Scaling limits in algebra, geometry, and probability

Proefschrift

ter verkrijging van
de graad van doctor aan de Universiteit Leiden,
op gezag van rector magnificus Prof. dr. ir. H. Bijl,
volgens besluit van het college voor promoties
te verdedigen op woensdag 23 februari 2022
klokke 10:00 uur

door

Elizaveta Arzhakova

geboren te Moskou, Rusland
in 1993

Promotores:

Prof. dr. E. A. Verbitskiy (Universiteit Leiden, Universiteit Groningen)
Prof. dr. W. Th. F. den Hollander

Co-promotor:

Dr. D. E. Terhesiu

Promotiecommissie:

Prof. dr. Ronald van Luijk
Prof. dr. Frans de Haas
Prof. dr. Beatrice de Tilière (University Paris Dauphine)
Prof. dr. Hanfeng Li (University at Buffalo)
Prof. dr. Sergey Shadrin (University of Amsterdam)

Contents

1 Introduction **5**

1.1 Dimer configurations and decimations 7

1.2 Spanning trees and determinantal point processes 9

1.3 Central Limit Theorem for dynamical systems 11

1.4 Dynamics on moduli space 12

2 Tropical limits of decimated polynomials **15**

2.1 Introduction 15

2.2 Decimation of polynomials 18

2.3 The scaling limit 19

3 Decimation limits of principal algebraic $\mathbb{Z}^d$-actions **27**

3.1 Introduction 27

3.2 Examples 31

3.3 Convex functions and Legendre duals 37

3.4 Amoebas and Ronkin functions 38

3.5 Decimation limits of polynomials 39

3.6 Decimations of principal actions and contracted ideals 43

3.7 Absolutely irreducible factorizations
and Gauss's Lemma 47

3.8 Decimated polynomials and decimated actions 49

3.9 Remarks and questions 55

3.10 Example of computing the decimation limit 59

4 On the determinantal process associated to spanning trees 65

4.1 Discrete determinantal processes 65

4.2 Projection operator 71

4.3 Graphs with abelian symmetries 75

4.4 Explicit expressions of correlation kernels 82

5 Rates of convergence in CLT for ergodic toral automorphisms 97

5.1 Introduction and main result 97

5.2 A brief survey of the methods of proof of CLT for dynamical systems 103

5.3 Stein's method for establishing CLT with rates of convergence . 109

5.4 Proof of CLT with rates of convergence for ergodic toral automorphisms 114

6 Connectivity of real isoperiodic sets on a torus with 3 poles 125

6.1 Introduction 125

6.2 Rigid forms 130

6.3 Proof of the connectivity of real isoperiodic sets in $\Omega\mathcal{S}_{1,3}$ 145

6.4 Appendix: connectivity of real isoperiodic sets in $\Omega\mathcal{S}_{1,2}$ 149

Bibliography 155

Samenvatting 164

Acknowledgements 166

Curriculum vitae 167

Chapter 1

Introduction

> *"The theory of dynamical systems is a major mathematical discipline closely intertwined with most of the main areas of mathematics. Its mathematical core is the study of the global orbit structure of maps and flows with emphasis on properties invariant under coordinate changes."*
> A. Katok, B. Hasselblat, Introduction to the modern theory of dynamical systems

A dynamical system is a system that undergoes time evolution. Formally speaking, it is a three-tuple consisting of:

1. A phase space X where each point describes some configuration of the system. This space is equipped with a measure μ together with a σ-algebra $\mathcal{B}$.
2. A parameter t that is often referred to as time. It typically belongs to the real numbers $\mathbb{R}$ (continuous time) or the integers $\mathbb{Z}$ (discrete time). However, this parameter can also be multi-dimensional (for example, algebraic $\mathbb{Z}^d$-actions).
3. The time evolution law, i.e., a family of transformations $\{T^t\}_t$ of the phase space X that are parametrised by time t. The law allows to obtain the state of the system X_t at any time t from the initial state X_0 by computing $T^t X_0$. The measure μ is called T-invariant if $\mu(T^{-t}A) = \mu(A)$ for any measurable $A \subset X$.

As the epigraph suggests, dynamical systems can be observed in many areas of mathematics and physics. The theory of dynamical systems unites these areas by studying similar problems associated to time evolution.

An important direction of the theory of dynamical systems is their classification, or, in other words, the identification of whether two dynamical systems are isomorphic. One can tackle the classification problem by studying invariants, i.e., properties or quantities associated to a system which are invariant under system isomorphisms. One of such invariant quantities is the Kolmogorov-Sinai entropy [58, 104]. It measures the limiting complexity of a system and is the complete invariant for Bernoulli systems. Let us discuss entropies of some solvable models investigated in the present thesis.

Consider the lattice $\mathbb{Z}^2$ as a graph, where the vertices are the nodes of the lattice, and the edges connect vertices which are adjacent in vertical and horizontal directions. A dimer configuration on $\mathbb{Z}^2$ is a subset of edges which covers each node exactly once. The set of all dimer configurations on $\mathbb{Z}^2$ is invariant under the $\mathbb{Z}^2$-action by shifts, and together with this action it defines a dynamical system. The dimer model has been extensively studied in the context of symbolic dynamics and statistical mechanics (see, for example, [55]). In particular, its topological entropy is given by

$$h_d = \frac{1}{4}\int_0^1 \int_0^1 \log\left(4 - 2(\cos 2\pi x_1 + \cos 2\pi x_2)\right)\, dx_1\, dx_2. \tag{1.1}$$

The uniform spanning forest model on $\mathbb{Z}^2$ is similar to the dimer model. It turns out that there exists a bijection between the set of spanning trees of an $n \times n$ box and the set of dimer matchings of a $(2n-1)\times(2n-1)$ box with a corner removed (the Temperley-Fisher bijection, see [107]). The entropy of the uniform spanning forest model h_s on $\mathbb{Z}^2$ is directly related to the dimer entropy [16]:

$$h_s = 4h_d = \int_0^1 \int_0^1 \log\left(4 - 2(\cos 2\pi x_1 + \cos 2\pi x_2)\right) dx_1\, dx_2. \tag{1.2}$$

Remarkably, a number of other systems share the same expression for entropy, for example, the sandpile model [24]. Principal algebraic actions considered in the present thesis also provide an example of a system with the same entropy. Denote the additive torus $\mathbb{R}/\mathbb{Z}$ by $\mathbb{T}$ and consider a group

$$X = \left\{ x \in \mathbb{T}^{\mathbb{Z}^2} : 4x^{n,m} - x^{n+1,m} - x^{n-1,m} - x^{n,m+1} - x^{n,m-1} = 0, (n,m) \in \mathbb{Z}^2 \right\}.$$

The group $\mathbb{T}^{\mathbb{Z}^2}$ admits an action of $\mathbb{Z}^2$ by shifts. Denote by α the restriction of this action to the compact group X. The principal algebraic action

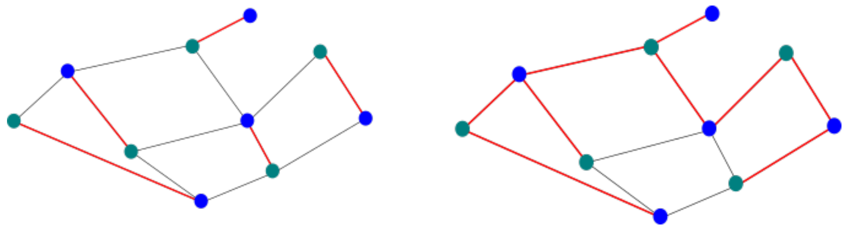


Figure 1.1: Left: a dimer configuration is a subset of edges covering each vertex exactly once. Right: a spanning tree is a subgraph that is a tree and connects all vertices.

(X, α) has the same entropy expression as (1.2) [64]. Moreover, this algebraic system is measure-theoretically isomorphic both to the double shift dimer model and to the uniform spanning forest model [99]. It is conjectured that the link between these models is, in fact, even stronger [100].

In the present thesis we investigate the limiting behaviour of some models. In particular, we consider the following problems:

- Existence of a scaling limit of principal actions (Chapters 2, 3).
- Existence and properties of a limiting measure or a limiting distribution (Chapters 4, 5).
- Properties of a foliation of a moduli space (Chapter 6).

1.1 Dimer configurations and decimations

Consider a planar weighted simple bipartite graph $G = (V, E)$, where V is the set of vertices and E is the set of unoriented edges of G. By definition, a *dimer configuration* of G is a subset of E that covers each vertex exactly once (see Figure 1.1).

One defines a *partition function* Z_G of a graph G as a formal sum of weights of dimer configurations, namely,

$$Z_G = \sum_{M \in \mathcal{M}(G)} \nu(M) = \sum_{M \in \mathcal{M}(G)} \prod_{e \in M} \nu(e),$$

where $\mathcal{M}(G)$ is the collection of all possible dimer configurations of G and $\nu(e)$ is the weight of $e \in E$. Suppose that G is an infinite weighted $\mathbb{Z}^2$-

periodic bipartite planar graph, and let $G_N = G/N\mathbb{Z}^2, N \geqslant 1$, be its finite periodic factor graph, which we view as a graph on the torus $\mathbb{T}^2$. It turns out that for any N, the partition function Z_{G_N} can be computed explicitly (see Section 6 in [93]).

In order to illustrate how Z_{G_N} is computed, let us consider the example of the honeycomb lattice (see Figure 1.2 for G_3). The graph G_N has N^2 black vertices, N^2 white vertices, and $3N^2$ edges. The positive weights a, b, c are assigned periodically to the edges of G_N as indicated in Figure 1.2. Denote by $N_b(M)$ and $N_c(M)$ the number of edges of type b and c, respectively, of a dimer configuration M of G_N. The partition function Z_{G_N} is then a sum of the weights of all possible dimer configurations of G_N:

$$Z_{G_N} = \sum_{M \in \mathcal{M}(G_N)} a^{3N^2 - N_b(M) - N_c(M)} b^{N_b(M)} c^{N_c(M)}.$$

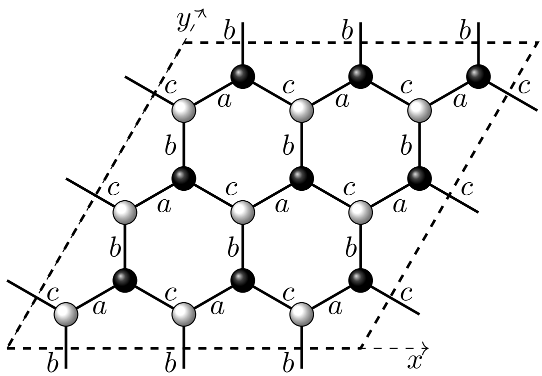


Figure 1.2: Periodic weighted hexagonal lattice.

It turns out that the partition function Z_{G_N} can be calculated by using the *characteristic polynomial* of two variables $P(z,w)$ that only depends on G_1. Namely,

$$Z_{G_N} = \frac{1}{2}\Big(-\mathcal{P}_N(1,1) + \mathcal{P}_N(-1,1) + \mathcal{P}_N(1,-1) + \mathcal{P}_N(-1,-1)\Big), \qquad (1.3)$$

where $\mathcal{P}_N(\pm 1, \pm 1) = \prod_{z^N = \pm 1} \prod_{w^N = \pm 1} P(z,w)$. In case of the hexagonal lattice $P(z,w) = a - bz - cw$ [55]. What is surprising is that not only the partition function Z_{G_N} (the weighted number of configurations), but also the coefficients of $Z_{G_N}(z,w)$ have a physical meaning. Namely, if $Z_{G_N}(z,w) = \sum_{\mathbf{n}} Z_{G_N}(\mathbf{n}) z^{n_1} w^{n_2}$, then $Z_{G_N}(\mathbf{n}) = \sum_{M \in \mathcal{M}_{\mathbf{n}}(G_N)} \nu(M)$, where $\mathcal{M}_{\mathbf{n}}(G_N)$, $\mathbf{n} = (n_1, n_2)$, is the collection of dimer configurations of G_N that have precisely

Nn_1 edges with weight b and Nn_2 edges with weight c. This means that $Z_{G_N}(\mathbf{n})$ is also a partition function. It turns out that the restricted partition functions $Z_{G_N}(\mathbf{n})$ have scaling limits: for $(s,t) \in \Delta_P$, i.e., in the Newton polytope of the polynomial P,

$$\lim_{N\to\infty} \frac{1}{N^2} \log Z_{G_N}([Ns],[Nt]) = -R_P^*(s,t), \tag{1.4}$$

where $[\cdot]$ is the integer part and $R_P^*(s,t)$ is the Legendre dual of the Ronkin function of P [55]. The coefficients of Z_{G_N} are directly related to the coefficients of the Nth decimation P_N of $P(z,w)$, which is defined as $P_N(z,w) = \prod_{q_1^N=1}\prod_{q_2^N=1} P(q_1 z, q_2 w)$. The latter can be generalised to polynomials with any number of variables. The N-th decimation f_N of a polynomial f is given by the formula

$$f_N(x_1,\dots,x_d) = \prod_{q_1^N=1} \cdots \prod_{q_d^N=1} f(q_1x_1,\dots,q_dx_d).$$

Note that f_N is a polynomial with integer coefficients with all powers of monomials pointwise divisible by N. A natural question is to understand whether the coefficients of f_N for arbitrary f also have scaling limits similar to (1.4). The methods used in [55] can only be applied for a class of polynomials in two variables that arise as characteristic polynomials of dimer models and cannot be generalised outside of this context. It is not difficult to demonstrate that (1.4) does not hold for arbitrary decimated polynomials (see, for instance, Example 3.2.3 of Chapter 3). However, in Chapters 2 and 3 we show that the *convex hulls* of decimated scaled coefficients always exist (Theorems 2.3.1 and 3.1.1). Chapter 2 is a geometric introduction to the problem of decimations of polynomials. It contains a proof of the existence of convex hulls of decimated scaled coefficients that relies on an analytic argument to bound Riemann sums. Chapter 3 views the problem of decimations from the algebraic point of view of $\mathbb{Z}^d$-actions. Even though the statement of its Theorem 3.1.1 of Chapter 3 coincides with the statement of Theorem 2.3.1 of Chapter 2, it features a different proof that only uses Mahler's estimates. Moreover, Chapter 3 provides a link between the decimation of a polynomial f and the decimation of a principal action associated to f stated in Theorem 3.8.12.

1.2 Spanning trees and determinantal point processes

Given a set E, a point process $X = \{0,1\}^E$ is called *determinantal* if the probability $\mathbb{P}(X(e_1) = 1,\dots,X(e_n) = 1)$ is given by a determinant of an

$n \times n$ matrix with entries given by a correlation function K. Namely, the (i, j)-th entry is equal to $K(e_i, e_j)$. Determinantal point processes (DPP) first emerged in 1960s in the framework of mathematical physics (see [30, 38] for the first well known examples) and in 1975 the general notion of DPP was first introduced for modelling fermion distributions at thermal equilibrium [79]. Even though DPPs have been studied for many years, most applications are unexplored yet promising (for instance, to statistics and machine learning).

In Chapter 4 we are concerned with a probabilistic problem of a DPP associated to uniform spanning forest measures. A spanning tree on a graph $G = (V, E)$ is a subset of edges $E' \subset E$ such that the graph (V, E') is a connected tree (see Figure 1.1). The subject of random spanning trees of a graph goes back to Kirchhoff in 1847, who showed its surprising relation to electrical networks. One of Kirchhoff's results expresses the probability that a uniformly chosen spanning tree contains a given edge in terms of the electrical current in the graph.The number of spanning trees $\tau(G)$ of a finite graph G can be calculated by applying the Kirchhoff's Matrix-Tree theorem: $\tau(G) = |V(G)|^{-1} \det' \Delta_G$, where $\det' \Delta_G$ is the product of all non-zero eigenvalues of the Laplacian of G.

Given a $\mathbb{Z}^d$-periodic infinite graph G, one can consider its approximation by finite graphs: $G_1 \subset G_2 \subset \ldots \subset G$, where $V(G_n) = \{v \in \mathbb{Z}^d : \max |v_i| \leqslant n\}$, with either free or wired boundary conditions. A number of properties of the spanning forest (a collection of spanning trees) structure on G has been established: for instance, the asymptotic limit of the number of spanning trees of G_n on $\mathbb{Z}^2$ is the entropy $\int_0^1 \int_0^1 \log\left(4 - 2(\cos 2\pi x_1 + \cos 2\pi x_2)\right) dx_1 dx_2$ as discussed above. Moreover, the probability measures associated to spanning trees on the finite graphs G_n converge to a uniform spanning forest measure on G that depends on the chosen boundary conditions. The resulting measures are determinantal (see Theorems 4.1.1 and 4.1.2 in Chapter 4, and [16, 89]), and the corresponding correlation kernels are expressed in terms of operator projections (Theorem 4.1.6 in Chapter 4). Despite the simple projection formulas of Theorem 4.1.6, the corresponding correlations kernels are difficult to compute explicitly even for simple graphs (see Example 4.1.3 of Chapter 4 for $\mathbb{Z}^d$ and [57] for computations on ladder-like graphs). Our result – Theorem 4.4.3, uses a functional-analytic generalisation of a simple linear algebra statement, which allows to effectively calculate correlation kernels of DPPs and correlations associated to uniform spanning forest measures on infinite graphs with $\mathbb{Z}^d$-symmetry. The goal of Chapter 4 is not only to provide a new method of computing the correlations, but also to demonstrate the application of the method. Therefore, Chapter 4 is equipped with a number of examples of computations for different $\mathbb{Z}^d$-periodic graphs.

1.3 Central Limit Theorem for dynamical systems

A probabilistic question closely related to the existence of a limiting measure is the question of existence of a limiting distribution. Consider a sequence of independent and identically distributed random variables $Y_1, Y_2, \dots$ with zero mean and variance $0 < \sigma^2 < \infty$. These satisfy the *Central Limit Theorem* (CLT)

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} Y_i \xrightarrow{d} \mathcal{N}(0, \sigma^2),$$

i.e., $\frac{1}{\sqrt{n}} \sum_{i=1}^{n} Y_i$ converges in distribution to a normal law. The same question can be asked in the framework of a dynamical system (X, μ, T) (where μ is T-invariant), for a sequence of centered functions $f \circ T^i$, $f \in L^2(X, \mu)$, that can be treated as random variables. We say that f satisfies the CLT if $\frac{1}{\sqrt{n}} \sum_{i=0}^{n-1} f \circ T^i \to \mathcal{N}(0, \sigma^2)$ for $0 < \sigma^2 < \infty$. There are relatively straightforward methods to prove the CLT for independent identically distributed random variables, but the variables $Y_i = f \circ T^i$ might not be independent. However, the CLT still holds in certain cases, for example, when the sequence of random variables possesses sufficiently strong mixing properties (see, for instance, [51, 97]) or forms a martingale (see [40]).

The models considered in Chapters 2, 3, 4 are the so-called solvable models, i.e, their free energy can be expressed in terms of some known function, usually a polynomial. The polynomials appearing in solvable models are typically non-expansive, i.e., some roots have unit absolute value. It is well known that expansive and non-expansive actions can have very different dynamical properties. In particular, in Chapter 5 we consider the CLT for ergodic (both expansive and non-expansive) toral automorphisms: the CLT has been established for ergodic toral automorphisms by using the martingale method (see [10, 62] for the CLT and its refinements for sufficiently smooth functions and [44] for general nilmanifords and Hölder functions).

When considering the CLT, we find that the difference between the hyperbolic (expansive) and non-hyperbolic (non-expansive) automorphisms be-

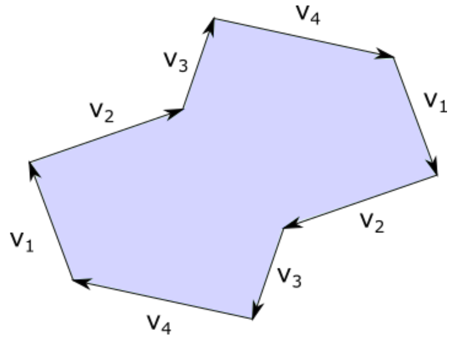


Figure 1.3: Flat surface - each pair of parallel sides is identified by translation. The resulting surface is a 2-torus (a sphere with 2 handles) with a conical singularity of angle 6π.

comes vivid. The hyperbolic toral automorphisms possess certain dynamical properties, such as a Markov partition and a spectral gap (in some Banach spaces), which allow one to prove the CLT in a variety of ways that do not apply in the context of the whole family of ergodic toral automorphisms (see Section 5.2 of Chapter 5 for detailed comparison of methods). In Chapter 5 we prove the CLT and obtain a new result – the rates of convergence – for Hölder functions in a special family of ergodic toral automorphisms, without applying the classic martingale method but rather by using Stein's method as in [49] and the mixing properties of Hölder functions [26, 90]. Moreover, it seems that the proof of the CLT and the rates of convergence in Chapter 5 can be generalised for the whole class of ergodic toral automorphisms, and also for the case of non-linear toral automorphisms and multivariate Hölder observables.

1.4 Dynamics on moduli space

In Chapter 5 we discussed dynamics on a torus, which is the simplest example of a flat surface, i.e., a surface obtained by pairwise identification of parallel sides of a collection of polygons in Euclidean plane (for an example less trivial than a torus, see Figure 1.3). Flat surfaces naturally appear in many areas of mathematics and physics, including billiards in rational polygons and electron transport in metal [85]. Despite the apparent simplicity of the definition, flat surfaces pose a number of mathematical questions that are still not answered, for example, questions on typical behaviour of a generic geodesic, ergodicity of the geodesic flow, and existence, number and length of closed geodesics [108]. It turns out that the dynamics on a flat surface is closely connected to the dynamics in the stratum of the moduli space of flat surfaces. Let us illustrate the latter statement with two examples. A famous theorem called Masur's criterion implies that if the vertical flow on a flat surface S is minimal but not ergodic, then the Teichmüller geodesic

$$\{g_t S\}_{t\in\mathbb{R}} = \left\{ \begin{pmatrix} e^t & 0 \\ 0 & e^{-t} \end{pmatrix} S \right\}$$

eventually leaves any fixed compact subset $K \subset \mathcal{M}_g$ in the moduli space and never visits it again [80]. Another example of the connection between the dynamics on a flat surface and on the moduli space is (a version of) Veech's dichotomy: if the $\mathrm{GL}^+(2,\mathbb{R})$-orbit of S is closed in the stratum $H(\kappa) \subset M_g$, then any directional flow on S is either completely periodic or uniquely ergodic. Therefore, it is reasonable to associate problems of dynamics on a flat surface with corresponding problems of dynamics on the moduli space $\mathcal{M}_g$ or in the stratum. In particular, a full understanding of the dynamics on $\mathcal{M}_g$ is crucial for the study of the dynamical properties of the geodesic flow on flat surfaces.

The moduli space $\mathcal{M}_g$ normally has a complicated topological structure being a non-compact orbifold, so it is often convenient to consider its compactification. For $g \geqslant 1$, consider the bundle $\Omega\mathcal{M}_{g,n} \to \mathcal{M}_{g,n}$ whose fiber over $(S, x_1, \dots, x_n)$ is the space of meromorphic forms ω on S having simple poles at $x_1, \dots, x_n$. The moduli space $\Omega\mathcal{M}_{g,n}$ appears naturally as a boundary component in the Deligne-Mumford compactification of moduli spaces $\Omega\mathcal{M}_g$ of holomorphic 1-forms. The moduli spaces of 1-forms have a natural period coordinate system, namely, to each pair $(S,\omega) \in \Omega\mathcal{M}_{g,n}$ one can associate a period map $p \in \mathrm{Hom}(H_1(\Sigma_{g,n^*}), \mathbb{C})$, where $p(\gamma) = \int_{f^*\gamma} \omega$ and f is a map from the reference surface Σ_{g,n^*} to S. Therefore period coordinates appear naturally as a coordinate system of $\Omega\mathcal{M}_{g,n}$. However, knowing the period coordinates of (S,ω) does not allow us to recover the pair (S,ω) even infinitesimally. In other words, it is always possible to find non-trivial isoperiodic deformations on $\Omega\mathcal{M}_{g,n}$ that give rise to the isoperiodic foliation on $\Omega\mathcal{M}_{g,n}$ [17]. When considering Deligne-Mumford compactification of $\Omega\mathcal{M}_g$, we see that the different isoperiodic foliations glue together and define a global algebraic foliation of a resulting compact space.

It is known that for degree at least 3 the isoperiodic sets of $\Omega\mathcal{M}_g, g \geqslant 2$, with no marked points are connected (see [17] for the proof and other dynamic properties, such as ergodicity of the foliation). However, it turns out that the same methods do not apply to the study of the isoperiodic foliations of moduli spaces of meromorphic 1-forms. Notably, not much is known about the isoperiodic foliations on $\Omega\mathcal{M}_{g,n}, n \geqslant 1$ (an overview can be found in the introduction of Chapter 6). The methods that were proposed to study the isoperiodic foliations of $\Omega\mathcal{M}_{g,n}, n \geqslant 1$, heavily depend on the values of g, n (known results are restricted to $n \leqslant 2$; see [18]). In Chapter 6 we propose a new geometric method of studying the isoperiodic sets of $\Omega\mathcal{M}_{g,n}$ that can be applied to $\Omega\mathcal{M}_{g,n}$ for any (small) value of g, n. We demonstrate the method by proving a new result, namely, that the real isoperiodic sets of $\Omega\mathcal{M}_{1,3}$ are connected (Theorem 6.1.4). The significance of the result of Chapter 6 is not only the novelty of the method: it seems that the general statements for arbitrary g, n can be proved by induction upon providing a sufficient induction base (similarly to [17]). Thus, Theorem 6.1.4 serves as the base of induction for our further research.

In conclusion, the present thesis addresses problems that arise in different mathematical areas - algebra, probability, geometry, statistical mechanics, graph theory. Nevertheless, the problems that we treat are similar in nature and are aimed at understanding limiting behaviour associated to dynamical systems. As this thesis demonstrates, the methods that apply in different mathematical contexts are diverse, and range from Diophantine approximations to combinatorics, linear algebra, functional analysis, and more.

Chapter 2

Tropical limits of decimated polynomials[1]

2.1 Introduction

Our reader at some point might have encountered the following mathematical puzzle:

Is it possible to tile the 8×8 *chessboard without two opposite corners with* 2×1 *dominoes?*

A simple parity argument shows that it is, indeed, not possible. A more difficult question is in how many ways we can tile the usual 8×8, and, more generally, $2n \times 2n$ chessboard with the 2×1 dominos?

In 1961, a Dutch physicist Piet Kasteleyn found complete solutions of several 'arrangement problems' of such nature [53]. In particular, he showed that the number of domino tilings of a chessboard of a size $2n \times 2n$ is given by

$$Z_n = \prod_{m=0}^{n-1} \prod_{k=0}^{n-1} \left(4 - 2\cos\left(\frac{2m+1}{2n+1}\pi\right) - 2\cos\left(\frac{2k+1}{2n+1}\pi\right)\right). \tag{2.1}$$

There is a one-to-one correspondence between the domino tilings and dimer configurations, or, in other words, the perfect matchings of a corre-

[1]This chapter is based on: E. Arzhakova, E. Verbitskiy, Tropical Limits of Decimated Polynomials. Arnold Math J. 5, 57–67 (2019). https://doi.org/10.1007/s40598-019-00108-9

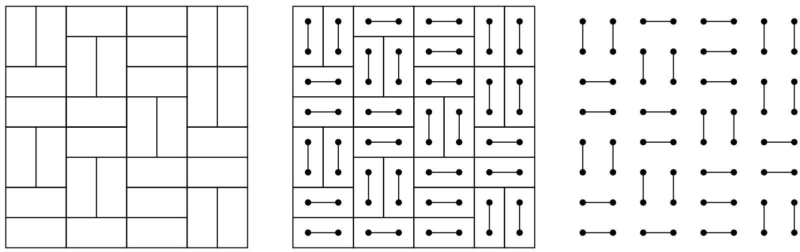


Figure 2.1: Correspondence between the domino tilings and the dimer matchings on boxes $2n \times 2n$.

sponding graph, as Figure 2.1 demonstrates. A *perfect matching* of a graph is a subset of its edges such that every vertex of the graph is incident to exactly one edge of the subset.

The method developed by Kasteleyn is not only applicable to counting the number of domino tilings, or equivalently, dimer configurations, but can also be used to compute the weighted sum of the form

$$Z_n = \sum_{M \in \mathcal{M}(G_{2n\times 2n})} w(M), \tag{2.2}$$

where $\mathcal{M}(G_{2n\times 2n})$ is a collection of all dimer configurations of the square box of size $2n \times 2n$. For any dimer configuration (matching) M, its weight is given by

$$w(M) = \prod_{e\in M} w(e), \quad w(e) = \begin{cases} u, & \text{for horizontal edges,} \\ v & \text{for vertical edges,} \end{cases}, \quad u, v > 0.$$

Obtaining explicit expressions for the *partition function* Z_n (2.2) is important in Statistical Physics, as it allows the computation of the *free energy* $F(u, v)$ given by

$$F(u, v) = -\lim_{n\to\infty} \frac{1}{n^2} \log Z_n, \tag{2.3}$$

and, by analysing the free energy, one is able to determine some important macroscopic properties of the systems in the thermodynamic limit. In particular, the singularities of the free energy function indicate the presence of the phase transitions.

