \mathbb{Z}$, which, after transformations very similar to what we have seen in the previous section, leads to

$$F_n(x;t) - F_n(x;0) = t\mathcal{S}^+(x)F_{n-1}(x;t) + t\mathcal{S}^0(x)F_n(x;t) + t\mathcal{S}^-(x)F_{n+1}(x;t) \tag{46}$$

for all $n \geq 1$. Note in particular that $F_n(x;0) = 0$ if the starting point P is not of the form $P = (p_1, n)$; if it does, then $F_n(x;0) = x^{p_1}$.
If, on the other hand, $n = 0$, then clearly the first sum on the right hand side of (45) vanishes, as no path can come from the negative half plane, leading to

$$F_0(x;t) - F_0(x;0) = t\mathcal{S}^0(x)F_0(x;t) + t\mathcal{S}^-(x)F_1(x;t). \tag{47}$$

We are ultimately interested in $F(x,y;t) = \sum y^n F_n(x;t)$, so we can multiply (46) by y^n, and take the sum over all $n \geq 1$. This, after some simple manipulations, leaves us with

$$F(x,y;t) - \varepsilon x^P - F_0(x;t) = \tag{48}$$

$$ty\mathcal{S}^+(x)F(x,y;t) + t\mathcal{S}^0(x)\left(F(x,y;t) - F_0(x;t)\right) + ty^{-1}\mathcal{S}^-(x)\left(F(x,y;t) - yF_1(x;t) - F_0(x;t)\right), \tag{49}$$

where ε is 0 if the starting point P has y-coordinate 0, and 1 otherwise. As $y\mathcal{S}^+(x) + \mathcal{S}^0 + y^{-1}\mathcal{S}^-(x) = \mathcal{S}(x,y)$, this simplifies to

$$F(x,y;t) - \varepsilon x^P - F_0(x;t) = \tag{50}$$

$$t\mathcal{S}(x,y)F(x,y;t) - t\left(\mathcal{S}^0(x)F_0(x;t) + \mathcal{S}^-(x)F_1(x;t) + y^{-1}\mathcal{S}^-(x)F_0(x;t)\right), \tag{51}$$

which by (47) reduces to

$$F(x,y;t)\left(1 - t\mathcal{S}(x,y)\right) - \varepsilon x^P - F_0(x;t) = -F_0(x;t) + \delta x^P - t\mathcal{S}^-(x,y)F_0(x;t), \tag{52}$$

where $\delta = 1$ if P has y-coordinate 0, and $\delta = 0$ otherwise. So, clearly, $\varepsilon + \delta = 1$ in every case, so altogether we arrive at

$$F(x,y;t)\left(1 - t\mathcal{S}(x,y)\right) = x^P - t\mathcal{S}^-(x,y)F_0(x;t). \tag{53}$$

We have therefore shown

Lemma 1.2.3. *The generating function $F(x, y; t)$ of walks in the half plane $\Lambda = \mathbb{Z} \times \mathbb{N}$, starting from a point P, satisfies*

$$F(x, y; t)\,(1 - t\mathcal{S}(x, y)) = x^P - t\mathcal{S}^-(x, y)F_0(x; t), \tag{54}$$

where $F_0(x, t)$ is the generating function of paths ending at height 0.

It may be worth pointing out that, if one is familiar with symbolic methods as presented, for example, in [1, I.], then this equation is clear: for the walks $\mathcal{W}$, we have

$$\mathcal{W} = \{P\}\dot{\cup}\,(\mathcal{W} \times \mathcal{S}) \setminus \left(\mathcal{W}_0 \times \mathcal{S}^-\right). \tag{55}$$

1.2.3 The Kernel Method[6]

Lemma 1.2.3 gives a functional relation between the generating functions of walks in the half plane and walks in the half plane ending on the line $\{y = 0\}$. While interesting to know, this does not immediately answer the question of what these functions actually look like. To get there, we need to invest a bit more work, and use the so-called kernel method. The latter is not only applicable to path counting problems; for a collection of examples see eg [22].
For the sake of simplicity, assume in the following that the starting point $P = (0, 0)$.
Consider again the functional equation. By defining the **kernel** of our equation, $\mathcal{K}(x, y) := 1 - t\mathcal{S}(x, y)$, and (54), it takes the form

$$\mathcal{K}(x, y)F(x, y; t) = 1 - t\mathcal{S}^-(x, y)F_0(x; t). \tag{56}$$

At this point, we can take note of a few things:

1. The functional equation looks almost the same as it does for the unrestricted walks. The bulk terms on the left hand side describe the unconstrained behaviour, whereas the right hand side - the only difference to the equation in the unrestricted case - makes a boundary term, discounting all disallowed paths.

2. Computing either of the two functions $F(x, y; t)$ or $F_0(x; t)$ is good enough for us, as we can always plug the result into our functional equation and obtain the other.

3. Essentially all terms that depend on y (with the exception of $\mathcal{S}^-(x, y) = y^{-1}\mathcal{S}^-(x)$) are on the left hand side. Consequently, we can play around with different values for y relatively easily, without changing the structure of the right hand side too much.

4. We have a product on the left hand side which gives us a lot of room to play around with; for example if $\mathcal{K}(x, y) = 0$, then the entire left hand side is 0.

There are essentially two ways to progress from here: one can try to eliminate containing y, in particular the kernel, or one can try to eliminate the boundary terms. We will discuss the first option here, because it is very well-suited to walks in the half plane, and arguably a lot more straightforward in this case. The second option will be essential in the following sections and discussed in-depth there.

[6]The structure of the following section was partly inspired by a talk given by A. Rechnitzer as part of the workshop "Lattice walks at the Interface of Algebra, Analysis and Combinatorics" at the Banff International Research Station on September 18, 2017.

Eliminating $\mathcal{K}(x, y)$:

In order to eliminate the left hand side of (56), it suffices to eliminate $\mathcal{K}(x, y)$. In particular, since $\mathcal{K}(x, y)$ is a Laurent polynomial in x, y, where y appears with degrees ± 1, $\mathcal{K}(x, y) = 0$ yields a quadratic equation in y. We can then solve this for y - concretely, we have:

$$\mathcal{K}(x, y) = 1 - t(y\mathcal{S}^+(x) + \mathcal{S}^0(x) + y^{-1}\mathcal{S}^-(x)) = 0 \quad \Leftrightarrow \tag{57}$$

$$y^2 \left(t\mathcal{S}^+(x)\right) + \left(t\mathcal{S}^0(x) - 1\right) y + t\mathcal{S}^-(x) = 0 \quad \Leftrightarrow \tag{58}$$

$$\frac{1 - t\mathcal{S}^0(x) \pm \sqrt{\left(t\mathcal{S}^0(x) - 1\right)^2 - 4t^2 S^+(x) S^-(x)}}{2t\mathcal{S}^+(x)} = y. \tag{59}$$

Computing the series expansion of the term under the square root yields an expression of the form $1 - t\mathcal{S}^0(x) + \mathcal{O}(t^2)$, which means that the solution $y^-(x, t)$ with the negative squareroot has a power series expansion in t with no constant coefficient. Therefore, if we fix x as a sufficiently small complex number, $F(x, y^-(x, t); t)$ is convergent as a formal power series in t, and we can make use of this substitution.
Using this, we obtain

$$0 = \mathcal{K}(x, y^-(x, t))F(x, y^-(x, t); t) = 1 - t\frac{\mathcal{S}^-(x)}{y^-(x, t)}F_0(x; t), \tag{60}$$

so we have

$$F_0(x; t) = \frac{y^-(x, t)}{t\mathcal{S}^-(x)} \tag{61}$$

for all x, t such that $y^-(x, t), \mathcal{S}^-(x)$ is neither 0 nor ∞. As $y^-(x, t) = \mathcal{O}(t)$ for fixed x, if we let x be in some open set U_1 with positive distance from 0, we can certainly find a neighbourhood V of 0 such that $0 < |y^-(x, t)| < \infty$ for $(x, t) \in U_1 \times V$. In particular, $F_0(x; t)$ is by construction (the only non-rational part is the square-root in $y^-(x; t)$) a well-defined algebraic function in both x and t on $U_1 \times V$. Now, consider again (56). On $U_1 \times V$, plugging in the above result for $F_0(x, t)$ we can compute $F(x, y; t)$ via

$$F(x, y; t) = \frac{1}{\mathcal{K}(x, y)}\left(1 - t\mathcal{S}^-(x, y)\frac{y^-(x, t)}{t\mathcal{S}^-(x)}\right) = \frac{1}{\mathcal{K}(x, y)}\left(1 - \frac{y^-(x, t)}{y}\right), \tag{62}$$

as long as $\mathcal{K}(x, y), y$ are both not 0. Again, this is certainly the case[7] for some open set U_2 with positive distance from 0 (possibly depending on the choice of U_1), and on $U_1 \times U_2 \times V$ we therefore know $F(x, y; t)$ to be a well-defined function, which by construction then also has to be algebraic. Fulfilling the relation (56) on an open set and by uniqueness of Laurent series, this function is indeed the generating function we were looking for.
We have thus shown:

Lemma 1.2.4. *The generating function $F(x, y; t)$ of walks in the half plane $\Lambda = \mathbb{Z} \times \mathbb{N}$ starting at the origin with step set $\mathcal{S}$, where for $s = (s_1, s_2) \in \mathcal{S}$ we have $|s_2| \leq 1$, fulfils*

$$\mathcal{K}(x, y)F(x, y; t) = 1 - t\mathcal{S}^-(x, y)F_0(x; t), \tag{63}$$

[7] Note that $\mathcal{K}(x, y)$, being a Laurent polynomial, is continuous; hence $\mathcal{K}^{-1}(0) \subset U_1 \times \mathbb{C}$ is closed in the subspace topology.

where $F_0(x;t)$ is the generating function of walks ending on the line $y = 0$. We have

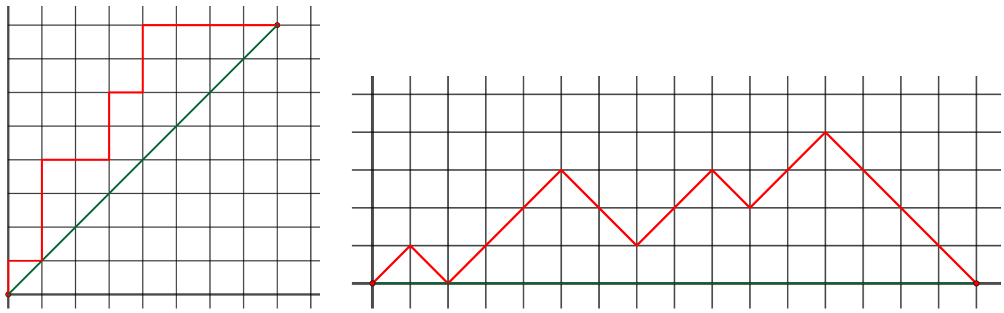
$$F_0(x;t) = \frac{y^-(x,t)}{t\mathcal{S}^-(x)}, \tag{64}$$

$$F(x,y;t) = \frac{1}{K(x,y)}\left(1 - \frac{y^-(x,t)}{y}\right). \tag{65}$$

In particular, both $F_0(x;t)$ and $F(x,y;t)$ are algebraic.

Example 1. Dyck Paths.

Consider the Ballot problem as described at the beginning of this section. We consider now the case that $m = n$, that is, both candidates will receive an equal amount of votes - we are interested in the probability that candidate A has at least as many votes as candidate B at any given moment in time. We have already seen that this can be visualised by paths from 0 to $P = (n, n)$, going up or to the right and never crossing the diagonal $x = y$. By clockwise rotation of this image by 45°t, we obtain a different interpretation: the number of possible countings with such an outcome is the number of paths from $(0, 0)$ to $(2n, 0)$, which never cross the x-axis. The allowed steps are $(1, 1)$ and $(1, -1)$, that is, we always go one step to the right, and - depending on which candidate gained an additional vote - either one unit up or down. For an illustration, see Fig. 2 below. Such paths are commonly called **Dyck-paths**.



c: An admissible counting for the ballot problem with $m = n$.

d: The corresponding Dyck path in the half plane.

Figure 2

In the notation from before, we have the step set $\mathcal{S} = \{(1,1),(1,-1)\}, \mathcal{S}(x,y) = xy + xy^{-1}$. Furthermore, we have $\mathcal{S}^+(x) = x, \mathcal{S}^0(x) = 0, \mathcal{S}^-(x) = x$, and $\mathcal{K}(x,y) = 1 - t(xy + xy^{-1})$. By equation (56), we have

$$\left(1 - t(xy + xy^{-1})\right) F(x,y;t) = 1 - txy^{-1}F_0(x;t). \tag{66}$$

Eliminating $\mathcal{K}(x,y)$ leaves us with

$$y^-(x,t) = \frac{1 - \sqrt{1 - 4t^2}}{2tx}. \tag{67}$$

As we are not interested in x (each step goes precisely one unit in the x-direction, thus it is redundant), and are interested in the total number of walks, we let $x = 1$, and obtain

$$y^-(1,t) = \frac{1-\sqrt{1-4t^2}}{2t} \quad \Rightarrow \tag{68}$$

$$F_0(1;t) = \frac{1-\sqrt{1-4t^2}}{2t^2}. \tag{69}$$

It is immediately obvious that there is no t in the generating function but only t^2. This makes sense, because in order to ever return to the line $y = 0$, we must have travelled an even number of steps, so all odd coefficients must be 0. Expanding this power series using the binomial theorem yields

$$a_0(1,2n) = \frac{1}{n+1}\binom{2n}{n}, \tag{70}$$

which is the n-th Catalan number. As there are $\binom{2n}{n}$ possible counting orders, corresponding to unrestricted paths from $(0,0)$ to $(2n,0)$ with the given steps (we can arbitrarily choose exactly n steps to go upwards, and the rest must then go downwards), we arrive at the resulting probabilty $P = \frac{1}{n+1}$.

A key condition for this technique to work it is that there can be no more than one boundary term. If there were two, then we would have a relation of three different functions; and after eliminating one we'd still be stuck with the other two. So, in that case, this method cannot be applied in this way. As we will be interested in studying walks in the quarter plane, where there clearly are two boundary conditions, we will have to choose a different approach.

2 Walks in the Quarter Plane I

Instead of the entire or the half plane as before, we now consider walks on the quarter plane, that is, on the lattice $\Lambda = \mathbb{N}\times\mathbb{N}$. We will restrict our step set $\mathcal{S}$ by only allowing *small steps*, so neither step goes more than unit distance in the direction of either axis, so $\mathcal{S} \subseteq \{-1,0,1\} \times \{-1,0,1\}$. In particular, we are interested which step sets $\mathcal{S}$ lead to a holonomic generating function. This question was posed in 2010 by M. Bousquet-Mélou and M. Mishna in [24], who conjectured that this is the case precisely if the group associated with the step set (which will be introduced very soon) is finite. This conjecture has since been proven [25, 26, 27, 28].
The methods used in its proof can be roughly divided into two parts; namely those using orbit sums and directly manipulating a functional equation, and those utilizing a more abstract approach via elliptic curves. In this section, the aim is to illustrate and explain the ideas underlying orbit sum methods, and showing that for all those walks with a finite group, the generating function must be holonomic.

2.1 The Group of a Walk

First, we are interested in a functional equation for our generating function, as we have already seen previously. Because directly computing such as we did for unrestricted walks and walks in the half plane turns out to be rather tedious here and can be done using the very same methods as in the half plane case, we will use the symbolic method instead.
Let $\mathcal{W}$ be the set of our walks, $\mathcal{W}_{0,y}, \mathcal{W}_{x,0}, \mathcal{W}_{0,0}$ be the set of paths ending on the x-axis, the y-axis or the origin respectively. Also, let $\mathcal{S}_x, \mathcal{S}_y$ the sets of steps with negative x and y component respectively. Then we have

$$\mathcal{W} = \{P\} \dot\cup (\mathcal{W} \times \mathcal{S}) \setminus (\mathcal{S}_y \times \mathcal{W}_{x,0} \cup \mathcal{S}_x \times \mathcal{W}_{0,y}). \tag{71}$$

In terms of generating functions, by inclusion-exclusion this gives us

$$F(x,y;t) = 1 + t\mathcal{S}(x,y)F(x,y;t) - t\left(F(x,0;t)\mathcal{S}_y(x,y) + tF(0,y;t)\mathcal{S}_x(x,y)\right) + t\varepsilon F(0,0;t), \tag{72}$$

where

$$\varepsilon = \mathcal{S}_{xy}(x,y) = \begin{cases} 0 & \text{if } (-1,-1) \notin \mathcal{S}, \\ x^{-1}y^{-1} & \text{if } (-1,-1) \in \mathcal{S}. \end{cases} \tag{73}$$

Writing this as a kernel equation, we have

$$F(x,y;t)\mathcal{K}(x,y;t) = 1 - tF(x,0;t)\mathcal{S}_y(x,y) - tF(0,y;t)\mathcal{S}_x(x,y) + t\varepsilon x^{-1}y^{-1}F(0,0;t). \tag{74}$$

The key difference to the half plane case is that here, if we eliminate the kernel, we still have multiple unknown functions, so this will generally not help us[8].
Remembering what we noticed about the kernel equation in section 1.2.3, this means that our next approach would be trying to eliminate the boundary terms. To illustrate the main idea of how this could (ideally) go, let us start with an example before proceeding in a more general setting.

[8] While it may look tempting, we cannot simply eliminate the kernel and then, roughly speaking, have all terms containing x belong to $F(x,0;t)$, and the terms containing y belong to $F(0,y;t)$: by eliminating the kernel, we write one of the variables as a power series of the other. So if we substitute x by an appropriate term $x(y)$ to eliminate the kernel, we will also have a power series of the form $F(x(y),0;t)$, which does in fact contain powers of y.

Example 1.

Consider again the functional equation for Dyck paths, as in the previous example on page 18, with $x = 1$. We have

$$\left(1 - t(y + y^{-1})\right) F(y;t) = 1 - ty^{-1}F_0(t). \tag{75}$$

Now, one could notice that the kernel seems to be very symmetrical - it is invariant under the transformation $y \mapsto y^{-1}$. The right hand side will not be changed all too much; by applying this transformation we get the very similar equation

$$\left(1 - t(y + y^{-1})\right) F(y^{-1};t) = 1 - tyF_0(t). \tag{76}$$

Using this system of two equations, we can now eliminate the boundary term $F_0(t)$, arriving at

$$yF(y;t) - y^{-1}F(y^{-1};t) = \frac{y - y^{-1}}{1 - t(y + y^{-1})}. \tag{77}$$

While at first it might seem as we still have one unknown too many, namely $F(y;t)$ and $F(y^{-1};t)$, this actually hardly matters: $F(y;t)$ is a power series in y, and as such it clearly contains no summands with powers of y^{-1}. Similarly, $F(y^{-1};t)$ will consist only of summands of powers of y^{-1}, but will have no powers of y. Therefore, if we consider the Laurent series expansion of the right hand side, all terms with a positive power of y make up $yF(y;t)$, and all terms with a negative power of y will make up $F(y^{-1};t)$. Put differently,

$$yF(y;t) = \left[y^{>0}\right] \frac{y - y^{-1}}{1 - t(y + y^{-1})}. \tag{78}$$

We can compute this by a partial fraction decomposition with respect to y, and obtain as a result that $F(y;t)$ is algebraic. For more details, see again [24].

In summary, what we did is essentially the following:

1. we found a transformation (here: $y \mapsto y^{-1}$) which leaves the kernel $\mathcal{K}(x, y)$ invariant, and does not affect the rest too much,
2. we iterated this, until we could eliminate the boundary terms solving a system of linear equations, where the Laurent series on the other hand split up nicely (in our case, all positive coefficients belonged to $F(y;t)$),
3. we could then compute $F(x, y;t)$ by a partial fraction decomposition.

So now let us see what happens in a more general setting. As we have done previously, we can now define Laurent polynomials $\mathcal{S}_x^+(y), \mathcal{S}_x^0(y), \mathcal{S}_x^-(y)$ by considering $\mathcal{S}(x, y)$ as a polynomial in x and taking the coefficients (monomials in y) of $x, 1, x^{-1}$ respectively. Similarly, we obtain $\mathcal{S}_y^+(x), \mathcal{S}_y^0(x), \mathcal{S}_y^-(x)$.
For the first step, we are interested in finding a - somewhat nice - change of variables leaving the kernel unaffected. We can write the kernel as

$$\mathcal{K}(x, y) = 1 - t\mathcal{S}(x, y) = 1 - t\left(\mathcal{S}_y^+(x)y + \mathcal{S}_y^0(x) + \mathcal{S}_y^-(x)y^{-1}\right). \tag{79}$$

One fairly intuitive transformation leaving the kernel invariant would be the mapping $y \mapsto$

$\frac{\mathcal{S}_y^-(x)}{\mathcal{S}_y^+(x)}y^{-1}$. Similarly, we also have

$$\mathcal{K}(x,y) = 1 - t\left(\mathcal{S}_x^+(y)x + \mathcal{S}_x^0(y) + \mathcal{S}_x^-(y)x^{-1}\right), \tag{80}$$

so a second transformation is given by $x \mapsto \frac{\mathcal{S}_x^-(y)}{\mathcal{S}_x^+(y)}x^{-1}$. Thus, we now have two plausible transformations to work with:

$$\Psi : \quad \begin{cases} x \mapsto x, \\ y \mapsto \frac{\mathcal{S}_y^-(x)}{\mathcal{S}_y^+(x)}y^{-1}, \end{cases} \tag{81}$$

$$\Phi : \quad \begin{cases} x \mapsto \frac{\mathcal{S}_x^-(y)}{\mathcal{S}_x^+(y)}x^{-1}, \\ y \mapsto y. \end{cases} \tag{82}$$

Clearly, Φ, Ψ are involutions, that is, $\Phi \circ \Phi = \Psi \circ \Psi = \mathrm{Id}$. Their product, $\Phi \circ \Psi$, however, generally is not.

Definition 2.1.1. *The* ***group of a walk*** *with step set $\mathcal{S}$, denoted by $G(\mathcal{S})$, is the cyclic group generated by Φ and Ψ as defined above.*

Lemma 2.1.1. *If $\Theta := \Psi \circ \Phi$ has order $n < \infty$ in $G := G(\mathcal{S})$, then G is finite and of order $2n$. In particular, any $g \in G$ has a unique representation of the form either $g = \Phi \circ \Theta^k$ or $g = \Theta^k$, where $0 \leq k < n$.*

Proof.
If $\Theta^n = \mathrm{Id}$, we have $\Phi \circ \Theta^{n-1} = \Phi \circ \Theta^{-1} = \Phi \circ (\Psi \circ \Phi)^{-1} = \Phi \circ \Phi \circ \Psi = \Psi$. As Ψ, Φ are involutions, any element of G can be written as an alternating product of Ψ, Φ. If this product ends with Φ, then clearly we can write it as Θ^k or $\Phi \circ \Theta^k$, depending on the first factor. If it ends with Ψ, then we can substitute $\Psi = \Phi \circ \Theta^{n-1}$ and are in the previous case. Thus, each element has a representation as claimed.
Assume $\Theta^k = \Theta^l$, or, equivalently, $\Phi \circ \Theta^k = \Phi \circ \Theta^l$, $0 \leq l \leq k < n$, then we have $\Theta^{k-l} = \mathrm{Id}$, and thus $n|(k-l)$, and because $0 \leq k-l < n$ we have $k = l$.
If $\Phi \circ \Theta^k = \Theta^l$, then $\Phi = \Theta^{l-k}$, and because $\Phi \circ \Theta^{n-1} = \Psi$, this would mean that both Ψ, Φ are generated by Θ. Thus G would be a cyclic group, but $\Phi^2 = \Psi^2 = \mathrm{Id}$, and Φ, Ψ are by definition always different elements $\neq \mathrm{Id}$, a contradiction. Thus the representation is unique.

□

As our strategy of eliminating the boundary terms will heavily revolve around iterating Φ, Ψ in order to get a solvable system of linear equations, we will naturally be very interested in the behaviour of this group. Note that, in the previous example, we had $\Psi = \mathrm{Id}$, so by construction the group had only 2 elements. This made our task remarkably simple - but this is not always the case. For what might happen, see the following examples.

Example 2.
Consider the step sets given in the following table.

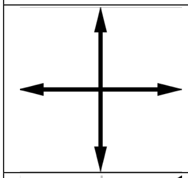
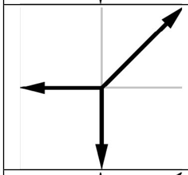
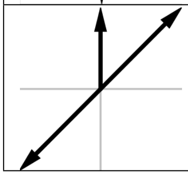
Steps	$\mathcal{S}(x,y)$	Φ	Ψ	Group order
	$x+y+x^{-1}+y^{-1}$	$x \mapsto x^{-1}$ $y \mapsto y$	$x \mapsto x$ $y \mapsto y^{-1}$	4
	$xy+x^{-1}+y^{-1}$	$x \mapsto x^{-1}y^{-1}$ $y \mapsto y$	$x \mapsto x$ $y \mapsto x^{-1}y^{-1}$	6
	$xy+x+x^{-1}y^{-1}$	$x \mapsto x^{-3}$ $y \mapsto y$	$x \mapsto x$ $y \mapsto y^{-2}(y+1)^{-1}$	∞

Table 2: Three step sets with corresponding transformations and group order

To see the corresponding group orders, we proceed for each step set separately:

1. $\mathcal{S} = x + y + x^{-1} + y^{-1}$.
   We have
   $$(x,y) \overset{\Phi}{\mapsto} (x^{-1},y) \overset{\Psi}{\mapsto} (x^{-1},y^{-1}) \overset{\Phi}{\mapsto} (x,y^{-1}) \overset{\Psi}{\mapsto} (x,y),$$
   thus we know that $(\Psi \circ \Phi)^2 = \mathrm{Id}$, so the order of the group is 4.

2. $\mathcal{S} = xy + x^{-1} + y^{-1}$.
   We have
   $$(x,y) \overset{\Phi}{\mapsto} \left(\frac{1}{xy}, y\right) \overset{\Psi}{\mapsto} \left(\frac{1}{xy}, \frac{1}{\frac{1}{xy}y}\right) = \left(\frac{1}{xy}, x\right) \overset{\Phi}{\mapsto} (y,x),$$
   and if we repeat this, we transform (y,x) back into (x,y). So we have $(\Psi \circ \Phi)^3 = \mathrm{Id}$, thus group order 6.

3. $\mathcal{S} = xy + x + x^{-1}y^{-1}$.
   We have
   $$(x,y) \overset{\Phi}{\mapsto} \left(\frac{1}{xy^2}, y\right) \overset{\Psi}{\mapsto} \left(\frac{1}{xy^2}, \frac{1}{\frac{1}{xy^2}\left(\frac{1}{xy^2}+1\right)y}\right) = \left(\frac{1}{xy^2}, \frac{x^2y^4}{(1+xy^2)y}\right) \mapsto \dots$$
   By a valuation argument, the group turns out to be infinite, see the proof of Theorem 2.2.1 on page 27.

As the groups of different walks can, as we have just seen, exhibit vastly different behaviour, it will be necessary to utilize different methods to study different step sets. As a consequence, it becomes necessary to take a step back, and try to find a way to classify all the different quarter plane walks.

2.2 Classification of Quarter Plane Walks

We will proceed similarly to [24].
Before considering the groups of walks in more detail, we can try to minimize the amount of work needed by finding some symmetries, and some trivial cases which will then not need further study. Steps will sometimes be referred to by their values in $\mathcal{S} = \{-1, 0, 1\} \times \{-1, 0, 1\}$, sometimes directly by their corresponding monomial, and sometimes by the corresponding cardinal directions (ie $(1, 1)$ is **NE**).
A priori, we have 256 choices for $\mathcal{S}$ - there are 8 different steps, each can either be in $\mathcal{S}$ or not. But we can reduce this number by a fair bit:

1. If $\mathcal{S}$ has no step in negative x-direction, that is, $\mathcal{S} \subseteq \{0, 1\} \times \{-1, 0, 1\}$, then instead of walks in the quarter plane, we can consider walks in the half plane $\{y \geq 0\}$. As any such walk can never enter the half plane $\{x < 0\}$, these are already necessarily restricted to the quarter plane. But by Lemma 1.2.4 on page 17, we already know that walks in the half plane will have algebraic generating functions. Thus it suffices to consider step sets $\mathcal{S}$ which have negative steps in both x- and y-directions.

2. In the same manner, it suffices to consider step sets $\mathcal{S}$ which have positive steps in both x- and y-directions. Put together with the previous point, this means that any admissible step set will contain each of the letters {**N,S,W,E**} somewhere in its representation via cardinal directions[9].

3. Similarly, if all steps are above the first diagonal $\{x = y\}$, then any walk in the half plane $\{y \geq 0\}$ will necessarily be in the quarter plane, so these cases are also not of interest to us. Analogously, the same applies if any steps are below the diagonal.

4. If $\mathcal{S}$ contains only steps below the second diagonal $\{x = -y\}$, then there is no possible first step from the origin. Thus, there is no walk, and the generating function $F(x, y; t) = 1$ is rational.

5. The question posed is clearly symmetric in x, y. Consequently, if one step set can be generated from another by interchanging x and y, we need not consider it separately.