It turns out [22, 55], that it is easier to compute a weighted partition function of the form (2.2) if we embed the $2n \times 2n$ square grid on a torus (see Fig. 2.2).

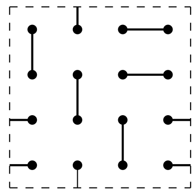


Figure 2.2: A dimer configuration (matching) on the 4×4 box on a torus.

The partition function $Z_n^{\mathbb{T}}$ for weighted dimer matchings on a torus possesses an elegant explicit expression:

$$Z_n^{\mathbb{T}} = \frac{1}{2}\left(-f_n(1,1) + f_n(-1,1) + f_n(1,-1) + f_n(-1,-1)\right), \tag{2.4}$$

where, for every integer $n \geqslant 1$,

$$f_n(x,y) = \prod_{m=0}^{n-1}\prod_{k=0}^{n-1} f\left(e^{2\pi i \frac{m}{n}}x, e^{2\pi i \frac{k}{n}}y\right). \tag{2.5}$$

with f being a Laurent polynomial in two variables x, y, namely,

$$f(x,y) = 4(u^2+v^2) - u^2(x+x^{-1}) - v^2(y+y^{-1}).$$

The limiting free energy F does not depend on whether we consider the partition function Z_n or $Z_n^{\mathbb{T}}$. In both cases,

$$\lim_{n\to\infty} \frac{1}{n^2}\log Z_n = \lim_{n\to\infty}\frac{1}{n^2}\log Z_n^{\mathbb{T}} = \int_0^1\int_0^1 \log|f(e^{2\pi i\theta_1}, e^{2\pi i\theta_2})|d\theta_1 d\theta_2. \tag{2.6}$$

More generally, expressions of a form (2.4) and (2.6) hold for all planar simple bipartite $\mathbb{Z}^2$-periodic graphs and some associated polynomials f. For example, for the honeycomb lattice, $f(x,y) = a+bx+cy$, where a, b, c are the weights of the horizontal, north-east, and south-east edges, respectively.

Kenyon, Okounkov, and Sheffield (see [55]) obtained a beautiful limiting shape result for the coefficients of $f_n(x,y)$ in (2.5) for the special 'dimer' polynomials f. In the present note we discuss the generalization of the result of Kenyon, Okounkov, and Sheffield to arbitrary Laurent polynomials in d variables, where $d \geqslant 2$.

2.2 Decimation of polynomials

A Laurent polynomial in d commuting variables $z_1, \ldots, z_d$, can be presented as a sum

$$f(\boldsymbol{z}) = \sum_{\boldsymbol{m}\in\mathbb{Z}^d} f^{\boldsymbol{m}} \boldsymbol{z}^{\boldsymbol{m}},$$

where we use the multi-index notation $\boldsymbol{m} = (m_1, \ldots, m_d)$. The sum has a finite number of terms: there are only finitely many $\boldsymbol{m}$'s with $f^{\boldsymbol{m}} \neq 0$. The multi-indices $\boldsymbol{m} \in \mathbb{Z}^d$ with $f^{\boldsymbol{m}} \neq 0$ are called the *exponent vectors.* The set of all exponent vectors is called *the support* of f and is denoted by $\mathsf{supp}(f)$.

Now, fix an arbitrary integer $n \geqslant 1$. Then, the n-th *decimation* of a Laurent polynomial $f(z_1, \ldots, z_d)$ is defined as

$$f_n(z_1, \ldots, z_d) = \prod_{k_1=1}^{n} \cdots \prod_{k_d=1}^{n} f\left(e^{2\pi i k_1/n} z_1, \ldots, e^{2\pi i k_d/n} z_d\right). \qquad (2.7)$$

Our interest in these polynomials arose when studying decimations (renormalization transformation) of the so-called principle algebraic actions – a natural class of algebraic dynamics, see [1] for more details. Such polynomials have been considered earlier by Purbhoo [92] who studied approximations of amoebas. In the present paper, we will discuss properties of the decimated polynomials.

For every n, a decimated polynomial f_n is again a Laurent polynomial

$$f_n(\boldsymbol{z}) = \sum_{\boldsymbol{m}\in\mathbb{Z}^d} f_n^{\boldsymbol{m}} \boldsymbol{z}^{\boldsymbol{m}}.$$

Moreover, the resulting exponent vectors of f_n are entry-wise divisible by n. In other words, f_n is a Laurent polynomial in $z_1^n, \ldots, z_d^n$.

Example 2.2.1. Consider a polynomial $f(x, y) = 1 - x - y$. Then, the first three decimations are (see Fig. 2.3):

1. $f_1(x, y) = 1 - x - y$;
2. $f_2(x, y) = 1 - 2x^2 - 2y^2 - 2x^2y^2 + y^4 + x^4$;
3. $f_3(x, y) = 1 - 3x^3 - 3y^3 + 3x^6 + 3y^6 - 3x^6y^3 - 3x^3y^6 - 21x^3y^3 - y^9 - x^9$.

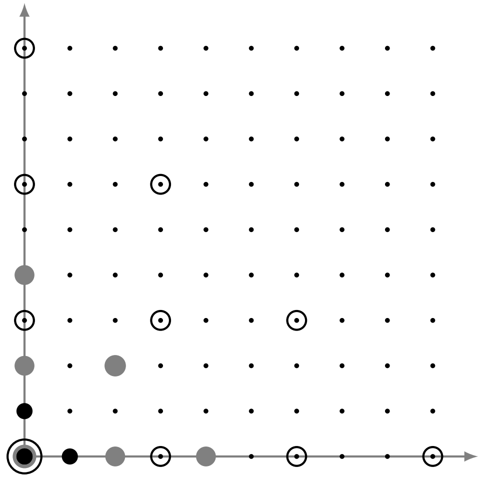


Figure 2.3: Black circles correspond to the exponent vectors of $f = f_1$, namely, $(0,0)$,$(1,0)$, and $(0,1)$; the exponent vectors of f_2 and f_3 are denoted by the grey and the white circles, respectively.

We remind the reader that the *Newton polytope* $\mathcal{N}(f)$ of a Laurent polynomial $f(z_1,\ldots,z_d)$ is a subset of $\mathbb{R}^d$ which is a convex hull of the exponent vectors of $f(z_1,\ldots,z_d)$. Note that the Newton polytopes of f_n and of f satisfy the following relation:

$$\mathcal{N}(f_n) = n^d \mathcal{N}(f).$$

Therefore, when n increases, the Newton polytope of f_n grows, and so do the absolute values of the non-zero coefficients $f_n^{\boldsymbol{m}}$. In fact, their growth rate is exponential in n. The natural question is whether there is a scaling limit of the coefficients of f_n. Namely, whether the limits

$$\lim_{n\to\infty} \frac{1}{n^d} \log |f_n^{\boldsymbol{m}_n}|, \tag{2.8}$$

exist for sequences $\boldsymbol{m}_n \in \mathsf{supp}(f_n)$ such that $\frac{\boldsymbol{m}_n}{n^d} \to \boldsymbol{u} \in \mathcal{N}(f)$.

2.3 The scaling limit

In tropical algebra, the standard addition and multiplication of real numbers are redefined as follows:

- tropical addition: $a \oplus b = \max\{a, b\}$;

- tropical multiplication: $a \odot b = a + b$.

Hence, $2 \oplus 5 = 5$ and $2 \odot 5 = 7$. The tropical operations allow to define tropical polynomials. For example, consider $f(x, y) = 2x^2 - 4x^2y + y$; its tropical analogue is then

$$F(x, y) = (2 \odot x \odot x) \oplus (-4 \odot x \odot x \odot y) \oplus (1 \odot y).$$

Using the tropical operations, defined above, one can easily evaluate F at any $(x, y) \in \mathbb{R}^2$

$$F(x, y) = \max\{2 + 2x, -4 + 2x + y, y\}.$$

Tropical geometry incorporates many facets of algebraic geometry and convex analysis [74].

Let us consider a Laurent polynomial $f(\boldsymbol{z}) = \sum_{\boldsymbol{m}} f^{\boldsymbol{m}} \boldsymbol{z}^{\boldsymbol{m}}$. The *tropicalization* of $f(\boldsymbol{z})$, denoted by $\mathsf{trop}(f)(\boldsymbol{t})$, is a function on $\mathbb{R}^d$ defined as follows: for any $\boldsymbol{t} = (t_1, \ldots, t_d) \in \mathbb{R}^d$, take

$$\begin{aligned} \mathsf{trop}(f)(\boldsymbol{t}) &= \bigoplus_{\boldsymbol{m} \in \mathrm{supp}(f)} \log|f^{\boldsymbol{m}}| \odot \boldsymbol{t}^{\boldsymbol{m}} \\ &= \max_{\boldsymbol{m} \in \mathrm{supp}(f)} \Big(\log|f^{\boldsymbol{m}}| + m_1 t_1 + \ldots m_d t_d \Big) \\ &= \max_{\boldsymbol{m} \in \mathrm{supp}(f)} \Big(\log|f^{\boldsymbol{m}}| + \langle \boldsymbol{m}, \boldsymbol{t} \rangle \Big), \end{aligned} \tag{2.9}$$

where $\langle \cdot, \cdot \rangle$ denotes the standard scalar product on $\mathbb{R}^d$.
Tropicalization of f is thus a tropical analogue of the Laurent polynomial $\sum_{\boldsymbol{m}} \log|f^{\boldsymbol{m}}| \boldsymbol{z}^{\boldsymbol{m}}$.

The *tropical variety* of $f(\boldsymbol{z})$ is the set of all points $\boldsymbol{t} \in \mathbb{R}^d$ such that the maximum in (2.9) is achieved at at least two monomials. Therefore, the tropicalization of f is a piecewise affine convex function on $\mathbb{R}^d$; each component of the complement of the tropical variety defines a domain where a certain monomial of f is maximal, c.f. (2.9). Figure 2.4 shows the tropical varieties of the first 4 decimations of a polynomial $1 + x + y$.

In a joint work with Doug Lind and Klaus Schmidt [1] (see Chapter 3), we established the following result on the existence of scaling limits of tropicalizations of the decimated polynomials f_n.

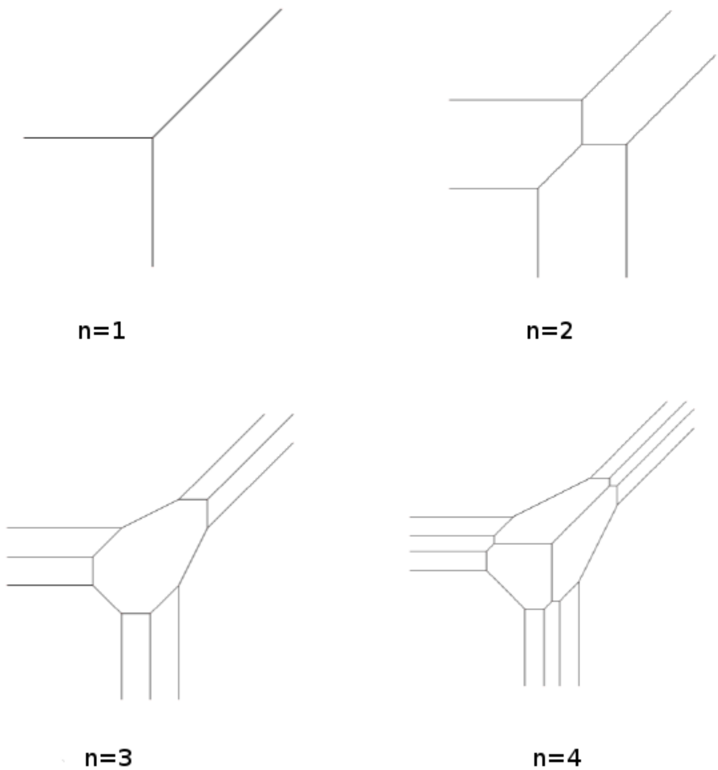


Figure 2.4: Tropical varieties of $\mathsf{trop}(f_n)$ for $f = 1 + x + y$ and $n = 1, 2, 3, 4$.

Theorem 2.3.1. For every non-zero Laurent polynomial $f(\boldsymbol{z})$ and all $\boldsymbol{t} \in \mathbb{R}^d$, there exists a limit

$$\lim_{n\to\infty} \frac{1}{n^d} \mathsf{trop}(f_n)(\boldsymbol{t}) = R_f(\boldsymbol{t}), \tag{2.10}$$

where $R_f : \mathbb{R}^d \to \mathbb{R}$ is the Ronkin function of f, given by

$$\begin{aligned} R_f(\boldsymbol{t}) &= \int_0^1 \cdots \int_0^1 \log |f(e^{t_1+2\pi i\theta_1}, \ldots, e^{t_d+2\pi i\theta_d})| d\theta_1 \ldots d\theta_d \\ &= \int_{\boldsymbol{t}\in\mathbb{T}^d} \log\big|f(e^{\boldsymbol{t}+2\pi i\boldsymbol{\theta}})\big| d\boldsymbol{\theta}. \end{aligned} \tag{2.11}$$

Sketch of the proof. (The full proof can be found in [1], Chapter 3.) The first observation relates the Ronkin functions of f and f_n; namely, $R_{f_n}(\boldsymbol{t}) = n^d R_f(\boldsymbol{t})$. Hence, it suffices to compare $\mathsf{trop}(f_n)$ and R_f.

The Ronkin function can be easily bounded from above: indeed, let us denote by $\sharp\mathsf{supp}(f_n)$ the number of integer points inside the support of f_n. Then,

$$\begin{aligned}
R_{f_n}(\boldsymbol{t}) &= \int_{\mathbb{T}^d} \log\Big|\sum_{\boldsymbol{m}} f_n^{\boldsymbol{m}} e^{\langle \boldsymbol{t},\boldsymbol{m}\rangle} e^{2\pi i\langle \boldsymbol{m},\boldsymbol{t}h\rangle}\Big| d\boldsymbol{t}h \\
&\leqslant \int_{\mathbb{T}^d} \log \sum_{\boldsymbol{m}\in\mathsf{supp}(f_n)} \Big| f_n^{\boldsymbol{m}} e^{\langle \boldsymbol{t},\boldsymbol{m}\rangle} e^{2\pi i\langle \boldsymbol{m},\boldsymbol{t}h\rangle}\Big| d\boldsymbol{t}h \\
&= \int_{\mathbb{T}^d} \log \sum_{\boldsymbol{m}\in\mathsf{supp}(f_n)} \Big| f_n^{\boldsymbol{m}} e^{\langle \boldsymbol{t},\boldsymbol{m}\rangle}\Big| \\
&\leqslant \log\left(\sharp\mathsf{supp}(f_n) \max_{\boldsymbol{m}\in\mathsf{supp}(f_n)} \exp\left(\log|f_n^{\boldsymbol{m}}| + \langle \boldsymbol{t},\boldsymbol{m}\rangle\right)\right) \\
&= \log\sharp\mathsf{supp}(f_n) + \max_{\boldsymbol{m}\in\mathsf{supp}(f_n)} \left(\log|f_n^{\boldsymbol{m}}| + \langle \boldsymbol{t},\boldsymbol{m}\rangle\right) \\
&= \log\sharp\mathsf{supp}(f_n) + \mathsf{trop}(f_n)(\boldsymbol{t}).
\end{aligned}$$

Hence, for all $n\in\mathbb{Z}$ and $\boldsymbol{t}\in\mathbb{R}^d$ one has

$$R_f(\boldsymbol{t}) = \frac{1}{n^d} R_{f_n}(\boldsymbol{t}) \leqslant \frac{1}{n^d}\log\sharp\mathsf{supp}(f_n) + \frac{1}{n^d}\mathsf{trop}(f_n)(\boldsymbol{t}).$$

Since the cardinality of the support of f_n grows at most as const $\cdot\, n^d$, we immediately conclude that

$$R_f(\boldsymbol{t}) \leqslant \liminf_{n\to\infty} \frac{1}{n^d}\mathsf{trop}(f_n)(\boldsymbol{t}). \tag{2.12}$$

Let us start the discussion of the lower bound of $R_f(\boldsymbol{t})$ with the following observation. Suppose that $\boldsymbol{z} = (z_1,\dots,z_d) \in (\mathbb{C}^*)^d$ is a d-tuple of non-zero complex numbers; denote by $t_j\in\mathbb{R}$ and $\varphi_j\in\mathbb{T}\sim[0,1)$ the modulus and the argument of z_j for every $j=1,\dots,d$, i.e., $z_j = e^{t_j+2\pi i\varphi_j}$. Note that

$$\frac{1}{n^d}|\log f_n(\boldsymbol{z})| = \frac{1}{n^d}\sum_{k_1=0}^{n-1}\cdots\sum_{k_d=0}^{n-1}\log\left|f\left(e^{t_1+2\pi i\varphi_1+2\pi i k_1/n},\dots,e^{t_d+2\pi i\varphi_d+2\pi i k_d/n}\right)\right|. \tag{2.13}$$

The expression on the right hand side is a Riemann sum for the integral (2.11) defining the Ronkin function $R_f(\boldsymbol{t})$. Note also, that despite the fact f may have zeros on a torus $\{\boldsymbol{z} : |\boldsymbol{z}| = e^{\boldsymbol{t}}\}$, the function $\log|f|$ is still integrable since the singularities are only logarithmic. One naturally expects that for almost all $\boldsymbol{z}$, the Riemann sums in (2.13) converge to $R_f(\boldsymbol{t})$. However, establishing such convergence turns out to be a rather intricate Diophantine problem [25, 65].

Fortunately, in order to establish the lower bound, one does not have to deal with a convergence problem in a complete generality. It suffices to prove that the Riemann sums are bounded from above by $R_f(\boldsymbol{t})$.

We say that the initial value $\boldsymbol{z} = (e^{t_1+2\pi i\varphi_1}, \dots, e^{t_d+2\pi i\varphi_d})$ is good if the points of the set

$$Q_n(\boldsymbol{z}) = \left\{ \left(e^{t_1+2\pi i\varphi_1+2\pi i k_1/n}, \dots, e^{t_d+2\pi i\varphi_d+2\pi i k_d/n} \right), \quad k_1, \dots k_d = 0, \dots n-1 \right\}$$

do not fall or come too close (depending on n) to the variety of f: $V_f = \{\boldsymbol{z} \in (\mathbb{C}^*)^d : f(\boldsymbol{z}) = 0\}$. For good points, it is relatively easy to show that the Riemann sums are close to the value of the integral $R_f(\boldsymbol{t})$. On the contrary, for the 'bad' initial values $\boldsymbol{z}$, the points in $Q_n(\boldsymbol{z})$, which are close to V_f, give a negative contribution to the sum (2.13). Hence, one is able to conclude that for all $\boldsymbol{z}$ with $|z_1| = e^{t_1}, \dots, |z_d| = e^{t_d}$,

$$\limsup_{n\to\infty} \frac{1}{n^d} \log |f_n(\boldsymbol{z})| \leqslant R_f(\boldsymbol{t}), \tag{2.14}$$

or, equivalently,

$$|f_n(\boldsymbol{z})| \leqslant \exp\left(n^d (R_f(\boldsymbol{t}) + o(1)\right). \tag{2.15}$$

The final part of the argument is based on a relatively simple statement from Fourier analysis: if the absolute value of a complex (trigonometric) polynomial is bounded from above by some constant M on a torus $\mathbb{T}_{\boldsymbol{t}} = \{\boldsymbol{z} : |\boldsymbol{z}| = e^{\boldsymbol{t}}\}$ then the absolute values of all of its monomials are also bounded from above by the same constant. Therefore, applying this result to f_n and the inequality (2.15), we conclude that that for all $\boldsymbol{m}$

$$|f_n^{\boldsymbol{m}} e^{\langle \boldsymbol{t}, \boldsymbol{m}\rangle}| \leqslant \exp\left(n^d (R_f(\boldsymbol{t}) + o(1)\right),$$

and, hence,

$$\mathsf{trop}(f_n)(\boldsymbol{t}) = \max_{\boldsymbol{m}} \left(\log |f_n^{\boldsymbol{m}}| + \langle \boldsymbol{t}, \boldsymbol{m}\rangle\right) \leqslant n^d (R_f(\boldsymbol{t}) + o(1)). \tag{2.16}$$

Combining the inequalities (2.12) and (2.16), we obtain the desired result. □

Remark 2.3.2. Theorem 2.3.1 provides some insight on the geometry of tropical varieties of f_n. In Figure 2.4, the similarity between the shapes of tropical varieties of decimations of $f = 1+x+y$ for various n is not accidental. Let us recall the notion of an amoeba of a Laurent polynomial f of d variables that was first suggested by Gelfand, Kapranov, and Zelevinsky in [36]. An *amoeba* of f denoted by $\mathcal{A}_f$ is an image of the variety V_f under the map $\mathrm{Log}\colon V_f \mapsto \mathbb{R}^d$ given by the formula $\mathrm{Log}(z_1, \dots, z_d) = (\log |z_1|, \dots, \log |z_d|)$.

The amoeba A_f is a closed subset of $\mathbb{R}^d$ with a non-empty convex complement (see Figure 2.5). The Ronkin function R_f is strictly convex over A_f and affine on each component of $\mathbb{R}^d \backslash A_f$ [83].

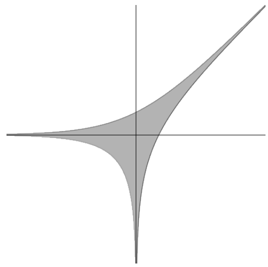


Figure 2.5: Amoeba A_f for $f = 1 + x + y$.

It is easy to see that A_f coincides with A_{f_n} for every positive n. According to Theorem 2.3.1, $\frac{1}{n^d}\mathsf{trop}(f_n)(\boldsymbol{t})$ is a sequence of piecewise-affine convex functions converging to the Ronkin function R_f which is affine on the complement of A_f and strictly convex inside A_f. Therefore, the outer boundary of the tropical varieties $\frac{1}{n^d}\mathsf{trop}(f_n)$ converge to the boundary of A_f as n approaches infinity.

2.3.1 Surface tension

The *Legendre transform* (or a dual) of a function $F : \mathbb{R}^d \to \mathbb{R}$ is defined as

$$F^*(\boldsymbol{t}) = \sup_{\boldsymbol{u} \in \mathbb{R}^d} \left(\langle \boldsymbol{t}, \boldsymbol{u} \rangle - F(\boldsymbol{u}) \right).$$

The Legendre transform F^* is always a convex function; moreover, for a convex closed function F, the Legendre transform is an involution

$$F^{**}(\boldsymbol{u}) = F(\boldsymbol{u}).$$

Suppose f is a Laurent polynomial. Then, the tropicalization of f is, in fact, a Legendre transform of the function F defined as follows:

$$F(\boldsymbol{u}) = \begin{cases} -\log|f^{\boldsymbol{m}}|, & \text{if } \boldsymbol{u} = \boldsymbol{m} \in \mathsf{supp}(f), \\ +\infty, & \text{otherwise}, \end{cases}.$$

Indeed, one has

$$F^*(\boldsymbol{t}) = \sup_{\boldsymbol{u} \in \mathbb{R}^d} \left(\langle \boldsymbol{t}, \boldsymbol{u} \rangle - F(\boldsymbol{u}) \right) = \sup_{\boldsymbol{m} \in \mathsf{supp}(f)} \left(\langle \boldsymbol{t}, \boldsymbol{m} \rangle - (-\log|f^{\boldsymbol{m}}|) \right) = \mathsf{trop}(f)(\boldsymbol{t}).$$

Applying the Legendre transform once again, we obtain $F^{**} = \mathsf{trop}(f)^*$. Since F, in general, is not convex, $F^{**} \neq F$. However, F^{**} is easy to describe; namely, $F^{**} = \mathsf{conv}(F)$, where $\mathsf{conv}(F)$ is the so-called *greatest convex minorant* or a *convex hull* of F: the largest convex function satisfying $\mathsf{conv}(F)(\boldsymbol{u}) \leqslant F(\boldsymbol{u})$ for all $\boldsymbol{u}$. Clearly, $\mathsf{conv}(F)(\boldsymbol{u}) = +\infty$ for al $\boldsymbol{u} \notin \mathcal{N}(f)$ and is finite on $\mathcal{N}(f)$.

Using the above arguments for the polynomials f_n and the result of the Theorem 2.3.1, one immediately obtains the following result:

Corollary 2.3.3. Let $\mathsf{conv}(F_n)$ be the greatest convex minorant of

$$F^{(n)} = \begin{cases} -\log|f_n^{\boldsymbol{m}}|, & \text{if } \boldsymbol{u} = \boldsymbol{m} \in \mathsf{supp}(f_n), \\ +\infty, & \text{otherwise.} \end{cases}$$

Then, for all $\boldsymbol{u} \in \mathbb{R}^d$,

$$\sigma_f(\boldsymbol{u}) := \lim_{n\to\infty} \frac{1}{n^d} \mathsf{conv}(F_n)(n^d \boldsymbol{u}) = -R_f^*(\boldsymbol{u}).$$

By analogy with [55], we refer to the function σ_f as to the *surface tension of* f.

Corollary 2.3.3 should be seen as a weaker, but at the same time, a more general version of the result established in [55] for polynomials appearing in dimer problems. It turns out that these polynomials are rather special in the following sense: for such polynomials, one is able to define the surface tension using the coefficients of f_n without the need to resort to convex hulls of the coefficients. In other words, some form of convexity is already present in the coefficients of f_n. It is, of course, very interesting to identify such polynomials. Okounkov and Kenyon [87] showed that for every Harnack curve, one can construct a polynomial of 2 variables, whose algebraic variety is the given Harnack curve, and for which the surface tension is well defined. At the present moment, it is not clear which conditions could play a similar role in higher dimensions.

Finally, we would like to remark that the use of tropical geometric methods to study limits of partition functions or similar quantities is very natural, and has been proposed in [34].

Chapter 3

Decimation limits of principal algebraic $\mathbb{Z}^d$-actions[1]

Abstract

Let f be a Laurent polynomial in d commuting variables with integer coefficients. Associated to f is the principal algebraic $\mathbb{Z}^d$-action α_f on a compact subgroup X_f of $\mathbb{T}^{\mathbb{Z}^d}$ determined by f. Let $N \geqslant 1$ and restrict points in X_f to coordinates in $N\mathbb{Z}^d$. The resulting algebraic $N\mathbb{Z}^d$-action is again principal, and is associated to a polynomial g_N whose support grows with N and whose coefficients grow exponentially with N. We prove that by suitably renormalizing these decimations we can identify a limiting behavior given by a continuous concave function on the Newton polytope of f, and show that this decimation limit is the negative of the Legendre dual of the Ronkin function of f. In certain cases with two variables, the decimation limit coincides with the surface tension of random surfaces related to dimer models, but the statistical physics methods used to prove this are quite different and depend on special properties of the polynomial.

3.1 Introduction

Let $d \geqslant 1$ and $f \in \mathbb{Z}[x_1^{\pm 1}, \ldots, x_d^{\pm 1}]$ be a Laurent polynomial with integer coefficients in d commuting variables. We write $f(x_1, \ldots, x_d) = f(\mathbf{x}) =$

[1]This chapter is based on: E. Arzhakova, D. Lind, K. Schmidt, E. Verbitskiy, Decimation limits of principal algebraic $\mathbb{Z}^d$-actions, arXiv:2104.04408

$\sum_{\mathbf{n}\in\mathbb{Z}^d} \widehat{f}(\mathbf{n})\mathbf{x}^{\mathbf{n}}$, where $\mathbf{x}^{\mathbf{n}} = x_1^{n_1}\dots x_d^{n_d}$ and $\widehat{f}(\mathbf{n}) \in \mathbb{Z}$ for all $\mathbf{n} \in \mathbb{Z}^d$ and is nonzero for only finitely many $\mathbf{n} \in \mathbb{Z}^d$.

Denote the additive torus $\mathbb{R}/\mathbb{Z}$ by $\mathbb{T}$. Use f to define a compact subgroup X_f of $\mathbb{T}^{\mathbb{Z}^d}$ by

$$X_f := \Big\{ t \in \mathbb{T}^{\mathbb{Z}^d} : \sum_{\mathbf{n}\in\mathbb{Z}^d} \widehat{f}(\mathbf{n}) t_{\mathbf{m}+\mathbf{n}} = 0 \quad \text{for all } \mathbf{m} \in \mathbb{Z}^d \Big\}. \tag{3.1}$$

By its definition this subgroup is invariant under the natural shift-action σ of $\mathbb{Z}^d$ on $\mathbb{T}^{\mathbb{Z}^d}$ defined by $\sigma^{\mathbf{n}}(t)_{\mathbf{m}} = t_{\mathbf{m}-\mathbf{n}}$. Hence the restriction α_f of σ to X_f gives an action of $\mathbb{Z}^d$ by automorphisms of the compact abelian group X_f. We call (X_f, α_f) the *principal algebraic $\mathbb{Z}^d$-action defined by f*.

Such $\mathbb{Z}^d$-actions serve as a rich class of examples and have been studied intensively. An observation of Halmos [46] shows that α_f automatically preserves Haar measure μ_f on X_f. It is known that the topological entropy of α_f coincides with its measure-theoretic entropy with respect to μ_f. For nonzero f this common value was computed in [64] to be the logarithmic Mahler measure of f, defined as

$$m(f) := \int_0^1 \cdots \int_0^1 \log |f(e^{2\pi i s_1}, \dots, e^{2\pi i s_d})| \, ds_1 \dots ds_d \tag{3.2}$$

(when $f = 0$ the entropy is infinite).

It will be convenient to identify the Laurent polynomial ring $\mathbb{Z}[x_1^{\pm1}, \dots, x_d^{\pm1}]$ with the integral group ring $\mathbb{Z}[\mathbb{Z}^d]$, where the monomial $\mathbf{x}^{\mathbf{n}}$ corresponds to $\mathbf{n} \in \mathbb{Z}^d$. Thus $f \in \mathbb{Z}[x_1^{\pm1}, \dots, x_d^{\pm1}]$ is identified with its coefficient function $\widehat{f} \colon \mathbb{Z}^d \to \mathbb{Z}$. When emphasizing the behavior of coefficients we will always use the notation $\widehat{f}$.

Fix a principal algebraic $\mathbb{Z}^d$-action (X_f, α_f). Let $N \geqslant 1$ and $r_N \colon \mathbb{T}^{\mathbb{Z}^d} \to \mathbb{T}^{N\mathbb{Z}^d}$ be the map restricting the coordinates of a point to only those in the sublattice $N\mathbb{Z}^d$. We call the image $r_N(X_f)$ the *Nth decimation* of X_f, although this is considerably more brutal that the term's original meaning since only every Nth coordinate survives. Clearly $r_N(X_f)$ is again a compact abelian group, and it is invariant under the natural shift action of $N\mathbb{Z}^d$ on $\mathbb{T}^{N\mathbb{Z}^d}$.

Using commutative algebra applied to contracted ideals in integral extensions, we show in §3.6 that $r_N(X_f)$ is a principal algebraic $N\mathbb{Z}^d$-action with some defining polynomial $g_N \in \mathbb{Z}[N\mathbb{Z}^d]$. Typically the support of g_N grows with N and its coefficient function $\widehat{g}_N$ grows exponentially in N. Our goal in this paper is to prove that with suitable renormalizations the concave hulls of the resulting functions converge uniformly on the Newton polytope of f to a continuous decimation limit D_f. Furthermore, D_f can be computed via Legendre duality using a well-studied object called the Ronkin function of f.

The analytical parts of our analysis apply to Laurent polynomials with complex coefficients. For such an $f \in \mathbb{C}[\mathbb{Z}^d]$ we define its Nth *decimation* f_N by

$$f_N(x_1, \dots, x_d) := \prod_{k_1=0}^{N-1} \cdots \prod_{k_d=0}^{N-1} f(e^{2\pi i k_1/N} x_1, \dots, e^{2\pi i k_d/N} x_d). \tag{3.3}$$

Since f_N is unchanged after multiplying each of its variables by an arbitrary Nth root of unity, it follows that it is a polynomial in the Nth powers of the x_i, i.e., that $f_N \in \mathbb{C}[N\mathbb{Z}^d]$. Decimations of polynomials have appeared in many contexts, including Purbhoo's approximations to shapes of complex amoebas [92], Boyd's proof that the Mahler measure of a polynomial is continuous in its coefficients [11], and dimer models in statistical physics [55].

For most $f \in \mathbb{Z}[\mathbb{Z}^d]$ the generator g_N of the Nth decimation of X_f coincides with f_N. But under special circumstances characterized in §3.6, involving the support of f and the Galois properties of the coefficients of the polynomials occurring in the factorization of f over the algebraic closure of the rationals, it can happen that g_N is a proper divisor of f_N. To give a simple example when $d = 1$, let $f(x) = x^2 - 2$. Then since f is already in $\mathbb{Z}[2\mathbb{Z}]$ we have that $g_2(x) = f(x)$, while $f_2(x) = f(x)f(-x) = f(x)^2$. Nevertheless even in these circumstances the renormalization behavior of the g_N can be determined from that of the f_N.

For $f \in \mathbb{C}[\mathbb{Z}^d]$ let $\operatorname{supp} f = \{\mathbf{n} \in \mathbb{Z}^d : \widehat{f}(\mathbf{n}) \neq 0\}$ denote its support. The *Newton polytope* $\mathcal{N}_f$ of f is the convex hull in $\mathbb{R}^d$ of $\operatorname{supp} f$. Since f_N is the product of N^d polynomials all of whose Newton polytopes are $\mathcal{N}_f$, it follows that $\mathcal{N}_{f_N} = N^d \mathcal{N}_f$.

The *Ronkin function* $R_f \colon \mathbb{R}^d \to \mathbb{R}$ of $0 \neq f \in \mathbb{C}[\mathbb{Z}^d]$ is defined by

$$R_f(u_1, \dots, u_d) := \int_0^1 \cdots \int_0^1 \log |f(e^{u_1} e^{2\pi i s_1}, \dots, e^{u_d} e^{2\pi i s_d})| \, ds_1 \dots ds_d. \tag{3.4}$$

This is a convex function on $\mathbb{R}^d$, and therefore has a Legendre dual R_f^* defined by

$$R_f^*(\mathbf{r}) := \sup\{\mathbf{r} \cdot \mathbf{u} - R_f(\mathbf{u}) : \mathbf{u} \in \mathbb{R}^d\},$$

which turns out to be a convex function on $\mathcal{N}_f$ (and is ∞ off $\mathcal{N}_f$).

To describe rescaling of polynomials $g \in \mathbb{C}[\mathbb{Z}^d]$ it is convenient to extend the domain of $\widehat{g}$ from $\mathbb{Z}^d$ to $\mathbb{R}^d$ by declaring its value to be 0 off $\operatorname{supp} g$.

Let $\varphi\colon \mathbb{R}^d \to \mathbb{C}$. For any $a > 0$ define the *rescaling operator* E_a on φ by $(E_a\varphi)(\mathbf{r}) = \varphi(a\mathbf{r})$ for all $\mathbf{r} \in \mathbb{R}^d$. When dealing with concave functions it is often convenient to use the extended range $\underline{\mathbb{R}} = \mathbb{R} \cup \{-\infty\}$, with the usual algebraic rules for handling $-\infty$ and with the convention that $\log 0 = -\infty$. Then $\log|\varphi|\colon \mathbb{R}^d \to \underline{\mathbb{R}}$, and we define its *concave hull* $CH(\log|\varphi|)$ to be the infimum of all affine functions on $\mathbb{R}^d$ that dominate $\log|\varphi|$.

Let $f \in \mathbb{C}[\mathbb{Z}^d]$ and f_N be its Nth decimation. Define the Nth *logarithmic rescaling* $L_N f$ of f by

$$L_N f := E_{N^d}\Big(\frac{1}{N^d}\log|\widehat{f_N}|\Big).$$

Clearly $L_N f(\mathbf{r}) = -\infty$ if $\mathbf{r} \notin \mathcal{N}_f$, and is finite at every extreme point of $\mathcal{N}_f$ and at only finitely many other points in $\mathcal{N}_f$. The Nth *renormalized decimation* $D_N f$ of f is the concave hull $CH(L_N f)$ of $L_N f$. By our previous remark, $D_N f$ equals $-\infty$ off $\mathcal{N}_f$ and is finite at every point of $\mathcal{N}_f$.

With these preparations we can now state one of our main results.

Theorem 3.1.1. Let $0 \neq f \in \mathbb{C}[\mathbb{Z}^d]$. Then the Nth renormalized decimations $D_N f$ of f are concave polyhedral functions on the Newton polytope $\mathcal{N}_f$ of f that converge uniformly on $\mathcal{N}_f$ as $N \to \infty$ to a continuous concave decimation limit function D_f (and off $\mathcal{N}_f$ they are equal to $-\infty$). Furthermore $D_f = -R_f^*$, where R_f^* is the Legendre dual of the Ronkin function R_f of f.

The proof of this theorem uses two main ideas: Mahler's fundamental estimate [77] relating the largest coefficient of a polynomial to its Mahler measure and support, and a method used by Boyd [11], applied to decimations along powers of 2, to prove that for polynomials whose support is contained in a fixed finite subset of $\mathbb{Z}^d$ the Mahler measure is a continuous function of their coefficients.

If $f \in \mathbb{Z}[\mathbb{Z}^d]$ the decimation limit of f contains dynamical information about α_f.

Corollary 3.1.2. Let $0 \neq f \in \mathbb{C}[\mathbb{Z}^d]$. Then the maximum value of the decimation limit D_f on the Newton polytope $\mathcal{N}_f$ equals the logarithmic Mahler measure $m(f)$ of f defined in (3.2). In particular, if $f \in \mathbb{Z}[\mathbb{Z}^d]$ then this maximum value equals the entropy of the principal algebraic $\mathbb{Z}^d$-action α_f.

Duality allows us to compute the decimation limit of a product of two polynomials. Suppose that $\varphi, \psi \colon \mathbb{R}^d \to \underline{\mathbb{R}}$ both have finite supremum. Define their *tropical convolution* $\varphi \circledast \psi$ by

$$(\varphi \circledast \psi)(\mathbf{r}) := \sup\{\varphi(\mathbf{s}) + \psi(\mathbf{r}-\mathbf{s}) : \mathbf{s} \in \mathbb{R}^d\}.$$

This is the tropical analogue of standard convolution, but using tropical (or max-plus) arithmetic in $\underline{\mathbb{R}}$.

Corollary 3.1.3. Let f and g be nonzero polynomials in $\mathbb{C}[\mathbb{Z}^d]$. Then $\mathsf{D}_{fg} = \mathsf{D}_f \circledast \mathsf{D}_g$.

Thus decimation limits live in the tropics.

3.2 Examples

Here we give some examples to illustrate the phenomenon we are investigating. They use either one or two variables, and for these we denote the variables by x and y rather than x_1 and x_2. Let $\Omega_N = \{e^{2\pi i k/N} : 0 \leqslant k < N\}$ denote the group of Nth roots of unity.

Example 3.2.1. Let $d = 1$ and $f(x) = x^2 - x - 1 = (x-\lambda)(x-\mu)$, where $\lambda = (1+\sqrt{5})/2$ and $\mu = (1-\sqrt{5})/2$. Then

$$\begin{aligned} f_N(x) &= \prod_{\omega \in \Omega_N} f(\omega x) = \prod_{\omega \in \Omega_N} (\omega x - \lambda)(\omega x - \mu) \\ &= (x^N - \lambda^N)(x^N - \mu^N) = x^{2N} - (\lambda^n + \mu^N)x^N + (-1)^N. \end{aligned}$$

Hence

$$(\mathsf{L}_N f)(r) = \begin{cases} 0 & \text{if } r = 0 \text{ or } 2, \\ \frac{1}{N}\log|\lambda^N + \mu^N| & \text{if } r = 1, \\ -\infty & \text{otherwise.} \end{cases}$$

Since $\mathsf{L}_N f(1) \to \log\lambda$ as $N \to \infty$, the concave hulls $\mathsf{D}_N f$ converge uniformly on $\mathcal{N}_f = [0,2]$ to the decimation limit

$$\mathsf{D}_f(r) = \begin{cases} r\log\lambda & \text{if } 0 \leqslant r \leqslant 1, \\ (2-r)\log\lambda & \text{if } 1 \leqslant r \leqslant 2, \\ -\infty & \text{otherwise,} \end{cases}$$

which is shown in Figure 3.1(a).

To compute the Ronkin function R_f, recall Jensen's formula that for every $\xi \in \mathbb{C}$ we have that

$$\int_0^1 \log |e^{2\pi i s} - \xi| \, ds = \max\{0, \log |\xi|\} := \log^+ |\xi|. \tag{3.5}$$

Thus

$$\begin{aligned} R_f(u) = \int_0^1 \log |f(e^u e^{2\pi i s})| \, ds &= \int_0^1 \log |e^u e^{2\pi i s} - \lambda| \, ds + \int_0^1 \log |e^u e^{2\pi i s} - \mu| \, ds \\ &= 2u + \log^+ |e^{-u}\lambda| + \log^+ |e^{-u}\mu|, \end{aligned}$$

whose polygonal graph is depicted in Figure 3.1(b). It is then easy to verify using the definition of Legendre transform that $D_f = -R_f^*$.

Finally, the decimation limits $D_{x-\lambda}$ and $D_{x-\mu}$ are computed similarly, and shown in Figures 3.1(c) and 3.1(d). It is easy to check using the definition of tropical convolution that $D_{x-\lambda} \circledast D_{x-\mu} = D_{(x-\lambda)(x-\mu)} = D_f$, in agreement with Corollary 3.1.3.

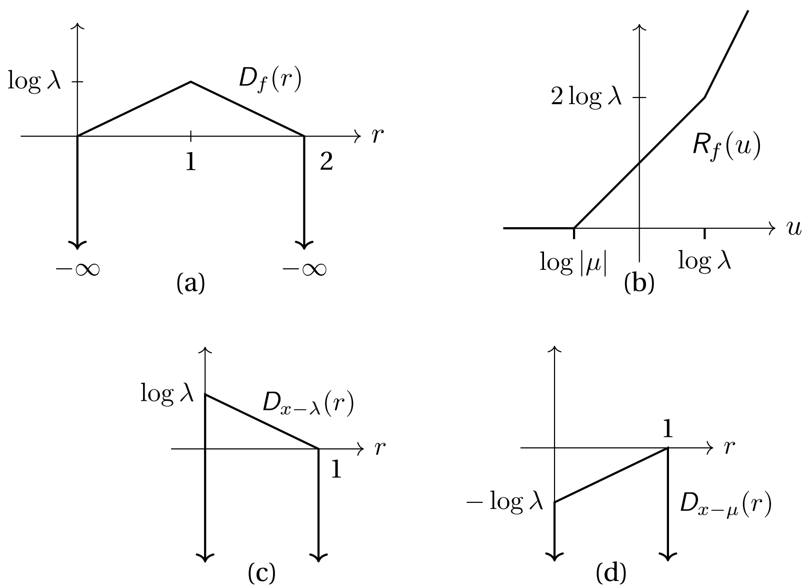


Figure 3.1: Graphs in Example 3.2.1

More generally, if $f(x) = \prod_{j=1}^m (x - \lambda_j)$ and $|\lambda_1| > |\lambda_2| > \cdots > |\lambda_m|$, then a computation similar to that in Example 3.2.1 shows that $(L_N f)(m) = 0$ and that $(L_N f)(k)$ converges to $\log |\lambda_1 \lambda_2 \ldots \lambda_{m-k}|$ for $k = 0, 1, \ldots, m-1$, and this gives uniform convergence of $D_N f$ to D_f on $\mathcal{N}_f = [0, m]$. However, if some roots of f have equal absolute value, then convergence is more delicate, or may even fail, as the next two examples show.

Example 3.2.2. Let $d = 1$ and $f(x) = x^4 - 4x^3 - 2x^2 - 4x + 1$, which is irreducible in $\mathbb{Z}[\mathbb{Z}]$. The roots of f are $\lambda = 1 + \sqrt{2} + \sqrt{2\sqrt{2}+2} \approx 4.611$, $\mu = 1 + \sqrt{2} - \sqrt{2\sqrt{2}+2} \approx 0.217$, and $1 - \sqrt{2} \pm i\sqrt{2\sqrt{2}-2} = e^{\pm 2\pi i\theta}$, where θ is irrational. Simple estimates show that $(\mathsf{L}_N f)(k)$ converges for $k = 0, 1, 3, 4$ with limits $0, \log\lambda, \log\lambda, 0$, respectively. However, the dominant term controlling the behavior of $(\mathsf{L}_N f)(2)$ is

$$\frac{1}{N}\log|2\lambda^N\cos(2\pi N\theta)|.$$

Since θ is irrational, the factor $\cos(2\pi N\theta)$ occasionally becomes very small, and so convergence is in question.

In fact, $(\mathsf{L}_N f)(2)$ does converge, but the proof requires a deep result of Gelfond [37, Thm. III, p. 28] on the diophantine properties of algebraic numbers on the unit circle. According to this result, if ξ is an algebraic number (such as $e^{2\pi i\theta}$ above) such that $|\xi| = 1$ and ξ is not a root of unity, and if $\varepsilon > 0$, then $|\xi^n - 1| > e^{-n\varepsilon}$ for all but finitely many n. From this it is easy to deduce that $|e^{2\pi i N\theta} - i| > e^{-N\varepsilon}$ for almost every N, and hence that $(1/N)\log|\cos(2\pi N\theta)| \to 0$ as $N \to \infty$. This convergence is illustrated in Figure 3.2(a).

Both $(\mathsf{L}_N f)(1)$ and $(\mathsf{L}_N f)(3)$ converge to $\log\lambda$, and clearly it holds that $\limsup_{N\to\infty}(\mathsf{L}_N f)(2) \leqslant \log\lambda$. Hence any lack of convergence of $(\mathsf{L}_N f)(2)$ would not affect the limiting behavior of the concave hull $\mathsf{D}_N f$, nor uniform convergence of $\mathsf{D}_N f$ to D_f on $[0, 4]$. Thus such diophantine issues are covered up by taking concave hulls.

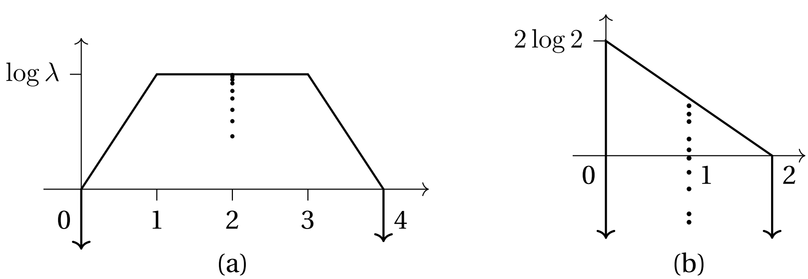
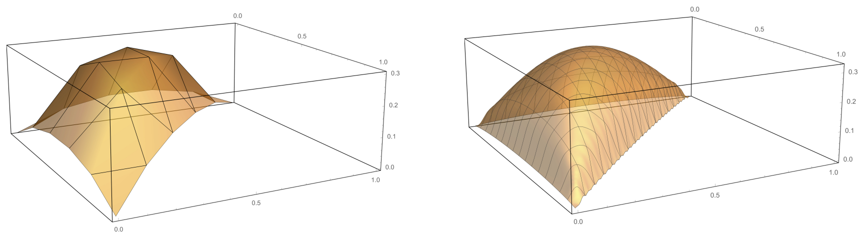


Figure 3.2: (a) Convergence in Example 3.2.2, and (b) lack of convergence in Example 3.2.3

The next example shows that if we allow the coefficients of f to be arbitrary complex numbers instead of integers, then $(\mathsf{L}_N f)(k)$ can fail badly to converge at some k.



(a) (b)

Figure 3.3: (a) Polyhedral approximation $D_5 f$, and (b) limiting smooth surface D_f for $f(x, y) = 1 + x + y$ in Example 3.2.4

Example 3.2.3. Let $d = 1$ and $f(x) = (x - 2e^{2\pi i\theta})(x - 2e^{-2\pi i\theta})$, where we will determine θ. Then $(L_N f)(0) = 2\log 2$ and $(L_N f)(2) = 0$ for all $N \geqslant 1$, while

$$(L_N f)(1) = \frac{1}{N}\log|2^N \cdot 2\cos(2\pi N\theta)|.$$

It is possible to construct an irrational θ and a sequence $N_j \to \infty$ such that $\frac{1}{N_j}\log|\cos(2\pi N_j\theta)| \to -\infty$ as $j \to \infty$ Hence using this value of θ to define f we see that $(L_N f)(1)$ does not converge, as depicted in Figure 3.2(b), although the concave hulls $D_N f$ do converge uniformly to D_f.

Using arguments similar to those above, it is possible to give an elementary direct proof of Theorem 3.1.1 in the case $d = 1$.

Example 3.2.4. Let $d = 2$ and $f(x, y) = 1 + x + y$. Then f_N is a polynomial in x^N and y^N of degree N^2. For example,

$$\begin{aligned} f_{\langle 5\rangle}(x, y) = \; & x^{25} + 5x^{20}y^5 + 5x^{20} + 10x^{15}y^{10} - 605x^{15}y^5 + 10x^{15} + 10x^{10}y^{15} \\ & + 1905x^{10}y^{10} + 1905x^{10}y^5 + 10x^{10} + 5x^5y^{20} - 605x^5y^{15} + 1905x^5y^{10} \\ & - 605x^5y^5 + 5x^5 + y^{25} + 5y^{20} + 10y^{15} + 10y^{10} + 5y^5 + 1. \end{aligned}$$

The Nth logarithmic rescaling $L_N f$ of f is finite at points in the unit simplex $\Delta = \mathcal{N}_f$ whose coordinates are integer multiples of $1/N$. Thus its concave hull $D_N f$ is a polyhedral surface over Δ, and as $N \to \infty$ these surfaces converge uniformly on Δ to the graph of the concave decimation limit D_f. Figure 3.3(a) shows the polyhedral surface $D_5 f$ corresponding to the calculation of $f_{\langle 5\rangle}$ above, and Figure 3(b) depicts the limiting smooth surface for D_f.

For this example it is possible to derive an explicit formula for D_f. Clearly

$\mathsf{D}_f(r,s)$ is symmetric in r and s, so we may assume that $s \leqslant r$. Let

$$\Delta_1 = \{(r,s) \in \Delta : s \leqslant r \text{ and } s \leqslant (1-r)/2\}, \tag{3.6}$$

$$\Delta_2 = \{(r,s) \in \Delta : s \leqslant r \text{ and } s \geqslant (1-r)/2\}. \tag{3.7}$$

For $(r,s) \in \Delta_1 \cup \Delta_2$ with $r+s<1$ define

$$b(r,s) = \csc[\pi(r+s)]\sin(\pi s).$$

Then it turns out that $0 \leqslant b(r,s) \leqslant 1$ for $(r,s) \in \Delta_1$ while $1 \leqslant b(r,s) < \infty$ for $(r,s) \in \Delta_2$.

Using Legendre duality and calculations of R_f by Lundqvist [70], we will show in Appendix A that if $(r,s) \in \Delta_1$ then

$$\mathsf{D}_f(r,s) = \sum_{n=1}^{\infty} \frac{(-1)^{n+1}}{\pi n^2} b(r,s)^n \sin[n\pi(1-r)] - s \log b(r,s), \tag{3.8}$$

while if $(r,s) \in \Delta_2$ then

$$\mathsf{D}_f(r,s) = \sum_{n=1}^{\infty} \frac{(-1)^{n+1}}{\pi n^2} b(r,s)^{-n} \sin[n\pi(1-r)] + (1-r-s)\log b(r,s). \tag{3.9}$$

We will prove in Corollary 3.1.2 that the maximum value of D_f equals the entropy of α_f, which is the logarithmic Mahler measure $\mathsf{m}(f)$ of f defined in (3.2). In this example, the maximum value is attained at $(1/3, 1/3)$, which is in both Δ_1 and Δ_2. Either formula therefore applies, and each gives Smyth's calculation [105] that

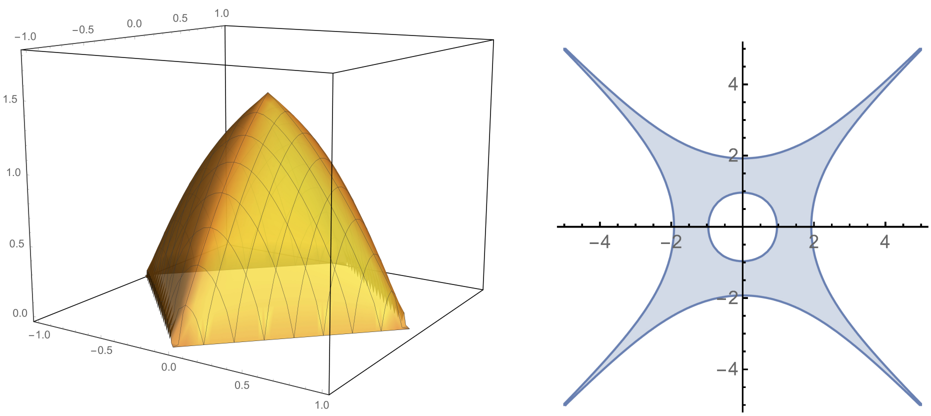
$$\mathsf{m}(1+x+y) = \mathsf{D}_f(1/3,1/3) = \frac{3\sqrt{3}}{4\pi}\sum_{n=1}^{\infty}\frac{\chi_3(n)}{n^2} = \frac{3\sqrt{3}}{4\pi} L(2,\chi_3) \approx 0.3230, \tag{3.10}$$

where χ_3 is the nontrivial character of $\mathbb{Z}/3\mathbb{Z}$ and $L(s,\chi_3)$ is the L-function associated with χ_3.

Unlike the previous example, some decimation limits exhibit non-smooth behavior.

Example 3.2.5. Let $d=2$ and $f(x,y) = 5+x+x^{-1}+y+y^{-1}$. The decimation limit D_f is depicted in Figure 3.4(a). The non-smooth peak at the origin is due to a "hole" in the amoeba of f, as defined in §3.4 and shown in Figure 3.4(b).

As in the previous example, the decimation limit describes the surface tension for a physical model, in this case dimer tilings of the square-octagon graph (see [55, Fig. 3].



(a) (b)

Figure 3.4: (a) The decimation limit for $f(x, y) = 5 + x + x^{-1} + y + y^{-1}$ from Example 3.2.5 , and (b) the "hole" in its amoeba causing the peak.

Remark 3.2.6. Dimer models have a long history in statistical physics. A particularly important instance involves $f(x, y) = 1 + x + y$ from Example 3.2.4, and has been studied in enormous detail by many authors, including Kenyon, Okounkov, and Sheffield [55].

To describe this model, let $\mathscr{H}$ denote the regular hexagonal lattice in $\mathbb{R}^2$. We can assign the vertices of $\mathscr{H}$ alternating colors red and black, much like a checkerboard. A *perfect matching* on $\mathscr{H}$ is an assignment of each red vertex to a unique adjacent black vertex, these forming an edge or *dimer*. A perfect matching is equivalent to a tiling of $\mathbb{R}^2$ by three types of lozenges, one type for each of the three edges incident to each vertex. Using a natural height function, such a lozenge tiling gives a surface, and the study of the statistical properties of such random surfaces has resulted in many remarkable discoveries (see Kenyon's survey [56], Okounkov's survey [86], or Gorin's detailed account of lozenge tilings [39]).

Kasteleyn discovered that by cleverly assigning signs to the edges of $\mathscr{H}$, he could compute the number of perfect matchings on a finite approximation using periodic boundary conditions by a determinant formula. Furthermore, this determinant can be explicitly evaluated to have the form of a decimation of $f(x, y) = 1 + x + y$. Each of the three terms of f correspond to one of the three types of lozenges in the random tiling. It then turns out that in the logarithmic scaling limit $D_f(r, s)$ counts the growth rate of perfect matchings for which the frequencies of the three lozenge types are r, s, and $1 - r - s$. As such, it is called the *surface tension* for this model.

The two-variable polynomials with integer coefficients arising from such dimer models, such as the preceding two examples, define curves of a very special type called Harnack curves. For these there are probabilistic interpretations of the coefficients of decimations. The additional structure enables one to show that the individual nonzero coefficients of f_N grow at a rate predicted by D_f. Example 3.2.3 shows this can fail if complex coefficients are allowed. But whether or not this is true for every polynomial in $\mathbb{Z}[\mathbb{Z}^d]$ for all $d \geqslant 1$ appears to be quite an interesting problem (see Question 3.9.3 for a precise formulation).

3.3 Convex functions and Legendre duals

We briefly review some basic facts about convex functions and their Legendre duals. Rockafellar's classic book [94] contains a comprehensive account of this theory.