Computing the number of interesting cases now becomes an exercise in generating polynomials, using inclusion-exclusion:

- The generating polynomial $P(z)$ of all possible step sets is $(1 + z)^8$, where the coefficients of z^k denote the number of step sets $\mathcal{S}$ with $|\mathcal{S}| = k$.

- The generating polynomial of step sets with no positive x-steps is given by $(1+z)^5$, because in each case 3 steps are prohibited. It is clearly the same as for no negative x-steps, or no positive/negative y-steps.

- The generating polynomial for step sets with no positive x- and y-steps is given by $(1+z)^3$, as there are only 3 allowed steps anymore. The same holds for x-positive, y-negative and vice versa.

[9] The letter does not have to be on its own - the set $\mathcal{S} = \{$**NE,SE,SW,NW**$\}$ is perfectly valid.

- The generating polynomial for step sets with neither positive nor negative x-steps is given by $(1+z)^2$, as there are only 2 allowed steps left. The same holds for y-steps.
- The generating polynomial for steps having neither positive nor negative x-steps and no positive y-step is given by $(1+z)$, as there is only one possible step left. The same holds for the symmetric cases.
- The generating polynomial for step sets having no positive or negative x- or y-steps is 1.

By inclusion-exclusion, we obtain the generating polynomial

$$P_1(z) = (1+z)^8 - 4(1+z)^5 + \left(4(1+z)^3 + 2(1+z)^2\right) - 4(1+z) + 1, \tag{83}$$

leaving us with a total number of 161 (by counting coefficients) different walks having steps in both positive and negative x- and y-directions.
If we want to subtract from these 161 those that lie above the first diagonal, first notice that they must include both $(1,1)$ and $(-1,-1)$ - otherwise, there would be no steps in positive x-, or negative y-direction. So there are three steps left that can be included in the step set (namely **W, NW, N**), so we have the generating polynomial of $z^2(1+z)^3$. Similarly, we get the same polynomial for the step sets below the diagonal. The only case we are counting twice is the step set $\{\mathcal{S} = \mathbf{SW,\ NE}\} = \{(-1,-1),(1,1)\}$, so we get our new counting polynomial

$$P_2(z) := P_1(z) - 2z^2(1+z)^3 + z^2, \tag{84}$$

leaving us with 146 cases.
Next, we want to discard those step sets where all steps lie below the second diagonal. Note that all those we have already discounted must include the steps **NW, SE**, because else there would be no positive x- or y-step. Thus, we are left with a counting polynomial of $z^2(1+z)^3$in the same manner as before, leading to[10]

$$P_3(z) := P_2(z) - z^2(1+z)^3, \tag{85}$$

and a total of 138 cases of interest.
Last, we want to take into account the symmetry. Since the transformation $\phi : (x \mapsto y, y \mapsto x)$ applied twice yields the identity, it is a bijection on all step sets. Also, step sets below the first diagonal will be mapped to sets above and vice versa; and sets below the second diagonal will stay below. Additionally, if a set $\mathcal{S}$ has both positive and negative x- and y- steps, this is preserved by ϕ. So ϕ respects the conditions imposed on our step sets, and is thus a bijection on all permissible step sets according to the restrictions above. Therefore, in order to get the cases which are different even after taking into account symmetry, we can simply count the symmetric permissible step sets, and divide the remainder by 2.

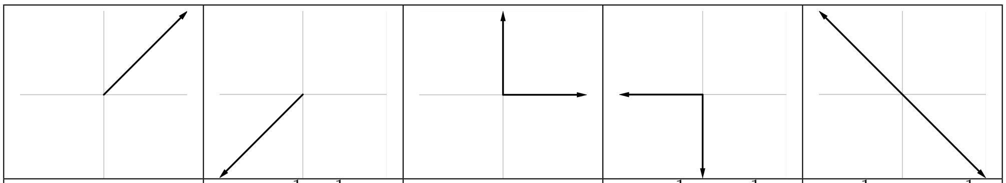
There are 5 parts a symmetric step set can consist of: $\{xy, x^{-1}y^{-1}, x+y, x^{-1}+y^{-1}, x^{-1}y+xy^{-1}\}$, as depicted in table 3. This yields a first counting polynomial

$$S(z) = (1+z)^2(1+z^2)^3. \tag{86}$$

The ones not having both positive and negative x- and y-steps are the empty step set, each of our 5 parts except $\{x^{-1}y + xy^{-1}\}$ on its own, as well as $\{\{\mathbf{NE,\ N,\ W}\}, \{\mathbf{SE,\ S,\ W}\}\}$. This leads us to

$$S_1(z) = S(z) - \left(1 + 2z + 2z^2 + 2z^3\right). \tag{87}$$

[10]In particular, all the cases we discard here were not already discarded in the previous step: in the latter each case included **NE**, which is not permissible here.



xy	$x^{-1}y^{-1}$	$x+y$	$x^{-1}+y^{-1}$	$x^{-1}y+xy^{-1}$
{**NE**}	{**SW**}	{**N,E**}	{**S,W**}	{**NW,SE**}

Table 3: The 5 building blocks of step sets with x, y-symmetry.

Conveniently, the step sets lying above the first diagonal are the same as those lying below, namely those that can be built out of $\{\{\mathbf{NW}\}, \{\mathbf{SE}\}\}$. As they must contain positive and negative steps in both directions, the only step set to discount ist {**NW,SE**}. Therefore, we have

$$S_2(z) = S_1(z) - z^2. \tag{88}$$

Finally, those sets that lie below the second diagonal can be built by $\{\{\mathbf{NW,\ SE}\}, \{\mathbf{S,W}\}, \{\mathbf{SW}\}\}$. They must contain {**NW,SE**}, or else they would have already been discounted, leading to

$$S_3(z) = S_2(z) - z^2(1+z)(1+z^2). \tag{89}$$

By counting coefficients, we arive at 20 permissible symmetric step sets. Consequently, the total number of cases of interest is given by $\frac{118}{2} + 20 = 79$. A counting polynomial for these cases is given by

$$\frac{1}{2}\left(P_3(z) + S_3(z)\right) = z^8 + 5z^7 + 16z^6 + 27z^5 + 23z^4 + 7z^3. \tag{90}$$

It turns out that symmetry is also useful when studying the group of these walks. We already argued that if there is an x, y-symmetry, then the corresponding walks are the same. If we just look at the groups generated by step sets, we can derive a stronger property:

Lemma 2.2.1 ([24, Lemma 2])**.** *Let $\mathcal{S}_1, \mathcal{S}_2$ be two step sets, and σ be an element of the symmetry group of the square D_4 such that $\sigma(\mathcal{S}_1) = \mathcal{S}_2$.*
Then $G(\mathcal{S}_1), G(\mathcal{S}_2)$ are isomorphic.

Proof. It suffices to show the statement for σ a generator of D_4. As D_4 is spanned by reflections along a vertical line and the first diagonal, we are left with two cases.

1. σ is the reflection along a vertical line.
   Then we have $\sigma(\mathcal{S}(x,y)) = \mathcal{S}(x^{-1}, y)$. This leads to

$$\mathcal{S}^+_{2,x}(y) = \mathcal{S}^-_{1,x}(y), \qquad \mathcal{S}^+_{2,y}(x) = \mathcal{S}^+_{1,y}(x^{-1}), \tag{91}$$

$$\mathcal{S}^-_{2,x}(y) = \mathcal{S}^+_{1,x}(y), \qquad \mathcal{S}^-_{2,y}(x) = \mathcal{S}^-_{1,y}(x^{-1}). \tag{92}$$

   Let δ be the transformation $(x,y) \mapsto (x^{-1}, y)$. Then, using the above, we have

$$(\delta \circ \Psi_2 \circ \delta)(x,y) = \delta\left(\frac{1}{x}, \frac{\mathcal{S}^-_{2,y}\left(\frac{1}{x}\right)}{\mathcal{S}^+_{2,y}\left(\frac{1}{x}\right)}\frac{1}{y}\right) = \delta\left(\frac{1}{x}, \frac{\mathcal{S}^-_{1,y}(x)}{\mathcal{S}^+_{1,y}(x)}\frac{1}{y}\right) = \Psi_1(x,y), \tag{93}$$

$$(\delta \circ \Phi_2 \circ \delta)(x,y) = \delta\left(\frac{\mathcal{S}^-_{2,x}(y)}{\mathcal{S}^+_{2,x}(y)}x, y\right) = \delta\left(\frac{\mathcal{S}^+_{1,x}(y)}{\mathcal{S}^-_{1,x}(y)}x, y\right) = \Phi_1(x,y). \tag{94}$$

   As $\delta \circ \delta = \mathrm{Id}$, this gives an isomorphism.

2. σ is the reflection along the first diagonal.
In this case, we have $\sigma\left(\mathcal{S}(x,y)\right) = \mathcal{S}(y,x)$, that is, σ swaps the coordinates. Similarly as before, we get

$$\mathcal{S}_{2,x}^{+}(y) = \mathcal{S}_{1,y}^{+}(y), \qquad \mathcal{S}_{2,y}^{+}(x) = \mathcal{S}_{1,x}^{+}(x), \tag{95}$$

$$\mathcal{S}_{2,x}^{-}(y) = \mathcal{S}_{1,y}^{-}(y), \qquad \mathcal{S}_{2,y}^{-}(x) = \mathcal{S}_{1,x}^{-}(x). \tag{96}$$

With $\delta : (x,y) \mapsto (y,x)$, we have

$$\left(\delta \circ \Psi_2 \circ \delta\right)(x,y) = \delta\left(y, \frac{\mathcal{S}_{2,y}^{-}(y)}{\mathcal{S}_{2,y}^{+}(y)} \frac{1}{x}\right) = \left(\frac{\mathcal{S}_{1,x}^{-}(y)}{\mathcal{S}_{1,x}^{+}(y)} \frac{1}{x}, y\right) = \Psi_1(x,y), \tag{97}$$

$$\left(\delta \circ \Phi_2 \circ \delta\right)(x,y) = \delta\left(\frac{\mathcal{S}_{2,x}^{-}(x)}{\mathcal{S}_{2,x}^{+}(x)} \frac{1}{y}, x\right) = \left(x, \frac{\mathcal{S}_{1,y}^{-}(x)}{\mathcal{S}_{1,y}^{+}(x)} \frac{1}{y}\right) = \Phi_1(x,y), \tag{98}$$

and in the same manner as before we have constructed an isomorphism.

□

This lemma will prove very useful as we continue to classify our walks by the order of their groups.

Theorem 2.2.1 ([24, Theorem 3])**.** *Of the 79 quarter plane walks, there are precisely 23 with a finite group, as depicted in Table 5 on page 80.*

Proof.
The finite groups:
To prove finiteness of a group turns out to be fairly simple in this case: one merely calculates iterates of $\Theta := \Psi \circ \Phi$ until one arrives back at the identity. In all these cases, the group is not only finite but reasonably small; there are only 2 sets $\mathcal{S}$ with order 8, 5 with order 6, and 16 with order 4. This has already been worked out for two cases in Example 2.1 on page 22, a table with all the other examples and corresponding groups can be found in [24, Table 8].
The infinite groups:
An approach as above is clearly doomed to fail in case of an infinite group. Instead, there are two different arguments that can be used; the first being in essence a divisibility argument, and the second working with fixed points and eigenvalues of the Jacobian.

1. **The evaluation method:**
This method works on the last 5 step sets depicted in Table 6 on page 81[11].
The underlying idea is to consider some valuation on the rational polynomials in x, y, and then find some evaluation which exhibits some kind of strictly monotonous behaviour under $\Theta := \Psi \circ \Phi$. If we succeed, then clearly the group cannot be finite.
To do so, we can assume that x, y are rational functions of a third variable z, that is, $x = f_1(z), y = f_2(z)$. By definition, the images of (x,y) under Θ^n are rational polynomials in x, y, so they lie in $\mathbb{C}(x,y) \subseteq \mathbb{C}(z)$. We can now consider the valuation ν in the rational function field $\mathbb{C}(z)/\mathbb{C}$ with respect to the ideal $z\mathbb{C}[z] \subset \mathbb{C}[z]$ (see eg [29, 1]). Letting $\nu(u,v) := (\nu(u), \nu(v))$, we have

$$\nu\left(\Phi(x,y)\right) = \left(\nu\left(\frac{\mathcal{S}_x^{-}(y)}{\mathcal{S}_x^{+}(y)} x^{-1}\right), \nu(y)\right) = \left(\nu\left(S_x^{-}(y)\right) - \nu\left(\mathcal{S}_x^{+}(y)\right) - \nu(x), \nu(y)\right), \tag{99}$$

$$\nu\left(\Psi(x,y)\right) = \left(\nu(x), \nu\left(\frac{\mathcal{S}_y^{-}(x)}{\mathcal{S}_y^{+}(x)} y^{-1}\right)\right) = \left(\nu(x), \nu\left(S_y^{-}(x)\right) - \nu\left(\mathcal{S}_y^{+}(x)\right) - \nu(y)\right). \tag{100}$$

[11] Interestingly, these step sets make up precisely for the singular walks.

We therefore need to compute $\nu(S^{+/-}_{x/y})$. Conveniently, in all 5 cases considered here we have $\mathcal{S}_x^-(y) = y, \mathcal{S}_y^-(x) = x$, so that part is rather simple.
As For $\mathcal{S}_x^+(y)$, note that in each case it contains y^{-1}. By the strict triangle inequality, we have $\nu(y^{-1} + y + 1) = \nu(y^{-1} + y) = \nu(y^{-1} + 1) = \nu(y^{-1})$ and thus $\nu\left(\mathcal{S}_x^+(y)\right) = \nu(y^{-1})$ for $\nu(y) > 0$, as $\nu(y) > \nu(1) > \nu(y^{-1})$. Similarly, we have $\nu\left(\mathcal{S}_x^+(y)\right) = \nu(y)$ for $\nu(y) < 0$.
$\mathcal{S}_y^+(x)$ also contains x^{-1} in each case, so we obtain

$$\nu\left(\mathcal{S}_x^+(y)\right) = \begin{cases} -\nu(y) & \text{if } \nu(y) > 0, \\ \nu(y) & \text{if } \nu(y) < 0, \end{cases} \tag{101}$$

$$\nu\left(\mathcal{S}_y^+(x)\right) = \begin{cases} -\nu(x) & \text{if } \nu(x) > 0, \\ \nu(x) & \text{if } \nu(x) < 0. \end{cases} \tag{102}$$

Thus, given $\nu(x) = a, \nu(y) = b$, we know that Φ, Ψ transform the valuations as follows:

$$(a, b) \overset{\Phi}{\mapsto} \begin{cases} (-a + 2b, b) & \text{if } b > 0, \\ (-a, b) & \text{if } b < 0, \end{cases} \tag{103}$$

$$(a, b) \overset{\Psi}{\mapsto} \begin{cases} (a, 2a - b) & \text{if } a > 0, \\ (a, -b) & \text{if } a < 0. \end{cases} \tag{104}$$

So all we need to do is to find $(a, b) \in \mathbb{Z}^2$ (allowing us to let $x = z^a, y = z^b$) such that the orbit under $\Phi \circ \Psi$ is infinite, and it turns out $(1, 2)$ has this property: by induction, it is easily shown that

$$\Theta^n(1, 2) = (2n + 1, 2n + 2), \tag{105}$$
$$(\Phi \circ \Theta^n)(1, 2) = (2n + 3, 2n + 2). \tag{106}$$

As a consequence, we know the groups corresponding to these step sets to be infinite.

2. **The fixed point method:**
Assume Θ has a fixed point P, and is well-defined on an open set containing P. Then, using the Taylor expansion, we have for u, v sufficiently small:

$$\Theta(P + (u, v)) = \Theta(P) + (u, v)\mathcal{J}_P + \mathcal{O}(uv) + \mathcal{O}(u^2) + \mathcal{O}(v^2), \tag{107}$$

where $\mathcal{J}_P$ denotes the Jacobian evaluated at P. As P is a fixed point, the first term on the right is the same as P. We can iterate this, leading to

$$\Theta^n(P + (u, v)) = P + (u, v)\mathcal{J}_P^n + \mathcal{O}(uv) + \mathcal{O}(u^2) + \mathcal{O}(v^2), \tag{108}$$

and if the order of Θ were finite, then we would for some $n > 0$ have

$$P + (u, v) = P + (u, v)\mathcal{J}_P^n + \mathcal{O}(uv) + \mathcal{O}(u^2) + \mathcal{O}(v^2) \tag{109}$$

for all u, v in a neighbourhood of P. For this to be the case, we would need $\mathcal{J}_P^n$ to be the identity matrix, because otherwise we could choose u, v sufficiently small that the quadratic terms cease to matter and arrive at a contradiction. Therefore, we know that if the order of Θ is n, then $\mathcal{J}_P^n = \mathrm{Id}$. This, however, would mean that any eigenvalue of $\mathcal{J}_P$ is an n-th root of unity. Thus we have a strategy for our proof: we merely need to show that $\mathcal{J}_P$ has no eigenvalues that are roots of unity.
In order to compute the eigenvalues, we can consider the characteristic polynomial $\chi(x) := \det|\mathcal{J}_P - x \cdot \mathrm{Id}|$. To illustrate this, consider the following example:

Example 1.
Consider the walk with the step set $\mathcal{S} = x + y + xy + x^{-1} + x^{-1}y^{-1}$.
We have

$$\Phi : (x, y) \mapsto \left(\frac{1}{xy}, y\right), \tag{110}$$

$$\Psi : (x, y) \mapsto \left(x, \frac{1}{x(x+1)} y^{-1}\right), \tag{111}$$

and thus we have

$$\Theta = \Psi \circ \Phi : (x, y) \mapsto \left(\frac{1}{xy}, \frac{x^2 y}{1 + xy}\right). \tag{112}$$

First we need to determine a fixed point of Θ. Let $x = a$ be an arbitrary parameter. Then we must have $\frac{1}{xy} = \frac{1}{ay} = a$, meaning $y = a^{-2}$. Plugging this into the second coordinate yields $a^3 - a - 1 = 0$.
Next, consider the Jacobian. We have

$$\mathcal{J}_{(a,a^{-2})} = \begin{pmatrix} -\frac{1}{x^2 y} & -\frac{1}{xy^2} \\ \frac{xy(xy+2)}{(xy+1)^2} & \frac{x^2}{(xy+1)^2} \end{pmatrix}_{x=a, y=a^{-2}} = \begin{pmatrix} -1 & -a^3 \\ \frac{1+2a^2}{a^6} & \frac{a^4}{1+a^2} \end{pmatrix}. \tag{113}$$

The characteristic polynomial is given by

$$\chi(x) = -a^7(x+1) + 2a^5(x^2 + x) + 4a^4 + a^3(x^2 + x) + 4a^2 + 1. \tag{114}$$

In order to check its roots, we eliminate a using the condition $a^3 - a - 1 = 0$.
By computing $a^7 = 1 + 2a + 2a^2, a^5 = 1 + a + a^2, a^4 = a + a^2, a^3 = 1 + a$, we can rewrite this polynomial as

$$a^2 f_2(x) + a f_1(x) = f_0(x), \tag{115}$$

and by squaring we obtain

$$a^4 f_2(x)^2 + 2a^3 f_1(x) f_2(x) + a^2 f_1(x)^2 = f_0(x)^2 \quad \Leftrightarrow \tag{116}$$

$$a^2\left(f_2(x)^2 + f_1(x)^2\right) + a\left(f_2(x)^2 + 2 f_1(x) f_2(x)\right) = f_0(x)^2 - f_1(x) f_2(x). \tag{117}$$

Using (115) to eliminate a^2 by substituting $\frac{f_0(x) - a f_1(x)}{f_2(x)}$ and eliminating the denominators, we arrive at an equation of the form

$$a g(x) = h(x) \quad \Leftrightarrow \tag{118}$$

$$a = \frac{h(x)}{g(x)}. \tag{119}$$

Using $a^3 - a - 1 = 0$, we can finally transform this into a polynomial in $\mathbb{Q}[x]$, which contains every root of $\chi(x)$. This polynomial itself is of degree 18 and somewhat unwieldy, but by comparing it to cyclotomic polynomials of smaller degree one quickly finds that $\mathcal{J}_{(a,a^2)}$ has no eigenvalues that are roots of unity.

A list of the different step sets to consider (here, Lemma 2.2.1 on page 26 turns out to be very useful in reducing the number of cases) as well as the corresponding fixed points and resulting polynomials is given in [24, Table 5].

As one of these two methods is applicable for each of the 56 remaining walks (see again [24, Theorem 3]), this concludes the proof.

□

2.3 Finite Group Implies Holonomy

The aim of this section is to prove that the generating function of all those walks with finite group order is holonomic. Therefore, in the following we will always assume that $\mathcal{S}$ is a step set such that $|G(\mathcal{S})| = 2n < \infty$.
First, we revisit our kernel equation (74)

$$F(x,y;t)\mathcal{K}(x,y;t) = 1 - tF(x,0;t)\mathcal{S}_y(x,y) - tF(0,y;t)\mathcal{S}_x(x,y) + t\varepsilon x^{-1}y^{-1}F(0,0;t). \tag{120}$$

Note that, to keep things short, we write here $\mathcal{S}_y$ and S_x instead of $\mathcal{S}_y^-, \mathcal{S}_x^-$ as in the previous section, because there will not need $\mathcal{S}_{x/y}^{+/0}$. Consequently, we have $\mathcal{S}_y(x,y) = y^{-1}\mathcal{S}_y(x)$, and $\mathcal{S}_x(x,y) = x^{-1}\mathcal{S}_x(y)$. Multiplying the previous equation by xy, this gives us

$$xy\mathcal{K}(x,y;t)F(x,y;t) = xy - tx\mathcal{S}_y(x)F(x,0;t) - ty\mathcal{S}_x(y)F(0,y;t) + t\varepsilon F(0,0;t). \tag{121}$$

We can additionally shorten this by setting $A(x) := tx\mathcal{S}_y(x)F(x,0;t), B(y) = ty\mathcal{S}_x(y)F(0,y;t)$, and ignore the remaining ts, arriving at

$$xy\mathcal{K}(x,y)F(x,y) = xy - A(x) - B(y) + \varepsilon F(0,0). \tag{122}$$

Now, remember Example 2.1 on page 21, which was in fact our main motivation for looking at the transformations Φ, Ψ and the group of a walk. Consider now the orbit of (x,y) under repeated application of Φ, Ψ. Let this orbit be

$$(x,y) \overset{\Phi}{\mapsto} (x_1,y) \overset{\Psi}{\mapsto} (x_1,y_1) \overset{\Psi}{\mapsto} (x_2,y_1) \overset{\Psi}{\mapsto} \cdots \overset{\Phi}{\mapsto} (x_n,y_{n-1}) \overset{\Psi}{\mapsto} (x_n,y_n) = (x,y). \tag{123}$$

Note in particular that, since $G(\mathcal{S})$ has order $2n$, the last transformation will always be Ψ, and that, according to Lemma 2.1.1 on page 22, this orbit includes all elements $g \in G(\mathcal{S})$. According to the same Lemma, the following definition makes sense:

Definition 2.3.1. *Let $g \in G = G(\mathcal{S})$. We define the* ***signum*** *of g to be*

$$\operatorname{sgn}(g) := \begin{cases} 1 & \text{if } g = \Theta^k, \\ -1 & \text{if } g = \Phi \circ \Theta^k. \end{cases} \tag{124}$$

We can then, as we did in Example 2.1, take alternating sums over the elements of the orbit and hope for something nice to happen. To abbreviate notation, let

$$g_i = \begin{cases} \Phi \circ \Theta^{(i-1)/2} & \text{if } i \text{ is odd}, \\ \Theta^{i/2} & \text{if } i \text{ is even}. \end{cases} \tag{125}$$

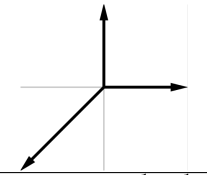
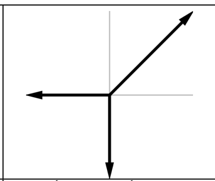
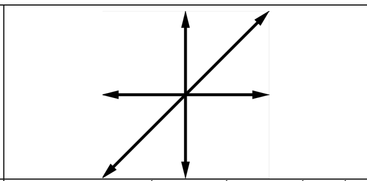
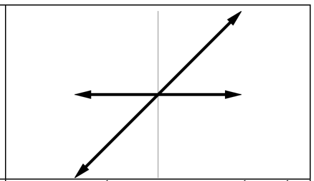
			
$x+y+x^{-1}y^{-1}$	$x^{-1}+y^{-1}+xy$	$x+y+x^{-1}+y^{-1}+x^{-1}y^{-1}$	$x+x^{-1}+xy+x^{-1}y^{-1}$

Table 4: The 4 exceptional step sets $\mathcal{S}$ where the orbit sum method does not work. For the first 3, using half orbit sums instead works out, the last one (the Gessel walk) requires a different approach.

We then have $g_i(x,y) = (x_{[(i+1)/2]}, y_{[i/2]})$. When looking at (123), it is quite noticeable that each x_i, y_i appears twice[12] in the orbit, and always consecutively. Clearly, the signum of two consecutive elements in the orbit sums to 0. So, if we substitute $(x,y) \mapsto g_i(x,y)$ into (121), this leads to

$$\mathcal{K}(x,y)\left(\sum_{i=1}^{2n} \operatorname{sgn}(g_i) g_i(x,y) F(g_i(x,y))\right) \tag{126}$$

$$= \sum_{i=1}^{2n} g_i(xy) - \sum_{i=1}^{2n} \operatorname{sgn}(g_i)\left(A(g_i(x)) + B(g_i(x))\right) + \varepsilon F(0,0)\left(\sum_{i=1}^{2n} \operatorname{sgn}(g_i)\right). \tag{127}$$

By the above remark, we know that the second and third sum on the right hand side are 0. We have therefore proven:

Lemma 2.3.1 ([24, Prop. 5 (Orbit sums)]). *If $G(\mathcal{S})$ is finite, then*

$$\mathcal{K}(x,y)\left(\sum_{g\in G} \operatorname{sgn}(g) g(xy) F(g(x,y))\right) = \sum_{g\in G} g(xy). \tag{128}$$

2.3.1 Holonomy using Orbit Sums

The previous Lemma enables us to prove holonomy for most of the walks with finite group.

Theorem 2.3.1 ([24, Prop. 8]). *Let $\mathcal{S}$ be a step set with finite group, and different from the 4 exceptional cases shown in Table 4.*
Then the generating function $F(x,y;t)$ is holonomic.

Proof.
First, consider $\mathcal{S} = x + x^{-1} + y^{-1} + xy + x^{-1}y$. Its group is of order 4, the orbit of (x,y) is given by

$$(x,y) \overset{\Phi}{\mapsto} (x^{-1}, y) \overset{\Psi}{\mapsto} \left(x^{-1}, \frac{1}{x+x^{-1}}y^{-1}\right) \overset{\Phi}{\mapsto} \left(x, \frac{1}{x+x^{-1}}y^{-1}\right)^{13}. \tag{129}$$

[12]Strictly speaking, it may be more often, as x_i may be equal to x_j even for $i \neq j$. But since then both x_i, x_j would appear twice consecutively, this does not change anything.