Let $\overline{\mathbb{R}}$ denote $\mathbb{R} \cup \{\infty\}$, with the standard conventions about arithmetic operations and inequalities involving ∞. Let $\varphi\colon \mathbb{R}^d \to \overline{\mathbb{R}}$ be a function, and define its *epigraph* by

$$\operatorname{epi}\varphi := \{(\mathbf{u}, t) : \mathbf{u} \in \mathbb{R}^d, t \in \mathbb{R}, \text{ and } t \geqslant \varphi(\mathbf{u})\} \subset \mathbb{R}^d \times \mathbb{R}.$$

Then φ is *convex* provided that $\operatorname{epi}\varphi$ is a convex subset of $\mathbb{R}^d \times \mathbb{R}$. Similarly, a function $\psi\colon \mathbb{R}^d \to \underline{\mathbb{R}}$ is *concave* if $-\psi\colon \mathbb{R}^d \to \overline{\mathbb{R}}$ is convex.

The *effective domain* of a convex function φ is defined by

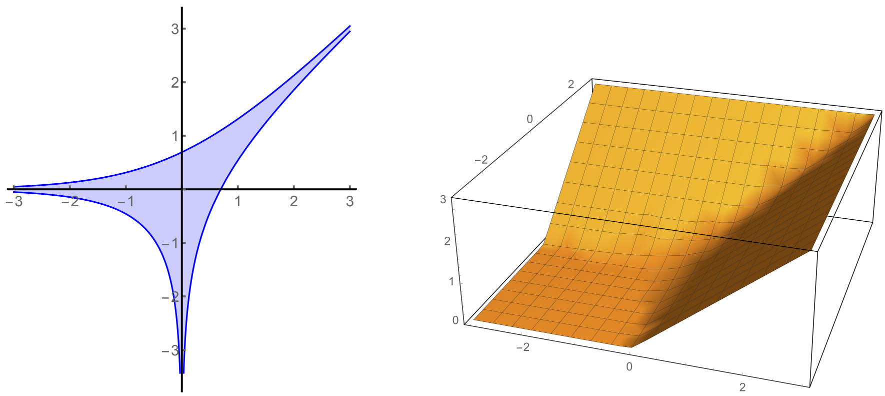
$$\operatorname{dom}\varphi := \{\mathbf{u} \in \mathbb{R}^d : \varphi(\mathbf{u}) < \infty\}.$$

By allowing φ to take the value ∞, we may assume that it is defined on all of $\mathbb{R}^d$, enabling us to combine convex functions without needing to take into account their effective domains. A convex function is *closed* if its epigraph is a closed subset of $\mathbb{R}^d \times \mathbb{R}$. This property normalizes the behavior of a convex function at the boundary of its effective domain, and holds for all convex (and concave) functions that arise here.

Suppose that $\varphi\colon \mathbb{R} \to \overline{\mathbb{R}}$ is convex. Its *Legendre dual* (or, more accurately, its *Legendre-Fenchel dual*) φ^* is defined for all $\mathbf{r} \in \mathbb{R}^d$ by

$$\varphi^*(\mathbf{r}) := \sup\{\mathbf{r} \cdot \mathbf{u} - \varphi(\mathbf{u}) : \mathbf{u} \in \mathbb{R}^d\}. \tag{3.11}$$

The Legendre dual φ^* is also a convex function, and provides an alternative description of $\operatorname{epi}\varphi$ in terms of its support hyperplanes. Furthermore, Legendre duality states that $\varphi^{**} = \varphi$ for closed convex functions.



(a) (b)

Figure 3.5: (a) The amoeba of $1 + x + y$, and (b) its Ronkin function

The Legendre dual of a concave function $\psi\colon \mathbb{R}^d \to \mathbb{R}$ is similarly defined as

$$\psi^*(\mathbf{r}) = \inf\{\mathbf{r} \cdot \mathbf{u} - \psi(\mathbf{u}) : \mathbf{u} \in \mathbb{R}^d\}. \tag{3.12}$$

Then $\varphi = -\psi$ is convex, and a simple manipulation shows that their Legendre duals are related by $\psi^*(\mathbf{r}) = -\varphi^*(-\mathbf{r})$.

3.4 Amoebas and Ronkin functions

Let $0 \neq f \in \mathbb{C}[\mathbb{Z}^d]$. Put $\mathbb{C}^* = \mathbb{C} \smallsetminus \{0\}$ and define $V(f) := \{\mathbf{z} \in (\mathbb{C}^*)^d : f(\mathbf{z}) = 0\}$. Let $\mathrm{Log}\colon (\mathbb{C}^*)^d \to \mathbb{R}^d$ be the map $\mathrm{Log}(z_1, \dots, z_d) = (\log|z_1|, \dots, \log|z_d|)$.

In 1993 Gelfand, Kapranov, and Zelevinsky [36] introduced the notion of the *amoeba* $\mathcal{A}_f$ of f, defined as

$$\mathcal{A}_f := \mathrm{Log}\big(V(f)\big) \subset \mathbb{R}^d.$$

The amoeba of $1 + x + y$ is depicted in Figure 3.5(a). The complement $\mathcal{A}_f^c = \mathbb{R}^d \smallsetminus \mathcal{A}_f$ of $\mathcal{A}_f$ consists of a finite number of connected components, all convex. The unbounded components are created by "tentacles" of $\mathcal{A}_f$. Unfortunately, biological amoebas look nothing like their mathematical namesakes.

Closely related to $\mathcal{A}_f$ is the Ronkin function R_f of f, introduced by Ronkin [95] in 2001, and defined earlier in (3.4). The Ronkin function of $1 + x + y$ is shown in Figure 3.5(b).

The Ronkin function of a polynomial f is known to be a convex function on $\mathbb{R}^d$ and affine on each connected component of $\mathcal{A}_f^c$ (for this and much more see [88]). Moreover, on each connected component of $\mathcal{A}_f^c$ the (constant) gradient of R_f is contained in $\mathcal{N}_f \cap \mathbb{Z}^d$, and the convex hull of these values equals $\mathcal{N}_f$. From this we conclude that the Legendre dual R_f^* of R_f has effective domain $\mathcal{N}_f$.

3.5 Decimation limits of polynomials

In this section we prove Theorem 3.1.1, one of our main results, and Corollaries 3.1.2 and 3.1.3. If $0 \neq f \in \mathbb{C}[\mathbb{Z}^d]$ we will show that the Nth renormalized decimation $D_N f = CH(L_N f)$ converges uniformly on $\mathcal{N}_f$ to a continuous concave limit function D_f, and that $D_f = -R_f^*$.

The first ingredient in our proof is the basic estimate of Mahler relating the largest coefficient of a polynomial to its Mahler measure and its support. Let us begin with some terminology. For $0 \neq g \in \mathbb{C}[\mathbb{Z}^d]$ define its *height* $H(g)$ by $H(g) = \max\{\, |\widehat{g}(\mathbf{k})| : \mathbf{k} \in \mathbb{Z}^d\}$. The *Mahler measure* of g is $M(g) = \exp\bigl(m(g)\bigr)$, where $m(g)$ is the logarithmic Mahler measure defined in (3.2).

Proposition 3.5.1 (Mahler [77])**.** Suppose that $0 \neq g \in \mathbb{C}[\mathbb{Z}^d]$ and that $\operatorname{supp} g \subset [0, C-1]^d \cap \mathbb{Z}^d$. Then

$$2^{-dC} H(g) \leqslant M(g) \leqslant C^d H(g). \tag{3.13}$$

Proof. Let $\mathbf{k} = (k_1, \dots, k_d) \in \operatorname{supp} g$. Then by [77, Eqn. (3)],

$$|\widehat{g}(\mathbf{k})| \leqslant \binom{C-1}{k_1}\binom{C-1}{k_2}\cdots\binom{C-1}{k_d} M(g).$$

Since each binomial coefficient is bounded above by 2^C, the first inequality in (3.13) follows.

The second inequality is a simple consequence of the triangle inequality, since

$$M(g) \leqslant \sum_{\mathbf{k}\in[0,C-1]^d\cap\mathbb{Z}^d} |\widehat{g}(\mathbf{k})| \leqslant |[0, C-1]^d \cap \mathbb{Z}^d| \cdot H(g) = C^d H(g). \qquad \square$$

Consider $(\mathbb{C}^*)^d$ as a group under coordinate-wise multiplication. Define the action of $\mathbf{z} \in (\mathbb{C}^*)^d$ on $f \in \mathbb{C}[\mathbb{Z}^d]$ by $(\mathbf{z} \cdot f)(x_1, \dots, x_d) = f(z_1 x_1, \dots, z_d x_d)$. This action is commutative since

$$\mathbf{z} \cdot (\mathbf{z}' \cdot f) = (\mathbf{z}\mathbf{z}') \cdot f = \mathbf{z}' \cdot (\mathbf{z} \cdot f),$$

and also $\mathbf{z}\cdot(fg) = (\mathbf{z}\cdot f)(\mathbf{z}\cdot g)$ for all $f, g \in \mathbb{C}[\mathbb{Z}^d]$. Hence the map $f \mapsto \mathbf{z}\cdot f$ is a ring isomorphism of $\mathbb{C}[\mathbb{Z}^d]$. Furthermore, $(\mathbf{z}\cdot f)\widehat{\ }(\mathbf{k}) = \mathbf{z}^{\mathbf{k}}\widehat{f}(\mathbf{k})$ for all $\mathbf{k} \in \mathbb{Z}^d$, and so $\mathcal{N}_{\mathbf{z}\cdot f} = \mathcal{N}_f$ for all $\mathbf{z} \in (\mathbb{C}^*)^d$.

Recall that Ω_N denotes the group of Nth roots of unity. For $\boldsymbol{\omega} \in \Omega_N^d \subset (\mathbb{C}^*)^d$ we call $\boldsymbol{\omega}\cdot f$ the *rotate of f by $\boldsymbol{\omega}$*. Then $f_N = \prod_{\boldsymbol{\omega}\in\Omega_N^d} \boldsymbol{\omega}\cdot f$ is the product of all rotates of f by elements in Ω_N^d.

If $g, h \in \mathbb{C}[\mathbb{Z}^d]$ then is it well known that $\mathcal{N}_{gh} = \mathcal{N}_f + \mathcal{N}_g$ (the Minkowski sum), and trivially $R_{gh} = R_g + R_h$. By our previous remarks,

$$\mathcal{N}_{f_N} = \sum_{\boldsymbol{\omega}\in\Omega_N^d} \mathcal{N}_{\boldsymbol{\omega}\cdot f} = \sum_{\boldsymbol{\omega}\in\Omega_N^d} \mathcal{N}_f = N^d \mathcal{N}_f.$$

Also, $R_{\boldsymbol{\omega}\cdot f} = R_f$, and hence $R_{f_N} = N^d R_f$.

For $\mathbf{u} \in \mathbb{R}^d$ put $e^{\mathbf{u}} = (e^{u_1}, \dots, e^{u_d})$. Then $(e^{\mathbf{u}}\cdot f)\widehat{\ }(\mathbf{k}) = e^{\mathbf{u}\cdot\mathbf{k}}\widehat{f}(\mathbf{k})$. Commutativity of the action of $(\mathbb{C}^*)^d$ on f then shows that $(e^{\mathbf{u}}\cdot f)_{\langle N\rangle} = e^{\mathbf{u}}\cdot(f_N)$. Also

$$R_f(\mathbf{u}) = \log M(e^{\mathbf{u}}\cdot f) = \frac{1}{N^d}\log M\big((e^{\mathbf{u}}\cdot f)_{\langle N\rangle}\big) = \frac{1}{N^d}\log M(e^{\mathbf{u}}\cdot f_N).$$

Observe that

$$\log H(e^{\mathbf{u}}\cdot f_N) = \max\{\mathbf{u}\cdot\mathbf{k} + \log|\widehat{f}_N(\mathbf{k})| : \mathbf{k}\in\mathbb{Z}^d\},$$

indicating a connection with Legendre duals.

Proof of Theorem 3.1.1. Let $0 \neq f \in \mathbb{C}[\mathbb{Z}^d]$. For $\mathbf{m} \in \mathbb{Z}^d$ let $g(\mathbf{x}) = \mathbf{x}^{\mathbf{m}} f(\mathbf{x})$. It is straightforward to verify that $(D_N g)(\mathbf{r}) = (D_N f)(\mathbf{r} - \mathbf{m})$ for all $\mathbf{r} \in \mathbb{R}^d$. Therefore by adjusting f by suitable monomial, we may assume that $\operatorname{supp} f \subset [0, B-1]^d \cap \mathbb{Z}^d$ for some $B \geqslant 1$. Then $\operatorname{supp}(e^{\mathbf{u}}\cdot f_N) \subset [0, N^d(B-1)]^d \cap \mathbb{Z}^d \subset [0, N^d B - 1]^d \cap \mathbb{Z}^d$ for every $\mathbf{u} \in \mathbb{R}^d$. By Prop.3.5.1,

$$\begin{aligned} R_f(\mathbf{u}) = \frac{1}{N^d}\log M(e^{\mathbf{u}}\cdot f_N) &\leqslant \frac{1}{N^d}\Big\{\log\big[(N^d B)^d\big] + \log H(e^{\mathbf{u}}\cdot f_N)\Big\} \\ &= \frac{\log\big[(N^d B)^d\big]}{N^d} + \frac{1}{N^d}\max_{\mathbf{k}\in\mathbb{Z}^d}\{\mathbf{u}\cdot\mathbf{k} + \log|\widehat{f}_N(\mathbf{k})|\}, \end{aligned}$$

where the error term $b_N := N^{-d}\log\big[(N^d B)^d\big] \to 0$ as $N \to \infty$, uniformly for $\mathbf{u} \in \mathbb{R}^d$.

An opposite inequality is based of the following fundamental observation, used both by Boyd [11] and Purbhoo [92] for different purposes. As we noticed before, f_N is a polynomial in the Nth powers of the variables. Therefore $E_N\widehat{f}_N$ is again a polynomial to which we can apply Prop. 3.5.1, but with improved constants since the support has now shrunk by a factor of N. This improvement is crucial.

Specifically,

$$\operatorname{supp}(e^{\mathbf{u}} \cdot f_N) \subset [0, N^d(B-1)]^d \cap (N\mathbb{Z}^d),$$

so that

$$\operatorname{supp}(e^{\mathbf{u}} \cdot E_N f_N) \subset [0, N^{d-1}(B-1)] \cap \mathbb{Z}^d.$$

Applying Prop. 3.5.1,

$$H(e^{\mathbf{u}} \cdot f_N) = H(e^{\mathbf{u}} \cdot E_N f_N) \leqslant 2^{dN^{d-1}B} M(e^{\mathbf{u}} \cdot E_N f_N) = 2^{dN^{d-1}B} M(e^{\mathbf{u}} \cdot f)^{N^d}.$$

Hence

$$\frac{1}{N^d} \log H(e^{\mathbf{u}} \cdot f_N) \leqslant \frac{dN^{d-1}B \log 2}{N^d} + \log M(e^{\mathbf{u}} \cdot f) = a_N + R_f(\mathbf{u}),$$

where again the error term $a_N := (dB\log 2)/N \to 0$ uniformly for $\mathbf{u} \in \mathbb{R}^d$. We can summarize these estimates as

$$\left| R_f(\mathbf{u}) - \frac{1}{N^d} \max_{\mathbf{k}\in\mathbb{Z}^d} \{\mathbf{u}\cdot\mathbf{k} + \log|\widehat{f}_N(\mathbf{k})\}\right| \leqslant \max\{a_N, b_N\} \to 0 \tag{3.14}$$

as $N \to \infty$ uniformly in $\mathbf{u} \in \mathbb{R}^d$.

Next we relate the first max occurring in (3.14) with the Nth normalized decimation $D_N f$. We have that

$$\begin{aligned}
\frac{1}{N^d} \max_{\mathbf{k}\in\mathbb{Z}^d} \{\mathbf{u}\cdot\mathbf{k} + \log|\widehat{f}_N(\mathbf{k})|\} &= \max_{\mathbf{k}\in\mathbb{Z}^d} \Big\{\mathbf{u}\cdot\Big(\frac{\mathbf{k}}{N^d}\Big) + \frac{1}{N^d}\log|\widehat{f}_N(\mathbf{k})|\Big\} \\
&= \max_{\mathbf{k}\in\mathbb{Z}^d} \Big\{\mathbf{u}\cdot\Big(\frac{\mathbf{k}}{N^d}\Big) + \frac{1}{N^d} E_{N^d} \log\Big|\widehat{f}_N\Big(\frac{1}{N^d}\mathbf{k}\Big)\Big|\Big\} \\
&= \max_{\mathbf{k}\in\mathbb{Z}^d} \Big\{\mathbf{u}\cdot\Big(\frac{\mathbf{k}}{N^d}\Big) + (D_N f)\Big(\frac{1}{N^d}\mathbf{k}\Big)\Big\} \\
&= \max_{\mathbf{r}\in\mathbb{R}^d} \{\mathbf{u}\cdot\mathbf{r} + D_N f(\mathbf{r})\} = -(D_N f)^*(-\mathbf{u}).
\end{aligned}$$

Hence by (3.14), $-(D_N f)^*(-\mathbf{u})$ converges to $R_f(\mathbf{u})$ uniformly for $\mathbf{u} \in \mathbb{R}^d$, or, equivalently,

$$(D_N f)^*(\mathbf{u}) \to -R_f(-\mathbf{u}) \quad \text{uniformly for } \mathbf{u} \in \mathbb{R}^d. \tag{3.15}$$

If φ and ψ are concave functions on $\mathbb{R}^d$ such that $|\varphi(\mathbf{u}) - \psi(\mathbf{u})| \leqslant \varepsilon$ for all $\mathbf{u} \in \mathbb{R}^d$, it is easy to check from the definitions that φ^* and ψ^* have the same effective domain, and that $|\varphi^*(\mathbf{r}) - \psi^*(\mathbf{r})| \leqslant \varepsilon$ for all $\mathbf{r} \in \operatorname{dom}\varphi^* = \operatorname{dom}\psi^*$. Applying this to (3.15) and using duality we finally obtain that $(D_N f)^{**} = D_N f \to -R_f^*$ uniformly on $\mathcal{N}_f$, completing the proof. □

Proof of Cor. 3.1.2: By Theorem 3.1.1, Legendre duality, and (3.12),

$$-m(f) = -R_f(0,0) = D_f^*(0,0) = \inf_{(r,s)\in\mathcal{N}_f} -D_f(r,s) = -\sup_{(r,s)\in\mathcal{N}_f} D_f(r,s).$$

We remark that differentiability of D_f at the maximum value is not assumed for Legendre duality to apply here, and Example 3.2.5 provides a case when differentiability fails. □

Proof of Cor. 3.1.3: Let f and g be nonzero polynomials in $\mathbb{C}[\mathbb{Z}^d]$. Clearly $R_{fg} = R_f + R_g$. By [94, Thm. 16.4], the Legendre dual of the sum $\varphi + \psi$ of two convex functions is their infimal convolution defined for $\mathbf{r} \in \mathbb{R}^d$ by $\inf\{\varphi(\mathbf{s}) + \psi(\mathbf{r} - \mathbf{s}) : \mathbf{s} \in \mathbb{R}^d\}$. Applying this with $\varphi = -R_f$ and $\psi = -R_g$, using Thm. 3.1.1, and taking negatives we obtain that $D_{fg} = D_f \circledast D_g$. □

Remark 3.5.2. Our estimate (3.14) can be expressed in the language of tropicalization of polynomials (see [74, §3.1] for background and motivation). Let $0 \neq g(\mathbf{x}) = \sum_{\mathbf{k}\in\mathbb{Z}^d} \widehat{g}(\mathbf{k})\mathbf{x}^{\mathbf{k}} \in \mathbb{C}[\mathbb{Z}^d]$. Define the *tropicalization* of g to be the function $\operatorname{trop} g \colon \mathbb{R}^d \to \mathbb{R}$ given by

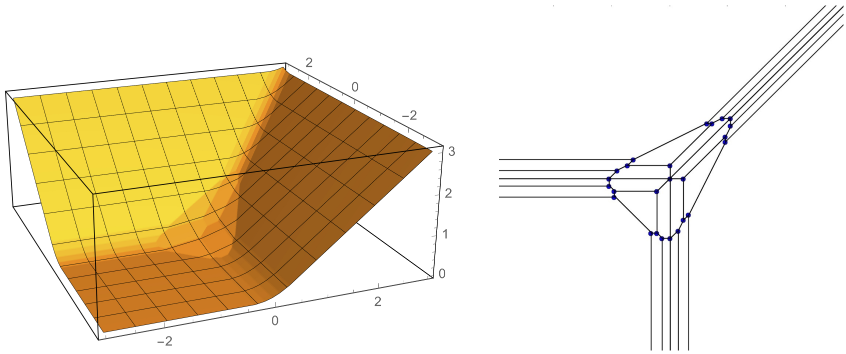
$$(\operatorname{trop} g)(\mathbf{u}) = \max_{\mathbf{k}\in\mathbb{Z}^d}\{\mathbf{u}\cdot\mathbf{k} + \log|\widehat{g}(\mathbf{k})|\},$$

which is a polyhedral convex function. Then by (3.14) we see that

$$\frac{1}{N^d} \operatorname{trop} f_N \to R_f \quad \text{uniformly on } \mathbb{R}^d, \tag{3.16}$$

so that the normalized tropicalization of f_N converges uniformly to the Ronkin function of f. Figure 3.6(a) depicts this polyhedral approximation for $f(x,y) = 1+x+y$ and $N = 5$ (compare with Figure 3.5(b)). The tropical variety of this polyhedral approximation is the projection to the plane of the vertices and edges of its graph, and is shown in 3.6(b). These tropical varieties converge in the Hausdorff metric to the amoeba of f as $N \to \infty$ (compare with Figure 3.5(a)).

Remark 3.5.3. In [92] Purbhoo used decimations for a different purpose, namely to find a computational way to detect whether or not a point is in the amoeba of a given polynomial. Call a polynomial *lopsided* if it has one coefficient whose absolute value strictly exceeds the sum of the absolute values of all the other coefficients. Let $f \in \mathbb{Z}[\mathbb{Z}^d]$ and $\mathbf{u} \in \mathbb{R}^d$. Clearly if $e^{\mathbf{u}} \cdot f$ is lopsided then $\mathbf{u} \notin \mathcal{A}_f$. Purbhoo used decimations to amplify size differences among the coefficients. More precisely, he proves that given $\varepsilon > 0$ there is an N_0, depending only on ε and the support of f, such that if the distance from $\mathbf{u}$ to $\mathcal{A}_f$ is greater than ε then $e^{\mathbf{u}}\cdot f_N$ is lopsided. Since f and f_N



(a) (b)

Figure 3.6: (a) Tropical approximation to the Ronkin function of $1 + x + y$, and (b) its corresponding tropical variety

have the same amoeba, this gives an effective algorithm for approximating the complement of $\mathcal{A}_f$.

One direct consequence of [92] is that the normalized tropicalizations in (3.16) converge to the Ronkin function off the amoeba of f, while our result is that this convergence is uniform on all of $\mathbb{R}^d$. Roughly speaking, Purbhoo is concerned with the coefficients of $e^{\mathbf{u}} \cdot f$ for points $\mathbf{u}$ off the amoeba, while our focus is on $\mathbf{u}$ within the amoeba.

Remark 3.5.4. Let F be a lower-dimensional face of the Newton polytope $\mathcal{N}_f$ of f, and put $f|_F = \sum_{\mathbf{n}\in F} \widehat{f}(\mathbf{n})\mathbf{x}^{\mathbf{n}}$. Clearly the restriction of D_f to F is just the decimation limit of $f|_F$, or in symbols $\mathsf{D}_f|_F = \mathsf{D}_{f|_F}$. By Corollary 3.1.2, this generalizes [64, Rem. 5.5], which gave a dynamical proof of the inequality due to Smyth [106, Thm. 2] that $\mathsf{m}(f) \geqslant \mathsf{m}(f_F)$ for every face F of $\mathcal{N}_f$.

3.6 Decimations of principal actions and contracted ideals

We return to decimations of principal algebraic $\mathbb{Z}^d$-actions, and in this section show that they are again principal. The proof uses machinery from commutative algebra, including contractions of ideals.

Suppose that X is a compact, shift-invariant subgroup of $\mathbb{T}^{\mathbb{Z}^d}$. Using Pontryagin duality we can obtain an alternative description of X as follows (for a comprehensive account see [98, Chap. II]).

As a discrete abelian group the Pontryagin dual of $\mathbb{T}^{\mathbb{Z}^d}$ is the direct sum of $\mathbb{Z}^d$ copies of $\mathbb{Z}$, which we suggestively write as $\bigoplus_{\mathbf{k}\in\mathbb{Z}^d}\mathbb{Z}\mathbf{x}^{\mathbf{k}} = \mathbb{Z}[\mathbb{Z}^d]$. The (additive) dual pairing between $\mathbb{T}^{\mathbb{Z}^d}$ and $\mathbb{Z}[\mathbb{Z}^d]$ is given by $\langle t, g\rangle = \sum_{\mathbf{k}\in\mathbb{Z}^d} t_{\mathbf{k}}\,\widehat{g}(\mathbf{k}) \in \mathbb{T}$. Multiplication by the inverses of each of the variables x_j on $\mathbb{Z}[\mathbb{Z}^d]$ gives a $\mathbb{Z}^d$-action that is dual to the natural shift action σ on $\mathbb{T}^{\mathbb{Z}^d}$ defined earlier.

Since X is shift-invariant, $\{g \in \mathbb{Z}[\mathbb{Z}^d] : \langle t, g\rangle = 0 \text{ for all } t \in X\}$ is an ideal $\mathfrak{a}$ in $\mathbb{Z}[\mathbb{Z}^d]$, and the dual group of X equals $\mathbb{Z}[\mathbb{Z}^d]/\mathfrak{a}$. Conversely, if $\mathfrak{a}$ is an arbitrary ideal in $\mathbb{Z}[\mathbb{Z}^d]$, then the compact dual group $X_{\mathfrak{a}}$ of $\mathbb{Z}[\mathbb{Z}^d]/\mathfrak{a}$ is a shift-invariant subgroup of $\mathbb{T}^{\mathbb{Z}^d}$. Thus there is a one-to-one correspondence between shift-invariant compact subgroups of $\mathbb{T}^{\mathbb{Z}^d}$ and ideals in $\mathbb{Z}[\mathbb{Z}^d]$. When $\mathfrak{a}$ is the principal ideal $\langle f\rangle$ generated by f, then $X_{\mathfrak{a}} = X_f$ as defined above, explaining the terminology "principal actions."

Fix $N \geqslant 1$ and recall the restriction map $r_N\colon \mathbb{T}^{\mathbb{Z}^d} \to \mathbb{T}^{N\mathbb{Z}^d}$ from §3.1. Let $f \in \mathbb{Z}[\mathbb{Z}^d]$. Then the Nth decimation $r_N(X_f)$ is a compact subgroup of $\mathbb{T}^{N\mathbb{Z}^d}$ that is invariant under the shift-action of $N\mathbb{Z}^d$. By our previous discussion, the dual group of $r_N(X_f)$ has the form $\mathbb{Z}[N\mathbb{Z}^d]/\mathfrak{a}_N$, where $\mathfrak{a}_N$ is an ideal in $\mathbb{Z}[N\mathbb{Z}^d]$. The following result identifies this ideal.

Lemma 3.6.1. Let $f \in \mathbb{Z}[\mathbb{Z}^d]$ and $N \geqslant 1$. Then the dual group of $r_N(X_f)$ is $\mathbb{Z}[N\mathbb{Z}^d]/\mathfrak{a}_N$, where $\mathfrak{a}_N = \langle f\rangle \cap \mathbb{Z}[N\mathbb{Z}^d]$.

Proof. Let $\mathfrak{b}_N = \{g \in \mathbb{Z}[N\mathbb{Z}^d] : \langle t, g\rangle = 0 \text{ for all } t \in r_N(X_f)\}$. If $g \in \mathfrak{a}_N$, then for every $t \in X_f$ we have that $0 = \langle t, g\rangle = \langle r_N(t), g\rangle$, so that $g \in \mathfrak{b}_N$. Conversely, if $g \in \mathfrak{b}_N$ and $t \in X_f$, then g annihilates the restriction of t to every coset of $N\mathbb{Z}^d$, and hence annihilates t, so that $g \in \mathfrak{a}_N$. □

The ideal $\langle f\rangle \cap \mathbb{Z}[N\mathbb{Z}^d]$ defining $r_N(X_f)$ is called the *contraction* of $\langle f\rangle$ to $\mathbb{Z}[N\mathbb{Z}^d]$. The main result of this section is that this contraction is always principal.

Proposition 3.6.2. Let $f \in \mathbb{Z}[\mathbb{Z}^d]$ and $N \geqslant 1$. Then the contracted ideal $\langle f\rangle \cap \mathbb{Z}[N\mathbb{Z}^d]$ is a principal ideal in $\mathbb{Z}[N\mathbb{Z}^d]$.

We begin by briefly sketching the necessary terminology and machinery from commutative algebra, all of which is contained in [5] or can be easily deduced from material there.