We can therefore apply Lemma 2.3.1 and obtain

$$xyF(x,y) - x^{-1}F(x^{-1},y) + \frac{x^{-1}y^{-1}}{x+x^{-1}}F\left(x^{-1}, \frac{1}{x+x^{-1}}y^{-1}\right) - \frac{xy^{-1}}{x+x^{-1}}F\left(x, \frac{1}{x+x^{-1}}y^{-1}\right) \tag{130}$$

$$= \frac{xy - x^{-1}y + \frac{x^{-1}y^{-1}}{x+x^{-1}} - \frac{xy^{-1}}{x+x^{-1}}}{\mathcal{K}(x,y)}. \tag{131}$$

As F is a power series in x, y, t, both sides of the equation certainly are power series in x, y, x^{-1}, y^{-1} (note that we can expand $\frac{1}{x+x^{-1}}$, leading to a valid substitution). However, we do know that $F(x,y)$ can only contain positive powers of y, as it is a power series, as opposed to a Laurent series. Only the first two summands on the left hand side contain positive powers of y at all, and in fact, they contain only positive powers of y, so we have

$$xyF(x,y) - x^{-1}F(x^{-1},y) = [y^{>0}]\frac{xy - x^{-1}y + \frac{x^{-1}y^{-1}}{x+x^{-1}} - \frac{xy^{-1}}{x+x^{-1}}}{\mathcal{K}(x,y)}. \tag{132}$$

We can now utilize the same argument for x, and arrive at

$$xyF(x,y) = [x^{>0}][y^{>0}]\frac{xy - x^{-1}y + \frac{x^{-1}y^{-1}}{x+x^{-1}} - \frac{xy^{-1}}{x+x^{-1}}}{\mathcal{K}(x,y)}. \tag{133}$$

It remains to show that this expression is holonomic. This follows eg from [30, p. 377], or [24, Prop. 1]. The necessary condition here is that we take the positive terms of a rational power series; if we only looked at the positive x- or y-terms, the result would still be algebraic, if we want both simultaneously, we stay holonomic.
It turns out that for the different walks with finite group, apart from the mentioned exceptional cases, the very same strategy is successful. In particular, in each of these cases, each of the elements in the orbit except (x,y) contains only negative powers of either x or y. Thus, by a similar argument as above only $xyF(x,y)$ remains on the left hand side. See also [24, Prop. 8]. □

2.3.2 Holonomy using Half Orbit Sums

In the previous section, the approach via orbit sums worked in particular because the only element of the orbit that was "nice enough" in the sense that something remained after extracting first positive coefficients of y, and then of x, was $F(x,y)$. If we look at the first 3 step sets in Table 4 on the preceding page, and the corresponding orbits in 5 on page 80, we can immediately deduce that summing over the whole orbit cannot possibly work, as we have $\Theta^3(x,y) = (y,x)$, which clearly cannot be eliminated in the same fashion as before. What we can do instead, however, is to instead only sum over half the orbit, and try to work from there. Analogously to Lemma 2.3.1, taking into account the non-cancelling boundary terms, we arrive at

$$xyF(x,y) - x^{-1}F(x^{-1}y^{-1},y) + y^{-1}F(x^{-1}y^{-1},x) \tag{134}$$

$$= \frac{1}{\mathcal{K}(x,y)}\left(xy - x^{-1} + y^{-1} - tx\mathcal{S}_y(x)F(x,0;t) - ty\mathcal{S}_x(y)F(y,0;t) + \varepsilon F(0,0;t)\right). \tag{135}$$

As all 3 walks considered here are symmetric in x, y, we know that $\mathcal{S}_x\mathcal{S}_y$, and $F(x,0;t) = F(0,x;t)$, so the latter simplifies to

$$xyF(x,y) - x^{-1}F(x^{-1}y^{-1},y) + y^{-1}F(x^{-1}y^{-1},x) \tag{136}$$

$$= \frac{1}{\mathcal{K}(x,y)}\left(xy - x^{-1} + y^{-1} - 2tx\mathcal{S}_y(x)F(x,0;t) + \varepsilon F(0,0;t)\right). \tag{137}$$

We can now extract the coefficients of y^0. For the left hand side, this is fairly straightforward: clearly, the first and third summands cannot contribute. The second one is relevant only at those points where the powers of y cancel, that is, if the power of $x^{-1}y^{-1}$ is the same as that of y. This means that the relevant coefficients are those of the walks ending on the diagonal $\{x = y\}$. These walks induce another generating function, which we will denote by $F_d(x;t)$. As a result, we have

$$[y^0]\left(xyF(x,y) - x^{-1}F(x^{-1}y^{-1},y) + y^{-1}F(x^{-1}y^{-1},x)\right) = -x^{-1}F_d(x^{-1};t). \tag{138}$$

Extracting the coefficient on the right hand side is a bit more tricky, as we have the term $\frac{1}{\mathcal{K}(x,y)}$. Things get a bit more technical here, but by what essentially boils down to a partial fraction decomposition, this can be done. The term appearing in the denominator, $\triangle(x)$, is defined by the discriminant of $\mathcal{K}(x,y)$ viewed as a quadratic polynomial in y, and will prominently feature in later sections. One can look at the zeros of the discriminant, and view them as a function of t, which can then be expressed as a Puiseux-series. By sorting the zeros corresponding whether the Puiseux-series has positive, negative or zero valuation in t, one can then factor the discriminant $\triangle(x) = \triangle_0\triangle^+(x)\triangle^-(x)$. As this procedure is quite technical and will not be used except here, this very rough outline shall suffice here. For details see [24, Lemma 7, 4.3-4.4].
Following through with all of this eventually leads to (see [24, 6]):

$$\sqrt{\triangle^-(x^{-1})}\left(\frac{x}{t} - x^{-1}\mathcal{S}_y^+(x)F_d(x^{-1})\right) \tag{139}$$

$$= \frac{1}{\sqrt{\triangle_0\triangle^+(x)}}\left(\frac{x(1-t\mathcal{S}_y^0(x))}{t} - x^{-1}\mathcal{S}^+(x) - 2tx\mathcal{S}_y^-(x)\mathcal{S}_y^+(x)F(x,0) + t\varepsilon\mathcal{S}_y^+(x)F(0,0)\right). \tag{140}$$

This equation turns out to be a lot more harmless than it might look at first glance, as the only power series remaining now (excepting the discriminants, which we can compute) are $F_d(x^{-1})$ on the left hand side, and $F(x,0)$ on the right hand side. In particular, the former contributes negative, the latter positive powers of x. Therefore, in all three cases it suffices to compute $\triangle^0, \triangle^+(x), \triangle^-(x)$ (which are always algebraic), and then extract the positive and negative coefficients of x in (139). This is done in detail, giving explicit representations as well as additional combinatorial identities, in [24, 6]. As a result, we have

Theorem 2.3.2 ([24, Prop. 13-15])**.** *Let $\mathcal{S}$ be one of the first three of the exceptional step sets with finite group as depicted in table 4.*
Then the generating function $F(x,y;t)$ is algebraic.

2.3.3 The Gessel Walk

The last of the exceptional step sets among those with finite group is the so-called *Gessel-walk*. According to [25], I. Gessel conjectured in an unpublished work, based on empirical observations, that the generating function counting those walks in the quarter plane with the corresponding step set is not just D-finite, but in fact even a hypergeometric series. This so-called Gessel conjecture has since been proven. Also, unlike all the other walks with a finite group, it could not be solved comparably easily using orbit or half orbit summation methods. While there are, by now, other proofs, such as eg [26], utilizing techniques similar to the analytic method in later sections, the method used to first show that the generating function of the Gessel walk is D-finite (and, in fact, even algebraic) in [25] relied heavily on computer algebra.

The general idea behind the proof, albeit simplified a lot, is to first consider the reduced kernel equation, which will also be the starting point in the next sections, i.e. to restrict oneself to the curve where the kernel vanishes. There, we can find a representation of y as a power series in x, t. Using this, all that is left over are the boundary terms $F(x, 0; t), F(0, y; t)$. One can then explicitly compute some coefficients of these series, and, using the reduced kernel equation as well as automated guessing, arrive at a potential holonomic or algebraic relation.

From there it remains to show that the candidate for the equation admits only one solution, and that this is fulfilled by $F(x, 0; t)$. Then, if we know both our boundary terms to be algebraic, it will follow that $F(x, y; t)$ is also algebraic.

It should be emphasized here that there is still a fair amount of work to be done and arguments to be made. However, in this case it also becomes very clear that the computer is an absolutely vital in this case, as the necessary computations far exceed the abilities of any human: according to [25, 3.1], the difference-operator which was guessed to to prove holonomy would, printed out, fill 500 pages of paper, and the minimal polynomial is even way beyond this.

This method stands in stark contrast to the ones previously mentioned, as in this case, we did not only prove the existence of some holonomy equation, but rather explicitly computed it. While they can sometimes be easily derived, as for the first three of the exceptional walks in [24], this is not the case for all of them. In particular, it is remarkable that such an absurdly complicated relation as for the Gessel walk has since been proven to exist entirely without the use of computer algebra in [26]. It is a moot point to argue whether a proof using computer algebra but leading to explicit results or a human proof merely showing the existence of a relation is in any way "nicer", but either of them are clearly worthy of note on their own, complementing each other.

2.4 Non-Holonomy using Iterated Orbit Sums

While it may seem that orbit summation methods are inherently built towards proving that some generating function is holonomic, this turns out to be not to be the case. In a paper by M. Mishna and A. Rechnitzer [27], what the authors call iterated orbit summation serves as a tool to prove that two of the walks with an infinite group have a non-holonomic generating function. Here, the idea will only be outlined very roughly, but as a very creative application of what we have done so far, this method definitely deserves notice.

The two step sets in question are given by the step sets $\mathcal{S} = \{\mathbf{NW,NE,SE}\}$, $\mathcal{T} = \{\mathbf{NW,N,SE}\}$. We will first sketch the idea to prove the statement for $\mathcal{S}$, as the x, y-symmetry makes some things a bit more intuitive. In this case, we have the kernel equation

$$(xy - tx^2y^2 - tx^2 - ty^2)F(x, y; t) = xy - tx^2F(x, 0; t) - ty^2F(0, x; t). \tag{141}$$

We can now once again consider $\mathcal{K}(x)$ as a polynomial in y, and get two solutions $y_{1,2}(x, t)$. In particular, one can see that one of these, let us just call it $y(x, t)$ from now on, is a power series in t without constant term – meaning we can plug it into another power series in t. For brevity's sake, in the following we will leave out the dependence on t in the corresponding functions. The idea now is that, as $\mathcal{K}(x, y(x)) = 0$, and by symmetry $F(x, 0) = F(0, x)$, we have

$$0 = xy(x) - tx^2 F(x, 0) - ty(x)^2 F(y(x), 0), \tag{142}$$

which reduces to

$$F(x, 0) = \frac{y(x)}{tx} - \frac{y(x)^2}{x^2} F(y(x), 0). \tag{143}$$

By iterating this identity[14], one can derive the formal representation

$$F(x, 0) = \frac{1}{x^2 t} \sum_{n=0}^{\infty} (-1)^n y^n(x) y^{n+1}(x), \tag{144}$$

where $y^n(x)$ denotes n-fold application of the mapping $x \mapsto y(x)$. By applying the functional equation, one can therefore obtain a formula for $F(1, 1)$ of the form

$$F(1, 1) = \frac{1 - 2tF(1, 0)}{1 - 3t}. \tag{145}$$

From here, all that remains to do is to show that this power series in t has an infinite amount of singularities, but is convergent elsewhere. This is done in [27, 2.3-2.4], and while not at all trivial, is not the focus here. The procedure for $\mathcal{T}$ works along much the same lines. The starting point is a bit more involved, as we do not have and thus cannot utilize an x, y-symmetry, so two recursions need to be considered separately (see [27, 3.1-3.2]).
The authors remark that in the finite group case this method breaks down due to the iterates repeating, i.e. one gets only a finite sum [27, 4.1]. The other direction, in a sense, however, is not entirely clear – an infinite group does not trivially imply an infinite orbit of $x \mapsto y(x)$. While a very interesting method to show non-holonomy and certainly a useful stretagy to be aware of, to the author's knowledge, this method has not yet been successfully utilized in order to show non-holonomy for the remaining walks with infinite group. We will therefore in the next section introduce some tools required for two known proofs that do positively answer the question about the non-holonomy of the walks with infinite group.

[14] thereby summing over the orbit of $x \mapsto y(x)$; hence the name *iterated orbit sum*

3 Geometric Prerequisites

If we remember the walks in the half plane, in particular to Example 1.2.3 on page 18, what we did there was in a sense the ideal scenario of what could happen: we were able to use an algebraic substitution in order to get rid of the kernel entirely. As mentioned at the beginning of the previous section, this approach does not work in the same manner for walks in the quarter plane, because we have an additional variable. Interpreting t as a (fixed) parameter for a moment, our solution of the equation $\mathcal{K}(x, y; t) = 0$ will not be a point, but instead a polynomial equation in x, y. However, while this is not intuitively as nice as in the one-dimensional case, there is a fair amount of theory concerning the solution of such equations. It is therefore not very far-fetched to hope that considering the structure of such solutions will lead to some insights about the generating function. In a somewhat more general setting, such methods were proposed by G. Fayolle, R. Iasnogorodski and V. Malyshev in [31], and they are applied to our case in [28, 32]. In particular, the non-holonomy of all the walks with infinite group will be shown, based on works of I. Kurkova and K. Raschel [28] and T. Dreyfus, C. Hardouin, J. Roques and M. Singer [32]. In both of these papers, the first step is to consider the Riemann surface associated with the kernel, which will (in most cases) lead to an elliptic curve. We can then lift the group of the walk to automorphisms on this curve, which enables us to use the very rich theory of such curves. It is particularly interesting, that from then on both works differ from how they approach such a curve. In [28], most of the work is done on $\mathbb{C}/\mathbb{Z}\omega_1 \times \mathbb{Z}\omega_2$, making use of explicit computations of the automorphisms. In [32], the emphasis lies more on the algebraic structure of the curve. Using methods from Galois theory of difference equations then yields results.
As it may not be immediately obvious how the kernel equation yields a Riemann surface, and how the mappings can be lifted, the beginning of this section will try to give an outline of the ideas used therein. Some of the more technical concepts can only be outlined here, and not all proofs can be given, but the aim is to make what happens seem at least somewhat intuitive and natural.

3.1 Covering Spaces, Riemann Surfaces

When trying to extend the logarithm to the complex plane $\mathbb{C}$, a rather well-known problem arises. Let us restrict ourselves to the unit circle $S := \{z \in \mathbb{C} : |z| = 1\}$ – if we have a continuation there, all that remains to be done is to scale the absolute value by the real part, where we already have a well-defined logarithm. Now, suppose we start at the point $p = 1$. A rather natural choice would be to go ahead and define $\log p = 0$, as, clearly, this satisfies $\exp(\log p) = \exp(0) = 1 = p$. We can then extend this logarithm along the unit circle, say, in a counter-clockwise direction[15], leading to $\log(e^{ti}) = ti$, which works marvellously until we arrive back at our original point p. Here, we would want our logarithm to be 0, but if we just go along with our – very natural – construction, continuity of log would give us $\log p = 2\pi i$. In hindsight, this is not very surprising, simply because exp is $2\pi i$-periodic – what we did was just picking a suitable pre-image, but at some point we transitioned from one fiber into another. This would leave us with a multi-valued function, which is not quite what we set out to do, but the best we can do on $\mathbb{C}$.
However, there is another way to think about this. We could pretend that, once we arrive back at our starting point, we have no conflicting value, we just *move onto another copy of* $\mathbb{C}$ instead. So, instead of a logarithm $S \to \mathbb{C}$, we get a function from some other set X to $\mathbb{C}$. This other set should certainly include the unit circle, and the transition should go smoothly in a way. This can

[15]Technically making use of the Implicit Function Theorem and a bit of complex analysis, which however shrouds the fairly intuitive idea.

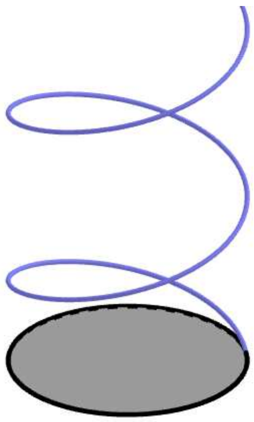


Figure 3: Covering space of the unit circle induced by the logarithm.

be easily achieved by turning the unit circle into an infinite spiral, as shown in Figure 3. Now let $X = \{(\cos t, \sin t, t) : t \in \mathbb{R}\}$ be this spiral, for convenience's sake treated as a subset of $\mathbb{R}^3$. Note that first we have a natural (continuous) map from X back to S, namely just the projection $\pi : X \to S : (x, y, z) \mapsto (x, y) \in \mathbb{R}^2 \cong \mathbb{C}$. But even nicer than that, for each point $q \in X$, we see that, locally, X is the same around q as S around $\pi(q)$. This is very useful to us, because, given any map $f : S \to \mathbb{C}$, we can then define $\widehat{f} : X \to C : \widehat{f}(q) := f(\pi(q))$, and f is continuous, then must $\widehat{f}$ be continuous. In particular, we can define $\widehat{\log} : X \to \mathbb{C} : (\cos t, \sin t, t) \mapsto ti + \left[\frac{t}{2\pi}\right] \cdot 2\pi i$, where $\left[\frac{t}{2\pi}\right]$ denotes the greatest integer less than or equal to $\frac{t}{2\pi}$. It can easily be checked that $\widehat{\log} : X \to \mathbb{C}$ satisfies $\exp \circ \widehat{\log} = \pi$, so it is an inverse of exp modulo the projection π. While $\widehat{\log}$ is not unique, as we can shift it by multiples by $2\pi i$, it becomes so if we fix its value on an arbitrary point, say, $\widehat{\log}(0, 0, 0) = 0$. As a stark contrast to before $\widehat{\log}$ is now a single-valued function on all of S.
Doing a similar construction for all of $\mathbb{C}$ instead of just the unit circle, is usually a very similar process, and works for many more functions than just log:

1. First we want to find a domain D where we certainly stay on a fixed branch of log, no matter which way we go. The most common choice here is $\mathbb{C} \setminus \mathbb{R}^+$ – as long as we do not cross the positive real axis, we cannot complete a circle, and thus cannot jump branches.

2. Second, we take one copy of D_i for each branch – in case of log, an infinite amount of them, so we have D_i for $i \in \mathbb{Z}$ – and see how they fit together. For this, label the lower side of the cut we made into $\mathbb{C}$ by b, and the upper side by a. When we go along any path that passes over the cut, then we switch from one branch to the next. As a consequence, we glue the copies such that a of D_{i+1} is glued together with b of D_i, obtaing a surface S.

The image obtained is very similar to that of figure 3; instead of a line circling upwards we have a ray from (but not containing) 0 parallel to $z = 0$ circling upwards. Again, around any point q, S is locally homeomorphic to $\pi(q) \in \mathbb{C}$, and we can define $\widehat{\log}$ as before.

The surface S we constructed is what is called a **covering** of $\mathbb{C} \setminus \{0\}$:

Definition 3.1.1. *Given a topological space X, a (complex)* ***covering space*** *is a topological space T together with a (holomorphic) continuous, surjective projection map $\pi \colon T \to X$, such that*

for all $p \in X$, there is a neighbourhood U such that $\pi^{-1}(U)$ consists of disjoing homeomorphic copies of U.
*If $q \in \pi^{-1}(p)$, then we say q **lies over** p. The disjoint copies of U in the pre-image are called **layers** of the covering.*

The main use of coverings here is that, as seen previously, we can pull back a continuous mapping f with domain X to a continuous mapping $\widehat{f}$ with domain T. There is, not surprisingly, a bit of theory on covering spaces, as well as their connection with analytic functions, a readable introduction is given eg in [34, ch. 4].

It is reasonable to assume that the construction we have done for the logarithm can be generalized to a broad range of analytic functions, enabling us to assign to each a topological space. And indeed, it turns out that this yields a connection between our kernel and an elliptic curve. To do so, some notation should be introduced. For a more thorough introduction to the following see eg [33] or [34].
We will be working over $\mathbb{C}$, and thus are interested in particular in spaces which locally look like $\mathbb{C}$. As an arbitrary topological space X does not come with a metric or anything similar equipped, the best we can do in this regard is a homeomorphism, which leads to the following definitions:

Definition 3.1.2. *A **complex chart** on a topological space X is an open set $U \subseteq X$ together with a homeomorphism $\phi : U \to V \subset \mathbb{C}$, such that ϕ is a homeomorphism.*

As charts need not be unique, and we want to avoid having everything to be dependent on the particular choice of charts, it makes sense to impose some kind of compatibility on them. In particular, we want changing between two charts to be a holomorphic function.

Definition 3.1.3. *Two complex charts $(U, \phi), (V, \psi)$ are called **compatible**, if $\phi \circ \psi^{-1}|_{\psi(U \cap V)}$, and $\psi \circ \phi^{-1}|_{\phi(U \cap V)}$ are holomorphic.*

Note that it would be sufficient for the purpose of the definition to require only one of the transition maps to be holomorphic, as the other direction follows (**REF**).
As we want our whole topological space to be equipped with such a structure, we define

Definition 3.1.4. *A set of complex charts (U_i, ϕ_i) on X is called a **complex atlas**, if (U_i, ϕ_i) and (U_j, ϕ_j) are compatible for all i, j, and $\bigcup U_i = X$.*
*A maximal complex atlas on is called a **complex structure** on X.*

Definition 3.1.5. *A **Riemann surface** is a topological Hausdorff space X equipped with a complex structure.*[16]

By considering chart expressions, we can extend holomorphy in a natural manner: we call a function f between two Riemann surfaces holomorphic, if the chart expression $\phi \circ f \circ \psi^{-1}$ is holomorphic. As two charts covering the same point have to be compatible, this is independent of the choice of charts.

Example 1.
The projective complex line, $\overline{\mathbb{C}}$, given by $\mathbb{C} \cup \{\infty\}$, or equivalently by $\{(x : y) \in \mathbb{C} \times \mathbb{C} \setminus \{(0,0\}\}/\sim$, where $(x_1 : y_1) \sim (x_2 : y_2) :\Leftrightarrow x_1 y_2 = x_2 y_1$, is a compact Riemann surface, the so-called **Riemann-sphere**. One can consider this space as two copies of $\mathbb{C}$ glued together –

[16] Some, but not all authors require a Riemann surface to have a second-countable topology, i.e. to have a countable basis. This will never be an issue here and is thus omitted.

one, here called $\mathbb{C}_1$, where $y \neq 0$, so we have a projection $\pi_1 : (x : y) \mapsto \frac{x}{y}$, and another one, which we will call $\mathbb{C}_2$ where $x \neq 0$, so we have a second projection $\pi_2 : (x : y) \mapsto \frac{y}{x}$.
We can cover $\overline{\mathbb{C}}$ by two charts: on $\mathbb{C}_1$, we use π_1, on $\mathbb{C}_2$, we use π_2. We need to show that the charts are compatible. For this purpose, let $0 \neq z \in \mathbb{C}$. By definition, $\psi(\mathbb{C}_1 \cap \mathbb{C}_2) = \mathbb{C} \setminus \{0\}$. So let $0 \neq z \in \mathbb{C}$. We have $\psi^{-1}(z) = (1 : z) = \left(\frac{1}{z} : 1\right)$, so $\phi \circ \psi^{-1}(z) = z^{-1}$. As $z \mapsto z^{-1}$ is holomorphic, $\overline{\mathbb{C}}$ is a Riemann surface.
By a stereographic projection, $\overline{\mathbb{C}}$ can be homeomorphically mapped onto the unit sphere in $\mathbb{R}^3$, hence it is compact – and thus also its name. The point at infinity is the image of the pole.

The question remains how to associate with an analytic function a Riemann surface. To this end, we want to consider different analytic continuations at one, lump them together if they are locally the same, and distinguish between them if they yield a different value. A useful tool to do so are function elements.

Definition 3.1.6. *A* ***function element*** *of an analytic function f with a power series expansion around $z \in \mathbb{C}$ is the pair (z, f).*
Let $(z_1, f), (z_2, g)$ be two function elements. We say $(z_1, f) \sim (z_2, g)$, or $(z_1, f), (z_2, g)$ are ***compatible****, if $z_1 = z_2$ and $f = g$ on their common radius of convergence. Clearly, $\sim$ induces an equivalence relation.*

We can then consider all function elements which we can obtain by compatible continuation along some curve. We will denote the set of all such function elements by $\mathcal{M}$. This set can then be equipped with a topology, where essentially all sets are open which have their base points form an open set and their functions being compatible. Around any function element (z, f), clearly f is a complex chart. Therefore, $\mathcal{M}$ is a Riemann surface.
In order to make this work, one still needs to pay attention to some details, a more thorough construction is given eg in [35, ch. 3], [34, ch. 4].[17]

Definition 3.1.7. *The Riemann surface $\mathcal{M}$, constructed in the previous manner from a function element (z_0, f), will be called the* ***Riemann surface associated to*** *f. The associated continuation of f is called the* ***complete analytic function*** *f.*

As for the logarithm above, we clearly have a projection $\pi : \mathcal{M} \to \mathbb{C} : (z, f) \mapsto z$. Since, by definition, the complete analytic function is a single-valued continuation of f on $\mathcal{M}$ (and, in fact, by definition the most general one), this is an extension of what we have done previously.

3.1.1 Riemann Surfaces of $\sqrt{p(x)}$

In particular, we will be interested in the Riemann surface corresponding to functions of the form $\sqrt{p(x)}$, $p(x)$ a polynomial. We proceed as we did for the logarithm: first, we find a domain where the function is single-valued, then we consider all the branches and figure out how to glue them. For ia more technical approach, see eg [34, ch. 4], [35, ch. 1]; for images see eg [35, Fig. 1-1 to 1-6].

[17] This construction can be generalized, leading to the concept of germs (see eg [34, ch. 4]). Essentially, a germ is the same as a function element, but we choose a slightly different equivalence relation on the set of all germs (not only those obtained by continuation along paths), and can use this to build a more general abstract Riemann surface; the one we are interested in is then merely an an equivalence class. Germs may be a bit more general, but also a lot less intuitive, and in our case yield the same.

- First assume that $p(x)$ is linear, so by a coordinate transformation we can assume $p(x) = x$. We have the two branches $\sqrt{x}$ and $-\sqrt{x}$, and as for the logarithm we change branches whenever we complete a turn around the 0. Thus we select the same domain as before, $\mathbb{C} \setminus \mathbb{R}^+$. However, differently from before, there are only 2 layers to cut, not an infinite amount: if we circle around 0 twice, we will have arrived at the original branch. So, again calling the domains D_1, D_2, and labeling the upper and lower parts of the cut a and b respectively, we glue b from D_1 to a from D_2, but now we also glue b from D_2 to a from D_1.
  The resulting surface can be better visualized after compactification: Instead of the complex plane C, we work on the projective line $\mathbb{PC}^1$, which we already know to be homeomorphic to a sphere (see 3.1 on page 38). Technically, what we do here is adding a point lying over the branch points 0 and ∞[18], which we cannot strictly do by meromorphic continuation, but have continuity: for each branch, as $x \to 0$, $\pm\sqrt{x}$ goes to 0 as well, and the same goes for infinity (see eg [34, 4.8]). A cut along the positive real axis corresponds to a cut from one pole to another. Topologically, both cut spheres are the same as a semi-circle, and if we glue two semi-circles together along the edges one arrives at a circle. As a result, we know that the Riemann surface obtained by continuing the function $y = \sqrt{x}$ is homeomorphic to a sphere; in particular it is of genus 1[19].

- If $p(x)$ is of the form $(x - a)(x - b)$, $a \neq b$, then again we need to find a domain where the continuation is independent of the path chosen. The issue here is that we have two branching points – informally, we can write $\sqrt{x-a}\sqrt{x-b}$, so if we go around either branch point once, the expression will change its sign. Consequently, our domain D must not allow us to go around one of the branch points, but not the other. Perhaps the most natural choice for D is therefore the complex plane with a line from a to b removed. Note that, on the Riemann sphere, this looks precisely the same as the previous case[20], and we again obtain a Riemann surface homeomorphic to a sphere.

- Now assume $p(x)$ is of the form $(x - a_1) \dots (x - a_n)$, with $a_i \neq a_j$ for $i \neq j$. The same considerations as before come into play; our domain cannot allow us to circle around one branch point only. Therefore, we simply cut into $\mathbb{C}$ a line from a_1 to a_2, from a_3 to a_4, and so on. If n is odd, then we add another cut from n to infinity. We thus obtain two spheres with $\left[\frac{n+1}{2}\right]$ cuts in it, and glue them together, arriving at a Riemann surface of genus $\left[\frac{n+1}{2}\right]$. Note that for this construction, it is essential that all a_i be disjoint, otherwise the surface might be different – in the case $y = \sqrt{x^2}$, for instance, we could never switch from one branch to another (even around 0, where the function values of both branches are the same, but they are not the same *in any neighbourhood*). Thus, we would have simply two disjoint copies of $\mathbb{C}$ here, with no branch points. However, in most cases of interest to us this will turn out to be so.

In fact, this method generalizes to higher roots; the only thing that changes is the number of layers of our covering surface. Similarly, computing the Riemann surface of any polynomial

[18] That ∞ is a branch point may not be entirely obvious, but follows from homogenization and a change in charts.

[19] However, as we added the points at 0 and ∞ later, the surface is not necessarily smooth there. This will be the main difference to the elliptic curve constructed later on, where we remove these singularities to make sure we obtain a smooth structure.

[20] In fact, if one compares the previous case, it actually is the same: loooking at the polynomial $y^2 = x$, then homogenization leads to $y^2 = xz$; so in the original case with 1 root we technically already had 2. All that changes here is that both are visible in the same affine plane.

is, if one is willing to put in a bit more effort, perfectly doable in a similar manner, see eg [37, §8].

3.1.2 A Riemann Surface Associated with the Kernel

We are now ready to return to our original goal, namely to work on the kernel equation 74 on page 20, as in [28, 2]. The kernel was defined as

$$\mathcal{K}(x, y; t) = 1 - t\mathcal{S}(x, y). \tag{146}$$

As $\mathcal{S}(x, y) = \sum_{(i,j)\in\mathcal{S}} x^i y^j$, the kernel is not yet a polynomial, but this can easily be remedied by multiplying it by xy. We then have

$$\mathcal{K}_2(x, y; t) = xy\mathcal{K}(x, y; t) = xy - txy\mathcal{S}(x, y). \tag{147}$$

From here on, we will strictly need only $\mathcal{K}_2(x, y; t)$; therefore, for the sake of brevity, we will write $\mathcal{K}(x, y; t)$ instead of $\mathcal{K}_2(x, y; t)$. Additionally, we will assume $t \in \mathbb{C}$ to be a constant in a neighbourhood of 0 (in particular, $|t| < |\mathcal{S}|$), so unless the particular choice of t matters, we will simply write $\mathcal{K}(x, y)$.

Lemma 3.1.1. *For all 79 step sets $\mathcal{S}$, $\mathcal{K}(x, y)$ is irreducible.*

Proof.
This follows from [31, Lemma 2.3.2]. If a walk is reducible, it (as well as its step set $\mathcal{S}$) is called singular[21] in [31]. But all these singular walks were already discarded during the classification of walks in section 2.2.

□

Similar to what we have done previously, we can now write

$$\begin{aligned}\mathcal{K}(x, y) &= y^2A_1(x) + yA_2(x) + A_3(x), &(148)\\ &= x^2B_1(y) + xB_2(y) + B_3(y). &(149)\end{aligned}$$

By the quadratic formula, we have

$$\begin{aligned} y &= \frac{1}{2A_1(x)}\left(-A_2(x) \pm \sqrt{A_3(x)^2 - 4A_1(x)A_3(x)}\right), &(150)\\ x &= \frac{1}{2B_1(y)}\left(-B_2(y) \pm \sqrt{B_3(y)^2 - 4B_1(y)B_3(y)}\right). &(151)\end{aligned}$$

To shorten the notation, let

$$\begin{aligned}\triangle_y &:= A_3(x)^2 - 4A_1(x)A_3(x), &(152)\\ \triangle_x &:= B_3(y)^2 - 4B_1(y)B_3(y). &(153)\end{aligned}$$

Keep in mind that, as we are working over $\overline{\mathbb{C}}$, the terms $\frac{1}{2A_1(x)}$, $\frac{1}{2B_1(y)}$ are single-valued holomorphic functions and thus does not really matter with regards to our covering space – what would be a pole when working over the complex numbers now simply maps to infinity.

[21] The word *singular* for a walk is ambiguously used in the literature. This definition of singular as in [31] is *not* the one that will generally be used here.

We already know that the Riemann surface of functions of the type $y = \sqrt{\triangle_x}$ depends only on the roots of $\triangle_x$. Therefore, in order to determine the genus of the Riemann surface $\mathcal{S}$ associated with the kernel, all we need to look at are the zeros of $\triangle_x, \triangle_y$. This essentially boils down to a rather lengthy computation, which fortunately has already been done in [31, 2.3.2].[22] By counting degrees, we know that $\triangle_x$ is a polynomial of degree 3 or 4.

Definition 3.1.8. *A step set $\mathcal{S}$ (or a corresponding walk) is called **singular**, if $\triangle_x$ or $\triangle_y$ is either reducible, or has double roots for arbitrary t.*

Theorem 3.1.1. *Out of all 56 walks with infinite group, 51 are non-singular (see Table 6 on page 81).*

Proof.
[31, Lemma 2.3.10] gives a number of criteria for the number of zeros of $\triangle_x$ dependent on the step set $\mathcal{S}$. Applying these to our walks yields the result. See also [28, 2].

□

By the properties of analytic continuation we have derived above[23], we have shown:

Theorem 3.1.2. *For any non-singular walk, the associated Riemann surface $\mathcal{S}$ is of genus 1, that is, a torus.*

In the latter theorem as well as the following, we assume that t is not one of the (finitely many) values such that any roots coincide. Also note that, technically, we have constructed two Riemann surfaces, by continuing the local representations of x and y respectively. However, again due to [34, p. 180], or alternatively an application of Lemma 3.2.5, we can see that the resulting surfaces are equivalent, so it does indeed make sense to talk about *the* associated Riemann surface. See also [31, 2.2.4].
This will prove to be extraordinarily useful, and furthermore emphasize the close relation to elliptic curves, which will be discussed in the next section.

3.2 Algebraic Curves

For a polynomial in 2 variables (such as, for instance, $\mathcal{K}(x, y)$), there is another, very natural way to assign to it a surface: we can simply look at its zero locus.

Definition 3.2.1. *The zero locus $\mathcal{C} \subseteq \mathbb{C}^2$ of a polynomial $p(x, y)$ is called an **affine (algebraic) curve**.*

Definition 3.2.2. *We call an affine curve $\mathcal{C}$ **smooth**, if the system $p(u, v) = \frac{\partial p}{\partial x}(u, v) = \frac{\partial p}{\partial y}(u, v) = 0$ has no solutions $(u, v) \in \mathbb{C}^2$.*

A smooth curve is in a natural way a Riemann surface: let $(u, v) \in \mathbb{C}$. As we assume $\mathbb{C}$ to be smooth, one of the partial derivatives must be non-zero at (u, v), we can assume that $\frac{\partial p}{\partial x}(u, v) \neq 0$. By the Implicit Function theorem, we write $x = \omega(y)$ in a neighbourhood U of (u, v) for some smooth function ω. In U, however, we can then select the chart $(x, y) \mapsto x$, which is easily seen to be a homeomorphism onto its image. After some computation one sees that all such charts are compatible, and thus we have a complex structure on $\mathbb{C}$ (this is worked out in more detail in

[22] As noted in [28, 2.1], the dependency on t does not change things.
[23] Some more technical work needs to be done as well; in particular that continuing the local zeros of $\mathcal{K}(x, y)$ truly yields a solution of $\mathcal{K}(x, y) = 0$ everywhere, and thus the conditions for $\triangle_x$, $\triangle_y$ are the same, see [34, p. 180].

[33, ch. 2]).
It turns out that in many cases it is a lot more convenient, however, to work in projective spaces, as projective space has, in a sense, "nicer" geometric properties. An affine curve, for instance, is usually not compact, and we do not have results such as eg Bézout's theorem. Curves are therefore often considered in $\mathbb{P}^2\mathbb{C} = \left(\mathbb{C}^3 \setminus \{(0,0,0)\}\right) / \sim$, where as before $(x_1 : y_1 : z_1) \sim (x_2 : y_2 : z_2) :\Leftrightarrow \exists \lambda \in \mathbb{C} : \lambda(x_1, y_1, z_1) = (x_2, y_2, z_2)$. In order for the zero locus of a polynomial to make sense in $\mathbb{P}^2\mathbb{C}$, it needs to be well-defined on equivalence classes. As such, we need our polynomial $p(x, y, z)$ to be homogeneous, as then, $0 = p(\lambda x, \lambda y, \lambda z) = \lambda^n p(x, y, z) = 0$ at any zero $(x : y : z) \in \mathbb{P}^2\mathbb{C}$.
Starting from an affine polynomial in two variables, the most intuitive way to arrive at a homogeneous polynomial $\overline{p}(x, y, z)$ is to homogenize, that is, adding powers of z to each monomial until the resulting polynomial is homogeneous.

Definition 3.2.3. *The zero locus $\overline{\mathcal{C}} \subseteq \mathbb{P}^2\mathbb{C}$ of a homogeneous polynomial $\overline{p}(x, y, z)$ is called a* ***projective*** *(algebraic) curve. Given a polynomial $p(x, y)$ the projective curve of p is understood to be the zero locus of the homogenization of p.*

Definition 3.2.4. *We call a projective curve $\overline{\mathcal{C}}$* ***smooth****, if the system $\overline{p}(u, v, w) = \frac{\partial \overline{p}}{\partial x}(u, v, w) = \frac{\partial \overline{p}}{\partial y}(u, v, w) = \frac{\partial \overline{p}}{\partial z}(u, v, w) = 0$ has no solutions in $\mathbb{P}^2\mathbb{C}$, or, equivalently, no solutions in $\mathbb{C}^3 \setminus \{(0,0,0)\}$.*

Given a projective curve, there are three immediate ways to get an affine curve, namely by dehomogenizing with respect to either x, y or z. It turns out that a projective curve $\overline{\mathcal{C}}$ is smooth if and only if each of these affine curves is smooth ([33, Lemma 3.5]). Note that any point in $\mathbb{P}^2\mathbb{C}$ is affine under some chart, and thus lies in one of the affine curves. By, again, some computations, one finds out that all charts are compatible, and therefore $\overline{\mathcal{C}}$ is a Riemann surface. Unlike in the affine case, $\overline{\mathcal{C}}$ is compact.
There is a vast amount of theory about algebraic curves, which can and shall not be explored here. However, one remarkable and beautiful result shows how deep the connection between projective curves and Riemann surfacess actually goes. What we have just considered is a homogeneous polynomial in 3 variables, and assigned to it a projective curve, which is a compact Riemann surface. Now, what happens if we add more variables? Intuitively, if we have one more variable, then the solution space will be larger by one dimension. So in order to be locally homeomorphic to $\mathbb{C}$, we would need one more polynomial equation, and if we intersect the resulting curves, the dimension should hopefully work out. One can do the computations, and see that $n - 1$ smooth, homogeneous polynomials in $n + 1$ variables define[24], by the intersection of their zero sets, something very close to a projective curve, called a **smooth intersection curve**. These curves can also be shown to be compact, and a Riemann surface is called **algebraic** if it can be generated as a smooth intersection curve.
In an astonishingly beautiful piece of mathematics, it turns out that *every* compact Riemann surface is algebraic. In particular, this means that by studying compact Riemann surfaces, one is studying zero-sets of groups of homogeneous polynomials. This is highly non-obvious, and requires a much deeper theory than is presented here. It can be found, for instance, in [36, ch. 2].

Associating a projective curve with a smooth polynomial is closely related to analytic continuation: locally around a point, we can always express one variable as a smooth function of the other, giving us a local solution of the polynomial. This is a function element. Analytically continuing this function element gives us the Riemann surface, on all points of which the polynomial must vanish. As the projective curve is connected (this is not entirely obvious, but holds true, see eg

[24] under certain conditions; see again eg [33, ch. 2]

[38, VII, 2.2, Lemma 1]), the analytic continuation will reach all non-singular points of the curve (compare [34, 4.13-4.14]). For a more thorough introduction into the both analytic continuation and algebraic curves and the relation between them, see [39]).

3.2.1 Elliptic Curves

There is a lot of literature about elliptic curves, and only the very basics can be outlined here. For a both readable and thorough introduction see eg [40]. Elliptic curves are not only interesting over $\mathbb{C}$, but also over other, in particular finite, fields. As these other cases will not be relevant later, however, we will restrict ourselves to elliptic curves over $\mathbb{C}$ here.

Definition 3.2.5. *An* ***Elliptic curve*** $\mathcal{C}$ *is a smooth projective curve of genus* 1*.*

The definition varies a bit; sometimes an elliptic curve is defined via its Weierstraß-equation (see below), and sometimes an elliptic curve is defined to have a particular base point, as is the case eg in [40]. This base point is needed to introduce the group structure on $\mathcal{C}$ later on. However, as we will by our kernel be able to naturally derive a smooth projective curve of genus 1, but not be given a particular point, we will proceed as in [41, 2.3.2] and fix a base point only when interested in the group structure.
The notion of the genus of a curve is technically a somewhat involved concept, as it is not immediately the same as the (more intuitive) topological genus of a surface. For a smooth curve, the genus is often defined as in [40, 5.4], by the Riemann-Roch theorem:

Theorem 3.2.1 (Riemann-Roch)**.** *Let* $\mathcal{C}$ *be a smooth projective curve and* ω *a canonical divisor of* $\mathcal{C}$*. Then there is an integer* g*, such that*

$$\deg D = l(\omega - D) + g - 1 \tag{154}$$

for every divisor D*.*
The integer g *is called the* ***genus*** *of* $\mathcal{C}$*.*

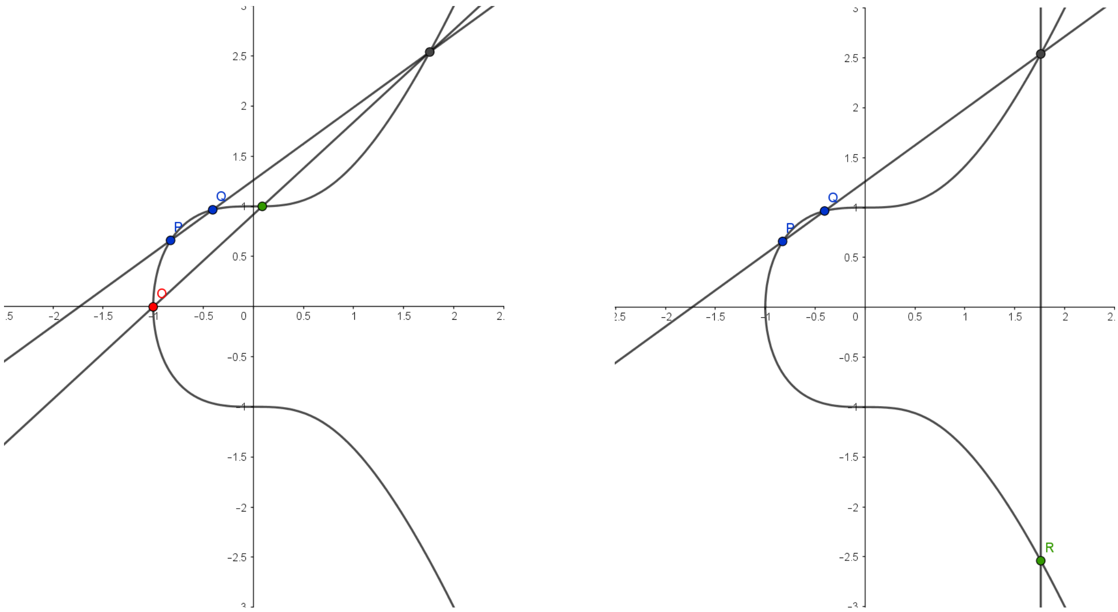
We will skip over precisely what (canonical) divisors are and quite generally what the above theorem actually means. All this, together with proofs, can be looked up eg in [29, 42]. It turns out, however, that the genus can be interpreted differently as well. In particular, the genus is the dimension of the vector space spanned by holomorphic differentials on $\mathcal{C}$ ([42, Prop. 6.38]), and it turns out to be just the same as the topological genus if we view the curve as a Riemann surface ([42, App. C]).
The Riemann-Roch theorem immediately enables us to give a more intuitive way to represent an elliptic curve ([40, Prop. 3.1, III.1]):

Lemma 3.2.1. *Let* $\mathcal{C}$ *be an elliptic curve. Then there are* $A, B \in \mathbb{C}$*, such that* $\mathcal{C}$ *is given (up to coordinate change) by the curve associated with the* ***Weierstraß equation***

$$y^2 = x^3 + Ax + B. \tag{155}$$

This turns out to be a useful tool to visualize elliptic curves, as well as for checking whether two curves are the same, i.e. one results from one another via a coordinate change. In particular, we have

Lemma 3.2.2. *Let* $\mathcal{C}$ *be a projective curve defined by a Weierstraß equation. Define the* ***discriminant*** $\triangle := -16(4A^3 - 17B^2)$*, and the* j***-invariant*** $j := -1728\frac{(4A)^3}{\triangle}$*. Let* $\mathcal{C}'$ *be another such curve. Then:*



(a) O chosen arbitrarily (b) O at infinity

Figure 4: The group law on $\mathcal{C} : y^2 = x^3 + 1$ for two choices of O; we have $P + Q = R$. The choice of O at infinity on the right hand side is probably the most common.

1. *$\mathcal{C}$ is smooth precisely if $\triangle \neq 0$.*
2. *$\mathcal{C}$ is isomorphic to $\mathcal{C}'$ precisely if $j = j'$.*

Proof.
By direct computation, see [40, Prop. III.1.4].

□

Another way of describing an elliptic curve, which will be of use later, is given by the Legendre form:

Lemma 3.2.3. *Let $\mathcal{C}$ be an elliptic curve. Then there is $\lambda \in \mathbb{C}$, such that $\mathcal{C}$ is given by an equation in* ***Legendre form***

$$y^2 = x(x-1)(x-\lambda). \tag{156}$$

Proof.
By factoring the Weierstraß equation and some coordinate changes, see eg [40, Prop. 1.7].

□

Similarly to before, one can compute the j-invariant given the Legendre form (see again [40, Prop. III.1.7]), and in particular it turns out that two equations in Legendre form with different values for λ may still describe the same curve[25].
One of, if not the main reason elliptic curves are so interesting is that they come equipped with a natural abelian group structure. For an illustration of the following definition, see Figure 4.

[25]This also shows that all elliptic curves over $\mathbb{C}$ can be parametrized with only one complex parameter, which may not be immediately obvious.

Definition 3.2.6. *Let $P, Q \in \mathcal{C}$, O be a fixed point*[26] *on $\mathcal{C}$. Then we let*

1. *$\overline{P+Q}$ be the third intersection point of the line going through P, Q (if $P = Q$, take a tangent) and the curve $\mathcal{C}$,*
2. *$P + Q := \overline{\overline{P+Q} + O}$.*

Lemma 3.2.4. *The operation defined above is an abelian group law on $\mathcal{C}$, with neutral element O. In particular, we have $P + Q + R = O$ for any three collinear points $P, Q, R \in \mathcal{C}$.*

Proof.
First note that, by Bézout's theorem, any line must have (counted with multiplicities) precisely 3 intersection points with $\mathcal{C}$, so the operations as above are well-defined. It remains to show they form an abelian group law.
Commutativity is clear, as a line going through two points clearly is the same no matter which point we choose first. That $O + P = P$ follows, as $\overline{O+P}$ is the third intersection point with $\mathcal{C}$ of the line through O, P, so $O + P = \overline{\overline{O+P} + O}$ must (again by Bézout) be P. Similarly, one has $P + \overline{P+O} = O$, so there is an inverse.
The one part of the proof that is a bit tedious is the associativity. This follows either by direct computation in coordinates, or more beautifully by an association of points with divisors, see [40, Prop. III.2.2].

□

The group structure is one reason why elliptic curves must be looked at in projective space to work, otherwise the group law might collapse, as Bézout's theorem is essential here. Also, if the curve was singular, there might be a double point on $\mathcal{C}$ without well-defined tangent, so again the addition would fail. While that the group laws can be directly expressed in functions on the coordinates, using no more then basic field operations (see [40, III.2.3]), so we do not need our field to be algebraically closed to make this work, the resulting formulae are not very nice to work with. They are sufficiently not-nice, in fact, that the iterated addition of points yields a discrete logarithm problem which is applicable in cryptography (though this is done with curves over finite fields, not over $\mathbb{C}$), see eg [40, IX].

It was already mentioned that the genus of a curve is the same as the dimension of holomorphic differentials, and the same as the topological genus of the corresponding Riemann surface. An explicit representation of such a differential form is given by $\omega = \frac{\mathrm{d}\,x}{y}$ (see [40, III.1.5]), and one can now look at path integrals on $\mathcal{C}$. In particular we would like to assign a value to the expression $\int_P^Q \omega$. We already know that our elliptic curve has a Legendre form, so we can write $y^2 = x(x-1)(x-\lambda)$, and then have integrals of the form[27] (this and the following is worked out in more detail in [40, VI])

$$\int_\gamma \frac{\mathrm{d}\,t}{\sqrt{t(t-1)(t-\lambda)}} \tag{157}$$

for some path γ connecting P and Q.
As we have seen, the denominator is single-valued on $\mathcal{C}$ interpreted as a Riemann surface, which has genus 1 and thus is a complex torus. By complex analysis, every integral along closed loop on a single-valued branch must be 0. Therefore, this integral can be non-zero only if our path cannot be contracted to 0 on the torus, that is, it contains – up to homotopy – at least one of

[26] This is the reason why some authors take an elliptic curve as the curve itself plus a fixed point; this gives us complete information about both the curve *and* the group law.
[27] These integrals are called *elliptic integrals*, and are in fact where the study of elliptic curves first arose.

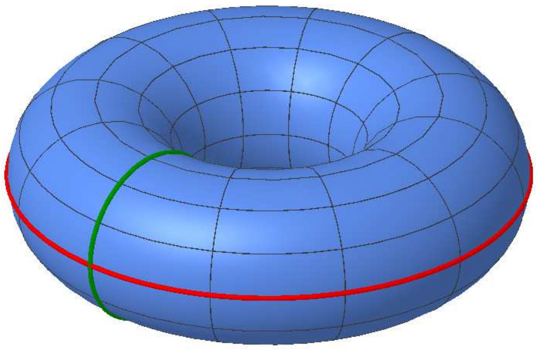


Figure 5: A choice of two not 0-homotopic loops on the torus, which are candidates for α and β below.

the loops shown in Figure 5 [28]. Call these two loops α and β respectively. So what does that mean for integral on $\mathcal{C}$ along a path γ connecting two points P, Q? First of all, clearly, γ is not unique. We could go directly from P to Q, but we could include α or β once, or even multiple times, and only then arrive at Q. Now, if we have two paths γ, γ_1 from P to Q, then we can look at how many times we have traveled – up to homotopy – along α and β, and if these numbers are the same, only then are γ and γ_1 homotopic. Assume now that γ can be constructed out of γ_1 by adding k times α, and l times β. We then have[29]

$$\int_\gamma \omega = \int_{\gamma_1 \circ \alpha^k \circ \beta^l} \omega = \int_\gamma \omega + k \int_\alpha \omega + l \int_\beta \omega. \tag{158}$$

As k, l certainly must be integers, we see that $\int_\gamma \omega$ is well-defined only up to integer multiples of

$$\omega_1 := \int_\alpha \omega, \tag{159}$$

$$\omega_2 := \int_\beta \omega. \tag{160}$$

In particular, the integral $\int_P^Q \omega$ is well-defined on $\mathbb{C}/\Lambda$, where Λ is the lattice $\omega_1\mathbb{Z} + \omega_2\mathbb{Z} \subset \mathbb{C}$. It can be shown (see [40, VI]) that ω_1, ω_2 are linearly independent over $\mathbb{R}$, so Λ is non-degenerate. We therefore have a mapping

$$\phi : \mathcal{C} \to \mathbb{C}/\Lambda : P \mapsto \int_O^P \omega. \tag{161}$$

Using translation invariance of ω, it can be checked that ϕ is even a group homomorphism.

On the other hand, assume now that we have a lattice Λ, and take a look at the space $\mathbb{C}/\Lambda$. In particular, by identifying edges, $\mathbb{C}/\Lambda$ can be viewed as a Riemann surface, so we can take a look at so-called *elliptic functions*, which are meromorphic functions $\mathbb{C}/\Lambda \to \mathbb{C}$ (see eg [33, 2.2],[34,

[28] This is equivalent to our path going over a branch cut, if one wants to go back to the construction of the covering surface with this in mind.

[29] If one wants to be very precise here, then we would need to take care of the order of paths, and define the homotopies more closely, but the idea should be clear.

ch. 3], [40, VI], where the following is worked out in detail). It turns out that there aren't very many of those, the most prominent being the Weierstraß $\wp$-function:

$$\wp(z) := \frac{1}{z^2} + \sum_{0\neq\omega\in\Lambda} \left(\frac{1}{(z-\omega)^2} - \frac{1}{\omega^2} \right). \tag{162}$$

As $\wp(z)$ can be checked to be meromorphic, so is $\wp'(z)$, and, surprisingly, *that is it*, there are essentially no more meromorphic functions on $\mathbb{C}/\Lambda$.

Theorem 3.2.2. *Any meromorphic function f on $\mathbb{C}/\Lambda$ can be written as*

$$f = Q(\wp(z), \wp'(z)), \tag{163}$$

where $\wp(z)$ is the Weierstraß $\wp$-function, and Q is a rational function.

Proof.
See eg [34, Th. 3.11.1] or [40, Th. VI.3.2].

□

In addition, by carefully rearranging and computing sums, one can show

Theorem 3.2.3. *We have*

$$\wp'(z)^2 = 4\wp(z)^3 - 60G_4\wp(z) - 140G_6, \tag{164}$$

where

$$G_4 = \sum_{0\neq\omega\in\Lambda} \omega^{-4}, \tag{165}$$

$$G_6 = \sum_{0\neq\omega\in\Lambda} \omega^{-6} \tag{166}$$

are the ***Eisenstein series*** *of weights 4 and 6 respectively.*

Proof.
See eg [34, Th. 3.10.4], [40, Th. VI.3.5].

□

In particular, this means that $\wp(z), \wp'(z)$ satisfy a Weierstraß equation. One can then check that the discriminant is non-zero, and we can therefore go in the other direction as before; we can now assign to a lattice an elliptic curve. In particular, we have a mapping

$$\psi : \mathbb{C}/\Lambda \to \mathcal{C} : z \mapsto (\wp(z) : \wp'(z) : 1). \tag{167}$$

Again, this mapping turns out to be a group homomorphism, see eg [40, Prop. VI.3.6]. Up to affine transformation of our lattice, and coordinate changes of our elliptic curve, this map turns out to be the inverse of $\phi : \mathcal{C} \to \mathbb{C}/\Lambda$ defined previously ([40, Prop. VI.5.2]).
As a result, we have the equivalence of the classes

$$\{\text{Elliptic curves over } \mathbb{C}, \text{ up to coordinate change}\} \leftrightarrow \{\text{Lattices } \Lambda \subset \mathbb{C}, \text{ up to homothety}\}.$$

We can transition between those classes by the group homomorphisms ϕ, ψ defined above.

3.2.2 An Elliptic Curve Associated with the Kernel

We now want to associate with the kernel, given as in (147), an elliptic curve. Unfortunately, just taking the curve in $\mathbb{P}^2\mathbb{C}$ as we have done previously will not usually work out, as in most cases one of the points at infinity $(0:1:0)$ or $(1:0:0:)$ turns out to be singular. What turns out to be the key idea here is looking at $\mathcal{K}(x,y)$ not in $\mathbb{P}^2\mathbb{C}$, but rather in $\mathbb{P}\mathbb{C}\times\mathbb{P}\mathbb{C}$[30]. However, this idea does come with a few technical hurdles. First of all, we defined a curve as an object in a projective space. It may not be immediately obvious that $\mathbb{P}\mathbb{C}\times\mathbb{P}\mathbb{C}$ is a projective space – but this just so happens to be the case. It can be proven that the Segre map

$$([x_0:x_1:\cdots:x_m],[y_0:y_1:\cdots:y_n]) \mapsto [(w_{ij}-x_iy_j)_{0\le i\le m, 0\le j\le n}] \tag{168}$$

is an embedding of $\mathbb{P}^m\times\mathbb{P}^n \to \mathbb{P}^{nm+n+m} = \mathbb{P}^{(n+1)(m+1)-1}$ ([43, 2.11]). Motivated by this, we can consider the curve of $\mathcal{K}(x,y)$ in $\mathbb{P}\mathbb{C}\times\mathbb{P}\mathbb{C}$, that is, the zero set of $\mathcal{K}(x,y)$ homogenized with respect to both x and y, such that the bidegree with respect to x and y is $(2,2)$. Let us denote this homogeneous polynomial by $\overline{\mathcal{K}(x,y)}$. Then, we can define at the discrimants $\overline{\triangle_x}$, $\overline{\triangle_y}$ as in section 3.1.2. The resulting expressions are given in [32, 2.4-2.5]. From here on, we proceed similarly as in [41, 2.4].

Lemma 3.2.5. *Assume one of $\overline{\triangle_x}, \overline{\triangle_y}$ has only simple zeros. Then so has the other, and $\overline{K(x,y)}$ is smooth.*

Proof. (See [41, Prop. 2.4.3])
Let $\mathcal{C}$ be the curve associated with $\overline{\mathcal{K}(x,y)}$, and let $p\in\mathbb{C}^2$. We will consider two cases:

1. $\overline{\triangle_y}(p)\neq 0$.
   Then, $p=:(x_0,x_1)\neq(0,0)$ (as $\overline{\triangle_y}$ is homogeneous with degree 2 in (x_0,x_1), i.e. else it would be zero, too). We know there we arrive at 2 different values for y and resulting points $P\in\mathcal{C}$, and claim that, for each of those, $\frac{\partial\overline{\mathcal{K}}}{\partial y}(P)\neq 0$.
   To see this, rewrite $\overline{\mathcal{K}(x,y)}$ as

$$\overline{\mathcal{K}(x,y)} = \left(2\overline{A_1(x)}y - \overline{A_2(x)}\right)^2 - \overline{\triangle_y}. \tag{169}$$

   Differentiating with respect to y leads to

$$\frac{\partial\overline{\mathcal{K}(x,y)}}{\partial y} = 4\overline{A_1(x)}\left(2\overline{A_1(x)}y - \overline{A_2(x)}\right). \tag{170}$$

   As $P\in\mathcal{C}$, from (169) and by the assumption $\overline{\triangle_y}\neq 0$ we know that the second factor on the right hand side is non-zero, so the only way P could be singular is if $\overline{A_1(p)}=0$. Here is when things get slightly technical. Note that we have so far pretended to be working in affine space, when actually we have $y\in\mathbb{P}\mathbb{C}$. As a consequence, we would have had to write $y=\frac{y_0}{y_1}$ (and get rid of denominators), which thus far did not matter. Similarly, we would have had to decide on whether we look at the operator $\frac{\partial}{\partial y_0}$ or $\frac{\partial}{\partial y_1}$. To prove smoothness of $\overline{K(x,y)}$, it suffices that *one* of these two derivatives is non-zero. So, our first try could be, if $\overline{A_1(p)}=0$, to switch y_0 for y_1, and we arrive at the condition $\overline{A_3(p)}=0$. If either of these is not satisfied, we have smoothness. If, however, $\overline{A_3(p)}=\overline{A_1(p)}=0$, then by $\overline{\triangle_y}\neq 0$ we know that $\overline{A_2(p)}\neq 0$. As we then have

$$\overline{\mathcal{K}(p,y)} = \overline{A_2(p)}y_0y_1 = 0, \tag{171}$$

[30] A way to arrive at this idea is by blowing up the singularity at infinity, see eg [41, Preface].

we know that our two choices of y are $(0:1)$ and $(1:0)$. At $(1:0)$, we have $\frac{\partial \overline{\mathcal{K}(p,y_0,y_1)}}{\partial y_1} = \overline{A_2(p)}y_0 = \overline{A_2(p)} \neq 0$, and at $(0:1)$, in the same manner we have $\frac{\partial \overline{\mathcal{K}(p,y_0,y_1)}}{\partial y_0} \neq 0$.