For brevity let $R = \mathbb{Z}[N\mathbb{Z}^d]$ and $S = \mathbb{Z}[\mathbb{Z}^d]$. Both R and S are unique factorization domains, and therefore both are integrally closed [5, Prop. 5.12]. Furthermore, S is integral over R since each variable x_j in S satisfies the monic polynomial $y^N - x_j^N \in R[y]$.

A prime ideal $\mathfrak{p}$ in an integral domain has *height one* if there are no prime ideals strictly between 0 and $\mathfrak{p}$. In a unique factorization domain the prime ideals of height one are exactly the principal ideals generated by irreducible elements. A proper ideal $\mathfrak{q}$ in an integral domain is *primary* if whenever $ab \in \mathfrak{q}$ then either $a \in \mathfrak{q}$ or $b^n \in \mathfrak{q}$ for some $n \geqslant 1$. In this case its radical $\{a : a^n \in \mathfrak{q} \text{ for some } n \geqslant 1\}$ is a prime ideal, say $\mathfrak{p}$, and then $\mathfrak{q}$ is called $\mathfrak{p}$-*primary*. Examples show that in general a power of a prime ideal need not be primary, that a primary ideal need not be the power of a prime ideal, and that even if an ideal has prime radical it need not be primary. The notion of primary ideal, although the correct one for decomposition theory, is quite subtle. However, in our situation things are much simpler.

Lemma 3.6.3. Let P be a unique factorization domain, and let $r \in P$ be irreducible. Then the principal ideal $\mathfrak{p} = \langle r \rangle$ is prime, and the $\mathfrak{p}$-primary ideals are exactly the powers $\mathfrak{p}^n$ of $\mathfrak{p}$ for $n \geqslant 1$.

Proof. It is clear that $\mathfrak{p}$ is prime. To prove that $\mathfrak{p}^n = \langle r^n \rangle$ is $\mathfrak{p}$-primary, suppose that $ab \in \mathfrak{p}^n$, but $a \notin \mathfrak{p}^n$. Then $r \mid b$, so $b^n \in \mathfrak{p}^n$, showing that $\mathfrak{p}^n$ is primary. Clearly the radical of $\mathfrak{p}^n$ is $\mathfrak{p}$, and so $\mathfrak{p}^n$ is $\mathfrak{p}$-primary.

Conversely, suppose that $\mathfrak{q}$ is a $\mathfrak{p}$-primary ideal. Since the radical of $\mathfrak{q}$ is $\mathfrak{p}$, it follows that $r^n \in \mathfrak{q}$ for some $n \geqslant 1$. Choose n to be the minimal such power. Then $\mathfrak{p}^n \subset \mathfrak{q}$, but there is an $a \in \mathfrak{q} \smallsetminus \mathfrak{p}^{n-1}$. Write $a = cr^m$, where $r \nmid c$. Choose a so that m is the maximal such power, where obviously $m \leqslant n-1$. Now $r^n \notin \mathfrak{q}$ by minimality of n, hence some power $c^k \in \mathfrak{q} \subset \mathfrak{p}$ since $\mathfrak{q}$ is primary. But this is absurd since $r \nmid c$ unless c is a unit. Thus $\mathfrak{q} = \mathfrak{p}^{n-1}$. □

If $\mathfrak{a}$ is an ideal in S, we denote its *contraction* $\mathfrak{a} \cap R$ to R by $\mathfrak{a}^{\mathsf{c}}$. If $\mathfrak{q}$ is a $\mathfrak{p}$-primary ideal in S, then $\mathfrak{p}^{\mathsf{c}}$ is prime and $\mathfrak{q}^{\mathsf{c}}$ is $\mathfrak{p}^{\mathsf{c}}$-primary in R.

One of the important results in commutative algebra, essential to developing a dimension theory using chains of prime ideals, is the so-called "Going Down" theorem [5, Thm. 5.16]. Its hypotheses are satisfied in our situation, and it says the following. Suppose that $\mathfrak{p}_0 \subsetneq \mathfrak{p}_1 \subsetneq \mathfrak{p}_2$ is a chain of prime ideals in R, and that there is a prime ideal $\mathfrak{q}_2$ in S with $\mathfrak{q}_2^{\mathsf{c}} = \mathfrak{p}_2$. Then there is a chain $\mathfrak{q}_0 \subsetneq \mathfrak{q}_1 \subsetneq \mathfrak{q}_2$ of prime ideals in S such that $\mathfrak{q}_j^{\mathsf{c}} = \mathfrak{p}_j$ for $j = 0, 1, 2$. From this it follows that prime ideals in S of height one contract to prime ideals in R of height one. In other words, if $h \in S$ is irreducible, then $\langle h \rangle_S \cap R$ is a principal ideal $\langle g \rangle_R$ in R generated by an irreducible polynomial g in R.

Proof of Prop. 3.6.2. First suppose that $f \in S$ is irreducible. As we just showed, there is an irreducible $g \in R$ such that $\langle f \rangle_S \cap R = \langle g \rangle_R$. Furthermore, if $n \geqslant 1$

then $\langle f^n\rangle_S$ is $\langle f\rangle_S$-primary, and so $\langle f^n\rangle_S \cap R$ is $\langle g\rangle_R$-primary, hence equals $\langle g^k\rangle_R$ for some $k \geqslant 1$.

The result is obvious if $f = 0$, so suppose that $0 \neq f \in S$, and let $f = f_1^{n_1} \cdots f_r^{n_r}$ be its factorization in S into powers of distinct irreducibles f_j. Then there are irreducible polynomials $g_j \in R$ and $k_j \geqslant 1$ such that $\langle f_j^{n_j}\rangle_S \cap R = \langle g_j^{k_j}\rangle_R$. Hence

$$\begin{aligned}\langle f\rangle_S \cap R &= \langle f_1^{n_1} \cdots f_r^{n_r}\rangle_S \cap R = \big(\langle f_1^{n_1}\rangle_S \cap \cdots \langle f_r^{n_r}\rangle_S\big) \cap R \\ &= \big(\langle f_1^{n_1}\rangle_S \cap R\big) \cap \cdots \cap \big(\langle f_r^{n_r}\rangle_S \cap R\big) \\ &= \langle g_1^{k_1}\rangle_R \cap \cdots \cap \langle g_r^{k_r}\rangle_R = \langle \mathrm{LCM}(g_1^{k_1}, \dots, g_r^{k_r})\rangle_R,\end{aligned}$$

proving that $\langle f\rangle_S \cap R$ is principal. □

Remarks 3.6.4. (1) It is possible for distinct principal prime ideals in S to contract to the same prime ideal in R. As a simple example, let $d = 1$, $N = 2$, $f_1(x) = x^2 - x - 1$, and $f_2(x) = x^2 + x - 1$. Then each is irreducible in S, but both $\langle f_1\rangle_S$ and $\langle f_2\rangle_S$ contract in $R = \mathbb{Z}[2\mathbb{Z}]$ to $\langle x^4 - 3x^2 + 1\rangle_R$, where $x^4 - 3x^2 + 1$ is irreducible in $\mathbb{Z}[2\mathbb{Z}]$ (but of course not in $\mathbb{Z}[\mathbb{Z}]$). In the proof this is accounted for by using the least common multiple LCM in the last line of the displayed equation above.

(2) A polynomial is *primitive* if the greatest common divisor of its coefficients is 1. If $0 \neq f \in S$ is a nonconstant primitive polynomial with factorization $f = f_1^{n_1} \cdots f_r^{n_r}$ into powers of distinct irreducible polynomials, then by Gauss's Lemma each f_j is primitive as well. Furthermore, $\langle f_j\rangle_S \cap R = \langle g_j\rangle_R$, where each g_j is nonconstant and primitive. It then follows from the proof that $\langle f\rangle_S \cap R$ is generated by a primitive element of R.

(3) There is a completely different proof of Prop. 3.6.2 using entropy that is valid for all polynomials in S except for those of a very special and easily determined form. Recall that the entropy of α_f is the logarithmic Mahler measure $\mathsf{m}(f)$ defined in (3.2). A *generalized cyclotomic polynomial* in S is one of the form $\mathbf{x}^{\mathbf{n}}c(\mathbf{x}^{\mathbf{k}})$, where c is a cyclotomic polynomial in one variable and $\mathbf{k} \neq \mathbf{0}$. Smyth [106] proved that $\mathsf{m}(f) = 0$ if and only if f is, up to sign, a product generalized cyclotomic polynomials. Assume that $f \in S$ is not such a polynomial, so that the entropy of α_f is strictly positive. A simple argument using cosets of $N\mathbb{Z}^d$ shows that $r_N(X_f)$ also has positive entropy. Now $r_N(X_f) = X_{\mathfrak{a}_N}$ by Lemma 3.6.1, where $\mathfrak{a}_N = \langle f\rangle_S \cap R$. But an ideal $\mathfrak{a}$ in R for which the shift action of $N\mathbb{Z}^d$ on $X_{\mathfrak{a}}$ has positive entropy must be principal [64, Thm. 4.2].

3.7 Absolutely irreducible factorizations and Gauss's Lemma

Suppose that $f \in \mathbb{Z}[\mathbb{Z}^d]$ is nonconstant and irreducible. Its factorization into absolutely irreducible polynomials in an extension field of $\mathbb{Q}$ will play a decisive role. A generalization of Gauss's Lemma to number fields enables us to deal with the algebraic properties of the coefficients of the factors.

Two polynomials in $\mathbb{C}[\mathbb{Z}^d]$ are *distinct* if one is not a nonzero scalar multiple of the other. An element $\varphi \in \mathbb{C}[\mathbb{Z}^d]$ is *adjusted* if $\mathbf{0}$ is an extreme point of its Newton polytope $\mathcal{N}_\varphi$, and is *monic* if it is both adjusted and $\widehat{\varphi}(\mathbf{0}) = 1$.

A polynomial in $\mathbb{C}[\mathbb{Z}^d]$ is *absolutely irreducible* if it is irreducible in the unique factorization domain $\mathbb{C}[\mathbb{Z}^d]$. Hence every non-unit $f \in \mathbb{C}[\mathbb{Z}^d]$ has some factorization $f = \varphi_1 \cdots \varphi_r$ into absolutely irreducible factors φ_j. The method of Galois descent [20] shows that, after multiplying the factors by suitable constants, there is a finite normal extension $\mathbb{K}$ of $\mathbb{Q}$ such that each $\varphi_j \in \mathbb{K}[\mathbb{Z}^d]$, and also that the coefficients of the φ_j generate $\mathbb{K}$, so that $\mathbb{K}$ is the splitting field of f. Furthermore an elementary argument shows that if f is adjusted, then we can multiply the φ_j by units in $\mathbb{K}[\mathbb{Z}^d]$ so that each φ_j is monic, $\mathcal{N}_{\varphi_j} \subset \mathcal{N}_f$, and $f = \widehat{f}(\mathbf{0})\varphi_1 \cdots \varphi_r$.

Remarks 3.7.1. (1) When $d = 1$ this factorization is into the linear factors guaranteed by the fundamental theorem of algebra.

(2) A simple sufficient condition for φ to be absolutely irreducible is that $\mathcal{N}_\varphi$ is not the nontrivial Minkowski sum of two integer polytopes (see [35] for applications of this idea).

(3) There are reasonably good factoring algorithms which, on input f, produce a monic irreducible polynomial in $\mathbb{Z}[x]$ with root θ and an absolutely irreducible $\varphi \in \mathbb{Q}(\theta)[\mathbb{Z}^d]$ such that $f = \sigma_1(\varphi)\sigma_2(\varphi)\cdots\sigma_r(\varphi)$, where the σ_j are all the distinct field embeddings of $\mathbb{Q}(\theta)$ into $\mathbb{C}$ (see [29] for an overview of these methods).

The following shows that, unlike factoring, divisibility is not affected when passing to an extension field.

Lemma 3.7.2. Suppose that $\mathbb{L}$ is an extension of the field $\mathbb{K}$ and that $f, g \in \mathbb{K}[\mathbb{Z}^d]$. Then f divides g in $\mathbb{K}[\mathbb{Z}^d]$ if and only if f divides g in $\mathbb{L}[\mathbb{Z}^d]$.

Proof. For the nontrivial direction, suppose there is an $h \in \mathbb{L}[\mathbb{Z}^d]$ such that $fh = g$. Equating coefficients of like monomials gives a system of $\mathbb{K}$-linear equations in the coefficients of h. Since this system has a solution over $\mathbb{L}$, Gaussian elimination shows that that this (unique) solution is actually over $\mathbb{K}$, and so $h \in \mathbb{K}[\mathbb{Z}^d]$. $\square$

Proposition 3.7.3. Let $f \in \mathbb{Z}[\mathbb{Z}^d]$ be nonconstant, adjusted, and irreducible in $\mathbb{Z}[\mathbb{Z}^d]$. Then there is a finite normal extension field $\mathbb{K}$ of $\mathbb{Q}$ and monic absolutely irreducible polynomials $\varphi_1, \dots, \varphi_r \in \mathbb{K}[\mathbb{Z}^d]$ such that $f = \widehat{f}(\mathbf{0})\varphi_1 \cdots \varphi_r$ and $\mathcal{N}_{\varphi_f} \subset \mathcal{N}_f$ for $1 \leqslant j \leqslant r$. Furthermore, the Galois group $\mathrm{Gal}(\mathbb{K} : \mathbb{Q})$ acts transitively on the set of factors φ_j, and these factors are pairwise distinct.

Proof. Our earlier discussion shows there is a factorization $f = \widehat{f}(\mathbf{0})\varphi_1 \cdots \varphi_r$ over the splitting field $\mathbb{K}$ of f, where each φ_j is monic and $\mathcal{N}_{\varphi_j} \subset \mathcal{N}_f$ for $1 \leqslant j \leqslant r$. Suppose that $\sigma \in \mathrm{Gal}(\mathbb{K} : \mathbb{Q})$. Since $\sigma(f) = f$, it follows that σ must permute the absolutely irreducible factors up to multiplication by units. But if $\sigma(\varphi_j) = c\mathbf{x}^{\mathbf{n}}\varphi_k$, then $\mathbf{n} = \mathbf{0}$ since the factors are adjusted and $c = 1$ since they are monic. Hence σ permutes the factors themselves. If there were a proper subset of factors that is invariant under $\mathrm{Gal}(\mathbb{K} : \mathbb{Q})$, then their product ψ would be in $\mathbb{Q}[\mathbb{Z}^d]$ since its coefficients are invariant under $\mathrm{Gal}(\mathbb{K} : \mathbb{Q})$. But then ψ would be a proper divisor of f in $\mathbb{Q}[\mathbb{Z}^d]$ by Lemma 3.7.2, contradicting irreducibility of f by Gauss's Lemma. A similar argument shows that each factor appears with multiplicity one. $\square$

We now give a brief sketch of the extension of Gauss's Lemma to number fields and the consequences we use. Let $\mathbb{K}$ be a finite extension of $\mathbb{Q}$, and $\mathcal{O}_{\mathbb{K}}$ be the ring of algebraic integers in $\mathbb{K}$. A *fractional ideal* $\mathfrak{a}$ in $\mathbb{K}$ is a nonzero $\mathcal{O}_{\mathbb{K}}$-sumbodule such that there is an integer b for which $b\mathfrak{a} \subset \mathcal{O}_{\mathbb{K}}$. Fractional ideals can be added and multiplied, with $\mathcal{O}_{\mathbb{K}}$ being the multiplicative identity. A fractional ideal contained in $\mathcal{O}_{\mathbb{K}}$ is an ideal in the usual ring-theoretic sense. The pivotal result is that the set of fractional ideals form a group, the set of principal fractional ideals (those of the form $\mathcal{O}_{\mathbb{K}}\beta$ for some $\beta \in \mathbb{K}$) form a subgroup, and the quotient of these groups is a finite abelian group called the class group which measures how far $\mathcal{O}_{\mathbb{K}}$ is from being a principal ideal domain.

Let $\varphi \in \mathbb{K}[\mathbb{Z}^d]$. Define the *content* $\mathfrak{c}_{\mathbb{K}}(\varphi)$ to be the fractional ideal in $\mathbb{K}$ generated by the coefficients of φ. Say that φ is *primitive* if $\mathfrak{c}_{\mathbb{K}}(\varphi) = \mathcal{O}_{\mathbb{K}}$. It is easy to check that although content depends on the ambient field $\mathbb{K}$, primitivity does not: if $\varphi \in \mathbb{K}[\mathbb{Z}^d]$ and $\varphi \in \mathbb{L}[\mathbb{Z}^d]$, then $\mathfrak{c}_{\mathbb{K}}(\varphi) = \mathcal{O}_{\mathbb{K}}$ if and only if $\mathfrak{c}_{\mathbb{L}}(\varphi) = \mathcal{O}_{\mathbb{L}}$ (see [75, Thm. 8.2]).

Theorem 3.7.4 (Gauss's Lemma for number fields)**.** Let $\mathbb{K}$ be a number field and $\varphi, \psi \in \mathbb{K}[\mathbb{Z}^d]$. The $\mathfrak{c}_{\mathbb{K}}(\varphi\psi) = \mathfrak{c}_{\mathbb{K}}(\varphi)\mathfrak{c}_{\mathbb{K}}(\psi)$. In particular, if $\varphi, \psi \in \mathcal{O}_{\mathbb{K}}[\mathbb{Z}^d]$ then $\varphi\psi$ is primitive if and only if both φ and ψ are primitive. If $\varphi, \psi \in \mathcal{O}_{\mathbb{K}}[\mathbb{Z}^d]$ are primitive, and if $\varphi = \beta\psi$ for some $\beta \in \mathbb{K}$, then β is a unit in $\mathcal{O}_{\mathbb{K}}$.

Remark 3.7.5. Suppose that $f \in \mathbb{Z}[\mathbb{Z}^d]$ is primitive and that $N \geqslant 1$. Let $\zeta_N = e^{2\pi i/N}$, which is a unit in $\mathbb{Q}(\zeta_N)$. Hence each rotate $\boldsymbol{\omega} \cdot f$, where $\boldsymbol{\omega} \in \Omega_N^d$, is primitive in $\mathbb{Q}(\zeta_N)[\mathbb{Z}^d]$. The preceding theorem then shows that the product f_N of these rotates is also primitive in $\mathbb{Q}(\zeta_N)[\mathbb{Z}^d]$, and hence in $\mathbb{Z}[\mathbb{Z}^d]$ (since primitivity is independent of ambient field), a fact we already observed in Remark 3.6.4(2).

3.8 Decimated polynomials and decimated actions

Let $f \in \mathbb{Z}[\mathbb{Z}^d]$ be irreducible. Here we explain the relationship between the Nth decimation f_N of f and the generator g_N of the contracted ideal $\langle f \rangle \cap \mathbb{Z}[N\mathbb{Z}^d]$ that defines the Nth decimation $r_N(X_f)$ of (X_f, α_f). Roughly speaking, g_N is a constant times the product of all distinct rotates by elements of Ω_N^d of the absolutely irreducible factors φ_j of f as described in Proposition 3.7.3. Each rotate appears with the same multiplicity e_N that can be computed from the φ_j. Thus $f_N = c\, g_N^{e_N}$, and an application of Gauss's Lemma shows that we may take $c = 1$. Furthermore, there is an integer $Q(f)$, that can also be computed from the φ_j, such that $f_N = g_N$ for all N relatively prime to $Q(f)$. Examples will illustrate the two sources of the multiplicity e_N.

In what follows we let $\zeta_N = e^{2\pi i/N}$, which is a generator of Ω_N.

Lemma 3.8.1. If $f \in \mathbb{Z}[\mathbb{Z}^d]$ then $f_N \in \mathbb{Z}[N\mathbb{Z}^d]$.

Proof. Since $f_N = \prod_{\boldsymbol{\omega} \in \Omega_N^d} \boldsymbol{\omega} \cdot f$, it follows that $f_N = \boldsymbol{\omega} \cdot f_N$ for every $\boldsymbol{\omega} \in \Omega_N^d$. Suppose that $\widehat{f_N}(\mathbf{k}) \neq 0$. Then since

$$\widehat{f_N}(\mathbf{k}) = (\boldsymbol{\omega} \cdot f_N)\widehat{\ }(\mathbf{k}) = \boldsymbol{\omega}^{\mathbf{k}}\, \widehat{f_N}(\mathbf{k}),$$

we see that $\boldsymbol{\omega}^{\mathbf{k}} = 1$ for every $\boldsymbol{\omega} \in \Omega_N^d$, and hence $\mathbf{k} \in N\mathbb{Z}^d$. Thus $f_N \in \mathbb{Q}(\zeta_N)[N\mathbb{Z}^d]$.

The Galois group $G := \mathrm{Gal}\big(\mathbb{Q}(\zeta_N) : \mathbb{Q}\big)$ acts on Ω_N^d coordinate-wise. If $\sigma \in G$, then $\sigma(\boldsymbol{\omega} \cdot f) = \sigma(\boldsymbol{\omega}) \cdot f$ since f has integer coefficients. Thus σ permutes the rotates of f, and so $\sigma(f_N) = f_N$ for every $\sigma \in G$. It follows that the coefficients of f_N are both rational and algebraic integers, and so $f_N \in \mathbb{Z}[N\mathbb{Z}^d]$. □

Lemma 3.8.2. Let $f \in \mathbb{Z}[\mathbb{Z}^d]$ and g_N be a generator of the contracted ideal $\langle f \rangle \cap \mathbb{Z}[N\mathbb{Z}^d]$. Then g_N divides f_N in $\mathbb{Z}[N\mathbb{Z}^d]$.

Proof. Since f is one of the factors in forming f_N, it follows that f divides f_N in $\mathbb{Q}(\zeta_N)[\mathbb{Z}^d]$. Hence f divides f_N in $\mathbb{Q}[\mathbb{Z}^d]$ by Lemma 3.7.2. The coefficients of f_N/f are both rational and algebraic integers, and so $f_N/f \in \mathbb{Z}[\mathbb{Z}^d]$. Hence $f_N \in \langle f\rangle \cap \mathbb{Z}[N\mathbb{Z}^d]$, and it is thus divisible by the generator g_N. □

Remark 3.8.3. Since the generator of a principal ideal is unique only up to units, it will be convenient to have a convention to pick a generator. In what follows we will assume that f is adjusted and that $\widehat{f}(\mathbf{0}) > 0$. Then clearly f_N has the same properties. By the previous lemma, we can also assume that g_N is adjusted, that $\mathcal{N}_{g_N} \subset \mathcal{N}_{f_N}$, and that $\widehat{g}_N(\mathbf{0}) > 0$.

Before continuing, we remark that if f is a constant integer n, then $f_N = n^{N^d}$ while $g_N = n$. Let us call a polynomial $f \in \mathbb{Z}[\mathbb{Z}^d]$ *nonconstant* if $|\operatorname{supp} f| > 1$, and it is these we now turn to.

Let $f \in \mathbb{Z}[\mathbb{Z}^d]$ be adjusted. Define its *support group* Γ_f to be the subgroup of $\mathbb{Z}^d$ generated by $\operatorname{supp} f$. It is easy to check that the support group is independent of which extreme point of $\mathcal{N}_f$ is used to adjust f. We say that f is *full* if $\Gamma_f = \mathbb{Z}^d$.

The following shows that in some cases, including $f(x,y) = 1+x+y$ from Example 3.2.4, $f_N = g_N$ for all $N \geqslant 1$.

Proposition 3.8.4. Let $f \in \mathbb{Z}[\mathbb{Z}^d]$ be adjusted, irreducible, and full. Further assume that f is absolutely irreducible in $\mathbb{C}[\mathbb{Z}^d]$. Then $f_N = g_N$ for every $N \geqslant 1$.

Proof. Since the map $f \mapsto \boldsymbol{\omega} \cdot f$ is a ring isomorphism of $\mathbb{C}[\mathbb{Z}^d]$, each $\boldsymbol{\omega} \cdot f$ is absolutely irreducible. Suppose that $\boldsymbol{\omega} \cdot f = \boldsymbol{\omega}' \cdot f$. Since $\boldsymbol{\omega}^{\mathbf{k}} \widehat{f}(\mathbf{k}) = (\boldsymbol{\omega}')^{\mathbf{k}} \widehat{f}(\mathbf{k})$, it follow that $\boldsymbol{\omega}^{\mathbf{k}} = (\boldsymbol{\omega}')^{\mathbf{k}}$ for all $\mathbf{k} \in \operatorname{supp} f$, hence for all $\mathbf{k} \in \Gamma_f = \mathbb{Z}^d$, and so $\boldsymbol{\omega} = \boldsymbol{\omega}'$. Thus the rotates of $\boldsymbol{\omega} \cdot f$ for $\boldsymbol{\omega} \in \Omega_N^d$ are pairwise distinct absolutely irreducible polynomials in $\mathbb{C}[\mathbb{Z}^d]$ whose product is f_N.

By Lemma 3.8.2, g_N divides f_N in $\mathbb{C}[\mathbb{Z}^d]$. Hence some rotate $\boldsymbol{\omega} \cdot f$ divides g_N. Since $g_N \in \mathbb{Z}[N\mathbb{Z}^d]$, it is invariant under all rotations in Ω_N^d. Hence g_N is divisible by all rotates $\boldsymbol{\omega} \cdot f$, and so g_N and f_N have the same absolute factorizations in $\mathbb{C}[\mathbb{Z}^d]$, and hence $f_N = c\, g_N$ for some constant $c \in \mathbb{C}$. Recalling our conventions in Remark 3.8.3, comparing constant terms shows that $c = \widehat{f}(\mathbf{0})^{N^d} / \widehat{g}_N(\mathbf{0}) \in \mathbb{Q}$. But f_N and g_N are both primitive in $\mathbb{Z}[N\mathbb{Z}^d]$, and so $c = \pm 1$, and our convention on positivity of constant terms then gives $c = 1$. □

The following example shows that when the polynomial is not full there can be multiplicity $e_N > 1$.

Example 3.8.5. Let $d = 2$ and $f(x, y) = 1+x+y^2$. Since $\mathcal{N}_f$ is not a nontrivial Minkowski sum of integer polytopes, we see that f is absolutely irreducible. Suppose that N is odd. Since $-1 \notin \Omega_N$, all rotates $\boldsymbol{\omega} \cdot f$ for $\boldsymbol{\omega} \in \Omega_N^2$ are distinct, and the same arguments as in the previous proposition show that $f_N = g_N$.

However, if N is even, then $-1 \in \Omega_N$ and the rotate of f by (ω_1, ω_2) equals that by $(\omega_1, -\omega_2)$. As we will see in Proposition 3.8.7, the product of the distinct rotates of f equals g_N, and so $f_N = g_N^2$ when N is even.

Next we characterize when rotates can coincide.

Lemma 3.8.6. Let $\varphi \in \mathbb{C}[\mathbb{Z}^d]$ be adjusted, and Γ_φ be its support group. Then the dual of the stabilizer group $S_N(\varphi) := \{\boldsymbol{\omega} \in \Omega_N^d : \boldsymbol{\omega} \cdot \varphi = \varphi\}$ is $\mathbb{Z}^d/(\Gamma_\varphi + N\mathbb{Z}^d)$. Two rotates of φ differ by a multiplicative unit in $\mathbb{C}[\mathbb{Z}^d]$ if and only if they are equal. If Γ_φ has finite index K in $\mathbb{Z}^d$, then $S_N(\varphi)$ is trivial for every N relatively prime to K.

Proof. Suppose that $\boldsymbol{\omega} \in S_N(\varphi)$. Since $\widehat{\varphi}(\mathbf{k}) = (\boldsymbol{\omega} \cdot \varphi)\widehat{\ }(\mathbf{k}) = \boldsymbol{\omega}^{\mathbf{k}} \widehat{\varphi}(\mathbf{k})$, it follows that $\boldsymbol{\omega}^{\mathbf{k}} = 1$ for every $\mathbf{k} \in \operatorname{supp} \varphi$. Hence $\boldsymbol{\omega}$ annihilates Γ_φ as well as $N\mathbb{Z}^d$, thus their sum. Conversely, every $\boldsymbol{\omega}$ annihilating $\Gamma_\varphi + N\mathbb{Z}^d$ must be in $S_N(\varphi)$. Hence the annihilator of $S_N(\varphi)$ equals $\Gamma_\varphi + N\mathbb{Z}^d$, and so its dual group is $\mathbb{Z}^d/(\Gamma_\varphi + N\mathbb{Z}^d)$.

The multiplicative units in $\mathbb{C}[\mathbb{Z}^d]$ have the form $c\,\mathbf{x}^{\mathbf{n}}$ for some $c \in \mathbb{C}$, so the second statement is obvious since φ is adjusted.