2. $\overline{\triangle_y}(p) = 0$, $\frac{\overline{\triangle_y}}{\partial x} \neq 0$.
In this case, there is only one corresonding value y, and point $P \in \mathcal{C}$. We claim that $\frac{\partial \overline{\mathcal{K}}}{\partial x}(P) \neq 0$.
Differentiating (169) with respect to x yields

$$\frac{\partial \overline{K(x,y)}}{\partial x} = 2\left(2\overline{A_1(x)}y - \overline{A_2(x)}\right)\frac{\partial\left(2\overline{A_1(x)}y - \overline{A_2(x)}\right)}{\partial x} - \frac{\partial \overline{\triangle_y}}{\partial x}. \tag{172}$$

However, at P, we have $\overline{\mathcal{K}(x,y)} = \overline{\triangle_y} = 0$, so from (169) we know that the first factor in (172) is 0. Thus $\frac{\partial \overline{K(x,y)}}{\partial x}(P) = \frac{\partial \overline{\triangle_y}}{\partial x}(P)$, which by assumption is non-zero.
By a similar computation for $\frac{\partial}{\partial y}$, we have that $\frac{\partial \overline{\mathcal{K}(x,y)}}{\partial y} = 0$, and as a consequence we know that $\overline{\mathcal{K}(x,y)}$ can be smooth only if $\overline{\triangle_y}$ has only simple zeros.

Put together, this shows the smoothness of $\overline{\mathcal{K}(x,y)}$, if one of the determinants has only simple roots – and it also shows that, if a determinant had a multiple zero, then $\overline{\mathcal{K}(x,y)}$ could not be smooth. Hence the statement follows.

□

In order to show that the curve of $\overline{\mathcal{K}(x,y)}$ in $\mathbb{PC} \times \mathbb{PC}$ is an elliptic curve, we thus need to show it has genus 1. This can be done by looking at the projection maps $(x,y) \to x : \mathbb{PC} \times \mathbb{PC} \to \mathbb{PC}$, and then applying the Riemann-Hurwitz formula. While this is, knowing the theory, rather straight-forward, the Riemann-Hurwitz formula leaves the scope of this work. This particular application is done in [41, 2.4.3], the Riemann-Hurwitz formula with proof can be looked up eg in [33, Th. 4.16], [29, Th. 3.4.13].
Note in particular that the condition for $\overline{\triangle_y}$ to only have simple zeros is the same we used in the definition of non-singular walks in section 3.1.2. As a result, we have proven

Theorem 3.2.4. *For every non-singular walk, the curve $\mathcal{E} := \{(x,y) \in \mathbb{PC} \times \mathbb{PC} : \overline{\mathcal{K}(x,y)} = 0\}$ is an elliptic curve.*

By construction, both the Riemann surface $\mathcal{S}$ and $\mathcal{E}$ are smooth, irreducible genus 1 surfaces on which the kernel vanishes, making them different models of the same surface in the sense of [39, Ch. 1, 6.6]. Therefore, we will not distinguish between $\mathcal{E}$ and $\mathcal{S}$ very strictly. When we are interested in the properties following from the construction as a Riemann surface, as in section 4.1, we will refer to the surface as $\mathcal{S}$, when we are interested in the properties following from the surface being an elliptic curve, as in section 4.2, we will refer to it as $\mathcal{E}$.

3.3 Lifting of Mappings

We will follow [32, 2], [28, 2-4], [31, 2-4]. Consider again the kernel equation 74 on page 20. On any of the surfaces constructed above, $\mathcal{K}(x,y)$ vanishes, leaving us with

$$0 = xy - txyF(x,0;t)\mathcal{S}_y(x) - txyF(0,y;t)\mathcal{S}_x(y) + t\varepsilon F(0,0;t). \tag{173}$$

As we are ultimately interested in holonomy, or hypertranscendency, of $F(x,0;t), F(0,y;t)$, which certainly is not dependent on the factors $txy\mathcal{S}_y(x,y), txy\mathcal{S}_x(x,y)$, we again try to keep the notation short and write[31]

$$0 = xy - F_1(x,t) - F_2(y,t) + \varepsilon F(t) \tag{174}$$

instead. What we aim to do is to lift all relevant mappings first to the surface, and study the properties there. In particular, the close relation between the two constructions done previously will become very obvious once we lift the maps to the surfaces' respective universal cover: both $\mathcal{E}$ and $\mathcal{S}$ are equivalent to $\mathbb{C}/\Lambda$ for some lattice Λ (for elliptic curves this has already been discussed, for a torus see eg [34, 3.5]), and it turns out that the universal cover is $\mathbb{C}$ in both cases[32]. This will prove immensely useful both for gaining a surprisingly simple form of the group of the walk, and in classifying the poles of $F_{1,2}$.
The following construction has been done in a somewhat more general setting in [31, ch. 2-4], and in the case applicable here for the Riemann surface in [28]. In [32], the surface is constructed as an elliptic curve rather than a Riemann surface, however, the following works out in much the same manner (and is, in fact, mostly done referring to the former works). Thus, we will not separately consider the covering as a Riemann surface or as an elliptic Curve separately.
Remember that we technically constructed the surface $\mathcal{S}$ twice; by analytically continuing local representations of x and y respectively. As we already know the resulting surfaces to be equivalent, however, we can work in just $\mathcal{S}$, but with different projection maps $h_x, h_y : \mathcal{S} \to \mathbb{C}$. Considering $\mathcal{S}$ as the elliptic curve, h_x would be nothing other than the first projection $h_x : (x,y) \mapsto x$, and $h_y : (x,y) \mapsto y$ the second projection. Clearly, h_x, h_y as coordinate maps are not independent of each other, as for all $s \in \mathcal{S}$ we must have $\mathcal{K}(h_x(s), h_y(s)) = 0$. As $x(y), y(x)$ are (except at the branch points) 2-valued functions, we cannot immediately use them to switch from one projection map to the other. However, if we take care of which branch we come from and take the corresponding solution, they allow moving from one chart to another. This is depicted in the following diagram:

$$\begin{tikzcd}
 & \mathcal{S} \arrow[dl, "h_x"'] \arrow[dr, "h_y"] & \\
\mathbb{C} \arrow[rr, shift left, "y(x)"] & & \mathbb{C} \arrow[ll, shift left, "x(y)"]
\end{tikzcd}$$

Figure 6: $\mathcal{S}$ seen as a cover of $\mathbb{C}$, with h_x, h_y being the projection maps.

Given any function $f(x)$ defined on some subset of $\mathbb{C}$, then we can clearly lift it to $\mathcal{S}$ by letting $\widehat{f} := f \circ h_x$ on some subset of $\mathcal{S}$. This is what we plan on doing both with $F_{1,2}$, and the group of the walk.

[31]Keep in mind that y cancels in $y\mathcal{S}_y(x,y)$, and thus the latter can be considered a function of x only.
[32]In fact, this goes to show once again that both constructions result in essentially the same surface, by identification of elliptic curves and lattices.

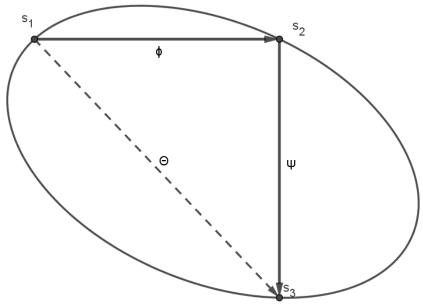


Figure 7: The QRT map, obtained via projection parallel to the axes on $\mathcal{S}$.

3.3.1 Lifting the Group of a Walk

Lifting the group of the walk, that is, the automorphisms Ψ, Φ as defined in section 2.2, turns out to be surprisingly intuitive. As Ψ, Φ leave $\mathcal{K}(x,y)$ invariant, they can be viewed as mappings $\mathcal{S} \to \mathcal{S}$. Let us look at Ψ here, Φ is considered entirely analogously. We then know that $\Psi(x,y) = (x,y')$, with y' being a rational function of x, y, and (x,y') being another zero of $\mathcal{K}(x,y)$. As $(x,y),(x,y') \in \mathcal{S}$, we know that y, y' must be the two branches of $y(x)$. Looking at $(x,y) = s, (x,y') = s' \in \mathcal{S}$, this means that $h_x(s) = h_x(s') = x$. But, unless x is a zero of $\triangle_x$, there are precisely two points in $\mathcal{S}$ with $h_x(s) = h_x(s') = x$, so all Ψ does is switch between the two cut copies of $\mathbb{C}$ we glued together. Let s' be this unique other point in $\mathcal{S}$, and s'' the point obtained analogously via h_y. We can then define[33]

$$\Psi: \quad \mathcal{S} \to \mathcal{S}: \quad s \mapsto \begin{cases} s' & \text{if } (\triangle_x \circ h_x)(s) \neq 0, \\ s & \text{if } (\triangle_x \circ h_x)(s) = 0, \end{cases} \tag{175}$$

$$\Phi: \quad \mathcal{S} \to \mathcal{S}: \quad s \mapsto \begin{cases} s'' & \text{if } (\triangle_y \circ h_y)(s) \neq 0, \\ s & \text{if } (\triangle_y \circ h_y)(s) = 0. \end{cases} \tag{176}$$

From this, $\Phi^2 = \Psi^2 = \mathrm{Id}$ follows immediately. As Φ, Ψ effectively act as a permutation of the layers of the covering $\mathcal{S}$ of $\mathbb{C}$, they are what is often considered a Galois automorphism of $\mathcal{S}$. There is a direct connection between such automorphisms and the "usual" Galois automorphisms of field extensions via looking at the corresponding function fields, which is outlined eg in [31, 2.4].

There is another, more geometrical way of viewing these automorphisms, as is indicated in Figure 7. Considering $\mathcal{K}(x,y)$ is quadratic in both x and y, given a point s_1, we can simply take the second intersection point of $\mathcal{S}$ and a line through s_1 parallel to the x-axis in order to obtain $s_2 := \Phi(s_1)$, and doing the same with a line parallel to the y-axis one obtains $s_3 := \Psi(s_2)$. Doing one after another, one arrives at what is called the QRT-map $\Theta := \Phi \circ \Psi$, and extensively studied in [41].

[33]The automorphisms will be called the same here independently on whether they are automorphisms on $\mathcal{S}$, on $\overline{\mathbb{C}}^2$ or on $\mathbb{C}/\Lambda$, the domain will be clear from the context.

Lifting the group to the universal cover

As $\mathcal{S}$ is a torus, it can be identified with $\mathbb{C}/\Lambda$ for some lattice $\Lambda = \omega_1\mathbb{Z} + \omega_2\mathbb{Z}$, such that $\frac{\omega_1}{\omega_2} \notin \mathbb{R}$; by the results of section 3.2.1 we already know how to obtain the constants[34] $\omega_{1,2}$ by integration over base cycles. By computation, we obtain

$$\omega_1 = i\int_{x_1}^{x_2} \frac{\mathrm{d}\,x}{\sqrt{-\triangle_x}}, \tag{177}$$

$$\omega_2 = \int_{x_2}^{x_3} \frac{\mathrm{d}\,x}{\sqrt{\triangle_x}}, \tag{178}$$

with roots x_i of $\triangle_x$ suitably chosen (see [28, 2.1-3.1], [31, 3]).
The universal cover of $\mathcal{S} \cong \mathbb{C}/\Lambda$ is $\mathbb{C}$ with the natural projection. The automorphism Θ must lift to an automorphism $\widehat{\Theta}$ on $\mathbb{C}$. In the following, we will need to utilize some results about the exact placement of the roots of the partial discriminants, which can be found in [28, 2-3],[31, 2.3-2.5]. In particular, we have (real) roots x_i such that

$$|x_1| < x_2 < 1 < x_3 < x_4, \tag{179}$$

$$|y_1| < y_2 < 1 < y_3 < y_4, \tag{180}$$

by potentially choosing x_4 or y_4 to be at infinity. When constructing $\mathcal{S}$, we cut $\overline{\mathbb{C}}$ x_1, x_2 and x_3, x_4, and in the same order for y respectively. We can then arrange a fundamental parallelogram such that by letting $\omega_3 := 2h_y^{-1}(y_4)$[35]

$$h_x^{-1}(x_1) = \frac{\omega_2}{2}, \qquad h_y^{-1}(y_1) = \frac{\omega_2 + \omega_3}{2}, \tag{181}$$

$$h_x^{-1}(x_2) = \frac{\omega_1 + \omega_2}{2}, \qquad h_y^{-1}(y_2) = \frac{\omega_1 + \omega_2 + \omega_3}{2}, \tag{182}$$

$$h_x^{-1}(x_3) = \frac{\omega_1}{2}, \qquad h_y^{-1}(y_3) = \frac{\omega_1 + \omega_3}{2}, \tag{183}$$

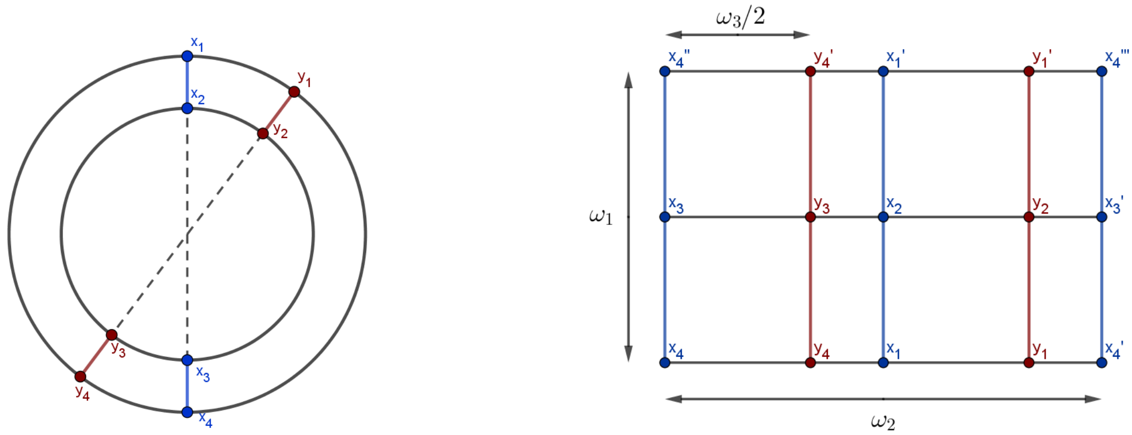
$$h_x^{-1}(x_4) = 0, \qquad h_y^{-1}(y_4) = \frac{\omega_3}{2}, \tag{184}$$

see Figure 8 on the next page.

Lemma 3.3.1. $\widehat{\Theta} : \mathbb{C} \to \mathbb{C}$ *can be chosen to be of the form* $\omega \mapsto \omega + z$ *for some* $z \in \mathbb{R}$.

Proof.
We know that Θ is as an automorphism on $\mathbb{C}/\Lambda$, and thus lifts to an automorphism on $\mathbb{C}$. An arbitrary automorphism f of $\mathbb{C}$ is an affine transformation: an automorphism has to be entire, so it has a globally converging power series and thus a singularity at ∞. If this singularity is essential, then by Picard's theorem we cannot be injective. In order for the singularity not to be essential, almost all coefficients of the power series must vanish, thus, f is a polynomial. To be injective, f must be linear, i.e. $f : z \mapsto az + b, 0 \neq a \in \mathbb{C}$.
Next, consider $\widehat{\Psi}$. We know that $\Psi^2 = \mathrm{Id}_{\mathbb{C}/\Lambda}$, and we know that $\widehat{\Psi}$ is an affine function. As a consequence, we immediately obtain that $\Psi(\omega) = a_1\omega + b_1$, where $a_1^2 = 1$. We also know that Ψ is not constant, and has fixed points on $\mathbb{C}/\Lambda$, therefore we must have $a_1 = -1$, and $\Psi(\omega) = -\omega + b_1$. Now consider the fixed points of Ψ. First, it needs to pointed out that a fixed point of Ψ does

[34]As mentioned in section 3.2.1, these are defined up to homothety; here we require ω_1 to be purely imaginary and ω_2 to be real, making them unique.
[35]By abuse of notation, h_x, h_y are now not the coordinate maps from $\mathcal{S}$ to $\mathbb{C}$, but rather from $\mathbb{C}/\Lambda \cong \mathcal{S}$.



(a) Gluing of the two cut copies of $\overline{\mathbb{C}}$.

(b) The fundamental rectangle. x_i, x_i' as well as y_i, y_i' are identified, as by traversing along a cut or a half circle twice, one arrives at the original point.

Figure 8: Gluing together the two cut copies of $\overline{\mathbb{C}}$, and the resulting fundamental rectangle [28, Fig. 4-5].

not have to one of $\widehat{\Psi}$ – it only has to be such modulo the projection map $\mathbb{C} \to \mathbb{C}/\Lambda$. Indeed, as Ψ has the 4 fixed points $x_{1,\dots,4}$, not all of these can be preserved by $\widehat{\Psi}$. But also notice that the lifting is unique only up to choosing the image of one point. We can therefore choose one of the x_i to be fixed, and by selecting (in accordance to [28, 31, 32]) this to be x_2, we have $b_1 = \omega_1 + \omega_2$, and thus $\widehat{\Psi} : \omega \mapsto -\omega + \omega_1 + \omega_2$.
By the same argument, we have Φ as a mapping of the form $\omega \mapsto -\omega + b_2$, except now we have fixed points $y_{1,\dots,4}$. Selecting now, for the same reasons as above, the fixed point of $\widehat{\Phi}$ to be y_2, we immediately have $b_2 = \omega_1 + \omega_2 + \omega_3$ in $\mathbb{C}/\Lambda$. Hence, we have $\widehat{\Phi} : \omega \mapsto -\omega + \omega_1 + \omega_2 + \omega_3$.
By composition, we have $\widehat{\Theta} = \widehat{\Psi \circ \Phi} = \widehat{\Psi} \circ \widehat{\Phi} : \omega \mapsto -(-\omega + \omega_1 + \omega_2 + \omega_3) + \omega_1 + \omega_2 = \omega - \omega_3$. So $z = -\omega_3$, and as ω_3 is real, we are done.[36]

□

As the equivalence between $\mathbb{C}/\Lambda$ and $\mathcal{E}$ is a group homomorphism, we therefore know Θ on $\mathcal{E}$ to be addition with some fixed point on $\mathcal{E}$, as is directly proven eg in [41, 2.5.2]. In addition, by retracing the construction of the universal cover, we can explicitely compute (note that $x(y_1)$ is unique)

$$\omega_3 = \int_{x(y_1)}^{x_1} \frac{\mathrm{d}\,x}{\sqrt{\triangle_x}}, \tag{185}$$

see again [28, 2-3],[31, 2.3-3] for details.
The above lemma immediately implies that the group of a walk, constricted to the elliptic curve $\mathcal{E}$, is finite if and only if $\frac{\omega_2}{\omega_3} \in \mathbb{Q}$. However, this does not immediately imply finiteness of the group on all of $\overline{\mathbb{C}}^2$. What does hold, however, is the following:

[36] The definition of $\Theta = \Psi \circ \Phi$ is consistent with the definition in previous sections, and how it was introduced in [24]. However, in [31, 28, 32], the corresponding map is selected to be $\Phi \circ \Psi$, so Θ^{-1}. This does not change anything in the following, as we will need translations by both $\pm\omega_3$ in any case.

Lemma 3.3.2 ([28, Remark 6],[32, Prop. 2.6])**.** *Let $\mathcal{S}$ be the step set of a walk with infinite group. Then there are only countably many $t \in \mathbb{C}$ such that the group on the surface associated with $\mathcal{K}(x, y; t)$ is finite.*

Proof.
Either using the fact that Θ is the addition of a point on $\mathcal{E}_t$ and showing that, for a fixed n, $\Theta^n = \mathrm{Id}$ for at most finitely many t, see [32], or using the condition that all depends on the fraction $\frac{\omega_2}{\omega_3}$, which will not assume the same value twice for different choices of t, as is shown in the proof of [44, Prop. 4].

□

3.3.2 Lifting of the Boundary Functions

The first thing we need to do is work out a suitable domain of F_1, F_2 in $\mathcal{S}$. For this, remember that we know $F(x, y; t)$ to be analytic for $|x|, |y| < 1$ (remember we chose t such that $|t| < \frac{1}{|\mathcal{S}|}$), and therefore $F_{1,2}$ are certainly convergent on $D_1 := \mathcal{S} \cap \{x < 1\}$ and $D_2 := \mathcal{S} \cap \{y < 1\}$ respectively[37], or, alternatively, on $\mathcal{D}_1 := \{s \in \mathcal{S} : |h_x(s)| < 1\}$ and $\mathcal{D}_2 := \{s \in \mathcal{S} : |h_y(s)| < 1\}$. By the way we chose the x_i, y_i, we know that only two of each are within the unit disk. As the corresponding cut therefore is contained in the interior of the unit disk, we know that on the two copies of $\overline{\mathbb{C}}$ we have one lifted unit circle each. As a result, we know that the set $\{s \in \mathcal{S} : |h_x(s)| = 1\}$ consists of two non-intersecting loops $\Gamma^x_{0,1}$ homological to the cut between x_1, x_2. The same applies for $\{s \in \mathcal{S} : |h_y(s)| = 1\}$, leading to loops $\Gamma^y_{0,1}$. It is shown in [44, Lemma 3] that one of the loops Γ^x_i, which we will denote by Γ^x_0, lies inside $\mathcal{D}_2$, while the other in its complement, and similarly, $\Gamma^y_0 \subset \mathcal{D}_1$, and Γ^y_1 lies completely outside $\mathcal{D}_1$. Consequently (see Fig. 9), $\Gamma^{x,y}_i$ do not intersect, and the interior of the segment of $\mathcal{S}$ bounded by Γ^x_0, Γ^y_0 consists of the set $\mathcal{D}_3 := \mathcal{D}_1 \cap \mathcal{D}_2 \neq \emptyset$. As $\mathcal{D}_3$ is open, we know that for all $s \in \mathcal{D}_3$, (174) holds:

$$0 = h_x(s)h_y(s) - F_1(h_x(s), t) - F_2(h_y(s), t) + \varepsilon F(t). \tag{186}$$

To once again shorten notation, we let $F_1(s) := F_1(h_x(s), t), F_2(s) := F_2(h_y(s), t)$. If $\mathcal{S}$ is identified with $\mathbb{C}/\Lambda$ or we take a point ω in the universal covering $\mathbb{C}$, we will simply write $F_{1,2}(\omega)$ instead.

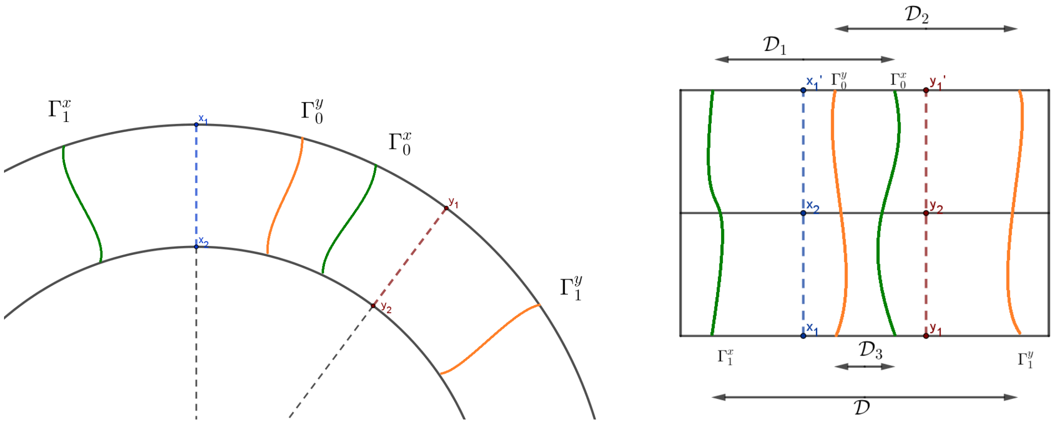
Lifting of the boundary functions to the universal cover

We know that $\Gamma^{x,y}_i$ are homological to the cut between x_1, x_2, so $\mathcal{D}_3$ forms a strip in a sense parallel to it, except its borders need not be straight, see again Fig. 9. Looking at the universal cover $\mathbb{C} \xrightarrow{\lambda} \mathbb{C}/\Lambda$, then

$$\widehat{\mathcal{D}_3} := \lambda^{-1}(\mathcal{D}_3) = \bigcup_{m,n \in Z} \mathcal{D}_3 + m\omega_1 + n\omega_2, \tag{187}$$

where $\mathcal{D}_3$ is seen as a subset of $\mathbb{C}/\Lambda$. Viewed geometrically, this is a union of copies the strip $\mathcal{D}_3$ continued up- and downwards, shifted to the right by integer multiples of ω_2. It is fairly intuitive, and can be checked computationally as in [28, 4.2], that each of these strips is bounded by corresponding segments of $\widehat{\Gamma^x_0} := \lambda^{-1}(\Gamma^x_0), \widehat{\Gamma^y_0} := \lambda^{-1}(\Gamma^y_0)$.
We know that we can therefore holomorphically continue $F_{1,2}$ on $\widehat{\mathcal{D}_3}$. But we can do even more.

[37] Here, $\mathcal{S}$ is viewed as a subset of $\mathbb{C}^2$.



(a) Possible placement of $\Gamma_{0,1}^{x,y}$ on $\mathcal{S}$.

(b) Corresponding placement of $\Gamma_{0,1}^{x,y}$ as well as $\mathcal{D}_{1,2}, \mathcal{D}_3 = \mathcal{D}_1 \cap \mathcal{D}_2, \mathcal{D} = \mathcal{D}_1 \cup \mathcal{D}_2$.

Figure 9: Possible placement of the pre-images of the unit circles in $\mathbb{C}$ on the covering space [28, Fig. 9].

Looking at just F_1, we could have holomorphically continued it on all of $\widehat{\mathcal{D}_1}$. The issue was that there is no immediate way to continue F_2 there, as it might not converge. However, (186) gives us a way to remedy that. We can simply define

$$F_2(\omega) := h_x(\omega)h_y(\omega) - F_1(\omega) + \varepsilon F(t), \quad \omega \in \widehat{\mathcal{D}_1}, \tag{188}$$

and analogously

$$F_1(\omega) := h_x(\omega)h_y(\omega) - F_2(\omega) + \varepsilon F(t), \quad \omega \in \widehat{\mathcal{D}_2}, \tag{189}$$

and thus have a continuation of the F_i to $\widehat{\mathcal{D}} := \widehat{\mathcal{D}_1} \cup \widehat{\mathcal{D}_2}$.

4 Walks in the Quarter Plane II

In this section, the main goal will be proving the other half of the conjecture by M. Bousquet-Mélou and M. Mishna stated in [24], namely that those walks with infinite group do not have holonomic generating functions. We will do so following the papers by I. Kurkova and K. Raschel [28] as well as by T. Dreyfus, C. Hardouin, J. Roques and M. F. Singer [32]. In both cases, the underlying idea is to count poles, and show that these do not allow for a holonomy equation. It is however interesting to see, how a different approach can be chosen for this purpose: the former electing to work via a meromorphic continuation of the boundary functions on $\mathbb{C}$, and study the poles there, while the latter instead works more algebraically, using Galois theory of difference equation and the function field of the elliptic curve. It is clearly impossible to claim that one method is in any way better than the other, and it only adds to the beauty of the problem that different approaches can be chosen. The results of the aforementioned papers are similar, bot not entirely the same: in [28], the conjecture is proven for all non-singular walks; those with finite group are furthermore classified into those with algebraic, and non-algebraic but holonomic generating function, using the sign of the introduced covariance of a walk. On the other hand, in [32], only walks with infinite group are treated, but they are also classified by whether or not their generating function is hyperalgebraic or simply non-holonomic. There also is a different paper by the same authors using similar methods for the remaining, singular, walks [45].

4.1 The Analytic Method

4.1.1 A Meromorphic Continuation on $\mathbb{C}$

In the following, we will proceed as in [31, 3.2],[28, 5.1].
From the previous section, we already know that we can meromorphically continue $F_{1,2}$ to $\widehat{\mathcal{D}} := \widehat{\mathcal{D}_1} \cup \widehat{\mathcal{D}_2}$. The idea is to find a way to suitably define $F_i(\widehat{\Theta}(\omega)) = F_i(\omega + \omega_3)$. If we can do this, then by the following Lemma we are already finished:

Lemma 4.1.1 ([31, Th. 3.2.1],[28, Lemma 5])**.**

$$\bigcup_{n\in\mathbb{Z}} \widehat{\mathcal{D}} + n\omega_3 = \mathbb{C}. \tag{190}$$

Proof.
Remember that the liftings Ψ, Φ on $\mathcal{S}$ are simply switching between the layers of the covering. By the construction of the $\Gamma_i^{x,y}$, namely, both of them corresponding to a unit circle on each copy of $\overline{\mathbb{C}}$, this means that $\Psi(\Gamma_i^x) = \Gamma_j^x$ for $i \neq j$. Similarly, $\Phi(\Gamma_i^y) = \Gamma_j^y$. One can easily seen that this property is inherited by the liftings, ie $\widehat{\Psi}(\widehat{\Gamma_i^x}) = \widehat{\Gamma_j^y}$, and $\widehat{\Phi}(\widehat{\Gamma_i^y}) = \widehat{\Gamma_j^y}$. As a consequence, we have that $\widehat{\Psi}(\widehat{\Gamma_1^x}) = \widehat{\Gamma_0^x} \subset \widehat{\mathcal{D}_y}$. As $\widehat{\mathcal{D}_y}$ is bounded by $\widehat{\Gamma_{0,1}^y}$, which are invariant under $\widehat{\Phi}$, we know that $\widehat{\Gamma_1^x} + \omega_3 = \widehat{\Psi} \circ \widehat{\Phi}(\widehat{\Gamma_1^x}) \subset \widehat{\mathcal{D}_y} \subset \widehat{\mathcal{D}}$. This, however, means that the borders of $\widehat{\mathcal{D}}, \widehat{\mathcal{D}} + \omega_3$ overlap. Iterating this, we see that the statement must hold.

□

Now, in order to come up with a suitable continuation, we will use (186), and again utilize the fact that h_x, and consequently also F_1, is invariant under $\widehat{\Psi}$.

First, we choose ω in a neighbourhood of $\widehat{\Gamma_1^x}$. Applying $\widehat{\Psi}$ and inserting into (186) – remember that by the proof of the previous lemma, we know that $\widehat{\Psi}(\omega)$ will be inside $\widehat{\mathcal{D}}$ – we have

$$0 = h_x(\widehat{\Psi}(\omega)h_y(\widehat{\Psi}(\omega) - F_1(\widehat{\Psi}(\omega)) - F_2(\widehat{\Psi}(\omega)) + \varepsilon F(t). \tag{191}$$

Using the aforementioned properties of h_x, F_1, and subtracting from this (186), we obtain

$$F_2(\widehat{\Psi}(\omega)) = F_2(\omega) + h_x(\omega)(h_y(\widehat{\Psi}(\omega)) - h_y(\omega)). \tag{192}$$

As h_y, and thus also F_2, is invariant under $\widehat{\Phi}$, and $\widehat{\Phi}(\omega)$ will still be inside $\widehat{\mathcal{D}}$ (again by the proof of the previous lemma), we know that $F_2(\widehat{\Psi}(\omega)) = F_2((\widehat{\Phi} \circ \widehat{\Psi})(\omega))$. Hence, by the previous equation we finally obtain

$$F_2(\omega + \omega_3) = F_2((\widehat{\Phi} \circ \widehat{\Psi})(\omega)) = F_2(\omega) + h_x(\omega) \left(h_y(\widehat{\Psi}(\omega) - h_y(\omega) \right). \tag{193}$$

In much the same fashion, we deduce

$$F_1(\omega - \omega_3) = F_1((\widehat{\Psi} \circ \widehat{\Phi})(\omega)) = F_1(\omega) + h_y(\omega) \left(h_x(\widehat{\Phi}(\omega) - h_x(\omega) \right) \tag{194}$$

for ω in a neighbourhood of $\widehat{\Gamma_1^y}$. As we already have meromorphic continuations of F_1, F_2 between $\widehat{\Gamma_1^y}$ and $\widehat{\Gamma_1^y} - \omega_3$ and between $\widehat{\Gamma_1^x}$ and $\widehat{\Gamma_1^x} + \omega_3$ respectively, we can use these parts and the above equations for a meromorphic continuation of $F_{1,2}$ on all of $\mathbb{C}$. As this continuation and the previously constructed one coincide on an (open) neighbourhood of $\widehat{\Gamma_1^y}$ and $\widehat{\Gamma_1^x}$ respectively, they must in fact be the same on all of $\widehat{\mathcal{D}}$. As h_x, h_y are ω_1-periodic, and the translation $\omega \mapsto \omega + \omega_3$ is perpendicular to ω_1 (remember that ω_3 is real), we have F_1, F_2 are ω_1-periodic on the whole of $\mathbb{C}$.
We still need to check that our continuation fulfils the lifted kernel equation (186) on all of $\mathbb{C}$. To see this, it suffices to show that the equation is satisfied after applying to ω both $\widehat{\Psi}$ and $\widehat{\Phi}$. First, by the definition of F_1, F_2, we see that the continuations are still invariant under $\widehat{\Psi}$ and $\widehat{\Phi}$ respectively. Therefore, we have

$$\begin{aligned}
&h_x(\widehat{\Psi}(\omega))h_y(\widehat{\Psi}(\omega)) - F_1(\widehat{\Psi}(\omega)) - F_2(\widehat{\Psi(\omega)}) + \varepsilon F(t) && (195)\\
=&h_x(\omega)h_y(\widehat{\Psi}(\omega)) - F_1(\omega) - F_2(\widehat{\Psi}(\omega)) + \varepsilon F(t) && (196)\\
\overset{(192)}{=}&h_x(\omega)h_y(\omega) - F_1(\omega) - F_2(\omega) + \varepsilon F(t) && (197)\\
=&0. && (198)
\end{aligned}$$

In the same manner we prove the lifted kernel equation for $\widehat{\Phi}(\omega)$, and consequently it must hold on all of $\mathbb{C}$. We have therefore shown

Theorem 4.1.1. *The functions F_1, F_2 as defined in section 3.3.2 can be meromorphically continued to all of $\mathbb{C}$. We have*

1. $F_1(\omega - \omega_3) = F_1(\omega) + h_y(\omega) \left(h_x(\widehat{\Phi}(\omega)) - h_x(\omega) \right)$,
2. $F_2(\omega + \omega_3) = F_2(\omega) + h_x(\omega) \left(h_y(\widehat{\Psi}(\omega)) - h_y(\omega) \right)$,
3. $h_x(\omega)h_y(\omega) - F_1(\omega) - F_2(\omega) + \varepsilon F(t) = 0$,

4. $F_1(\omega+\omega_1) = F_1(\omega)$,
$F_2(\omega+\omega_1) = F_2(\omega)$.

For some additional properties that are not explicitly listed here, see [28, Theorem 4] or [32, Remark 2.7].

In the next step, we now want to see how all this connects to our original boundary functions $F(x,0), F(0,y)$. The spaces we have worked and constructed so far are as follows: We started from $\mathbb{C}$, and then looked at the Riemann surface/algebraic curve where the kernel vanishes. From there, we then proceeded to the universal cover, which again is $\mathbb{C}$. Summed up, we have

$$\mathbb{C} \xrightarrow{\lambda} \mathbb{C}/\Lambda \cong \mathcal{S} \xrightarrow{h_x,h_y} \mathbb{C}. \tag{199}$$

As the following might otherwise be notationally confusing, we will write F_1 for the original function F_1 defined via the kernel equation in (174), and $\widehat{F_1}$ for the meromorphic continuation of $\overline{F_1}$ on the whole universal cover $\mathbb{C}$.

We started out by being interested in the behavior of F_1 as a function on $\mathbb{C}$. To study this, we can essentially go back down first from $\mathbb{C}$ to $\mathbb{C}/\Lambda \cong \mathcal{S}$, and then from $\mathcal{S}$ to $\mathbb{C}$ again. First, we note that composition of $\widehat{F_1}$ with λ induces a multi-valued function $\overline{F_1}$ on $\mathbb{C}/\Lambda \cong \mathcal{S}$. One can construct all branches $\overline{F_1}$ by continuation along paths: fix any point $s \in \mathcal{S}$ and a preimage $\widehat{s} \in \lambda^{-1}(s)$. Now for any other point $t \in \mathcal{S}$, select *any* path $\gamma : s \to t$ in $\mathcal{S}$, lift it to a path $\widehat{\gamma}$ on the universal cover $\mathbb{C}$, such that $\widehat{\gamma} : \widehat{s} \to \widehat{t}$. Note that, as $\widehat{\gamma}$ is a lifting of γ, we must have $\lambda(\widehat{t}) = t$. We can then choose $\overline{F_1}(t) = \widehat{F_1}(\widehat{t})$.

Given the multi-valued function $\widehat{F_1}$, composition with h_x now gives a multi-valued function F_1 on $\mathbb{C}$, again by pathwise continuation. The aim of the following is to see on what domain this function is single-valued, and what its branches are. To do so notice that we can pass from one branch of $\overline{F_1}$ to another only by leaving one fundamental parallelogram and transitioning into another. As by Theorem 4.1.1, $\widehat{F_1}$ is ω_1-periodic, transitions upwards (that is, along circles homotopic to either of the cuts of the gluing of $\mathcal{S}$, see again Fig. 8) do not at all matter, so we can just look at one horizontal slice of the universal covering. However, where things can happen is by going sideways, and transitioning over one of the cuts x_1x_2 or x_3x_4. It is therefore a reasonable idea to look at segments of the form

$$\mathfrak{D}_k := \left[\frac{k}{2}, \frac{k+1}{2}\right)\omega_2 \times [0,1)\omega_1. \tag{200}$$

This is one half of the fundamental parallelogram; staying in between the two cuts. Indeed, the behavior of $\widehat{F_1}$ on each of these slices determines a branch of $\overline{F_1}$, and thus also a branch of F_1 (notice how each $\mathfrak{D}_k$ corresponds precisely to one copy of $\mathbb{PC}$ glued together in the construction of $\mathcal{S}$!). By this, we immediately obtain that F_1 is a single-valued function on $\mathbb{PC}\backslash([x_1,x_2]\cup[x_3,x_4])$, as leaving out these cuts keeps us from transitioning from one $\mathfrak{D}_k$ to another. We can even do slightly more: we have already established in the proof of Lemma 3.3.1 that the lifting $\widehat{\Psi}$ of Ψ corresponds to the mapping $\omega \mapsto -\omega+\omega_1+\omega_2$, and by the previous Theorem 4.1.1 $\widehat{F_1}$ is invariant under both $\widehat{\Psi}$ and transition by ω_1, we have

$$\widehat{F_1}(\omega) = \widehat{F_1}(\widehat{\Psi}(\omega)) = \widehat{F_1}(-\omega+\omega_2+\omega_1) = \widehat{F_1}(-\omega+\omega_2). \tag{201}$$

This immediately goes to show that it suffices to consider the $\mathfrak{D}_k$ with $k > 0$, as the negative indices do not give us new branches. In addition, as $\mathfrak{D}_0$ and $\mathfrak{D}_1$ result in the same branch, we know that the resulting branch of F_1 (which corresponds with the very first single-valued function F_1 defined on the unit circle in (174) actually extends to a single-valued map not only on $\mathbb{PC} \setminus ([x_1, x_2] \cup [x_3, x_4])$ but even on $\mathbb{PC} \setminus [x_3, x_4]$. By the choice of the x_i, $[x_3, x_4]$ lies outside the unit disk, so this branch must coincide with the original F_1 on the entire unit disk.

Now, remember that we have (see (174))

$$F_1(x) = \mathcal{S}_y(x) F(x, 0). \tag{202}$$

Again, (see (174),(186)) both $\mathcal{S}_y, F_(x, 0)$ are both ω_1-periodic. Also, since $\mathcal{S}_y \circ h_x$ is precisely the lifting of the rational (thus meromorphic and single-valued function) $\mathcal{S}_y$ on $\mathbb{C}$, we can now simply let

$$F(x, 0) = \frac{F_1(x)}{S_y(x)}. \tag{203}$$

This is now a meromorphic continuation of $x \mapsto F(x, 0)$ on $\mathbb{C} \setminus ([x_1, x_2] \cup [x_3, x_4]$, with the branches defined by the values of $\frac{F_1}{\mathcal{S}_y}$ on the $\mathfrak{D}_k$ as explained above. In particular, we know that the branch corresponding to $\mathfrak{D}_{0,1}$ of this meromorphic continuation of F_1 must again coincide with the "original" F_1 on at least the entire unit disk, so its series representation around 0 must be the same, and is single-valued on $\mathbb{PC} \setminus [x_3, x_4]$.
Similarly, the branches of $y \mapsto F(0, y)$ on $\mathbb{C} \setminus ([y_1, y_2] \cup [y_3, y_4])$ correspond to the values of $\frac{F_2}{\mathcal{S}_x}$ on

$$\mathfrak{L}_k := \frac{\omega_3}{2} + \left[\frac{k}{2}, \frac{k+1}{2}\right) \omega_2 \times [0, 1)\omega_1 = \frac{\omega_3}{2} + \mathfrak{D}_k \subset \mathbb{C}. \tag{204}$$

In summary, we have now shown

Theorem 4.1.2. *For any fixed t with $|t| < \frac{1}{|\mathcal{S}|}$, the boundary functions $x \mapsto F(x, 0; t), y \mapsto F(0, y; t)$ can be meromorphically continued to multi-valued functions F_x, F_y on all of $\mathbb{C}$. Their branches are defined by the liftings $\widehat{F_1}, \widehat{F_2}$ on the rectangles $\mathfrak{D}_k, \mathfrak{L}_k, k > 0$, after division by $\mathcal{S}_y(x)$ and $\mathcal{S}_x(y)$ respectively.*
The branches are single-valued on $\mathbb{C} \setminus ([x_1, x_2] \cup [x_3, x_4])$. On the first branches, given by $\mathfrak{D}_1, \mathfrak{L}_1$ we have $F_x(x) = F(x, 0; t)$ and $F_y(y) = F(0, y; t)$ on the unit disk.

4.1.2 Proof Strategy

The main idea of [28] is to prove that, for all t such that $\frac{\omega_2}{\omega_3} \notin \mathbb{Q}$, the liftings F_x, F_y have an infinite number of singularities, which lie dense on a certain curve inside the fundamental parallelogram. This immediately shows that $F_{x,y}$ cannot be holonomic by Theorem 1.1.2 on page 8. A priori it might not be clear that this already implies that $x \mapsto F(x, 0; t), y \mapsto F(0, y; t)$ is non-holonomic, but this turns out to be the case: it can be shown by looking at local series representations that analytic continuation commutes with taking a derivative, i.e. the derivative of an analytic continuation is the analytic continuation of the derivative [46, 5.7.2]. From this and the Principle of Permanence, which in this case states that a functional relations between germs of analytic functions is preserved by analytic continuation, one obtains that any holonomy equation which is satisfied by $x \mapsto F(x, 0; t)$ or $y \mapsto F(0, y; t)$ would also be satisfied on the entirety of $F_{x,y}$ respectively [46, 5.9].
This means that any branch of $F_{x,y}$ can only have finitely many poles, namely at the the zero-set of the coefficients (see the proof of Theorem 1.1.2). However, if the poles lie dense on some curve in the fundamental parallelogram, this clearly is not the case – hence, $F(x, 0; t), F(0, y; t)$ cannot

be holonomic. As holonomic functions are closed under specialization of variables (Lemma 1.1.6), this in turn implies that $F(x, y; t)$ is non-holonomic. If one succeeds in showng the above, one therefore has proven

Theorem 4.1.3. *For all non-singular walks with infinite group, the generating function $F(x, y; t)$ is non-holonomic.*

The main idea of the proof is nicely illustrated in [28, 7.1]: assume the group of a walk is infinite, and select t such that $\frac{\omega_2}{\omega_3} \notin \mathbb{Q}$ (this is possible by Lemma 3.3.2 on page 55). Let $\Pi := \mathfrak{D}_0 \cup \mathfrak{D}_1$, so an entire fundamental parallelogram. By Theorem 4.1.1 we know that

$$F_1(\omega - \omega_3) - F_1(\omega) = h_y(\omega)\left(h_x(-\omega + \omega_2 + \omega_3) - h_x(\omega)\right). \tag{205}$$

Denote by f_x the function on the right hand side of this equation. Furthermore, we will assume that there is a point $\omega_0 \in \Pi$, such that:

1. $f_x(\omega_0) = \infty$,
2. $F_1(\omega_0) \neq \infty$,
3. $\forall \omega \in \Pi : (\operatorname{Im}\omega = \operatorname{Im}\omega_0 \wedge f_x(\omega) = \infty) \Rightarrow \omega = \omega_0$,

ie ω_0 is the unique pole of f_x on some horizontal line of Π. We will now utilize the functional equation for F_1 above to translate ω_0 to the right n times, $n \in \mathbb{Z}$. As $\frac{\omega_2}{\omega_3}$ is irrational, we will never hit the same point of the fundamental parallelogram (which has length ω_2) again. This means that the points we are hit are dense on the line $\{\operatorname{Im}\omega = \operatorname{Im}\omega_0\} \subset \Pi$. But it is easy to show that each of these points is a pole of F_1:

$$F_1(\omega_0 - n\omega_3) = F_1(\omega_0) + f_x(\omega_0) + \sum_{k=1}^{n-1} f_x(\omega_0 + k\omega_3). \tag{206}$$

By assumption, $F_1(\omega)$ is finite, and so is each of the summands, and therefore the entire (finite) sum. $f_x(\omega_0)$, however, is infinite – therefore, $\omega_0 - n\omega_3$ is a pole of F_1. As we have $F(x, 0; t) = \frac{F_1(x)}{\mathcal{S}_y(x)}$, and $\mathcal{S}_y$ has at most two zeros, the poles of $F(x, 0; t)$ must still be dense on $\{\omega \in \Pi : \operatorname{Im}\omega = \operatorname{Im}\omega_0\}$. An argument for F_2 can be made in the very same fashion. Unfortunately, it turns out that things are not quite so simple: in most cases, an ω_0 as above does not exist. It is fairly easy to determine which points are poles of f_x. Remember that h_x, h_y are the x- and y-coordinates of a point $x, y \in \mathcal{E} \cong \mathcal{S}$. So for a point $(x, y) \in \mathcal{E}$, we have

$$f_x = y(\Phi(x) - x), \tag{207}$$

where $\Phi(x)$ denotes the x-coordinate of $\Phi(x, y)$, and $x, \Phi(x)$ are given by the solutions of $\mathcal{K}(x, y) = 0$ (150). Consequently, the only way for f_x to be ∞ is that either $y, x, \Phi(x)$ is ∞. This leads to the following possibilities

1. $\underline{y = \infty}$:
   By direct computation we get the two solutions[38] for x

$$x = x^\circ := \lim_{y\to\infty} \frac{1}{2B_1(y)}\left(-B_2(y) + \sqrt{B_3(y)^2 - 4B_1(y)B_3(y)}\right), \tag{208}$$

$$\vee \quad x = x^\bullet := \lim_{y\to\infty} \frac{1}{2B_1(y)}\left(-B_2(y) - \sqrt{B_3(y)^2 - 4B_1(y)B_3(y)}\right). \tag{209}$$

[38] If we were to consider the homogenized kernel equation $\overline{\mathcal{K}}$ instead, and work over $\mathbb{PC}$, we would not need to bother with the limits, as then inserting ∞ into our equation would make perfect sense. However, we will follow [28] here. Note that compared to the latter, the roles of x and y are reversed.

2. $\underline{x = \infty}$:
In the same manner as above, we have

$$y = y^* := \lim_{x \to \infty} \frac{1}{2A_1(x)} \left(-A_2(x) + \sqrt{A_3(x)^2 - 4A_1(x)A_3(x)} \right), \tag{210}$$

$$\vee \quad y = y^\star := \lim_{x \to \infty} \frac{1}{2A_1(x)} \left(-A_2(x) - \sqrt{A_3(x)^2 - 4A_1(x)A_3(x)} \right). \tag{211}$$

3. $\underline{\Phi(x) = \infty}$:
We already know that $y^*, y^\star$ correspond to $x = \infty$. As Φ is a horizontal switch, we now need the *other* y-coordinates leading to $x^*, x^\star$. By some computation, we work out the signs of the root, arriving at

$$x = x^* := \lim_{y \to y^*} \frac{1}{2B_1(y)} \left(-B_2(y) + \sqrt{B_3(y)^2 - 4B_1(y)B_3(y)} \right), \tag{212}$$

$$\vee \quad x = x^\star := \lim_{x \to x^\star} \frac{1}{2B_1(y)} \left(-B_2(y) + \sqrt{B_3(y)^2 - 4B_1(y)B_3(y)} \right). \tag{213}$$

The more technical part of the proof now consists of determining the exact position of these 6 points for each of the walks, and then finding a curve on which the singularities are dense. This still is not an entirely straightforward computation, requiring a few ideas and intermediate results to get a handle of the precise structure of the poles, but will not be elaborated upon here. Instead, we will content ourselves with noting that this works out in all cases, and the results as well as technical details can be found in [28, Lemma 17].

4.2 The Algebraic Method

In this section, we will follow the approach in [32].
Instead of working directly with meromorphic functions on $\mathbb{C}$, one can make use of a duality which features rather prominently in the study of irreducible algebraic curves, namely that between the curve and its function field. To give an idea of how why the following definition is sensible, suppose we want to have coordinate maps on a curve $\mathcal{C} := \{f = 0\}$ for some irreducible, smooth polynomial f. The most natural way to do so would be by working with polynomials in x, y[39] – but it does not make sense to consider any polynomial, as any multiple of f will be 0 on all of $\mathcal{C}$. Therefore, we can consider the ring $\mathbb{C}[x, y]$, where x, y satisfy the relation $f(x, y) = 0$ – formally speaking, we have $\mathbb{C}[x, y] / \langle f(x, y) \rangle$, where $\langle f(x, y) \rangle$ is the ideal generated by $f(x, y)$. This ring is a domain (since we required f to be irreducible), and we therefore have a quotient field.

Definition 4.2.1. *Given an irreducible (projective) algebraic curve $\mathcal{C} : f(x, y) = 0$ over $\mathbb{C}$, the **function field** of $\mathcal{C}$ is the quotient field*

$$F(\mathcal{C}) := \mathbb{C}\,(x, y) \,/\, \langle f(x, y) \rangle\,. \tag{214}$$

It turns out that if f, g are different irreducible polynomials, one will arrive at a non-isomorphic function field as long as there is no birational mapping from one curve to the other, and we therefore have a correspondence

$$\{\text{function fields over } \mathbb{C}\} \leftrightarrow \{\text{projective, irreducible algebraic curves over } \mathbb{C}\},$$

[39] Technically, since we are working with projective curves here, we do have a third variable, which is obtained by homogenization. This makes the construction a bit more technical; see eg [29, App. B].

where the function fields are unique up to isomorphism, and the curves up to birational equivalence [39, II.1.7, Remark 1].
Given such a function field, consisting of rational functions on $\mathcal{C}$, one can then look at poles of zeros of any element $f \in F(\mathcal{C})$. Formally speaking, we have

$$\mathrm{Div}(f) := \mathrm{Div}_0(f) - \mathrm{Div}_\infty(f) = \sum_{P\in\mathcal{C}, P \text{ is a zero of } f} n_P P - \sum_{P\in\mathcal{C}, P \text{ is a pole of } f} n_P P, \tag{215}$$

While it may not immediately obvious how to determine the multiplicity of a zero (or, by inverting, of a pole), the geometry is quite clear here: given any irreducible element $g \in \mathbb{C}[x, y]/\langle f\rangle$, we can again consider it as an algebraic curve. Now, if g has a zero at P in the funciton field, then f, g must at P. n_P is then just the multiplicity of this intersection. This easily extends to all of $F(\mathcal{C})$: if g is not irreducible, then we sum over multiplicities of the irreducible components. If $g = \frac{g_1}{g_2}$, then we simply let $n_P^g := n_P^{g_1} - n_P^{g_2}$. The one thing to take care of is to not forget about the points at infinity, as we are still working in projective space. These can be treated by a chart transition, i.e. homogenizing, and then dehomogenizing with respect to x, y. From this construction, by Bézout's Theorem, one immediately obtains:

1. $\mathrm{Div}(f)$ is a finite sum,
2. the n_P sum up to 0,
3. the mapping $\nu_P : f \mapsto n_P$ for a fixed P is a valuation, the so-called **valuation at P**.

One can define all of this purely algebraically, without the need for a geometric intuition, leading to the concept of places, which is vital especially when $\mathbb{C}$ is not algebraically closed. In fact, the reason why the above construction was fairly intuitive is that any place in $F(\mathcal{C})$ necessarily corresponds to a point on $\mathcal{C}$, but this shall not be elaborated upon here. For more about function fields in general, see eg [3, 29], for the correspondence between curves and function fields see eg [29, App. B].
The main idea of the following is now to look at a possible holonomy equation not on $\mathbb{C}$, but on the function field $F(\mathcal{E})$ of the elliptic curve $\mathcal{E}$ associated with the kernel. First, we can see that $\Theta := \Phi \circ \Psi$ as defined in (81) is an automorphism on $F(\mathcal{E})$. Furthermore, we know that Θ corresponds to addition by a point. Now, let us consider a potential holonomy equation for $F(x, 0)$. By lemma 1.1.5, we know that there is then a holonomy equation using only derivatives with respect to x. This derivative uniquely defines a derivation δ on $F(\mathcal{E})$ (see [29, Prop. 4.1.4]). As addition by a point is just a translation on $\mathbb{C}$, and the derivative with respect to x is invariant under translation, by an application of [29, Lemma 4.1.3], the same applies to the induced derivation, so δ commutes with Θ.
This particular structure leads us to the following

Definition 4.2.2 ([32, Def. 3.4])**.** *A field F with a derivation δ and an automorphism τ such that*

$$\delta \circ \tau = \tau \circ \delta \tag{216}$$

*is called a $\delta\tau$-**field**.*
*Its **constants** are the constants with respect to τ, meaning $\{c \in F : \tau(c) = c\}$.*

Letting $\tau := \Theta$, and δ the same as before, we have just seen that $(F(\mathcal{E}), \delta, \tau)$ is a $\delta\tau$-field. in order to save time, we will in the following use $\tau = \Theta^{-1}$, which clearly works just as well.

By arguments similar as in the previous section, to consider holonomy of $F(x,0)$ it suffices to consider holonomy of $F_1(x)$. Again by Theorem 4.1.1, we have

$$F_1 \circ \tau - F_1 = b, \tag{217}$$

where b is some rational function that we can compute. Now, application of Galois theory of difference equations tells us that if F_1 is not hyperalgebraic, then can find $n \in \mathbb{N}$ and $c_i \in \mathbb{C}$ such that

$$\delta^n(b) + c_{n-1}\delta^{n-1}(b) + \cdots + c_0 b = h \circ \tau - h \tag{218}$$

for some rational function h. This type of equation is called a **telescoper equation**. It is particularly interesting that this equation is now a linear differential equation, rather than the polynomial differential equation we want $F_1(x)$ to satisfy. So what we want to do is to find some property of g which makes the existence of a telescoper equation impossible. It turns out that such an equation exists precisely if all orbit sums of g are 0 – and this is something that can be checked for our walks.
Consequently, in the following we will first introduce some basic notions about the Galois theory of difference equations, in particular formalizing and proving the above statements. Afterwards, we will then apply this to F_1, F_2 and therefore again show that walks with infinite groups are not holonomic – and, in all except 9 cases, even hypertranscendental.

4.2.1 Galois Theory of Difference Equations

We will follow [32, App. A,B] rather closely, only accruing the results needed for an application in the above context. For a wider overview about this topic, see eg [47, 48].

Definition 4.2.3. *A **difference ring (field)** is a ring (field) R together with an automorphism $\tau : R \to R$. A **difference ideal** is an ideal of R which is closed under τ.*
*Its **constants** R^τ are the elements $\{c \in R : \tau(c) = c\}$.*

The name stems from the simple example $R = \mathbb{C}[x], \tau : x \mapsto x + 1$. In this case, the constants are just $\mathbb{C}$ itself, as for any non-constant polynomial $p \in \mathbb{C}[x]$, we have $p(x+1) \neq p(x)$.

Lemma 4.2.1. *The field $\mathcal{M}(\mathcal{E})$ of meromorphic functions on $\mathcal{E}$ together with the automorphism $\tau := \Theta$ as defined in previous sections forms a difference field. If the group (i.e. the order of Θ) is infinite, then its field of constants is $\mathbb{C}$.*

Proof.
The only non-obvious statement is the one concerning the constants. Assume f is an elliptic function invariant under the transformation $\tau : \omega \mapsto \omega + \omega_3$. Then, as the orbit of τ of any point is dense on the horizontal line through the point in the fundamental parallelogram, τ would only depend on the real part of ω. However, in section 3.2.1 we have seen that elliptic functions are generated by the Weierstraß $\wp$-function and its derivative. As $\wp$ can easily be seen to exhibit the same behaviour with respect to the edges of the fundamental parallelogram, and the same consequently applies for $\wp'$, f must therefore be constant.

□

What we now want to consider are linear difference equations, i.e. equations of the form

$$\tau^n(y) + c_{n-1}\tau^{n-1}(y) + \cdots + c_0 y = 0, \tag{219}$$

where the c_i lie in some difference field K. Analogously as for differential equation, we can write this as a system of first-order linear difference equations, letting $y_0 := y, y_1 := \tau(y)$ etc., $Y := (y_0, \dots, y_n)^T$, we arrive at

$$\tau(Y) = \begin{pmatrix} 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & \ddots & 0 \\ \vdots & \vdots & \vdots & \ddots & 0 \\ 0 & 0 & \dots & 0 & 1 \\ -c_0 & -c_1 & -c_2 & \dots & -c_{n-1} \end{pmatrix} Y = A_L Y, \tag{220}$$

where $A_L \in GL_n(K)$ is the **companion matrix** of the difference equation (219). Continuing as we would for ordinary differential equations, we define a **fundamental solution matrix** to be any $U \in GL_n(K)$, such that $\tau(U) = A_L U$.
When studying algebraic equations using Galois theory, this is the point where we would like to jump into a splitting field and then look at automorphisms fixing our base field. However, here it is not entirely clear what our quotient field should be. The following – albeit slightly unintuitive – definition gives a proper substitute:

Definition 4.2.4. *For a difference field (K, τ), together with a linear difference equation $\tau(Y) = A_L Y$, a **Picard-Vessiot ring** R is a $K - \tau-$algebra, such that*

1. *R contains no non-trivial τ-ideals,*
2. *There is a $U \in GL_n(R) : \tau(U) = A_L(U)$,*
3. *$R = K\left[U, \frac{1}{\det U}\right]$.*

The second condition loosely translates to the property that an algebraic equation is solvable over the splitting field, and the third to its minimality (compare [48, Def. 1.5]). The first is needed in order to retain some nice algebraic properties. It is not at all immediately obvious that a Picard-Vessiot ring always exists, but this turns out to be the case.