Suppose that Γ_φ has finite index K in $\mathbb{Z}^d$. If N is relatively prime to K, then multiplication by N on $\mathbb{Z}^d/\Gamma_\varphi$ is injective, hence surjective. Thus modulo Γ_φ every element in $\mathbb{Z}^d$ is a multiple of N, and hence $\Gamma_\varphi + N\mathbb{Z}^d = \mathbb{Z}^d$. □

Proposition 3.8.7. Let $f \in \mathbb{Z}[\mathbb{Z}^d]$ be adjusted and irreducible, and further assume that f is absolutely irreducible in $\mathbb{C}[\mathbb{Z}^d]$. Then $f_N = g_N^{e_N}$, where $e_N = |S_N(f)| = |\mathbb{Z}^d/(\Gamma_f + N\mathbb{Z}^d)|$.

Proof. Recall our conventions in Remark 3.8.3. Since g_N divides f_N, it must be divisible by at least one (absolutely irreducible) rotate of f. Invariance of g_N by every rotate in Ω_N^d shows that g_N is therefore divisible by the product h of all the distinct rotates of f. The arguments in Lemmas 3.8.1 and 3.8.2 apply to show that $h \in \langle f \rangle \cap \mathbb{Z}[N\mathbb{Z}^d]$. Thus g_N divides h in $\mathbb{C}[\mathbb{Z}^d]$ as well, and so $g_N = c\,h$ for some $c \in \mathbb{C}$. Evaluating constant terms shows that $c \in \mathbb{Q}$. Since g_N is irreducible in $\mathbb{Z}[N\mathbb{Z}^d]$, it is primitive. Each rotate of f is primitive in $\mathbb{Q}(\zeta_N)[\mathbb{Z}^d]$, and so h is primitive by Theorem 3.7.4. Hence $c = \pm 1$, and then $c = 1$ follows from our sign conventions. By Lemma 3.8.6, each rotate of f is repeated exactly e_N times, and so $f_N = g_N^{e_N}$. $\square$

When f is absolutely irreducible, the only source of multiplicity $e_N > 1$ is its support group. However, if f has several absolutely irreducible factors, a new source of multiplicity can occur, namely that one factor could rotate to another factor. This possibility is illustrated in the following three examples.

Example 3.8.8. Let $d = 1$ and

$$f(x) = 1 - 2x^2 = (1 + \sqrt{2}x)(1 - \sqrt{2}x) = \varphi_1(x)\varphi_2(x).$$

Let $\sigma \in \mathrm{Gal}(\mathbb{Q}(\sqrt{2}) : \mathbb{Q})$ be given by $\sigma(\sqrt{2}) = -\sqrt{2}$. Then $\sigma(\varphi_1) = \varphi_2 = (-1) \cdot \varphi_1$. Now f_N is the product of $\zeta_N^j \cdot \varphi_k$ for $0 \leqslant j < N$ and $k = 1, 2$. If N is odd, then $-1 \notin \Omega_N$ and so all $2N$ factors are distinct. Our earlier arguments then show that $f_N = g_N$. However, if N is even, then $-1 \in \Omega_N$, and the set of rotates of φ_1 coincide with set of those of φ_2, and so $f_N = g_N^2$ for even N. Here f is an irreducible polynomial with a pair of roots whose ratio is a nontrivial root of unity.

The commingling of absolutely irreducible factors under rotations can happen in more subtle ways.

Example 3.8.9. Let $d = 1$ and $f(x) = 1 - 2x + 4x^2 - 3x^3 + x^4$, which is full and irreducible in $\mathbb{Z}[\mathbb{Z}]$. Let $\lambda = (1 + \sqrt{5})/2$, $\mu = (1 - \sqrt{5})/2$, and $\zeta = \zeta_5$. The absolutely irreducible factorization of f is

$$f(x) = (1 - \zeta\lambda x)((1 - \zeta^4\lambda x)(1 - \zeta^2\mu x)(1 - \zeta^3\mu x) = \varphi_1(x)\varphi_2(x)\varphi_3(x)\varphi_4(x).$$

Note that $\zeta^3 \cdot \varphi_1 = \varphi_2$ and that $\zeta \cdot \varphi_3 = \varphi_4$. If N is relatively prime to 5, then $\zeta \notin \Omega_N$, and so all $4N$ rotates are distinct and $f_N = g_N$ as before. However, if $5 \mid N$ then $\zeta \in \Omega_N$ and each rotate is repeated twice, and so $f_N = g_N^2$ in this case.

What is driving this example is the inclusion $\mathbb{Q}(\sqrt{5}) \subset \mathbb{Q}(\zeta)$, and so the Galois automorphism $\sqrt{5} \mapsto -\sqrt{5}$ of $\mathbb{Q}(\sqrt{5})$ is the restriction of the automorphism $\zeta \mapsto \zeta^2$ of $\mathbb{Q}(\zeta)$.

Remark 3.8.10. Irreducible polynomials in $\mathbb{Z}[x]$ having distinct roots whose ratio is a root of unity, such as those in the previous two examples, are called *degenerate*. Such polynomials have an extensive literature (see for instance [32, §1.1.9]), and appear in the celebrated Skolem-Mahler-Lech Theorem that the set of indices at which a recurring sequence of integers vanishes is, modulo a finite set, the union of arithmetic progressions [7].

There is a simple way to detect whether $f(x) \in \mathbb{Z}[x]$ is degenerate. Introduce a new variable t, and compute the resultant $g(x) \in \mathbb{Z}[x]$ of the polynomials $f(tx)$ and $f(t)$ with respect to t, which can be done efficiently using rational arithmetic. The roots of $g(x)$ are the ratios of all pairs of roots of f. Thus $f(x)$ is degenerate if and only if $g(x)$ contains a nontrivial cyclotomic factor. Applying this to $f(x)$ from the previous example gives

$$g(x) = (x-1)^5(x^4-4x^3+6x^2+x+1)(x^4+x^3+6x^2-4x+1)(x^4+x^3+x^2+x+1).$$

The last factor reveals that $f(x)$ has two roots whose ratio is a nontrivial 5th root of unity.

Example 3.8.11. Let $d = 2$ and $f(x, y) = 1 - x - y - xy + x^2 + y^2$, which is full and irreducible in $\mathbb{Z}[\mathbb{Z}^2]$. Let $\zeta = \zeta_3$. The absolutely irreducible factorization of f is

$$f(x, y) = (1 + \zeta x + \zeta^2 y)(1 + \zeta^2 x + \zeta y) = \varphi_1(x, y)\varphi_2(x, y).$$

Here φ_1 is mapped to φ_2 by the element σ in $\mathrm{Gal}(\mathbb{Q}(\zeta) : \mathbb{Q})$ mapping ζ to ζ^2, and also $\sigma(\varphi_1) = \varphi_2 = (\zeta, \zeta^2) \cdot \varphi_1$. By the now familiar arguments, if N is relatively prime to 3 then $\zeta \notin \Omega_N$, and so all rotates are distinct and hence $f_N = g_N$. However, if 3 divides N, then distrinct rotates are repeated twice, and so $f_N = g_N^2$. For instance

$$f_3 = (1 + 3x^3 + 3y^3 + 3x^6 - 21x^3y^3 + 3y^3 + x^9 + 3x^3y^6 + 3x^6y^3 + y^9)^2 = g_3^2.$$

With these examples in mind, we come to the main result of this section.

Theorem 3.8.12. Let $f \in \mathbb{Z}[\mathbb{Z}^d]$ be irreducible, which we may assume is adjusted with positive constant term. For every $N \geqslant 1$ there is an irreducible $g_N \in \mathbb{Z}[N\mathbb{Z}^d]$ and $e_N \geqslant 1$ such that

$$\langle f \rangle_{\mathbb{Z}[\mathbb{Z}^d]} \cap \mathbb{Z}[N\mathbb{Z}^d] = \langle g \rangle_{\mathbb{Z}[N\mathbb{Z}^d]} \quad \text{and} \quad f_N = g_N^{e_N}.$$

The multiplicity e_N can be computed from the absolutely irreducible factorization of f in $\mathbb{C}[\mathbb{Z}^d]$. If the support of f generates a finite-index subgroup of $\mathbb{Z}^d$, then there is an integer $Q(f)$, which can also be computed from the absolutely irreducible factors of f, such that $e_N = 1$ for every N that is relatively prime to $Q(f)$. Finally,

$$\langle f^k \rangle_{\mathbb{Z}[\mathbb{Z}^d]} \cap \mathbb{Z}[N\mathbb{Z}^d] = \langle g_N^k \rangle_{\mathbb{Z}[N\mathbb{Z}^d]}$$

for every $k \geqslant 1$.

Proof. Recall our conventions in Remark 3.8.3. Let $\mathbb{K}$ be the splitting field of f, and $f = \widehat{f}(\mathbf{0})\varphi_1 \cdots \varphi_r$ be the factorization of f using monic absolutely irreducible $\varphi_j \in \mathbb{K}[\mathbb{Z}^d]$ from Proposition 3.7.3. Let $\Phi = \{\varphi_1, \dots, \varphi_r\}$. Since the φ_j are monic, $\mathrm{Gal}(\mathbb{K} : \mathbb{Q})$ permutes the elements of Φ, and this action is transitive by irreducibility of f.

Now fix $N \geqslant 1$. Then $\mathbb{K}(\zeta_N)$ is a normal extension of $\mathbb{Q}$. Let $G = \mathrm{Gal}(\mathbb{K}(\zeta_N) : \mathbb{Q})$. Consider the set $\Omega_N^d \times \Phi$. The group Ω_N^d acts on this set via $\boldsymbol{\omega}' \cdot (\boldsymbol{\omega}, \varphi_j) = (\boldsymbol{\omega}'\boldsymbol{\omega}, \varphi_j)$. The group G also acts on this set via $\sigma \cdot (\boldsymbol{\omega}, \varphi_j) = (\sigma(\boldsymbol{\omega}), \sigma(\varphi_j))$. More precisely, $\sigma \in G$ acts of the first coordinate using its restriction to $\mathbb{Q}(\zeta_N)$ and on the second coordinate using its restriction to $\mathbb{K}$. These actions combine to give an action of the semidirect product $G \ltimes \Omega_N^d$ defined using the action of G on Ω_N^d, so that $\sigma\boldsymbol{\omega} = \sigma(\boldsymbol{\omega})\sigma$.

Define an equivalence relation $\sim$ on $\Omega_N^d \times \Phi$ by $(\boldsymbol{\omega}, \varphi_j) \sim (\boldsymbol{\omega}', \varphi_k)$ if and only if $\boldsymbol{\omega} \cdot \varphi_j = \boldsymbol{\omega}' \cdot \varphi_k$. It is routine to verify that $G \ltimes \Omega_N^d$ preserves equivalence classes. Since $\mathrm{Gal}(\mathbb{K} : \mathbb{Q})$ acts transitively on Φ, it follow that $G \ltimes \Omega_N^d$ acts transitively on $\Omega_N^d \times \Phi$. Hence all equivalence classes have the same cardinality, say $e_N \geqslant 1$. Pick one representative $(\boldsymbol{\omega}, \varphi_j)$ from each equivalence class, and let $\widetilde{g}_N$ be the product of the corresponding polynomials $\boldsymbol{\omega} \cdot \varphi_j$.

Observe that by its construction $\widetilde{g}_N$ is invariant under $G \ltimes \Omega_N^d$. Invariance under Ω_N^d implies that $\widetilde{g}_N \in \mathbb{K}(\zeta_N)[N\mathbb{Z}^d]$, and invariance under G further implies that $\widetilde{g}_N \in \mathbb{Q}[N\mathbb{Z}^d]$. Then transitivity of $G \ltimes \Omega_N^d$ on $\Omega_N^d \times \Phi$ shows that $\widetilde{g}_N$ is irreducible in $\mathbb{Q}[N\mathbb{Z}^d]$.

We have that $f_N = \widehat{f}(\mathbf{0})^{N^d}\, \widetilde{g}_N^{e_N}$. Let q be the least positive integer such that $q\widetilde{g}_N \in \mathbb{Z}[N\mathbb{Z}^d]$, so that $g_N := q\widetilde{g}_N$ is primitive. Then

$$f_N = \big(\widehat{f}(\mathbf{0})^{N^d}/q^{e_N}\big) g_N^{e_N}.$$

But both f_N and $g_N^{e_N}$ are primitive with positive constant terms, and hence $f_N = g_N^{e_N}$.

We now turn to computing e_N. Each of the absolutely irreducible factors φ_j has the same support since they are all Galois conjugates. Let Γ_φ denote the common support group of each. By Lemma 3.8.6, each contributes multiplicity $|\mathbb{Z}^d/(\Gamma_\varphi + N\mathbb{Z}^d)|$. Further multiplicity arises if one factor can be rotated by an element of Ω_N^d to another. This property divides Φ into equivalence classes, with all classes having the same cardinality s. It then follows that $e_N = |\mathbb{Z}^d/(\Gamma_\varphi + N\mathbb{Z}^d)|s$.

Next, we determine sufficient conditions on N so that $e_N = 1$. Assume that Γ_f has finite index in $\mathbb{Z}^d$. Clearly $\Gamma_f \subset \Gamma_\varphi$, and so Γ_φ also has finite index. By Lemma 3.8.6, if N is relatively prime to the index $[\mathbb{Z}^d : \Gamma_\varphi]$ of Γ_φ, then $|\mathbb{Z}^d/(\Gamma_\varphi + N\mathbb{Z}^d)| = 1$.

To analyze when one φ_j can rotate to another, we need to consider the group $\Omega_{\mathbb{K}}$ of roots of unity in the splitting field $\mathbb{K}$ of f. This is a finite cyclic group, and so equals Ω_n for some $n \geqslant 1$. Now $[\mathbb{Q}(\zeta_n) : \mathbb{Q}] = \varphi(n)$, where φ denotes the Euler function. Since $\mathbb{Q}(\zeta_n) \subset \mathbb{K}$, it follows that $\varphi(n) \leqslant [\mathbb{K} : \mathbb{Q}]$. A simple argument shows that $\varphi(n) \geqslant \sqrt{n}/2$ for all $n \geqslant 1$, and so $n \leqslant 4[\mathbb{K} : \mathbb{Q}]^2$. Hence if N is relatively prime to $(4[\mathbb{K} : \mathbb{Q}]^2)!$, then $\Omega_N \cap \Omega_{\mathbb{K}} = \{1\}$. For such an N suppose that $\boldsymbol{\omega} \cdot \varphi_i = \varphi_j$ for some $\boldsymbol{\omega} \in \Omega_N^d$. For each $\mathbf{k} \in \operatorname{supp} \varphi_i = \operatorname{supp} \varphi_j$ we have that $\boldsymbol{\omega}^{\mathbf{k}} \widehat{\varphi}_i(\mathbf{k}) = \widehat{\varphi}_j(\mathbf{k})$, and so

$$\boldsymbol{\omega}^{\mathbf{k}} = \widehat{\varphi}_j(\mathbf{k}) / \widehat{\varphi}_i(\mathbf{k}) \in \Omega_N \cap \Omega_{\mathbb{K}} = \{1\}.$$

But this implies that $\varphi_i = \varphi_j$.

Putting these together, we let $Q(f) = [\mathbb{Z}^d : \Gamma_\varphi](4[\mathbb{K} : \mathbb{Q}]^2)!$, and conclude that if N is relatively prime to $Q(f)$ then $e_N = 1$. □

3.9 Remarks and questions

Here we make some further remarks and ask several questions related to decimations.

3.9.1 More general lattices

Let us call a finite-index subgroup of $\mathbb{Z}^d$ a *lattice*. We have used the sequence $\{N\mathbb{Z}^d\}$ of lattices to define decimation, but these definitions easily extend to all lattices. Let $\Lambda \in \mathbb{Z}^d$ be a lattice, and let Ω_Λ denote the dual group of $\mathbb{Z}^d/\Lambda$, which has cardinality $[\mathbb{Z}^d : \Lambda]$, the index of Λ in $\mathbb{Z}^d$. Define $f_{\langle\Lambda\rangle} = \prod_{\boldsymbol{\omega} \in \Omega_\Lambda} \boldsymbol{\omega} \cdot f$, and

$$\mathsf{L}_\Lambda f = \mathsf{E}_{[\mathbb{Z}^d : \Lambda]}\Big(\frac{1}{[\mathbb{Z}^d : \Lambda]} \log |\widehat{f_{\langle\Lambda\rangle}}| \Big). \tag{3.17}$$

For a sequence $\{\Lambda_N\}$ of lattices, let us say $\Lambda_N \to \infty$ if for every $r > 0$ we have that $\{\mathbf{n} \in \Lambda_N : \|\mathbf{n}\| < r\} = \{\mathbf{0}\}$ for all large enough N.

Question 3.9.1. Let $0 \neq f \in \mathbb{C}[\mathbb{Z}^d]$, and let $\{\Lambda_N\}$ be a sequence of lattices with $\Lambda_N \to \infty$. Do the concave hulls $\mathsf{CH}(\mathsf{L}_{\Lambda_N})$ converge uniformly on $\mathcal{N}_f$ to D_f?

Our methods for $N\mathbb{Z}^d$ do not extend directly to this more general situation. We made essential use of the property of f_N that it is a polynomial in the Nth powers of the variables, enabling us to apply the Mahler estimates to the polynomial $\mathsf{E}_N \widehat{f_N}$ of lower degree, gaining a crucial improvement. There is no corresponding argument for general lattices.

3.9.2 Partial decimation

By taking different sequences of lattices, we can in effect decimate along lower rank subgroups. The following example illustrates this idea.

Let $d = 2$ and $f(x,y) = 1 + x + y$. We will use the sequence of lattices $\Lambda_N = N\mathbb{Z} \oplus \mathbb{Z}$, which corresponds to decimating with respect to x. Using the notation from the previous section, $\Omega_{\Lambda_N} = \Omega_N \times \{1\}$, and so

$$f_{\langle \Lambda_N \rangle}(x,y) = \prod_{\omega \in \Omega_N} (1 + \omega x + y) = (1+y)^N \pm x^N.$$

It is well-known that the growth rate of the binomial coefficients can be computed using Stirling's approximation to be

$$\frac{1}{N} \log \binom{N}{pN} \approx \eta(p) := -p \log p - (1-p)\log(1-p)$$

for $0 \leqslant p \leqslant 1$. Hence the decimation limit $\mathsf{D}_f^{(1)}(r,s)$ of f with respect to x using this sequence of lattices is the concave hull of the curve $\big(0, p, \eta(p)\big)$ for $0 \leqslant p \leqslant 1$ together with the point $(1,0,0)$, as shown in Figure 3.7(a).

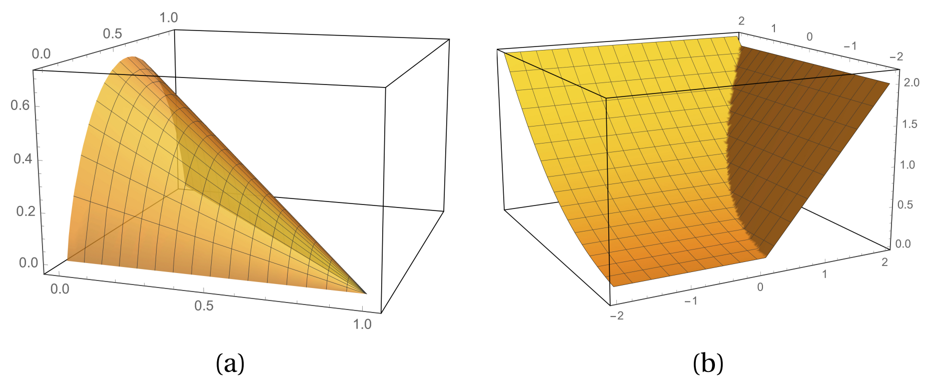


Figure 3.7: (a) The partial decimation limit of $1 + x + y$, and (b) its partial Ronkin function

Define the partial Ronkin function of f with respect to x to be

$$\mathsf{R}_f^{(1)}(u,v) = \int_0^1 \log |f(e^u e^{2\pi i \theta}, e^v)|\, d\theta.$$

Figure 3.7(b) shows this in our case. One can show that here $\mathsf{D}_f^{(1)} = -\big(\mathsf{R}_f^{(1)}\big)^*$ on $\mathcal{N}_f$.

This example suggests a more general phenomenon. Let $\mathcal{C}(\mathbb{Z}^d)$ denote the set of subgroups of $\mathbb{Z}^d$. We can give a topology to $\mathcal{C}(\mathbb{Z}^d)$ by declaring two subgroups to be close if they agree on a large ball around $\mathbf{0}$. For example, in this topology $\Lambda_N \to \{\mathbf{0}\}$ means $\Lambda_N \to \infty$ from §3.9.1, and in the above example $N\mathbb{Z} \oplus \mathbb{Z} \to 0 \oplus \mathbb{Z}$. This is a special case of the Chabauty topology on the set $\mathcal{C}(G)$ of closed subgroups of a locally compact group G. This topology is named after Claude Chabauty, who in 1950 introduced it [19] to generalize Mahler's compactness criterion [76] for lattices in $\mathbb{R}^d$ to lattices in locally compact groups. The Chabauty space $\mathcal{C}(G)$ has been investigated by many authors, for instance by Cornulier [21] when G is abelian. Even for familiar groups their Chabauty space can be intricate to analyze. For example, Hubbard and Pourezza [50] used a tricky argument to prove that $\mathcal{C}(\mathbb{R}^2)$ is homeomorphic to the four-dimensional sphere.

Let K be a compact subgroup of $\mathbb{T}^d$, and let μ_K denote normalized Haar measure on K. For $\mathbf{s} \in K$ we let $e^{2\pi i \mathbf{s}}\mathbf{u}$ mean $(e^{2\pi i s_1}u_1, \dots, e^{2\pi i s_d}u_d)$. We then define the Ronkin function of f with respect to K to be

$$\mathsf{R}_f^{(K)}(\mathbf{u}) = \int_K \log|f(e^{2\pi i \mathbf{s}}\mathbf{u})|\, d\mu_K(\mathbf{s}).$$

Question 3.9.2. Is there a limiting shape for decimations corresponding to a sequence of lattices $\{\Lambda_N\}$ in $\mathbb{Z}^d$ converging to a non-trivial subgroup $\Gamma \in \mathcal{C}(\mathbb{Z}^d)$?

3.9.3 Exponential size of decimation coefficients

In Example 3.2.3 we saw that if $f \in \mathbb{C}[\mathbb{Z}]$ is allowed to have complex coefficients, then some of the coefficients of f_N may have exponential size drastically different from that predicted by D_f. However, if $f \in \mathbb{Z}[\mathbb{Z}]$ is restricted to have integer coefficients, then this behavior cannot happen, as indicated by Example 3.2.2. More precisely, using the diophantine results of Gelfond mentioned there, one can show that if $f \in \mathbb{Z}[\mathbb{Z}]$ has $\operatorname{supp} f = \{0, 1, \dots, r\}$ and $\varepsilon > 0$, then for all sufficiently large N we have that $|\widehat{f}_N(kN)|$ is between $e^{N(\mathsf{D}_f(k)\pm\varepsilon)}$ for each $0 \leqslant k \leqslant r$ for which $\widehat{f}_N(kN) \neq 0$.

This raises the intriguing question of whether this extends to $f \in \mathbb{Z}[\mathbb{Z}^d]$ for $d \geqslant 2$, i.e., do all nonzero coefficients of f_N have the approximate exponential size predicted by D_f. The following gives a precise quantitative formulation.

Question 3.9.3. Let $f \in \mathbb{Z}[\mathbb{Z}^d]$. Fix $\mathbf{r}_0 \in \mathcal{N}_f$, and let $\varepsilon > 0$. Are there $\delta > 0$ and $N_0 \geqslant 1$ such that if $N \geqslant N_0$ and $\mathbf{r} \in N^{-d}\mathbb{Z}^d \cap \mathcal{N}_f$ with $\|\mathbf{r} - \mathbf{r}_0\| < \delta$, and if $\mathsf{L}_N f(\mathbf{r}) \neq -\infty$, then $|\mathsf{L}_N f(\mathbf{r}) - \mathsf{D}_f(\mathbf{r})| < \varepsilon$? Can δ and N_0 be chosen uniformly for $\mathbf{r}_0 \in \mathcal{N}_f$?

Some evidence for a positive answer comes from polynomials in two variables related to dimer models, as discussed in Remark 3.2.6. Using the additional machinery afforded by the physical interpretation of the related partition function and the resulting subadditivity, the exponential size of the coefficients can be shown to obey the estimates in the question. In particular, this applies to $f(x, y) = 1 + x + y$, although we do not know of any direct argument for this.

3.9.4 Continuity of $\exp[\mathsf{D}_f]$ in the coefficients of f

Start by fixing a cube $B_n = \{-n, \dots, n\}^d \subset \mathbb{Z}^d$. We can identify a polynomial $f \in \mathbb{C}[\mathbb{Z}^d]$ whose support is in B_n with its coefficient function $\widehat{f} \in \mathbb{C}^{B_n}$. Boyd [11] showed that the function $\mathbb{C}^{B_n} \to [0, \infty)$ given by $\widehat{f} \mapsto \mathsf{M}(f) = \exp[\mathsf{m}(f)]$ is continuous in the coefficients of f.

Recalling that $\mathsf{m}(f)$ is the maximum value of D_f, this suggests looking at $\exp[\mathsf{D}_f]$, which is a nonnegative upper semicontinuous function on B_n (the discontinuities occur at the boundary of $\mathcal{N}_f \subset B_n$). A function $\varphi \colon B_n \to \mathbb{R}$ is upper semicontinuous if and only if its subgraph $\{(\mathbf{u}, t) \in B_n \times \mathbb{R} : t \leqslant \varphi(\mathbf{u})\}$ is closed in $B_n \times \mathbb{R}$. The space $\mathrm{USC}(B_n)$ of all upper semicontinuous functions on B_n carries a natural topology by declaring two elements to be close if their subgraphs are close in the Hausdorff metric on closed subsets of $B_n \times \mathbb{R}$ (see [6] for details).

Question 3.9.4. Is the map $\widehat{f} \to \exp[\mathsf{D}_f]$ from $\mathbb{C}^{B_n}$ to $\mathrm{USC}(B_n)$ continuous?

3.9.5 Nonprincipal actions

Decimation makes sense for every algebraic $\mathbb{Z}^d$-action (indeed for every algebraic action of a countable residually finite group). Suppose that $\mathfrak{a}$ is an ideal in $\mathbb{Z}[\mathbb{Z}^d]$, and let $X_{\mathfrak{a}}$ be the dual group of $\mathbb{Z}[\mathbb{Z}^d]/\mathfrak{a}$ as described in §3.6. The commutative algebra there shows that the Nth decimation $r_N(X_{\mathfrak{a}})$ is defined by the contracted ideal $\mathfrak{a} \cap \mathbb{Z}[N\mathbb{Z}^d]$. However, there is no obvious replacement for g_N to measure growth when $\mathfrak{a}$ is not principal,

Question 3.9.5. If $\mathfrak{a}$ is a nonprincipal ideal in $\mathbb{Z}[\mathbb{Z}^d]$, are there objects related to the contractions $\mathfrak{a} \cap \mathbb{Z}[N\mathbb{Z}^d]$ which can be normalized to converge to a limiting object?

If $\mathfrak{a}$ is not principal, then the $\mathbb{Z}^d$-shift action on $X_{\mathfrak{a}}$ has zero entropy. However, by restricting the shift to iterates close to lower dimensional subspaces of $\mathbb{R}^d$ the action can have positive entropy [13, §6]. This suggests that the partial decimations from §3.9.2 may play a role here.

Examining concrete examples may shed some light on this question. These include the case of commuting toral automorphisms (see [52, §6] for many such examples), the $\mathbb{Z}^2$-action defined by multiplication by 2 and by 3 on $\mathbb{T}$ (corresponding to $\mathfrak{a} = \langle x-2, y-3\rangle$), and the so-called space helmet example [31, Example 5.8] (corresponding to $\mathfrak{a} = \langle 1+x+y, z-2\rangle$).

An important example of a different character is due to Ledrappier [61], which corresponds to the nonprincipal ideal $\langle 1+x+y, 2\rangle \subset \mathbb{Z}[\mathbb{Z}^2]$. This example has zero entropy as a $\mathbb{Z}^2$-action, but strictly positive entropy along every 1-dimensional subspace of $\mathbb{R}^2$ (see [13, Example 6.4] for the explicit description). Another curious feature of this example is decimation self-similarity. Because $(1+x+y)^{2^n} = 1+x^{2^n}+y^{2^n}$ when taken mod 2, the 2^nth decimation of the example, when rescaled by 2^n, is just the original action.

3.10 Example of computing the decimation limit

There are few explicit calculations of the logarithmic Mahler measure, or more generally of the Ronkin function, of polynomials in $\mathbb{Z}[\mathbb{Z}^d]$ when $d \geqslant 2$. Depending on the relative sizes of the coefficients, evaluation of the integrals involved typically requires the torus to be subdivided into a large number of subregions with complicated boundaries, and so simple formulas in terms of familiar functions are rare.

Here we treat the case $f(x,y) = 1+x+y$ from Example 3.2.4, where these calculations can be carried out, resulting in the formulas (3.8) and (3.9) for D_f.