Theorem 4.2.1. *A Picard-Vessiot ring R exists for any linear difference equation and difference field F. If K^τ is algebraically closed, then any two Picard-Vessiot rings are τ-isomorphic. In that case, we have $K^\tau = R^\tau$.*

Proof.
See [48, 1.1].

□

As this will apply for the case we are interested in and turn out to be very useful, we will in the following always assume that K^τ is algebraically closed, and $\mathbb{C} \subseteq K$.
Having now found a suitable substitute for our splitting field, we can define the Galois group as usual.

Definition 4.2.5. *Let R be a Picard-Vessiot ring of a differential equation $\tau(Y) = A_L Y$ in some differential field K. Then the **Galois group** of R (or of $\tau(Y) = A_L Y$) is*

$$\mathcal{G} := \{\sigma \in \mathrm{Aut}_K(R) : \tau \circ \sigma = \sigma \circ \tau\}, \tag{221}$$

where $\mathrm{Aut}_K(R)$ denotes the K-algebra automorphisms of R.

This Galois group inherits many of the properties of a “normal” Galois group. For instance, subgroups of $\mathcal{G}$ correspond to intermediate difference rings F with $K \subseteq F \subseteq R$, and if $\sigma(z) = z$

for all $\sigma \in \mathcal{G}$, then this implies $z \in K$. Similarly, normal subgroups correspond to what is comparable to a normal extension, in the sense that for any element $z \in F \setminus K$ we can find a $\sigma \in \mathcal{G}$ such that $\sigma \in \mathrm{Aut}_K(F), \sigma(z) \neq z$. A much more thorough description of the Galois correspondence and its similarity with usual Galois groups and proofs of these statements can be found for instance in [47, 10-13],[48, 1.2-1.3].
Also, in the same fashion that the Galois group of a splitting field maps zeros of polynomials to zeros, $\mathcal{G}$ maps fundamental solutions to others. The proof is the same as in the field extension case. Let U be a fundamental solution of $\tau(U) = A_L U$, then we have

$$\tau(\sigma(U)) = \sigma(\tau(U)) = \sigma(A_L U) = A_L \sigma(U), \tag{222}$$

as $A_L \in GL_n(K)$ is fixed by σ. This, however, implies that

$$\tau(U^{-1}\sigma(U)) = U^{-1} A_L A_L^{-1} U = U^{-1}\sigma(U), \tag{223}$$

which means that $U^{-1}\sigma(U)$ is fixed by τ, which means $U^{-1}\sigma(U) := D_\sigma \in GL_n(R^\tau) = GL_n(K^\tau)$ (by the same argument, this holds true for any two fundamental solutions). We can therefore define a mapping

$$\rho : \mathcal{G} \to GL_n(K^\tau) : \sigma \mapsto D_\sigma. \tag{224}$$

Lemma 4.2.2. *The mapping ρ defined above is an injective group homomorphism.*

Proof.
First, we show that ρ is a group homomorphism.

1. $\underline{\rho(\sigma_1)\rho(\sigma_2) = \rho(\sigma_1 \circ \sigma_2)\text{:}}$
   <u>We need to show</u>

   $$U^{-1}(\sigma_1 \circ \sigma_2)(U) = U^{-1}\sigma_1(U)U^{-1}\sigma_2(U). \tag{225}$$

   By multiplication with $U\sigma_1(U^{-1})$, this is equivalent to

   $$\sigma_1(U^{-1}\sigma_2(U)) = U^{-1}\sigma_2(U), \tag{226}$$

   but as $U^{-1}\sigma_2(U) = \rho(\sigma_2) \in GL_n(K^\tau)$, it is invariant under $\mathcal{G}$ (so in particular under σ_1) and the statement follows.

2. $\underline{\rho(\sigma^{-1}) = \rho(\sigma)^{-1}\text{:}}$
   <u>We have</u>

   $$\rho(\sigma)^{-1}\rho(\sigma) = \sigma(U)^{-1}UU^{-1}\sigma(U) = \sigma(U)^{-1}\sigma(U) = \mathrm{id}\,. \tag{227}$$

To prove that ρ is injective, all we need to show is that its kernel is trivial. Suppose we have a σ such that

$$\rho(\sigma) = U^{-1}\sigma(U) = \mathrm{id}\,. \tag{228}$$

This is equivalent to $\sigma(U) = U$, i.e. σ fixes U. By definition 4.2.4, this implies that σ fixes all of $R = K\left[U, \frac{1}{\det U}\right]$ (keep in mind that the determinant of a matrix is a polynomial function of its entries). Therefore, $\sigma = \mathrm{id}$.

□

Theorem 4.2.2. *The subgroup $\rho(\mathcal{G}) \subseteq GL_n(K^\tau)$ is a **linear algebraic group**, i.e. a subgroup of $GL_n(K^\tau)$ which is defined by polynomial equations*[40].

Proof.
See [48, 1.2-1.3].

□

Theorem 4.2.2 allows us to view $\mathcal{G}$ as an algebraic variety over K^τ. As such, we can assign to it a dimension – there are multiple equivalent ways to do so, maybe the most intuitive ones being the maximal length of a chain of distinct, nonempty subvarieties, or the maximal length of a chain or distinct prime ideals in the coordinate ring. A number of these equivalences and examples can be found in [43, 11]. We will not go into these here, but the following statement turns out to be exceedingly useful for us:

Theorem 4.2.3. *If $\rho(\mathcal{G})$, viewed as an algebraic variety over K^τ, is connected, then we have*

$$\operatorname{trdeg}\left(\operatorname{Quot}(R)/K\right) = \dim \rho(\mathcal{G}). \tag{229}$$

Proof.
See [48, 1.2-1.3].

□

The importance of the above theorem should be clear: we are interested in finding out whether or not there is some kind of algebraic relation between functions, and the existence of the latter is now immediately tied to the geometric properties of the Galois group $\mathcal{G}$[41].
We will now, still following [32, App.], turn to two examples which will play a major part in what is to follow.

Example 1.
Consider a difference system $\tau(Y) = A_L Y$, given by

$$A_L := \begin{pmatrix} 1 & b \\ 0 & 1 \end{pmatrix}, b \in K. \tag{230}$$

Then, for any z with $\tau(z) - z = b$, or equivalently, $\tau(z) = z + b$, we find that $\begin{pmatrix} 1 & z \\ 0 & 1 \end{pmatrix}$ is a fundamental solution, as

$$A_L Y = \begin{pmatrix} 1 & b \\ 0 & 1 \end{pmatrix}\begin{pmatrix} 1 & z \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & b+z \\ 0 & 1 \end{pmatrix} = \tau\begin{pmatrix} 1 & z \\ 0 & 1 \end{pmatrix}. \tag{231}$$

We can check that the Picard-Vessiot ring corresponding to (230) is isomorphic to $K\left[\begin{pmatrix} 1 & z \\ 0 & 1 \end{pmatrix}\right] \cong K[z]$.
Our aim is now to describe the structure of $\rho(\mathcal{G})$. To see this, we compute D_σ. To shorten the notation, let $C_z := \begin{pmatrix} 1 & z \\ 0 & 1 \end{pmatrix}$. We then have, again by a short computation,

$$D_\sigma = C_z^{-1} C_{\sigma(z)} = C_{-z} C_{\sigma(z)} = C_{\sigma(z)-z}. \tag{232}$$

We notice here that D_σ is again of the shape C_d for some $d \in K^\tau$ (in particular, $\sigma(z) - z$ is in K^τ, i.e. invariant under τ, as $\tau(\sigma(z)) - \sigma(z) = \sigma(b) = b = \tau(z) - z$). We can therefore

[40] Equivalently, as phrased in [32, 48], it is Zariski-closed in $GL_n(K^\tau)$.
[41] If one delves a bit closer into the underlying theory, this is not entirely surprising, see eg [43, Def. 11.5].

identify $\rho(\mathcal{G})$ with some subgroup of

$$\left\{ \begin{pmatrix} 1 & d \\ 0 & 1 \end{pmatrix}, d \in K^\tau \right\}, \tag{233}$$

defined by polynomial equations. By $C_{z_1} C_{z_2} = C_{z_1+z_2}$, the latter group is isomorphic to a subgroup of the additive group $(K^\tau, +)$. Consequently, a polynomial $P(x) = a_n C_d^n + \dots + a_1 C_d + a_0 C_d^0 = 0$ translates to (keep in mind that $C_d^0 = \mathrm{id} = C_0$) an equation of the form $na_n d + \dots + a_1 d = 0 \Leftrightarrow d(na_n + \dots + a_1) = 0$. If $d \neq 0$, this equation is fulfilled if and only if $na_n + \dots + a_1 = 0$, i.e. it is either true for all, or for no values of d. As a result, we know that $\rho(\mathcal{G}) = \{0\} \vee \rho(\mathcal{G}) = K^\tau$.
Note in particular that due to $C_{\sigma(z)} = C_{\sigma(z)-z} C_z = \rho(\sigma) C_z$, the identification of $\rho(\mathcal{G})$ with the additive group $(K^\tau, +)$ agrees with the isomorphism $R = K[C_z] \cong K[z]$, and if $\rho(\sigma) = C_d$, then we have $\rho(\sigma) : z \mapsto z + d$ in $K[z]$.

Example 2.
We proceed with a generalization of the previous example. Let now

$$A_L := \begin{pmatrix} B_0 & 0 & \dots & 0 \\ 0 & B_1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & B_n \end{pmatrix}, \tag{234}$$

with the B_i again of the form $B_i = \begin{pmatrix} 1 & b_i \\ 0 & 1 \end{pmatrix}, b_i \in K$. In the same manner as before, a solution of this matrix system turns out to be equivalent to a solution of the equations $\tau(z_i) - z_i = b_i$. We obtain a Picard-Vessiot ring $R \cong K[z_0, \dots, z_n]$, and $\rho(\mathcal{G})$ is a linear algebraic subgroup of

$$\left\{ \begin{pmatrix} C_0 & 0 & \dots & 0 \\ 0 & C_1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & C_n \end{pmatrix}, C_i = \begin{pmatrix} 1 & d_i \\ 0 & 1 \end{pmatrix}, d_i \in K^\tau \right\}. \tag{235}$$

It turns out that the latter group is simply the direct sum of the group in the previous example, therefore we know that $\rho(\mathcal{G})$ is some algebraic subgroup of $(\bigoplus_{i=0}^n K^\tau, +)$, and in the same fashion as before, if $\rho(\sigma) = (C_{d_i})$, then we have $\rho(\sigma) : z_i \mapsto z_i + d_i$. As K^τ is a field, we can view $(\bigoplus_{i=0}^n K^\tau, +)$ as a vector space over K^τ, and the linear algebraic subgroups can be seen to be subspaces. Since any subspace is contained in some hyperplane, we have therefore shown

Lemma 4.2.3. *For a Picard-Vessiot extension of (234), $\rho(\mathcal{G})$ is either isomorphic to $\bigoplus_{i=0}^n K^\tau$, or we can find $c_0, \dots, c_n \in K^\tau$ which are not all 0 such that*

$$\rho(\mathcal{G}) \subseteq \{(d_0, \dots, d_n) \in \bigoplus_{i=0}^n K^\tau : \sum_{i=0}^n c_i d_i = 0\}. \tag{236}$$

This is sufficient for us to prove the existence of a telescoper equation for an algebraically dependent solution of (234).

Theorem 4.2.4. *Let $z_0, \dots, z_n$ be algebraically dependent solutions of (234). Then there is a $g \in K$, $c_0, \dots, c_i \in K^\tau$, not all of which are zero, such that*

$$\tau(g) - g = c_0 b_0 + \dots + c_n b_n, \tag{237}$$

where the b_i are the same as in (234).

Proof.
By the above results on $\rho(\mathcal{G})$, it is connected as a variety. As $R = K[z_0, \dots, z_n]$, and by assumption there is an algebraic relation between the z_i, we therefore know that $\operatorname{trdeg}(\operatorname{Quot}(R)/K) < n+1$. By Theorem 4.2.3, $\rho(\mathcal{G})$ is thus not isomorphic to $\bigoplus_{i=0}^{n} K^\tau$ but rather to a proper subgroup, by Lemma 4.2.3 we can therefore find $c_0, \dots, c_n \in K^\tau$, not all 0, such that for all $\sigma \in \mathcal{G}$,

$$c_0 d_0 + \cdots + c_n d_n = 0, \tag{238}$$

where $\rho(\sigma) = (C_{d_i}) \leftrightarrow (d_i)$. Now consider the sum $\sum_{i=0}^{n} c_i z_i$. Any σ leaves this sum invariant, as

$$\rho(\sigma)\left(\sum_{i=0}^{n} c_i z_i\right) = \sum_{i=0}^{n} c_i(z_i + d_i) = \sum_{i=0}^{n} c_i z_i + \sum_{i=0}^{n} c_i d_i = \sum_{i=0}^{n} c_i z_i. \tag{239}$$

Therefore, we know that $\sum_{i=0}^{n} c_i z_i$ lies in K. Let $\sum_{i=0}^{n} c_i z_i =: g \in K$. Now, applying τ to g, we obtain

$$\tau(g) = \tau\left(\sum_{i=0}^{n} c_i z_i\right) = \sum_{i=0}^{n} c_i \tau(z_i) = \sum_{i=0}^{n} c_i z_i + \sum_{i=0}^{n} c_i b_i. \tag{240}$$

By subtraction, the claim follows.

□

4.2.2 Existence of a Telescoper Equation

We will now proceed by working explicitly in $F(\mathcal{E})$, with τ being the addition by a point $P \in \mathcal{E}$, and δ the derivation on $F(\mathcal{E})$, which we already know to commute with τ. The goal will be to give a criterion for the existence of a telescoper equation, as is outlined in [32, App. B].
Before we can proceed, we need to be able to look at local representations of a function in $F(\mathcal{E})$. When we want such a representation for some meromorphic function in $\mathbb{C}$, let us assume around 0, then all we need to do is look at a Laurent series expansion of f, given by $f(x) = \sum_{k=-n}^{\infty} c_k x^k$. For $f \in F(\mathcal{E})$ we essentially want to do the same thing, except we do not have a natural "base function" such as x at our disposal. To make up for this, we need to introduce the concept of uniformizers.

Definition 4.2.6. *A* ***local parameter*** *or* ***uniformizer*** *at a point $Q \in \mathcal{E}$ is a function $u_Q \in F(\mathcal{E})$ such that $\nu_Q(u_Q) = 1$.*
We say a set of local parameters is ***coherent****, if $u_{Q+P} = \tau(u_Q)$.*
Given a fixed set of local parameters $\{u_Q : Q \in \mathcal{E}\}$, we can define the local representation of $f \in F(\mathcal{E})$ as

$$f = \sum_{k=-n}^{\infty} c_{k,Q} u_Q^k = \tilde{f} + \sum_{k=-n}^{-1} c_{k,Q} u_Q^k, \tag{241}$$

where $\tilde{f}$ is the ***regular part of*** *f.*

Uniformizers always exist (eg [29, Th. 1.1.6]). We can, by simply selecting one local parameter in each orbit of τ and defining the rest to be such obtain a coherent set of uniformizers as well. We will for the following choose one such set of coherent local parameters, which is otherwise arbitrary, but stay fixed.
We will be very interested in whether or not the irregular parts of functions cancel over the orbits of τ, hence the next definition:

Definition 4.2.7. *Given $f \in F(\mathcal{E})$, we define the* ***j-th orbit residue*** *of f at Q as*

$$\operatorname{ores}_{Q,j}(f) := \sum_{n\in\mathbb{Z}} c_{Q+nP,-j} \tag{242}$$

for any $j > 0$.

Put in words, the j-th orbit residue of f at Q is the sum of the coefficients of the j-th order poles of f at all points in the orbit of Q under τ. By definition, we have $\operatorname{ores}_{Q,j}(f) = \operatorname{ores}_{\tau(Q),j}(f)$.

Lemma 4.2.4. *Let $f \in F(\mathcal{E})$. Then*

$$h(x) := \sum_{k=1}^{t} f(x + n_i P), \tag{243}$$

is regular at $x = Q$, where the n_i are those integers such that $Q + n_i P$ is a pole of f precisely if for all $Q \in \mathcal{E}, j > 0$ we have

$$\operatorname{ores}_{Q,j} = 0. \tag{244}$$

Proof.
This essentially follows from the definition. The first statement means that there are no poles in the sum, the second means that all possible poles cancel. Clearly, the two are the same.

□

The next steps are a bit technical, and will therefore be only outlined. As we are interested in the potential cancelling of poles, the following definitions are still somewhat intuitive:

Definition 4.2.8. *We define*

1. *the* ***polar dispersion of f at Q*** $\operatorname{pdisp}_Q(f)$ *for Q a pole of f to be the maximum n such that $Q + nP$ is another pole of f,*
2. *the* ***polar disperson of f*** $\operatorname{pdisp}(f)$ *to be the maximum polar dispersion of f at Q for all $Q \in \mathcal{E}$, i.e. the maximum distance (in terms of τ) between two poles of f,*
3. *the* ***weak polar dispersion of f*** $\operatorname{wpdisp}(f)$ *to be the same as the maximum polar disperson of f, except we only count poles with order at least 2.*

By [29, 1.3], all these numbers are finite. The next lemma essentially tells us, that we will in the following be able to restrict ourselves to functions with somewhat nice behaviour, in that the polar dispersions are rather small.

Lemma 4.2.5 ([32, Lemma B.14])**.** *For any $f \in F(\mathcal{E})$, there are $f^*, g \in F(\mathcal{E})$ such that* $\operatorname{wpdisp}(f^*) = 0, \operatorname{pdisp}(f^*) \leq 1, f^* = f + \tau(g) - g$.

Proof.
See [32, Lemma B.14]. The idea is to repeatedly apply the Riemann-Roch Theorem in order to get rid of unwanted poles.

$\square$

Using the above, we can now prove

Lemma 4.2.6 ([32, Prop. B.8])**.** *Let $f \in F(\mathcal{E})$. Then we have*

$$\mathrm{ores}_{Q,j}(f) = 0 \tag{245}$$

for all $Q \in \mathcal{E}, j > 0$ precisely if there are $Q \in \mathcal{E}, e \in F(\mathcal{E})$ and $g \in \mathcal{L}(Q + (Q -_{\mathcal{E}} P))$[42]*, such that*

$$f = \tau(e) - e + g, \tag{246}$$

where $\mathcal{L}(Q+(Q-_{\mathcal{E}}P)$ denotes the ***Riemann-Roch space*** *associated with the divisor $Q+(Q-_{\mathcal{E}}P)$, i.e. all functions that have poles at most at $Q, Q-P$ with multiplicity* 1.

Proof.
First, given that the second condition holds, and that $\mathrm{ores}(\tau(e)) = \mathrm{ores}(e)$, we have $\mathrm{ores}(f) = \mathrm{ores}(g)$ with $g \in \mathcal{L}(Q + (Q -_{\mathcal{E}} P))$. The first condition then follows by noting that all poles of a compact Riemann surface under a given differential form must sum up to 0, see [32, Lemma B.15].
Given the first condition, by Lemma 4.2.5 we can assume without loss of generality that we have $\mathrm{wpdisp}(f) = 0$ (and consequently no pole of order > 1, as else the polar dispersion could not work out), and $\mathrm{pdisp}(f) \leq 1$ (and consequently all poles must appear in pairs $Q_i, Q_i + P$, only one pair in every orbit of τ). Let us assume we have r such pairs. If $r \leq 1$, we are done.
So assume $r > 2$, then by Riemann-Roch there is a $h \in \mathcal{L}(Q_1 + Q_2)$, h not constant. Thus, h has a pole at either Q_1 or Q_2, assume it is Q_1. Then, we can choose a scalar $a \in \mathbb{C}$ such that $\tilde{f} := f - (\tau(h) - h)$ has no pole at Q_1 (note that poles of $\tau(h)$ can only appear at $Q_1 - P, Q_2 - P$). As $\tilde{f}$ has no new pole compared to f, but rather one – at Q_1 – less, we know that $\mathrm{pdisp}(\tilde{f}) \leq 1, \mathrm{wpdisp}(\tilde{f}) = 0$. As one easily sees that $0 = \mathrm{ores}_{Q_1,1}(f) = \mathrm{ores}_{Q_1,1}(\tilde{f})$, we know that $\tilde{f}$ cannot possibly have a pole at $Q_1 - P$. Therefore, $\tilde{f}$ now fulfils the same role as f previously, but with one less pair of poles. Iterating this procedure yields the statement.

$\square$

Lastly, we need

Lemma 4.2.7 ([32, Prop. B.5])**.** *Let $f \in F(\mathcal{E})$. Then we can find $Q \in \mathcal{E}, e \in F(\mathcal{E})$ and $g \in \mathcal{L}(Q + (Q -_{\mathcal{E}} P))$ such that*

$$f = \tau(e) - e + g \tag{247}$$

precisely if there is a $h \in F(\mathcal{E})$ and a non-zero operator $L \in \mathbb{C}[\delta]$ such that

$$L(f) = \tau(h) - h. \tag{248}$$

Proof.
First, assume we have found a $h \in F(\mathcal{E})$ and an operator $L \in \mathbb{C}[\delta]$ as required. As before, we can assume without loss of generality that $\mathrm{wpdisp}(f) = 0, \mathrm{pdisp}(f) \leq 1$. First, we show that f

[42] We have two different additive structures here, the addition of divisors, denoted by a simple $+$, as well as the subtraction of points on $\mathcal{E}$, denoted here by the subscript $-_{\mathcal{E}}$. If these do not appear simultaneously, they are otherwise both written as $+/-$.

can, as in the proof of Lemma 4.2.6, have only one pair of poles in each orbit of τ, each pair separated by $\pm P$. To do so, first notice that, if f has a pole at some point Q, then the entire left hand side will have a pole at Q (we can see this by applying δ to the local representation of f at Q), and so either $\tau(h)$ or h must have a pole at Q as well. As the arguments are the same in either case, we will assume that h has a pole at Q. Now assume that $Q + rP$, $Q - sP$ are also poles of h, $r, s \in \mathbb{N}$, namely the poles at maximum distance to Q of h. Then we know that $Q + rP$ certainly cannot be a pole of $\tau(h)$, and $Q - (s+1)P$ is a pole of $\tau(h)$ but not of h, so $Q + rP, Q - (s+1)P$ are both poles of $\tau(h) - h$. Computation (or, alternatively, [32, Lemma 3.2]) shows that these must then also be poles of f. However, as $\mathrm{pdisp}(f) \leq 1$, this implies that $r = s = 0$. Consequently, f has poles only in pairs, one pair in each orbit, and from here on we proceed to construct e as in the proof of Lemma 4.2.6.
Now assume the first condition holds, i.e. we have $Q \in \mathcal{E}, e \in F(\mathcal{E})$ and $g \in \mathcal{L}(Q + (Q -_{\mathcal{E}} P))$ such that $f = \tau(e) - e + g$. First, consider the poles of $\delta(g)$. Again by local representations of g, it becomes clear that $\delta(g)$ has poles at most at $Q, Q - P$, and these are of order at most 2. We can therefore choose an $h \in \mathcal{L}(2Q)$ such that the second order pole of $\delta(g) + f$ cancels. In fact, doing some additional work, [32, Lemma B.15] shows that $\delta(g) - (\tau(f) - f)$ already cancels both second order poles, so we know that $\delta(g) - (\tau(f) - f)) \in \mathcal{L}(Q + (Q -_{\mathcal{E}} P))$. By Riemann-Roch, the dimension of the latter vector space is 2, and it already contains $1, g$, therefore there must be constants c, d such that

$$\delta(g) - (\tau(f) - f) - cg - d = 0. \tag{249}$$

Applying δ and rearranging yields

$$\delta^2(g) - c\delta(g) = \tau(\delta(f)) - \tau(f). \tag{250}$$

We therefore have found an operator $L := \delta^2 - c\delta$ and a function $\tilde{h} := \delta(f) \in F(\mathcal{E})$, such that

$$L(g) = \tau(\tilde{h}) - \tilde{h}. \tag{251}$$

However, applying L to f together with $f = \tau(e) - e + g$ yields

$$L(f) = L(\tau(e) - e + g) = L(\tau(e)) - L(e) + L(g) = \tau(L(e)) - L(e) + \tau(\tilde{h}) - \tilde{h}, \tag{252}$$

and by letting $h := L(e) - \tilde{h}$, this reads as

$$L(f) = \tau(h) - h, \tag{253}$$

and hence we are done. □

Combining Lemmas 4.2.6,4.2.7, we can immediately see that an operator L as above exists precisely if all orbit residues are 0. We can slightly reformulate this, keeping in mind Lemma 4.2.4, and obtain

Lemma 4.2.8 ([32, Prop. B.2])**.** *For any $b \in F(\mathcal{E})$, there exists a non-zero $L \in \mathbb{C}[\delta]$ and a $g \in F(\mathcal{E})$ such that*

$$\tau(g) - g = L(b) \tag{254}$$

precisely if

$$h(x) := \sum_{k=1}^{t} b(x + n_k P) \tag{255}$$

is regular, where for $x = Q$, n_k are those integers such that b has a pole at $Q + n_k P$.

With this, we have all the tools necessary to prove hypertranscendence.

4.2.3 Proof Strategy

In section 4.1.2 we have already argued that, if F_1, F_2 are holonomic, then so must be their meromorphic continuations $F_{x,y}$ to $\mathbb{C}$. The same argument holds for $F_{1,2}$ being hyperalgebraic; so in the following we can restrict ourselves to working with $F_{x,y}$. We will discuss F_x here, as F_y works out in much the same way.
First of all, we know from Theorem 4.1.1, that F_x is a doubly periodic meromorphic function, and thus, by Theorem 3.2.2 a rational function of $\wp, \wp'$, and therefore lies in the function field $F(\mathcal{E})$. Clearly, so do its derivatives. Now, assume there is some algebraic relation between $F_x, \frac{\partial}{\partial \omega} F_x, \ldots, \frac{\partial^n}{\partial^n \omega}$. Let $\tau := \Theta$ be the automorphism $\omega \mapsto \omega + \omega_3$ of $\mathbb{C}/\Lambda$. Then we have, again by Theorem 4.1.1, that

$$\tau(F_x) - F_x = b \tag{256}$$

for some $b \in F(\mathcal{E})$, and as $\frac{\partial}{\partial \omega} =: \delta$ and τ commute, we also have

$$\tau(\delta^k F_x) - \delta^n F_x = \delta^k(\tau F_x) - \delta^k F_x = \delta^k b. \tag{257}$$

We have already seen in Lemma 4.2.1 that $\mathcal{M}(\mathcal{E})$ (which is isomorphic to $F(\mathcal{E})$), as stated above) is, together with τ, a difference ring, and if we select t such that the order of τ is infinite, then the only constants are $\mathbb{C}$.
We can now consider the very same linear difference system as in (234), with $b_i := \delta^i b$. We are then in the exact case of Example 4.2.1, and can therefore apply Theorem 4.2.4. We therefore know that, if $F_x, \ldots, \delta^n F_x$ are algebraically dependent, we have a telescoper equation, i.e. $c_0, \ldots, c_n \in \mathbb{C}$, not all zero, and a $g \in K(\mathcal{E})$ such that

$$\tau(g) - g = c_0 b + \cdots + c_n \delta^n b. \tag{258}$$

We can now utilize the results from section 4.2.2. Clearly, $L := \sum_{k=0}^{n} c_k \delta^k$ is an operator as in Lemma 4.2.7, so by this and Lemma 4.2.6, we know that we must have

$$\operatorname{ores}_{Q,j}(b) = 0 \tag{259}$$

for all $Q \in \mathcal{E}, j > 0$. By Lemma 4.2.8, this is equivalent to the function

$$h(x) := \sum_{k=1}^{t} b(x + n_k P) \tag{260}$$

being regular, where the n_k are such that the sum runs over all poles in a given orbit of τ. So, if we can either prove that the orbit residue of some order at some point is non-zero, or, equivalently, that the latter function h is not regular somewhere, then we are done. Keep in mind that, starting at (258), we can conveniently forget about F_x, which we do not know, and instead only examine b, which we *do* know and can explicitly compute.

Example 3.
To see an example[a] where all this works very, very smoothly, consider the walk with the steps $\{\mathbf{NW},\mathbf{NE},\mathbf{SE},\mathbf{S}\}$.

We can compute $b = h_y(\omega)\left(h_x(\widehat{\Phi}(\omega)) - h_x(\omega)\right)$, and arrive (in an affine chart) at

$$b(x,y) = x\left(\frac{x^2+x}{y(x^2+1)} - y\right). \tag{261}$$

One can explicitly compute the poles of this rational function[b] (keep in mind that we can only count those poles which do, in fact, lie on $\mathcal{E}$): poles occur when either $x = \infty, y = \infty$, or $(x^2+1) = 0$. Computing the resulting second coordinates on $\mathcal{E}$, we therefore have

$$\mathcal{P} = \{(\infty, \pm i), (\pm i, \infty), (\pm i, \pm it + t)\} \tag{262}$$

for the set $\mathcal{P}$ of poles. We want to find a pole Q which does not have any other in its orbit under τ, which would immediately entail $\mathrm{ores}_{Q,j}(b) \neq 0$, and therefore F_x could not be hyperalgebraic. The order of the pole is entirely irrelevant here, as it cannot cancel no matter the order.
One thing to notice here which makes computations a lot easier is that the automorphism $\sigma : \mathbb{C} \to C : i \mapsto -i$ commutes with τ. We now want to utilize this in order to show that there is no pole in the orbit of $(\infty, 1)$ under τ other than itself. To do so, let us define $P_{1,2} := (\infty, \pm i), Q_{1,2} := (\pm i, \infty), R_{1,2} = (\pm i, \pm it + t)$. We proceed in several steps.

1. $P_{1,2}$ are in different orbits.
   Assume $P_2 = \tau^n P_1$. Then we would have

   $$P_1 = (\infty, i) = \sigma(\infty, -i) = \sigma(\tau^n(\infty, i)) = \tau^n(\sigma(\infty, i)) = \tau^n P_2 = \tau^n(\tau^n P_1), \tag{263}$$

   which would mean that $\tau^{2n} P_1 = P_1$, which is a contradiction as τ is addition by a non-torsion point (the group is infinite).

2. $P_{1,2}, Q_{1,2}$ are in different orbits.
   We will show that P_1, Q_1 are in different orbits, the other cases can be worked out analogously. Remember that $\tau = \Theta = \Psi \circ \Phi$. We have $\Psi(P_i) = P_j$ and $\Phi(Q_i) = Q_j$, and $\Phi^2 = \Psi^2 = \mathrm{id}$.
   Now, assume that $Q_1 = \tau^k P_1$. We can immediately check that $P_2 = \tau^k Q_2$ as well. Therefore, we have

   $$\tau^{-(k+1)} P_1 = \tau^{-k-1}\Psi(P_2) = (\Psi \circ \Phi)^{-k-1}\Psi(P_2) = (\Phi \circ \Psi)^{k+1}\Psi(P_2) \tag{264}$$
   $$= \Phi \circ (\Psi \circ \Phi)^k (P_2) = \Phi \tau^k P_2 = \Phi(Q_2) = Q_1. \tag{265}$$

   Consequently, we have $\tau^{-k-1} P_1 = \tau^k P_1 = Q_1$, which again is a contradiction to τ being addition by a non-torsion point.

3. $P_{1,2}, R_{1,2}$ are in different orbits.
   We have that $\Psi(R_i) = Q_i$. Again, we will only show that P_1, R_1 are in different orbits, the other cases work in the same fashion.
   Assume $R_1 = \tau^k P_1$. Then we have $Q_1 = \Psi(R_1) = \Psi\tau^k P_1 = \Psi\tau^k\Psi P_2$. But we can write $\Psi\tau^k\Psi$ as $\Psi \circ (\Psi \circ \Phi)^k \circ \Psi = \Psi^2 \circ (\Phi \circ \Psi)^k = (\Phi \circ \Psi)^k = \tau^{-k}$, therefore we have $Q_1 = \tau^{-k}(P_2)$, in contradiction to 2.

Consequently, P_1 is the sole pole in its orbit under τ. Therefore, we know that a telescoper equation cannot exist, and thus F_x has to be hypertranscendental, and so must be the entire generating function $F(x, y; t)$.

[a] This example comes in part from a talk by M. F. Singer as part of the workshop "Lattice walks at the Interface of Algebra, Analysis and Combinatorics" at the Banff International Research Station on September 21, 2017, as well as [32, Ex. 4.1].

[b] Alternatively, one could use the representation $b = h_y(\omega)\left(h_x(\widehat{\Phi}(\omega)) - h_x(\omega)\right)$, and notice that the only possible poles are at $h_y(\omega), h_x(\omega), h_x(\widehat{\Phi}(\omega))$.

While the computations are not always so simple, they can be done, as in [32, 5]. In all but 9 cases (see 6 on page 81), it turns out that there orbit residues do not sum to zero, and therefore we have a hypertranscendental generating function. We have therefore shown

Theorem 4.2.5 ([32, Th. 1])**.** *For any non-singular walk with infinite group other than the* 9 *exceptional walks, the generating function is hypertranscendental.*

However, for these 9 exceptional cases, we have so far neither shown that they are hyperalgebraic or non-holonomic. It turns out that both statements hold, and we will now give a sketch as to why this is the case.

Theorem 4.2.6 ([32, Th. 2])**.** *For the* 9 *non-singular exceptional walks, the generating function is hyperalgebraic.*

Proof.
We will sketch a proof as in [32, Prop. 6.2].
If the orbit residues are 0, then we already know there is a non-zero operator $L \in \mathbb{C}[\delta]$ as well as a $g \in F(\mathcal{E})$, such that

$$L(b) = \tau(g) - g. \tag{266}$$

By computation, one sees that $L(F_1) - g$ is both ω_1- and ω_3-invariant, by Theorem 3.2.2 it can therefore be written as a rational polynomial of some Weierstraß $\wp$-function and its derivative. As those are hyperalgebraic (see Theorem 3.2.3), and we know that g is hyperalgebraic (by construction) then so is $L(F_1)$. As hyperalgebraic functions are closed under differential operators by the same argument as in the proof of Lemma 1.1.4, F_1 must be hyperalgebraic as well. Since the same can be said for F_2, the result must hold for the entire generating function $F(x, y; t)$. For details see [32, Prop. 6.2].

□

It remains to show that the 9 exceptional walks are not holonomic.

Theorem 4.2.7 ([32, Th. 7.1])**.** *For the* 9 *non-singular exceptional walks, the generating function is non-holonomic.*

Proof.
We will outline the proof as in [32, Th. 7.1]. We know that, if $F(x, y; t)$ is holonomic, so is F_1. In this case, it could only have finitely many poles, so the set of poles in $\mathbb{C}/\Lambda$ would have to be finite. As we have seen previously, for the exceptional walks we can find a non-zero operator L and a $g \in F(\mathcal{E})$ such that $h := L(F_1) - g$ is τ-invariant. If F_1 has finitely many poles, so does h. We can now lift h to the universal cover $\mathbb{C}$. Then the set of poles must be $\omega_1\mathbb{Z} + \omega_2\mathbb{Z}$-invariant, simply because h was build in $F(\mathcal{E})$. However, we also know that h is τ-invariant, i.e. the set of poles must also be invariant. As in each fundamental parallelogram we know there to be only finitely many poles, and ω_1, ω_3 are independent, this implies that h has no poles, and is therefore constant.
This is lead to a contradiction by showing that

1. F_1 has no poles in $\mathcal{D}$ (for the definition see section 3.3),
2. g has a pole in $\mathcal{D}$,

for details see [32, 7.1-7.3].

□

4.3 Singular Walks

So far, we have only considered the case where the curve associated with the polynomial $\mathcal{K}(x, y)$ is non-singular, i.e. the discriminants had no double roots. If this is not the case, then the resulting surface will not be of genus 1, but instead of genus 0, i.e. birationally equivalent to the projective line, meaning there is a rational map $\phi : \mathbb{P}^1 \to \mathcal{C}$. One can show that the automorphism Θ then corresponds to a dilation by a $q \in \mathbb{R}$ with $|q| \neq 1$. Then, methods as before can be utilized, pulling back the boundary terms to functions satisfying $f(q\omega) - f(\omega) = g(\omega)$, and then looking at orbits to determine the existence of a telescoper equation. This is done in [45] – which is, in fact, a follow-up on the paper [32] by the same authors. Doing this is not trivial, but the ideas are mainly the same as in the genus 1 case, and therefore it shall not be elaborated upon here. After working through all the necessary steps, one arrives at

Theorem 4.3.1 ([45, Main Thm.])**.** *For all singular walks, the corresponding generating functions are hypertranscendental.*

This completes the proof that the generating function of a walk is holonomic if and only if the group of the walk is finite.

4.4 Comparison between the Methods

Both the algebraic and the analytic method studied above do have a lot in common. Both methods work by studying poles, and in the ω_3-periodicity of F_1 as in Theorem 4.1.1 plays an essential part in the proof. In fact, the function f_x defined in section 4.1.2 is the very same as the b defined in section 4.2.3, so in both cases we studied the poles of the same function. Not only that, we even looked for poles on the same curve, namely on the orbits of $\Theta = \tau$.
What is in a sense the difference between the two methods, despite the technical framework used, is that the algebraic method as presented here applies more theory, in particular the Galois theory of difference equation, and therefore studies the singularities a bit more closely. Compared to the analytic method, where the argument is essentially about counting singularities, i.e. there must be an infinite amount, in the algebraic method the order of a pole is also taken into account, which leads to a more precise knowledge of the behaviour of poles – and, consequently, to a somewhat stronger result: instead of showing non-holonomy, the functions are additionally classified depending on whether they are hyperalgebraic or hypertranscendental. In the cases where the orbit residues do not cancel, which are the 9 exceptional walks, one would therefore expect the methods to boil down to the same argument – and indeed, as can be seen in the proof of Theorem 4.2.7, in this case the argumentation is exactly the same as in the analytic method: one shows that there must be infinitely many singularities, which is a contradiction.
That one can get more accurate results when utilizing a more involved machinery is not really unexpected. However, it should also be pointed out that the algebraic method heavily relies on the underlying results obtained using the analytic approach; in particular the lifting and

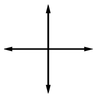
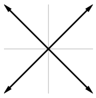
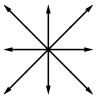
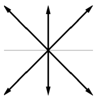
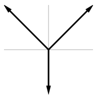
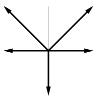
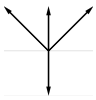
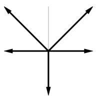
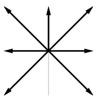
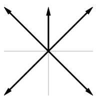
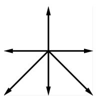
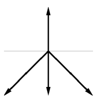
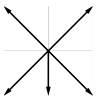
meromorphic continuation of mappings as well as computational results. In a way, one could therefore see the algebraic approach to be a – maybe unexpected – application of new theory, which complements the existing methods and leads to even stronger results.
In any case, there lies a certain beauty in the fact that one can treat this problem in either fashion, and that, while the underlying idea of examining poles along the orbit of an automorphism remains the same, this can be achieved using wildly different tools and methods.

5 Conclusions, Outlook

In the previous sections, we have now classified the generating functions of walks with small steps in the quarter plane according to whether or not they are holonomic and in particular shown the conjecture by M. Bousquét-Melou and M. Mishna in [24] that holonomy is equivalent to a finite group. The tools necessary to do so covered a remarkably broad scope of mathematics, and it also has become quite clear that in this case there is no one "right" way to prove the statement, but rather a certain idea, which can then be looked at from a multitude of different points of view and treated with corresponding methods. It is fascinating and beautiful how a problem which starts in a very discrete setting can be solved using theory about Riemann surfaces or Galois theory and Elliptic curves.
While what we originally set out to do is by now finished, the same can in no way be said about the study of lattice walks in general. There is a number of questions that could be asked, for instance

- Can we find an explicit formula for the coefficients of the generating function?
- What happens if we look at walks in other subsets of the plane?
- What happens if we impose other restrictions, eg self-avoidance?
- What happens if we allow for more general step sets?
- What happens in higher dimensions?

When trying to generalize the previous methods, there is a variety of ways one can run into trouble. For example, the kernel equation will have more boundary terms, and the resulting surface might be of a higher genus, which do not have as nice an algebraic structure as elliptic curves or the projective line. Therefore, the study of automorphisms and orbits will be a lot harder, if at all possible. Also, the number of possible walks increases drastically – even if one is interested only in walks with small steps, in the three-dimensional case, there are a priori $2^{26} - 1$ walks to consider. Similarly, if one allows larger steps, the number of possibilities rapidly increases.
Nonetheless, some progress has been made. One of the first model which was shown not to be D-finite was the so-called Knight's Walk in the quarter plane, given by the steps $\{(2, -1), (-1, 2)\}$ does not lead to a D-finite generating function, as proven in [49]. Another step towards classification of walks with larger steps was made in [50], proposing a generalization of orbit sum methods for walks with larger steps, and proving holonomy or non-holonomy for several thousands of them. While existing methods have been refined , other approaches, for example via boundary problems and integral representations [52] and invariants [53] have been tried as well, and it remains open to see if there is progress in this direction. There has also been work on asymptotic result in higher dimensions, connecting lattice paths to discrete partial equations and boundary problems, as for instance in [54], as well as approaches towards changing the shape of the allowed region, most notably considering walks not in, but rather avoiding the quarter plane, as in [55],[56]. This is, of course, not a complete list of progress that has been made since, but just naming a few examples of work in this direction, and goes to show that research about lattice paths is far from at its end, and there are is likely a lot more similarly beautiful theory waiting to be found.

	Steps	Orbit of (x,y)	Group order
1		$(x,y),(x^{-1},y),(x^{-1},y^{-1}),(x,y^{-1})$	4
2		$(x,y),(x^{-1},y),(x^{-1},y^{-1}),(x,y^{-1})$	4
3		$(x,y),(x^{-1},y),(x^{-1},y^{-1}),(x,y^{-1})$	4
4		$(x,y),(x^{-1},y),(x^{-1},y^{-1}),(x,y^{-1})$	4
5		$(x,y),(x^{-1},y),\left(x^{-1},y^{-1}\frac{1}{x+x^{-1}}\right),\left(x,y^{-1}\frac{1}{x+x^{-1}}\right)$	4
6		$(x,y),(x^{-1},y),\left(x^{-1},y^{-1}\frac{1}{x+x^{-1}}\right),\left(x,y^{-1}\frac{1}{x+x^{-1}}\right)$	4
7		$(x,y),(x^{-1},y),\left(x^{-1},y^{-1}\frac{1}{x+1+x^{-1}}\right),\left(x,y^{-1}\frac{1}{x+1+x^{-1}}\right)$	4
8		$(x,y),(x^{-1},y),\left(x^{-1},y^{-1}\frac{1}{x+1+x^{-1}}\right),\left(x,y^{-1}\frac{1}{x+1+x^{-1}}\right)$	4
9		$(x,y),(x^{-1},y),\left(x^{-1},y^{-1}\frac{x+x^{-1}}{x+1+x^{-1}}\right),\left(x,y^{-1}\frac{x+x^{-1}}{x+1+x^{-1}}\right)$	4
10		$(x,y),(x^{-1},y),\left(x^{-1},y^{-1}\frac{x+x^{-1}}{x+1+x^{-1}}\right),\left(x,y^{-1}\frac{x+x^{-1}}{x+1+x^{-1}}\right)$	4
11		$(x,y),(x^{-1},y),(x^{-1},y^{-1}(x+1+x^{-1})),(x,y^{-1}(x+1+x^{-1}))$	4
12		$(x,y),(x^{-1},y),(x^{-1},y^{-1}(x+1+x^{-1})),(x,y^{-1}(x+1+x^{-1}))$	4
13		$(x,y),(x^{-1},y),\left(x^{-1},y^{-1}\frac{x+1+x^{-1}}{x+x^{-1}}\right),\left(x,y^{-1}\frac{x+1+x^{-1}}{x+x^{-1}}\right)$	4

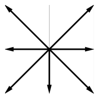
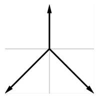
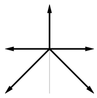
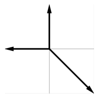
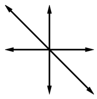
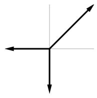
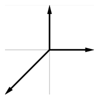
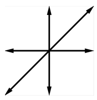
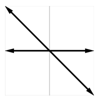
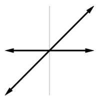
14		$(x,y), (x^{-1},y), \left(x^{-1}, y^{-1}\frac{x+1+x^{-1}}{x+x^{-1}}\right), \left(x, y^{-1}\frac{x+1+x^{-1}}{x+x^{-1}}\right)$	4
15		$(x,y), (x^{-1},y), \left(x^{-1}, y^{-1}(x+x^{-1})\right), \left(x, y^{-1}(x+x^{-1})\right)$	4
16		$(x,y), (x^{-1},y), \left(x^{-1}, y^{-1}(x+x^{-1})\right), \left(x, y^{-1}(x+x^{-1})\right)$	4
17		$(x,y), (x^{-1}y,y), (x^{-1}y,x^{-1}),$ $(y^{-1},x^{-1}), (y^{-1},y^{-1}x), (x,y^{-1}x)$	6
18		$(x,y), (x^{-1}y,y), (x^{-1}y,x^{-1}),$ $(y^{-1},x^{-1}), (y^{-1},y^{-1}x), (x,y^{-1}x)$	6
19		$(x,y), (x^{-1}y^{-1},y), (x^{-1}y^{-1},x),$ $(y,x), (y,x^{-1}y^{-1}), (x,x^{-1}y^{-1})$	6
20		$(x,y), (x^{-1}y^{-1},y), (x^{-1}y^{-1},x),$ $(y,x), (y,x^{-1}y^{-1}), (x,x^{-1}y^{-1})$	6
21		$(x,y), (x^{-1}y^{-1},y), (x^{-1}y^{-1},x),$ $(y,x), (y,x^{-1}y^{-1}), (x,x^{-1}y^{-1})$	6
22		$(x,y), (yx^{-1},y), (yx^{-1},yx^{-2}), (x^{-1},yx^{-2}),$ $(x^{-1},y^{-1}), (xy^{-1},y^{-1}), (xy^{-1},x^2y^{-1}), (x,y^{-1}x^2)$	8
23		$(x,y), (x^{-1}y^{-1},y), (x^{-1}y^{-1},x^2y), (x^{-1},x^2y),$ $(x^{-1}y^{-1}), (xy,y^{-1}), (xy,x^{-2}y^{-1}), (x,x^{-2}y^{-1})$	8

Table 5: The 23 step sets corresponding to a finite group [24, Tab. 3]. Step sets 19,20,21,23 are the exceptional cases – the first three can be solved using half orbit sums, the last one using a guess-prove method.

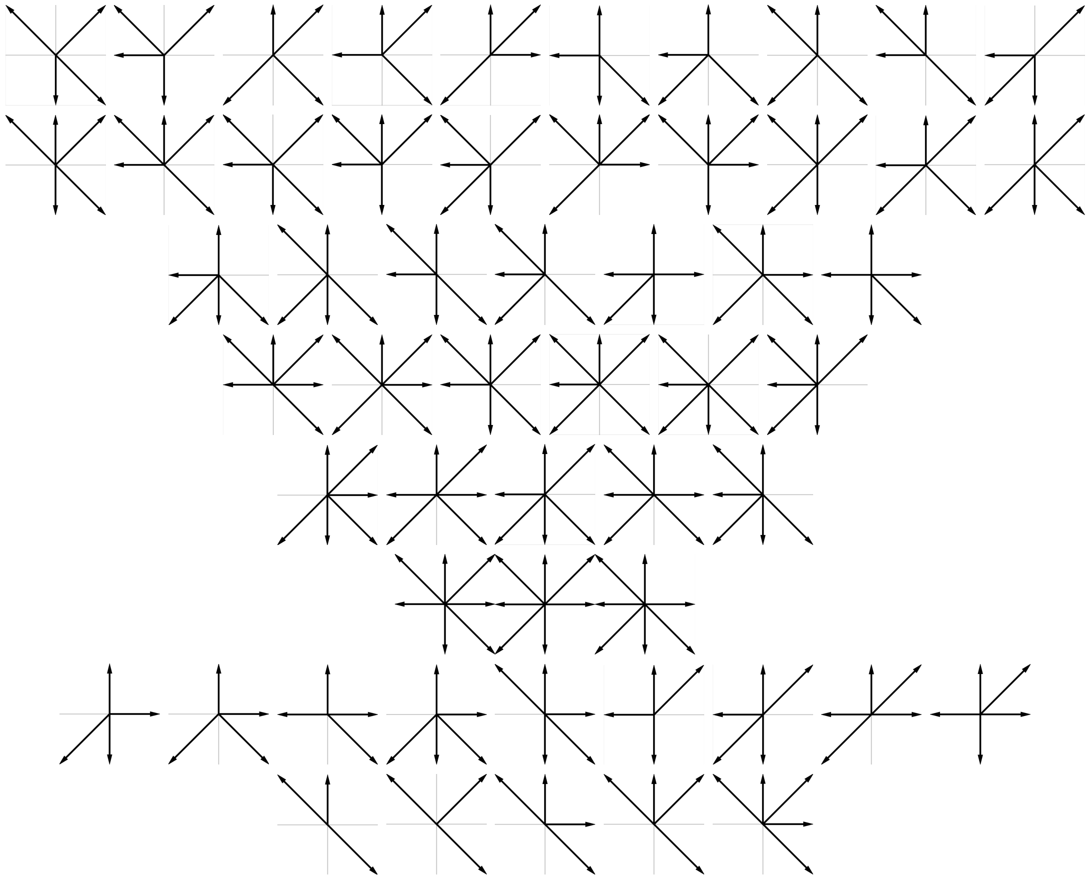


Table 6: The 56 step sets corresponding to an infinite group [28, Fig. 17], [24, Tab. 4], [32, Fig. 1-2]. The second to last row consists of the 9 non-singular, exceptional walks; the last row consists of the 5 singular walks.

References

[1] P. Flajolet, R. Sedgewick, *Analytic Combinatorics*, Cambridge University Press, 2009

[2] H. Wilf, *generatingfunctionology*, Academic Press Inc., 1990

[3] M. Rosen, *Number Theory in Function Fields*, Springer, 2002

[4] M. Kauers, P. Paule, *The Concrete Tetrahedron*, Springer, 2011

[5] R. Stanley, *Enumerative Combinatorics, Vol. 2*, Cambridge University Press, 1999

[6] S. Gerhold, *On Some Non-Holonomic Sequences*, The Electronic Journal of Combinatorics 11 (2004), # R87

[7] P. Flajolet, S. Gerhold, B. Salvy, *On the non-holonomic character of logarithms, powers, and the n-th prime function*, The Electronic Journal of Combinatorics 11(2) (2005), #A2

[8] M. Kauers, *The Holonomic Toolkit* in *Computer Algebra in Quantum Field Theory*, pp. 119-144, Springer, 2013

[9] W. Wasow, *Asymptotic expansions for ordinary differential equations*, Dover, 1987

[10] P. Henrici, *Applied and Computational Complex Analysis*, Vol. 2, John Wiley & Sons, 1977

[11] L. Lipshitz, *D-Finite Power Series*, Journal of Algebra 122, pp. 353-373, 1989

[12] K. Humphreys, *A history and a survey of lattice path enumeration*, Journal of Statistical Planning and Inference 140, pp. 2237–2254, 2010

[13] A. W. F. Edwards, *Pascal's Problem: The 'Gambler's Ruin'*, InternationalStatisticalReview 51, pp. 73-79, 1983

[14] E. Győri, G. O. H. Katona, L. Lovász, *Horizons of Combinatorics*, Springer, 2008

[15] W. A. Whitworth, *Arrangements of m things of one sort and n things of another sort under certain conditions of priority*, Messenger of Mathematics Vol. 8, pp. 105-114, 1879

[16] C. Krattenthaler, *Lattice Path Enumeration*, Handbook of Enumerative Combinatorics, Chapman and Hall/CRC, 2015

[17] G. H. Weiss, *Aspects and Applications of the Random Walk*, North-Holland, 1994

[18] F. Flory, *Principles of Polymer Chemistry*, Cornell University Press, 1953

[19] B. Malkiel, *A Random Walk Down Wall Street*, W. W. Norton & Company, 1973

[20] P. Mörters, Y. Peres, *Brownian Motion*, Cambridge University Press, 2010

[21] M. Wallner, *Combinatorics of Lattice Paths and Tree-like Structures*, Diss., Technical University of Vienna, 2016

[22] C. Banderier, P. Flajolet, *Basic analytic combinatorics of directed lattice paths*, Theoretical Computer Science 281, pp. 37-80, 2002

[22] H. Prodinger, *The Kernel Method: A Collection Of Examples*, Séminaire Lotharingien de Combinatoire [electronic only] (50), 2004

[24] M. Bousquet-Mélou, M. Mishna, *Walks with small steps in the quarter plane*, Contemporary mathematics, American Mathematical Society, 2010, 520, pp. 1-40. ffal-00333741v3f

[25] A. Bostan, M. van Hoeij, M. Kauers, *The complete generating function for Gessel walks is algebraic*, Proceedings of the American Mathematical Society, 138 (9), pp. 3063–3078, 2009

[26] A. Bostan, I. Kurkova, K. Raschel, *A Human Proof of Gessel's Lattice Path Conjecture*, Transactions of the American Mathematical Society 369, 1365-1393, 2017

[27] M. Mishna, A. Rechnitzer, *Two non-holonomic lattice walks in the quarter plane*, Theoretical Computer Science, 410, pp. 3616-3630, 2009

[28] I. Kurkova, K. Raschel, *On the functions counting walks with small steps in the quarter plane*, Publications Mathématiques de l'IHÉS, 116, pp. 69-114, 2012

[29] H. Stichtenoth, *Algebraic Function Fields*, Springer, 2009

[30] L. Lipshitz, *The Diagonal of D-Finite Power Series Is D-Finite*, Journal of Algebra 113, pp. 373-378, 1988

[31] G. Fayolle, R. Iasnogorodski, V. Malyshev, *Random Walks in the Quarter Plane*, Springer, 1999

[32] T. Dreyfus, C. Hardouin, J. Roques, M. F. Singer, *On the Nature of the Generating Series of Walks in the Quarter Plane*, Inventiones mathematicae, Vol. 213, Issue 1, pp. 205-236, 2018

[33] R. Miranda, *Algebraic Curves and Riemann Surfaces*, American Mathematical Society, 1995

[34] G. A. Jones, D. Singermann *Complex Functions. An algebraic and geometric viewpoint*, Cambridge University Press, 1987

[35] G. Springer, *Introduction to Riemann Surfaces*, Adison-Wesley Publishing Company, 1957

[36] P. Griffiths, J. Harris, *Principles of Algebraic Geometry*, John Wiley & Sons, 1978

[37] O. Forster, *Lectures on Riemann Surfaces*, Springer, 1981

[38] I. R. Shafarevich, *Basic Algebraic Geometry 2*, Springer, 1996

[39] I. R. Shafarevich, *Algebraic Geometry I. Algebraic Curves, Algebraic Manifolds and Schemes*, Springer, 1994

[40] J. H. Silverman, *The Arithmetic of Elliptic Curves*, Springer, 2009

[41] J. J. Duistermaat, *Discrete Integrable Systems*, Springer, 2010

[42] F. Kirwan, *Complex Algebraic Curves*, Cambridge University Press, 1992

[43] J. Harris, *Algebraic Geometry. A First Course*, Springer, 1992

[44] K. Raschel, *Counting walks in a quadrant: a unified approach via boundary value problems*, DMTCS Proceedings of AofA 2012, pp. 109-124

[45] T. Dreyfus, C. Hardouin, J. Roques, M. F. Singer, *Walks in the Quarter Plane, Genus Zero Case*, Journal of Combinatorial Theory A., Vol. 174, 2020

[46] M. E. Goldstein, W. H. Braun, *Advanced Methods for the Solution of Differential Equations*, NASA SP-316, Washington, D. C., 1973

[47] C. Hardouin *Galois Theories of Linear Difference Equations: An Introduction*, Mathematical Surveys and Monographs of the AMS (accessed on 16/7/2020 via https://perso.math.univ-toulouse.fr/hardouin/files/2020/03/notescolombiacharlotte.pdf)

[48] M. van der Put, M. F. Singer, *Galois Theory of Difference Equations*, Springer, 1997

[49] M. Bousquet-Mélou, M. Petkovsek, *Walks confined in a quadrant are not always D-Finite*, Theoret. Comput. Sci. 307, 257-276, 2003

[50] A. Bostan, M. Bousquet-Mélou, S. Melczer, *Counting walks with large steps in an orthant*, ArXiv:1806.00968, 2018

[51] I. Kurkova, K. Raschel, *New steps in walks with small steps in the quarter plane*, Annals of Combinatorics 19, 461-511, 2015

[52] K. Raschel, *Counting walks in a quadrant: a unified approach via boundary value problems*, Journal of the European Mathematical Society 14, 749-777, 2012

[53] O. Bernardi, M. Bousquet-Mélou, K. Raschel, *Counting quadrant walks via Tutte's invariant method*, ArXiv 1708.08215; extended abstract in DMTCS Proceedings of FPSAC 2016, 203-214

[54] B. Bogosel, V. Perrollaz, K. Raschel, A. Trotignon, *3D positive lattice walks and spherical triangles*, Journal of Combinatorial Theory, Series A 172, 2020

[55] M. Bousquet-Mélou, *Plane lattice walks avoiding a quadrant*, J. Combin. Theory Ser. A, 144, 37-79, 2016

[56] K. Raschel, A. Trotignon, *On walks avoiding a quadrant*, Electronic Journal of Combinatorics 26, Paper 3.31, 1-34, 2019