Smyth [105] first computed the logarithmic Mahler measure $m(f) = R_f(0,0)$ to have the value in (3.10). Twenty years later Maillot [78, §7.3], aided by Cassigne, computed the entire Ronkin function $R_f(u,v)$, providing in his long memoir a concrete example of the canonical height of a hypersurface. Their result involves the Bloch-Wigner dilogarithm function, which is an alternative formulation of the series representation in our formulas. Lundqvist [70] gave the formulas for the partial derivatives of R_f we use here. He also investigated the polynomial $1+x+y+z$, and showed that the second order partial derivatives of its Ronkin function can be expressed in terms of standard elliptic functions.

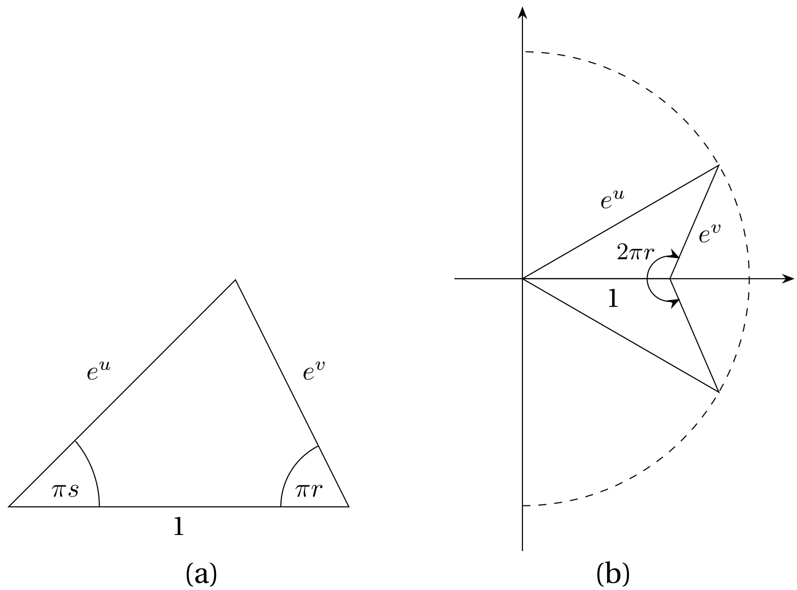


Figure 3.8: Determining partial derivatives from angles and sides

Let $\Delta = \mathcal{N}_f$ be the unit simplex, and denote its interior by Δ°. Let $\mathcal{A}_f$ be the amoeba of f, as shown in Figure 3.5, and $\mathcal{A}_f^\circ$ be its interior. To evaluate $R_f^*(r,s)$ for $(r,s) \in \Delta^\circ$, we need to know the value of $(u,v) \in \mathcal{A}_f^\circ$ at which the partial derivatives of $R_f(u,v)$ with respect to u and v equal r and s, respectively. Fortunately, there is a simple relationship that was established by Lundqvist [70], whose treatment we follow.

Lemma 3.10.1. Let $(u,v) \in \mathcal{A}_f^\circ$, so that 1, e^u, and e^v form the sides of a nondegenerate triangle. Let πr and πs be the angles in this triangle shown in Figure 3.8(a). Then

$$\frac{\partial R_f}{\partial u}(u,v) = r \quad \text{and} \quad \frac{\partial R_f}{\partial v}(u,v) = s. \tag{3.18}$$

Proof. We will compute the partial derivatives by differentiating the integrand in

$$\begin{aligned} R_f(u,v) &= \int_0^1 \int_0^1 \log|1 + e^u e^{2\pi i\theta} + e^v e^{2\pi i\varphi}|\, d\theta\, d\varphi \\ &= \operatorname{Re}\Big[\int_0^1 \int_0^1 \log(1 + e^u e^{2\pi i\theta} + e^v e^{2\pi i\varphi})\, d\theta\, d\varphi\Big]. \end{aligned}$$

In the last line log represents a local inverse to exp, which is well-defined up to the addition of an integral multiple of $2\pi i$. After taking partial derivatives, we will get a result that is independent of this multiple.

By symmetry, it suffices to compute $\partial R_f/\partial u$. Differentiating the integrand gives

$$\frac{\partial R_f}{\partial u}(u,v) = \mathrm{Re}\Big[\int_0^1\int_0^1 \frac{e^u e^{2\pi i\theta}}{1+e^u e^{2\pi i\theta}+e^v e^{2\pi i\varphi}}\,d\theta\,d\varphi\Big].$$

Rewriting the integrals as contour integrals, we see that

$$\begin{aligned}\int_0^1\int_0^1 \frac{e^u e^{2\pi i\theta}}{1+e^u e^{2\pi i\theta}+e^v e^{2\pi i\varphi}}\,d\theta\,d\varphi &= \frac{1}{(2\pi i)^2}\int_{|z|=e^u}\int_{|w|=e^v}\frac{1}{1+z+w}\,dz\frac{dw}{w}\\ &= \frac{1}{2\pi i}\int_{|w|=e^v}\Big[\frac{1}{2\pi i}\int_{|z|=e^u}\frac{dz}{z-(-1-w)}\Big]\frac{dw}{w}.\end{aligned}$$

The inner integral is the winding number of the circle of radius e^u around $-1-w=-1-e^v e^{2\pi i\varphi}$, and so has value 1 if $|1+e^v e^{2\pi i\varphi}|<e^u$ and 0 if $|1+e^v e^{2\pi i\varphi}|>e^u$ (these are mistakenly reversed in [70]). A glance at Figure 3.8(b) shows that the value is 1 for an interval of φ of length $2\pi r$, and 0 otherwise. Since $(1/2\pi i)(dw/w)$ is normalized Lebesgue measure $d\varphi$, we obtain that $(\partial R_f/\partial u)(u,v)=r$. □

To compute the decimation limit D_f, we need to express u and v in terms of r and s. Let $a=e^u$ and $b=e^v$ be the sides of the triangle in Figure 3.8(a). By the law of sines,

$$\frac{a}{\sin\pi r}=\frac{b}{\sin\pi s}=\frac{1}{\sin\pi(1-r-s)}=\frac{1}{\sin\pi(r+s)},$$

and hence

$$a=a(r,s)=e^{u(r,s)}=\frac{\sin\pi r}{\sin\pi(r+s)}, \tag{3.19}$$

$$b=b(r,s)=e^{v(r,s)}=\frac{\sin\pi s}{\sin\pi(r+s)}. \tag{3.20}$$

For $(u,v)\in\mathcal{A}_f^\circ$ it follows from the definition (3.11) that

$$-R_f^*(u,v)=\inf_{(r,s)\in\Delta^\circ} R_f(u,v)-ru-sv,$$

and by calculus the infimum is attained at the (u,v) given by (3.18). Thus for $(r,s)\in\Delta^\circ$ we have that

$$D_f(r,s)=-R_f^*\big(u(r,s),v(r,s)\big)=R_f\big(u(r,s),v(r.s)\big)-r\,u(r,s)-s\,v(r,s), \tag{3.21}$$

where $u(r,s)$ and $v(r,s)$ are determined by (3.19) and (3.20).

Remark 3.10.2. Observe that the functions $u(r,s)$ and $v(r,s)$ in (3.19) and (3.20) are real analytic on Δ°. Also, $R_f(u,v)$ is real analytic on $\mathcal{A}_f^\circ$. Together these show that $D_f(r,s)$ is real analytic on Δ°.

It remains to compute $R_f(u,v)$. By symmetry it suffices to assume that $u \geqslant v$. Using Jensen's formula (3.5), we see that

$$\begin{aligned} R_f(u,v) &= \int_0^1 \int_0^1 \log|1 + e^u e^{2\pi i\theta} + e^v e^{2\pi i\varphi}|\, d\theta\, d\varphi \\ &= u + \int_0^1 \int_0^1 \log|e^{-u} + e^{v-u} e^{2\pi i\varphi} + e^{2\pi i\theta}|\, d\theta\, d\varphi \\ &= u + \int_0^1 \log^+ |e^{-u} + e^{v-u} e^{2\pi i\varphi}|\, d\varphi. \end{aligned}$$

Note that $|e^{-u} + e^{v-u} e^{2\pi i\varphi}| \geqslant 1$ if and only if $|1 + e^v e^{2\pi i\varphi}| \geqslant e^u$, and another glance at Figure 3.8(b) shows this occurs exactly when $-\pi(1-r) \leqslant 2\pi\varphi \leqslant \pi(1-r)$. Hence

$$\begin{aligned} R_f(u,v) &= u + \int_{-\frac{1}{2}(1-r)}^{\frac{1}{2}(1-r)} \log|e^{-u} + e^{v-u} e^{2\pi i\varphi}|\, d\varphi \\ &= u - (1-r)u + \int_{-\frac{1}{2}(1-r)}^{\frac{1}{2}(1-r)} \log|1 + e^v e^{2\pi i\varphi}|\, d\varphi \\ &= r\,u + \int_{-\frac{1}{2}(1-r)}^{\frac{1}{2}(1-r)} \log|1 + e^v e^{2\pi i\varphi}|\, d\varphi. \end{aligned}$$

First suppose that $e^v < 1$, which corresponds to $(r,s) \in \Delta_1^\circ$, where Δ_1 is defined in (3.6). The series expansion of $\log(1+z)$ for $1+z$ in the domain of integration converges uniformly, and the imaginary part vanishes by symmetry. Hence

$$\begin{aligned} R_f(u,v) &= r\,u + \int_{-\frac{1}{2}(1-r)}^{\frac{1}{2}(1-r)} \sum_{n=1}^{\infty} \frac{(-1)^{n+1}}{n} e^{nv} e^{2\pi i n\varphi}\, d\varphi \\ &= r\,u + \sum_{n=1}^{\infty} \frac{(-1)^{n+1}}{n} e^{nv} \frac{1}{\pi n} \sin[\pi n(1-r)]. \end{aligned}$$

Recalling that $e^{v(r,s)} = b(r,s) = (\sin \pi s)/\sin[\pi(r+s)]$, we conclude that

$$\begin{aligned} D_f(r,s) &= R_f\big(u(r,s), v(r,s)\big) - r\,u(r,s) - s\,v(r,s) \\ &= \sum_{n=1}^{\infty} \frac{(-1)^{n+1}}{\pi n^2} b(r,s)^n \sin[\pi n(1-r)] - s\,\log[b(r,s)]. \end{aligned} \tag{3.22}$$

Now suppose that $e^v > 1$, which corresponds to $(r, s) \in \Delta_2^\circ$, where Δ_2 is defined by (3.7). Then $\log|1 + e^v e^{2\pi i\varphi}| = v + \log|1 + e^{-v} e^{-2\pi i\varphi}|$. Calculating as before,

$$\begin{aligned} R_f(u, v) &= r\,u + (1 - r)v + \int_{-\frac{1}{2}(1-r)}^{\frac{1}{2}(1-r)} \sum_{n=1}^{\infty} \frac{(-1)^{n+1}}{n} e^{-nv} e^{-2\pi i n\varphi}\, d\varphi \\ &= r\,u + (1 - r)v + \sum_{n=1}^{\infty} \frac{(-1)^{n+1}}{\pi n^2} b(r, s)^{-n} \sin[\pi n(1 - r)]. \end{aligned}$$

Thus for $(r, s) \in \Delta_2^\circ$ we find that

$$D_f(r, s) = \sum_{n=1}^{\infty} \frac{(-1)^{n+1}}{\pi n^2} b(r, s)^{-n} \sin[\pi n(1 - r)] + (1 - r - s) \log[b(r, s)]. \quad (3.23)$$

Finally, note that on the overlap $\Delta_1 \cap \Delta_2$ inside Δ°, we have that $b(r, s) = 1$ and so the series in (3.6) and (3.7) converge and agree, hence give the value of $D_f(r, s)$ by continuity of the Legendre transform.

Chapter 4

On the determinantal process associated to spanning trees[1]

4.1 Discrete determinantal processes

Let E be a finite or countable set and $X \subset E$ be a random subset of E. Then, one can associate X to a random variable $\tilde{X}$ with values in $\{0,1\}^E$ where $\tilde{X}(e) = 1$ if $e \in X$. The random variable $\tilde{X}$ is called a point process. A point process $\tilde{X}$ is called *determinantal* (abbreviated to the Determinantal Point Process, or DPP) if there exists a function $\mathbb{K}\colon E \times E \to \mathbb{C}$ such that for any finite collection of distinct points $\{e_1, \ldots, e_n\} \in E$, the probability $\mathbb{P}(e_1, \ldots, e_n \in X)$ is given by the following determinant:

$$\mathbb{P}(e_1, \ldots, e_n \in X) = \mathbb{P}(\tilde{X}(e_i) = 1, i = 1, \ldots, n) = \det\big[\mathbb{K}(e_i, e_j)\big]_{i,j=1}^{n}. \quad (4.1)$$

The function $\mathbb{K}$ is called the *correlation kernel* of the determinantal point process $\tilde{X}$.

Analogously, a probability measure $\mathbb{P}$ on $\{0,1\}^E$ is called determinantal if there exists a function $\mathbb{K} : E \times E \to \mathbb{C}$ such that (4.1) is valid for all $\{e_1, \ldots, e_n\} \in E$ where the points e_i are all distinct. A natural question is to determine which functions $\mathbb{K}$ are correlation kernels, i.e., for which $\mathbb{K}$ the expressions in (4.1) define a probability measure on $\{0,1\}^E$. To address to this question we recall the three most common ways to define determinantal measures and their correlation kernels:

1. **Fourier transform:** Suppose $E = \mathbb{Z}^d$ and $f : \mathbb{T}^d \to [0,1]$ is an L^2-

[1] This chapter is based on: E. Arzhakova, T. Shirai, E. Verbitskiy, On the determinantal process associated to spanning trees, in progress

integrable function. Define the correlation kernel $\mathbb{K}$ on $\mathbb{Z}^d \times \mathbb{Z}^d$ by

$$\mathbb{K}(n,m) = \widehat{f}(n-m), \quad n,m \in \mathbb{Z}^d,$$

where $\widehat{f}(k)$ is the k-th Fourier coefficient of f given by

$$\widehat{f}(k) = \int_{\mathbb{T}^d} f(\theta) e^{-2\pi i \langle k,\theta \rangle} d\theta, k \in \mathbb{Z}^d.$$

The kernel $\mathbb{K}$ defines a stationary translation invariant determinantal measure on $\mathbb{Z}^d$ [72].

2. **Projection:** Consider the Hilbert space $\mathcal{H} = \ell^2(E)$ with the scalar product $\langle \cdot, \cdot \rangle$. Note that $\mathcal{H}$ is generated by the indicator functions $\{\mathbf{1}_n, n \in E\}$ given by

$$\mathbf{1}_n(m) = \begin{cases} 1, & m = n, \\ 0, & m \neq n. \end{cases}$$

Suppose H is a closed subspace of $\mathcal{H}$. Denote by P_H the orthogonal projection operator onto H. Then the kernel $\mathbb{K}(n,m) = \langle P_H \mathbf{1}_n, \mathbf{1}_m \rangle$ defines a determinantal measure.

3. **Positive contraction:** Again, consider the Hilbert space $\ell^2(E)$, and suppose $Q : \ell^2(E) \to \ell^2(E)$ is a positive contraction, i.e.,

$$0 \leqslant \langle Qu, u \rangle \leqslant \langle u, u \rangle \quad \forall u \in \ell^2(E).$$

Then the kernel $\mathbb{K}(n,m) = \langle Q\mathbf{1}_n, \mathbf{1}_m \rangle$, $n, m \in E$, defines a determinantal measure on $\{0,1\}^E$.

Note that here is a natural bijection between the DPPs generated by projections (2) and those generated by positive contractions (3). Clearly, any projection operator is a positive contraction, so, $(2) \subset (3)$. In the other direction, consider a positive contraction A that acts on $\ell^2(E)$; then, the operator

$$\widehat{A} := \begin{pmatrix} A & \sqrt{A(I-A)} \\ \sqrt{A(I-A)} & I-A \end{pmatrix}$$

is a projection operator on $\ell^2(E_1) \oplus \ell^2(E_2)$ where $E_1 = E_2 = E$ since $\widehat{A}$ is idempotent ($\widehat{A}^2 = \widehat{A}$) and self-adjoint ($\widehat{A}^* = \widehat{A}$). Moreover, $\widehat{A}$ is a dilation of A; in other words, $P_{\ell^2(E_1)}\widehat{A}u = Au$ [73] and $P_{\ell^2(E_2)}\widehat{A}u = (I-A)u$. We have shown that each positive contraction A generates a projection operator.

The third class of examples (3) also contains the first class of examples (1). Indeed, since $\ell^2(\mathbb{Z}^d) \cong L^2(\mathbb{T}^d)$ via the Fourier transform, the operator $\widehat{Q} : L^2(\mathbb{T}^d) \to L^2(\mathbb{T}^d)$ which acts by the formula $\widehat{Q}h = f \cdot h$, is also a positive contraction on $L^2(\mathbb{T}^d)$ since $f : \mathbb{T}^d \to [0,1]$. It follows that the adjoint operator Q on $\ell^2(\mathbb{Z}^d)$ is also a positive contraction on $\ell^2(\mathbb{Z}^d)$.

In the current work we focus on two primary examples of discrete determinantal point processes which are *Uniform Spanning Trees process* and *Uniform Spanning Forest process.*

Let $G = (V,E)$ be a finite connected graph. A *spanning tree* $T \subset E$ is a subset of edges of G such that the graph (V,T) is connected and does not have cycles. Denote by $\mathcal{T}(G)$ the set of all spanning trees of G. Since G is finite, $|\mathcal{T}(G)| < \infty$, and by a well-known Kirkhoff's Matrix Tree theorem, $|\mathcal{T}(G)|$ is equal to the determinant of the (reduced) Laplacian matrix of G. Let us consider the random variable T assuming values in $\mathcal{T}(G)$ with uniform probabilities. Burton and Pemantle [16] have proved the *Transfer Current Theorem,* showing that the edges of uniformly chosen spanning trees form a determinantal process.

Theorem 4.1.1 ([16])**.** Let $G = (V,E)$ be a finite graph. Denote by $\mathbb{P}$ be the uniform spanning tree measure on G which we view as a probability measure on $\{0,1\}^E$. This measure describes the probability that edges $e_1,\dots,e_n$ are present in a random spanning tree which is chosen uniformly. Choose an arbitrary orientation of the edges of the graph G and let $\mathbb{K}(e_i,e_j)$ be equal to the expected signed number of crossings of $e_i = \overrightarrow{xy}$ by a simple random walk on G started at s and stopped when it hits t with $e_j = \overrightarrow{st}$. Then, the uniform spanning tree measure on G is given by the determinant

$$\mathbb{P}(e_1,\dots,e_n \in T) = \det\big[\mathbb{K}(e_i,e_j)\big]_{i,j=1}^n, \tag{4.2}$$

i.e., $\mathbb{P}$ is a determinantal probability measure.

In [89] Pemantle considered uniform spanning trees on certain infinite graphs, namely, on integer lattices $\mathbb{Z}^d$ with $d \geqslant 2$. He showed that the uniform measures $\mathbb{P}_n$ on spanning trees of finite boxes $B_n = [-n,n]^d \cap \mathbb{Z}^d$ converge weakly as $n \to \infty$ to the limiting measure $\mathbb{P}$. If $d \leqslant 4$ the limiting measure $\mathbb{P}$ is concentrated on spanning trees; otherwise, it concentrates on *spanning forests,* i.e., on collections of disjoint spanning trees that span the whole lattice. This limiting measure is referred to as *uniform spanning tree measure* or *uniform spanning forest measure,* respectively.

As discussed above, Burton and Pemantle [16] established the determinantal structure of uniform measures on spanning trees of finite graphs, c.f, Theorem 4.1.1. They also showed that the determinantal structure is present on some infinite graphs, namely, on $\mathbb{Z}^d$ lattices and some $\mathbb{Z}^d$-periodic graphs:

Theorem 4.1.2 ([16]). Let $G = (V, E)$ be a $\mathbb{Z}^d$-periodic, D-regular, connected graph. Denote by $\mathbb{P}$ be the uniform spanning forest (USF) measure on G which is viewed as a probability measure on $\{0,1\}^E$. Then $\mathbb{P}$ is determinantal with the correlation kernel $\mathbb{K}$ given in Theorem 4.1.1.

Let us illustrate the computation of the correlation kernel that features in the two theorems discussed above:

Example 4.1.3. In order to compute the correlation kernel $\mathbb{K}$ for the USF measure on $\mathbb{Z}^d$ explicitly one can use the Green's function $g\colon \mathbb{Z}^d \to \mathbb{R}$ of the simple random walk. The expression of the Green's function depends on the dimension d: for $d = 2$, it is given by

$$\begin{aligned} g(n) &= \sum_{k \geqslant 0} \Big(\mathbb{P}(X_k = n | X_0 = 0) - \mathbb{P}(X_k = 0 | X_0 = 0) \Big) \\ &= \int_{\mathbb{T}^2} \frac{e^{-2\pi i \langle n, \theta \rangle} - 1}{1 - \frac{1}{2}\cos 2\pi\theta_1 - \frac{1}{2}\cos 2\pi\theta_2} d\theta, \end{aligned}$$

and for $d \geqslant 3$ it is given by

$$g(n) = \sum_{k \geqslant 0} \mathbb{P}(X_k = n | X_0 = 0) = \int_{\mathbb{T}^d} \frac{e^{-2\pi i \langle n, \theta \rangle}}{1 - \frac{1}{d}\sum_{j=1}^d \cos 2\pi\theta_j} d\theta.$$

Then [16, Theorem 4.2], [71, Proposition 10.15], the kernel $\mathbb{K}$ calculated on the edges $e = \overrightarrow{xy}$ and $\tilde{e} = \overrightarrow{zw}$ is given by the expression

$$\mathbb{K}(e, \tilde{e}) = \frac{1}{2d} \Big[g(z - x) - g(z - y) - g(w - x) + g(w - y) \Big]. \tag{4.3}$$

Note that explicit computation of the Green's function for an arbitrary graph G is not trivial, e.g., even for $\mathbb{Z}^2$ one has to modify the standard definition to accommodate for non-integrability of the denominator. Fortunately, in $\mathbb{Z}^2$ the required modification is straightforward, but even for simple graphs the calculation of the Green's function becomes cumbersome. Therefore, it is natural to consider other methods to calculate the correlation kernel $\mathbb{K}$. In the following Section we recall a method presented in [12] that calculates $\mathbb{K}$ as a projection in a certain Hilbert space.

4.1.1 Hilbert space approach

In [12], Benjamini et al. further developed the electrical network approach that we discussed above and obtained a description of the determinantal structures of the USF measures on general finite and infinite graphs in terms of certain projection operators.

Let us first consider a finite connected undirected graph $G = (V, E)$. For each undirected edge $[e] \in E$ select an arbitrary orientation denoted by e; then, we denote the reversed orientation by $-e$. We obtain a collection or oriented edges $\overline{E}$ containing each edge from E with two possible orientations. In other words, $\overline{E} = \cup_{[e]\in E}\{e, -e\}$. Throughout the paper, we will use the following notation. If the collection of edges is denoted by a capital letter $E, F, \ldots$, then the edges are assumed to be undirected; and for the set of all possible directed graphs we will use $\overline{E}, \overline{F}, \ldots$. For $e \in \overline{E}$ denote by $o(e)$ the origin of e and by $t(e)$ the terminus of e. For the reversed edge $-e$ we naturally have $o(-e) = t(e)$ and $t(-e) = o(e)$.

Denote by $\ell^2(V)$ the real Hilbert space of functions $f\colon V \to \mathbb{R}$ with the standard inner product given by

$$\langle f_1, f_2\rangle = \sum_{v\in V} f_1(v) f_2(v). \tag{4.4}$$

Denote by $\ell^2_-(\overline{E})$ the space of antisymmetric (i.e., $\varphi(-e) = -\varphi(e)$) real functions on $\overline{E}$ with the standard inner product given by

$$\langle \varphi_1, \varphi_2\rangle = \frac{1}{2}\sum_{e\in\overline{E}} \varphi_1(e)\varphi_2(e). \tag{4.5}$$

Define the coboundary operator $d\colon \ell^2(V) \to \ell^2_-(\overline{E})$ and the divergence operator $d^*\colon \ell^2_-(\overline{E}) \to \ell^2(V)$ as

$$df(e) = f(o(e)) - f(t(e)), \quad d^*\theta(v) = \sum_{o(e)=v} \theta(e). \tag{4.6}$$

It is easy to check that d and d^* are adjoint: $\langle df, \theta\rangle = \langle f, d^*\theta\rangle$ for all $f \in \ell^2(V)$ and $\theta \in \ell^2_-(\overline{E})$. Consider a subspace $\operatorname{Im} d$ defined as

$$\operatorname{Im} d = d\,\ell^2(V) = \text{span}\{\, d\,\mathbb{1}_v\}_{v\in V} \subset \ell^2_-(\overline{E}).$$

Finally, denote by $P_{\overline{\operatorname{Im} d}}$ the orthogonal projection from $\ell^2_-(\overline{E})$ onto the closed subspace $\overline{\operatorname{Im} d}$.

Theorem 4.1.4 ([12]). Let $\mathbb{P}$ be the uniform spanning tree measure for $G = (V, E)$. Then $\mathbb{P}$ is determinantal with

$$\mathbb{P}(e_1, \ldots, e_k \in T) = \det[(P_{\overline{\operatorname{Im} d}} \mathbb{1}_{e_i}, \mathbb{1}_{e_j})]_{i,j=1,\ldots,k},$$

where for $e \in \overline{E}$, $\mathbb{1}_e \in \ell^2_-(\overline{E})$ is an antisymmetric indicator function of the directed edge e (i.e., $\mathbb{1}_e(e) = 1, \mathbb{1}_e(-e) = 1$ and $\mathbb{1}_e(j) = 0$ for all $j \in \overline{E}$ such that $j \neq e, -e$).

As was mentioned above, Pemantle [89] constructed a USF measure $\mathbb{P}$ on $\mathbb{Z}^d$ by considering weak limits of uniform spanning tree measures on finite graphs. This approach can be extended to a larger class of infinite connected locally finite graphs $G = (V, E)$. Consider a sequence of finite subsets of vertices $\{V_n\}$ of V such that $V_1 \subset V_2 \subset \ldots$ and $\cup_1^\infty V_n = V$. Using the sequence $\{V_n\}$ we can define two sequences of finite graphs converging to $G = (V, E)$. Firstly, let $G_n^F = (V_n, E_n)$ be the graph obtained by deleting all vertices in V which lie outside of V_n (free boundary conditions), and secondly, let $G_n^W = (V_n \cup \{\varnothing\}, \tilde{E}_n)$ be the graph obtained by contracting all vertices in V which lie outside of V_n into one vertex $\varnothing$ (wired boundary conditions). Denote by $\mathbb{P}_n^F$ and $\mathbb{P}_n^W$ the corresponding uniform spanning tree measures on $G_n^F = (V_n, E_n)$ and $G_n^W = (V_n \sqcup \{\varnothing\}, \tilde{E}_n)$, respectively. The measures $\mathbb{P}_n^F$ and $\mathbb{P}_n^W$ have the following monotonicity property: for a fixed collection of edges B and for all sufficiently large n, one has

$$\mathbb{P}_n^F(B \subseteq T) \geqslant \mathbb{P}_{n+1}^F(B \subseteq T), \quad \text{and} \quad \mathbb{P}_n^W(B \subseteq T) \leqslant \mathbb{P}_{n+1}^W(B \subseteq T),$$

which allows one to define limiting measures

$$\mathbb{P}^F = \lim_{n\to\infty} \mathbb{P}_n^F, \quad \mathbb{P}^W = \lim_{n\to\infty} \mathbb{P}_n^W,$$

called the *free* and the *wired* uniform spanning forest measures, respectively. In general, the measures $\mathbb{P}^F$ and $\mathbb{P}^W$ do no necessarily coincide.

It turns out that both measures $\mathbb{P}^F$ and $\mathbb{P}^W$ are determinantal with kernels similar to that in Theorem 4.1.4. To describe the corresponding kernels we introduce additional notations. Let us call *stars* the functions of the form $d\mathbb{1}_e$ where $\mathbb{1}_e$ is an indicator function of an oriented edge $e \in \overline{E}$; denote by $\bigstar = \overline{d\ell^2(V)}$ the closed infinite-dimensional subspace of $\ell^2_-(\overline{E})$ spanned by stars [12]. If a collection of oriented edges $(e_1, \ldots, e_n)$ forms an oriented cycle, then a function $\sum_{i=1}^n \mathbb{1}_{e_i}$ is called a *cycle*. Cycles span a subspace of $\ell^2_-(\overline{E})$ denoted by $\diamondsuit$. The subspaces $\bigstar$ and $\diamondsuit$ of $\ell^2_-(\overline{E})$ are orthogonal. In [12] it is shown that $\mathbb{P}^W = \mathbb{P}^F$ if and only if $\bigstar \oplus \diamondsuit = \ell^2_-(\overline{E})$.

Remark 4.1.5. Let G be a graph equipped with an action of an abelian finitely-generated group Γ (see Sections 3 and 4 below). Then $\mathbb{P}^F = \mathbb{P}^W$ [33].

Theorem 4.1.6 ([12])**.** For an infinite connected graph G, the uniform spanning forest measures $\mathbb{P}^F$ and $\mathbb{P}^W$ corresponding to the free and the wired boundary conditions, are determinantal:

$$\mathbb{P}^W(e_1,\ldots,e_k \in T) = \det[\langle P_\bigstar \mathbb{1}_{e_i}, \mathbb{1}_{e_j}\rangle]_{1\leqslant i,j\leqslant k},$$

$$\mathbb{P}^F(e_1,\ldots,e_k \in T) = \det[\langle P_\Diamond^\perp \mathbb{1}_{e_i}, \mathbb{1}_{e_j}\rangle]_{1\leqslant i,j\leqslant k},$$

where $P_\bigstar$ is an orthogonal projection on the closure of $\bigstar$ and $P_\Diamond^\perp$ is an orthogonal projection on the closure of the orthogonal complement of $\Diamond$.

Remark 4.1.7. Note that both $\mathbb{P}^W(e_1,\ldots,e_k \in T)$ and $\mathbb{P}^F(e_1,\ldots,e_k \in T)$ do not depend on the choice of orientation.

Theorem 4.1.6 shows that the correlation kernel can be expressed in terms of projection operators. However, finding explicit expressions of projections in infinite-dimensional Hilbert spaces is a challenging task. In [71, Chapter 4] one can find computations of correlation kernels for lattices $\mathbb{Z}^d$ using graph Laplacians; this proof requires a number of technical steps. The same construction can, in principle, be used for other graphs, however, the proofs become more complex, in particular, for graphs where the simple random walk is recurrent.

In the following section we show how to compute a projection operator $P_\bigstar$ of Theorem 4.1.6 explicitly for graphs with abelian symmetry groups.

4.2 Projection operator

We start the discussion with a well-known fact on computation of projections in finite-dimensional spaces. Let us consider a k-dimensional subspace $W \subset \mathbb{R}^n$ with $k < n$, and let P_W be the orthogonal projection from $\mathbb{R}^n$ onto W. Suppose W is spanned by linearly independent vectors $w_1,\ldots,w_k$.

Denote by A the $n \times k$ matrix, made of column vectors $w_1,\ldots,w_k$:

$$A = [w_1;\ldots;w_k] \in \mathbb{R}^{n\times k}.$$

Since the vectors w_i are linearly independent, the $k \times k$ matrix A^TA is invertible. A standard exercise in Linear Algebra shows that the projection operator P_W is given by the matrix product

$$P_W = A(A^TA)^{-1}A^T \tag{4.7}$$

In view of Theorem 4.1.6, one would be interested whether a similar expression is true in greater generality. The following result is a natural generalization of (4.7) to the infinite-dimensional separable Hilbert spaces.

Theorem 4.2.1. Suppose X, Y are separable Hilbert spaces, and $A\colon X \to Y$ is a bounded linear operator. Let $A^*\colon Y \to X$ be the adjoint operator of A and let $P_A : Y \to Y$ be the orthogonal projection onto the closed subspace $\overline{\operatorname{Im} A} \subset Y$. Then

$$\lim_{\epsilon \to 0} A(A^*A + \epsilon I)^{-1}A^* = P_A,$$

where $I : X \to X$ is the identity operator, and the convergence is understood in the strong operator topology.

We start the proof of Theorem 4.2.1 with the following proposition:

Proposition 4.2.2. Under conditions of Theorem 4.2.1, for any $\epsilon > 0$ the operators $P(\epsilon) : Y \to Y$ given by

$$\mathbb{P}(\epsilon) = A(A^*A + \epsilon I)^{-1}A^*,$$

are well-defined and have uniformly bounded norms.

Proof. The operator $(A^*A+\epsilon I)^{-1}$ is non-negative and self-adjoint, and therefore, has a unique non-negative self-adjoint square root $B(\epsilon) = (A^*A + \epsilon I)^{-1/2}$. Let $C(\epsilon) = AB(\epsilon) = A(A^*A + \epsilon I)^{-1/2}$, then $C^*(\epsilon) = B(\epsilon)A^*$, and since $P(\epsilon) = C(\epsilon)C^*(\epsilon)$, one has

$$\begin{aligned} ||\mathbb{P}(\epsilon)|| &= ||C(\epsilon)C^*(\epsilon)|| = ||C^*(\epsilon)C(\epsilon)|| \\ &= ||(A^*A + \epsilon I)^{-1/2}A^*A(A^*A + \epsilon I)^{-1/2}||, \end{aligned} \tag{4.8}$$

therefore, $||\mathbb{P}(\epsilon)|| = ||g(A^*A)||$ where $g(x) = \frac{x}{x+\epsilon}, \epsilon > 0$. Note that the operator A^*A is bounded and positive-definite: therefore, its spectral radius $\lambda_\infty(A^*A)$ coincides with the norm of A^*A. The spectral mapping theorem implies that $||\mathbb{P}(\epsilon)|| = ||g(A^*A)|| = g(\lambda_\infty(A^*A)) < 1$; one concludes that $||\mathbb{P}(\epsilon)|| < 1$ for every $\epsilon > 0$. $\square$

Remark 4.2.3. The operator $P(\epsilon), \epsilon > 0$ is a positive contraction; for discrete (finite or countable) Y it defines a determinantal point process with the correlation kernel $\mathbb{K}_\epsilon(e_i, e_j) = \langle P_\epsilon \mathbb{1}_{e_i}, \mathbb{1}_{e_j}\rangle, e_i, e_j \in Y$.

We present the reader with two proofs of Theorem 4.2.1:

Analytic proof of Theorem 4.2.1. Let us start by showing that $\lim_{\epsilon \to 0} \mathbb{P}(\epsilon)Ax = Ax$ for every $x \in X$. Applying $\mathbb{P}(\epsilon)$ to Ax, we see that

$$\begin{aligned}\mathbb{P}(\epsilon)Ax &= A(A^*A+\epsilon I)^{-1}A^*Ax = A(A^*A+\epsilon I)^{-1}[(A^*A+\epsilon I)-\epsilon I]x \\ &= Ax - \epsilon A(A^*A+\epsilon I)^{-1}x.\end{aligned} \tag{4.9}$$

However, the norm of an operator $\epsilon A(A^*A+\epsilon I)^{-1}$ can be bounded as follows:

$$\begin{aligned}||\epsilon A(A^*A+\epsilon I)^{-1}|| &= ||\epsilon^{1/2}A(A^*A+\epsilon I)^{-1/2}\epsilon^{1/2}(A^*A+\epsilon I)^{-1/2}|| \\ &\leqslant \epsilon^{1/2}||A(A^*A+\epsilon I)^{-1/2}|| \cdot ||\epsilon^{1/2}(A^*A+\epsilon I)^{-1/2}||.\end{aligned} \tag{4.10}$$

The norm of the operator $C(\epsilon) = A(A^*A+\epsilon I)^{-1/2}$ is bounded. Indeed, since $C^*(\epsilon)C(\epsilon) = P(\epsilon)$, one has $||C(\epsilon)||^2 = ||C^*(\epsilon)C(\epsilon)|| = ||\mathbb{P}(\epsilon)|| \leqslant 2$ by Proposition 4.2.2. The norm of the operator $\epsilon^{1/2}(A^*A+\epsilon I)^{-1/2}$ is bounded by 1. Therefore, $||\epsilon A(A^*A+\epsilon I)^{-1}|| \leqslant \epsilon^{1/2} \cdot ||C(\epsilon)|| \cdot 1 \leqslant \sqrt{2\epsilon} \to 0$ as $\epsilon \to 0$. Thus, by (4.9)

$$\mathbb{P}(\epsilon)Ax - Ax = \epsilon A(A^*A+\epsilon I)^{-1}x, \text{ with } ||\epsilon A(A^*A+\epsilon I)^{-1}|| \leqslant \sqrt{2\epsilon}$$

Hence, $\mathbb{P}(\epsilon)Ax \to Ax$, and thus $\mathbb{P}(\epsilon) \to I$ on $\overline{\mathrm{Im}A}$. Moreover, $\mathbb{P}(\epsilon) = 0$ on $(\overline{\mathrm{Im}A})^\perp$. Therefore, one concludes that

$$\lim_{\epsilon \to 0} A(A^*A+\epsilon I)^{-1}A^* = \mathrm{Proj}_{\overline{\mathrm{Im}A}} = P_A.$$

□

Spectral proof of Theorem 4.2.1. Note that $Y = \overline{\mathrm{Im}A} \oplus (\overline{\mathrm{Im}A})^\perp$. Let us consider a sequence of operators $A(A^*A+\epsilon I)^{-1}A^*Ax$ and prove that it converges to Ax.

Each self-adjoint operator $K\colon H \to H$ on a separable Hilbert space H admits a sum representation $K = \sum_{m\in\mathbb{N}} q_m\langle x, w_m\rangle w_m$ where $\{w_m\}_{m\in\mathbb{N}}$ is a countable orthonormal basis of H and $\{q_m\}_{m\in\mathbb{N}}$ is the spectrum of K. Note that if $K = A^*A$ for some operator A then its spectrum is non-negative and real. Then, one can define a family of spectral projections $\{E_{\lambda_n}\}_{n\in[0,\infty)}$ by

$$E_{\lambda_n}(K)(x) = \sum_{q_m \in (-\infty, \lambda_n]} q_m\langle x, w_m\rangle w_m.$$

This family is called the resolution of unity of the operator K; it implies that $K = \int_{-\infty}^{\infty} \lambda dE_\lambda(K)$. If K is non-negative, clearly, $K = \int_0^{\infty} \lambda dE_\lambda(K)$.

Denote by $\mathrm{Spec}A^*A \subset \mathbb{R}$ the spectrum of the operator A^*A (note that the spectrum of a self-adjoint operator is real). Then, the spectrum of the inverse operator $\mathrm{Spec}(A^*A)^{-1}$ on its range $\mathrm{Im}A^*A$ is equal to $(\mathrm{Spec}A^*A\backslash\{0\})^{-1}$. Moreover, if w is an eigenvector of A^*A with a positive eigenvalue λ_e then w is an eigenvector of $(A^*A+\epsilon I)^{-1}$ with an eigenvalue $\frac{1}{\lambda+\epsilon}$. We conclude that

$$\mathrm{Spec}(A^*A+\epsilon I)^{-1}A^*A = \left\{\frac{\lambda}{\lambda+\epsilon}\right\}_{\lambda\in\mathrm{Spec}A^*A}.$$

By the property of the resolution of unity it follows that on $(\ker A^*A)^{\perp}$

$$(A^*A+\epsilon)^{-1}A^*A = \int_{(0,\infty)} \frac{\lambda}{\lambda+\epsilon} dE_{\lambda}(A^*A).$$

If $x \in \ker A^*A$ then $(A^*A+\epsilon)^{-1}A^*A(x) = 0$. Therefore, on the whole H_1,

$$A(A^*A+\epsilon)^{-1}A^*A = A\int_{(0,\infty)} \frac{\lambda}{\lambda+\epsilon} dE_{\lambda}(A^*A).$$

Note that $\ker A^*A = \ker(A^*A)^{-1} = \ker A$. We leave the proof of this fact as an exercise for the reader. Then, $A(E_{\{0\}}(A^*A)) = 0$; so,

$$A(A^*A+\epsilon)^{-1}A^*A = A\int_{[0,\infty)} \frac{\lambda}{\lambda+\epsilon} dE_{\lambda}(A^*A).$$

When $\epsilon \to 0$, the expression $\frac{\lambda}{\lambda+\epsilon} \to 1$, so

$$\lim_{\epsilon\to 0} A(A^*A+\epsilon)^{-1}A^*A = A\int_0^{\infty} dE_{\lambda}(A^*A).$$

By the definition of the resolution of unity, $\int_0^{\infty} dE_{\lambda}(A^*A) = I$. Then, for $y = Ax+z, x\in H_1, z\in(\mathrm{Im}A)^{\perp}$, it holds that

$$\lim_{\epsilon\to 0} A(A^*A+\epsilon)^{-1}A^*y = Ax.$$

It follows that for any $y \in H_2$,

$$\lim_{\epsilon\to 0} A(A^*A+\epsilon)^{-1}A^*y = P_A y,$$

where $P_A(y)$ is a projection of y onto $\overline{\mathrm{Im}(A)}$.

□

In the following sections, we will apply Theorem 4.2.1 to computation of projections in relation to Theorem 4.1.6. Namely, we will show that for infinite graphs with *symmetry* one can obtain explicit expressions of correlation kernels.

4.3 Graphs with abelian symmetries

In this section we discuss infinite graphs that possess sufficiently rich symmetry groups, for instance, periodic graphs. Consider a connected unoriented graph $G = (V, E)$ where $V = V(G)$ is the set of vertices and $E = E(G)$ is the set of unoriented edges. We remind the reader of the notation used in Section 1: $\overline{E}$ is the set of oriented edges: each unoriented edge $[e] \in E$ is included with two possible orientations $e, -e$ into $\overline{E}$; $o(e)$ and $t(e)$ are the origin and the terminus of $e \in \overline{E}$, respectively.

Definition 4.3.1. A mapping $g : G \to G$ consists of a pair of mappings $g_V : V \to V$ and $g_{\overline{E}} : \overline{E} \to \overline{E}$. The mapping $g : G \to G$ is called a *graph automorphism* if

1. $o(g_{\overline{E}}(e)) = g_V(o(e))$, i.e., the origin of the edge is mapped onto the origin of the edge,

2. $t(g_{\overline{E}}(e)) = g_V(t(e))$, i.e., the terminus of the edge is mapped onto the terminus of the edge,

3. $-g_{\overline{E}}(e) = g_{\overline{E}}(-e)$ $(e \in \overline{E})$, i.e., the image of the reverse of an edge is the reverse of the image of the edge.

It is easy to check that due to the conditions above, any graph automorphism $g : G \to G$ naturally induces a mapping on undirected edges $g_E : E \to E$ on E by $g_E([e]) := [g_{\overline{E}}(e)]$

Let Γ be a discrete countable group which acts on a graph $G = (V, E)$ by graph automorphisms:

$$\Gamma \ni \gamma \mapsto g^\gamma = (g_V^\gamma, g_E^\gamma) \in \mathsf{Aut}(G).$$

Definition 4.3.2. A pair $(G_0 = (V_0, E_0), \pi : G \to G_0)$ is called *a quotient of G with respect to* Γ denoted by G/Γ if for every $v_0 \in V_0$, $\pi^{-1}(v_0) = \{g_V^\gamma(v) | \gamma \in \Gamma, v \in \pi^{-1}(v_0)\}_\gamma := \Gamma v$ and for every $e_0 \in E_0$, $\pi^{-1}(e_0) = \{g_E^\gamma(e) | \gamma \in \Gamma, e \in \pi^{-1}(e_0)\} := \Gamma e$. In other words, π is a covering map of G_0, i.e., for every $\gamma \in \Gamma$ one has $\pi \circ g^\gamma = \pi$.

From now on, we will simplify the notation and use the same letter γ to denote the graph automorphism g^γ corresponding to $\gamma \in \Gamma$.

Let $G = (V, E)$ be an infinite graph and let Γ be a finitely generated abelian group of automorphisms of G that satisfies the following conditions:

1. The group Γ acts freely on G: namely, $\gamma v \neq v$ and $\gamma e \neq \pm e$ unless $\gamma = \mathsf{id}$.
2. The quotient graph G/Γ is a finite graph $G_0 = (V_0, E_0)$.

We would need one more algebraic object associated with a countable additive group Γ : namely, the group ring $\mathbb{Z}\Gamma$, which we view as the ring of Laurent polynomials in variable $\boldsymbol{x}$. This ring is formed by all expressions of the form

$$f = \sum_{\gamma\in\Gamma} f_\gamma \cdot \boldsymbol{x}^\gamma,$$

where $f_\gamma \in \mathbb{Z}$ for all $\gamma \in \Gamma$ are such that the set $\{\gamma \in \Gamma : f_\gamma \neq 0\}$ is finite. For the Laurent polynomial $f = \sum_{\gamma\in\Gamma} f_\gamma \cdot \boldsymbol{x}^\gamma$, the adjoint of f is defined as $f^* = \sum_{\gamma\in\Gamma} f_\gamma \cdot \boldsymbol{x}^{-\gamma}$. The sum of two polynomials $f = \sum_{\gamma\in\Gamma} f_\gamma \cdot \boldsymbol{x}^\gamma$ and $h = \sum_{\gamma\in\Gamma} h_\gamma \cdot \boldsymbol{x}^\gamma$ is defined as $f + h = \sum_{\gamma\in\Gamma}(f_\gamma + h_\gamma) \cdot \boldsymbol{x}^\gamma$, and the product is defined as $f \cdot h = \sum_{g\in\Gamma}(f\cdot h)_\gamma \cdot \boldsymbol{x}^\gamma$ with

$$(f \cdot h)_\gamma = \sum_{\gamma'\in\Gamma} f_{\gamma-\gamma'} h_{\gamma'}.$$

We will write 1 for $\boldsymbol{x}^0$.

Let $\mathscr{F} = (\mathscr{F}_V, \mathscr{F}_{\overline{E}})$ be an arbitrary fundamental set of G with respect to the Γ-action. It means that

$$V = \coprod_{\gamma\in\Gamma} \gamma(\mathscr{F}_V), \quad \overline{E} = \coprod_{\gamma\in\Gamma} \gamma(\mathscr{F}_{\overline{E}}).$$

Let us fix an orientation $\mathscr{F}_{E^o}$ on $\mathscr{F}_E$: note that it defines an orientation E^0 on E. Take $\pi\colon G \to \mathscr{F}$ as a covering map. For any $e \in \mathscr{F}_{E^o}$ fix $v, v' \in \mathscr{F}_V, e' \in \overline{E}$ and $\gamma \in \Gamma$ (γ can be $\mathbf{1}$) such that $\pi(e') = e$, $o(e') = v \in \mathscr{F}_V$ and $t(e') = \gamma v' \in V$ (naturally, $\mathscr{F}_{\overline{E}} = \mathscr{F}_{E^o} \sqcup -\mathscr{F}_{E^o}$). Let us define an $|E_0| \times |V_0|$ "edge-vertex" incidence matrix $\partial_{\mathscr{F}} = (\partial_{\mathscr{F}_{E^o}}(e, v))$ where $e \in \mathscr{F}_{E^o}, v \in \mathscr{F}_V$ and its adjoint $\partial^*_{\mathscr{F}} = (\partial^*_{\mathscr{F}}(v, e))$ where $v \in \mathscr{F}_V, e \in \mathscr{F}_{E^o}$, by the formulas

$$\begin{aligned}
\partial_{\mathscr{F}}(e, v) &= \mathbb{1}_v(o(e')) - \sum_{\gamma\in\Gamma\colon\, \gamma v = t(e')} \boldsymbol{x}^\gamma, \\
\partial^*_{\mathscr{F}}(v, e) &= \mathbb{1}_v(o(e')) - \sum_{\gamma\in\Gamma\colon\, \gamma v = t(e')} \boldsymbol{x}^{-\gamma}.
\end{aligned} \tag{4.11}$$

Remark 4.3.3. The size of the sets $\mathscr{F}_V, \mathscr{F}_{E^o}$ does not depend on $\mathscr{F}$; therefore, the size of the matrices $\partial_{\mathscr{F}}, \partial^*_{\mathscr{F}}$ also does not depend on the choice of $\mathscr{F}$. The set $\mathscr{F}_V$ contains exactly one element of each orbit of Γ acting on

V; the same holds for $\mathscr{F}_{E^o}$ with respect to E^0. One can check that the expression $\mathbb{1}_v(o(e')) - \sum_{\gamma\in\Gamma\colon \gamma v = t(e')} \boldsymbol{x}^\gamma$ depends only on the orbits containing v and e, but not explicitly on the choice of elements; therefore, the matrices $\partial_{\mathscr{F}}, \partial^*_{\mathscr{F}}$ depend on $\mathscr{F}$ only up to reordering of elements in $\mathscr{F}_V$ and $\mathscr{F}_{E^o}$. Therefore, hereafter we sometimes omit the index $\mathscr{F}$ and simply write ∂, ∂^*.

Definition 4.3.4. The graph Laplacian of $\mathscr{F}$ is a $|V_0| \times |V_0|$ matrix with elements in $\mathbb{Z}\Gamma$ given by

$$\Delta_{\mathscr{F}} = \partial^*_{\mathscr{F}} \partial_{\mathscr{F}}.$$

The entries of $\Delta_{\mathscr{F}}$ are given by

$$\begin{aligned}\Delta_{\mathscr{F}}(i,j) &= \deg v_i \mathbb{1}_{v_i}(v_j) \\ &\quad - \sum_{e\in\mathscr{F}_{E^o}, \gamma\in\Gamma} \boldsymbol{x}^\gamma \left(\mathbb{1}_{v_i}(o(e))\mathbb{1}_{\gamma v_j}(t(e)) + \mathbb{1}_{v_j}(o(e))\mathbb{1}_{-\gamma v_i}(t(e))\right) \\ &= \deg v_i \mathbb{1}_{v_i}(v_j) - \sum_{\gamma\in\Gamma\colon\, v_i\sim\gamma v_j} \boldsymbol{x}^\gamma.\end{aligned}$$

The choice of another fundamental domain $\mathscr{F}$ will correspond to the change of basis of matrix $\Delta_{\mathscr{F}}$ and, therefore, the determinant $\det \Delta_{\mathscr{F}}$ does not depend on the choice of $\mathscr{F}$.

4.3.1 Examples

In order to demonstrate the computation of the introduced notions, we will now consider a number of simple examples.

Example 4.3.5 ($\mathbb{Z}^2$-lattice, Figure 4.1)**.** For the lattice $\mathbb{Z}^2$ the quotient graph consists of one vertex and two loops corresponding to two basis vectors in $\mathbb{Z}^2$. Thus, $\mathscr{F}_V = (0,0)$ and $\mathscr{F}_{E^o} = \{e_1 = \{(0,0),(1,0)\}, e_2 = \{(0,0),(0,1)\}\}$.

Therefore,

$$\partial_{\mathscr{F}} = \begin{pmatrix} 1 - \boldsymbol{x}^{(1,0)} \\ 1 - \boldsymbol{x}^{(0,1)} \end{pmatrix}, \quad \partial^*_{\mathscr{F}} = \begin{pmatrix} 1 - \boldsymbol{x}^{-(1,0)}, & 1 - \boldsymbol{x}^{-(0,1)} \end{pmatrix}$$

If we take $x_1 = \boldsymbol{x}^{(1,0)}$ and $x_2 = \boldsymbol{x}^{(0,1)}$, then for any $\mathbf{n} = (n_1, n_2) \in \mathbb{Z}^2$, we can write $\boldsymbol{x}^{\mathbf{n}}$ as $x_1^{n_1} x_2^{n_2}$ (multi-index notation), and then the incidence matrices

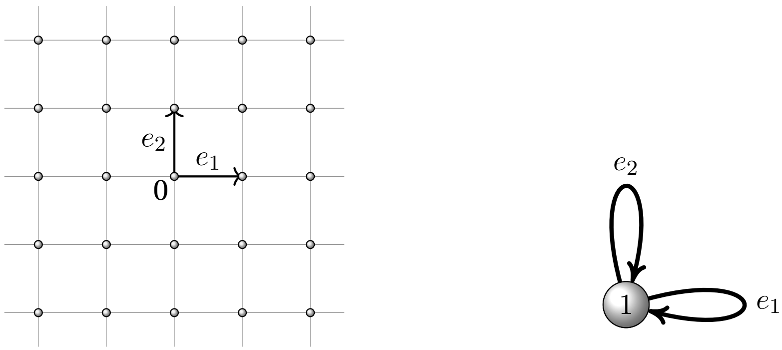


Figure 4.1: $\mathbb{Z}^2$-lattice and the corresponding quotient graph

can be rewritten as

$$\partial_{\mathscr{F}} = \begin{pmatrix} 1-x_1 \\ 1-x_2 \end{pmatrix}, \quad \partial^*_{\mathscr{F}} = \begin{pmatrix} 1-x_1^{-1} & 1-x_2^{-1} \end{pmatrix}.$$

Therefore, the Laplacian is

$$\Delta_{\mathscr{F}} = \partial^*_{\mathscr{F}}\partial_{\mathscr{F}} = \sum_{i=1}^{2}(1-x_i^{-1})(1-x_i) = 4-(x_1+x_1^{-1}+x_2+x_2^{-1}).$$

Example 4.3.6 (Ladder graph, Figure 4.2)**.** We consider the ladder graph $G=(V,E)$ where $V=\mathbb{Z}\times B$ with $B=\{1,2\}$, hence $V=\{(k,1),(k,2):k\in\mathbb{Z}\}$ and $E^o=\{e_{k,p},p=1,2,3,k\in\mathbb{Z}\}$. Here for $k\in\mathbb{Z}$, the edges are

$$e_{k,1}=\{(k,1),(k+1,1)\},\ e_{k,2}=\{(k,2),(k+1,2)\},\ e_{k,3}=\{(k,1),(k,2)\}.$$

Modulo the $\mathbb{Z}$-symmetry, one has $\mathscr{F}_V=\{1,2\}$ and $\mathscr{F}_{E^o}=\{e_1,e_2,e_3\}$.

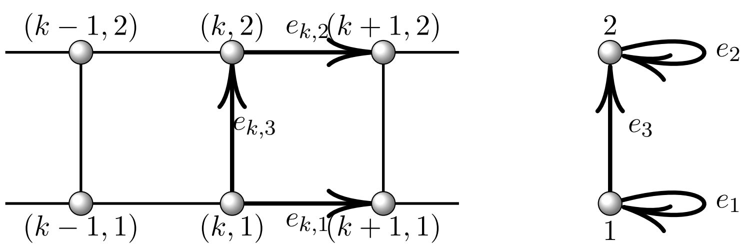


Figure 4.2: Ladder graph and its quotient.

Therefore,

$$\partial_{\mathscr{F}} = \begin{pmatrix} 1-x & 0 \\ 0 & 1-x \\ 1 & -1 \end{pmatrix},\ \partial^*_{\mathscr{F}} = \begin{pmatrix} 1-x^{-1} & 0 & 1 \\ 0 & 1-x^{-1} & -1 \end{pmatrix},$$

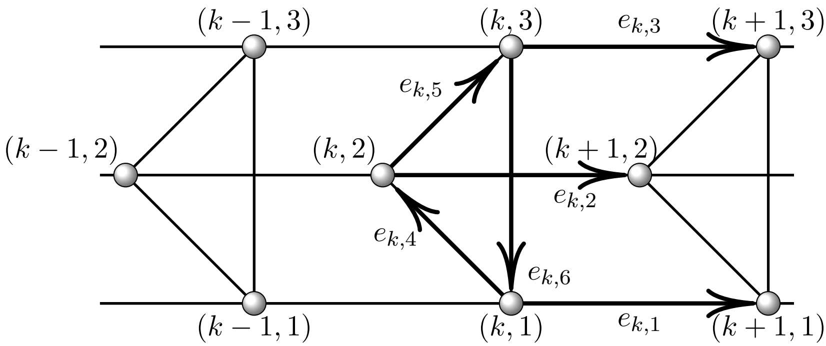


Figure 4.3: A three-ladder graph

$$\text{and } \Delta_{\mathscr{F}} = \begin{pmatrix} 3-x-x^{-1} & -1 \\ -1 & 3-x-x^{-1} \end{pmatrix}.$$

Example 4.3.7 (Three-ladder graph, Figure 4.3)**.** We consider the three-ladder graph $G=(V,E)$ (see Figure 4.3) where $V=\mathbb{Z}\times B$ with $B=\{1,2,3\}$ and $E^o=\{(e_{k,p}, p=\{1,2,3\}, k\in\mathbb{Z}\}$ where

$$e_{k,1}=\{(k,1),(k+1,1)\}, e_{k,2}=\{(k,2),(k+1,2)\}, e_{k,3}=\{(k,3),(k+1,3)\},$$

$$e_{k,4}=\{(k,1),(k,2)\}, e_{k,5}=\{(k,2),(k,3)\}, e_{k,6}=\{(k,3),(k,1)\}.$$

Let $\mathscr{F}_V^1 = \{(0,1),(0,2),(0,3)\}$ and $\mathscr{F}_{E^o}^1 = \{e_{0,1},e_{0,2},e_{0,3},e_{0,4},e_{0,5},e_{0,6}\}$. Then (see Figure 4.4)

$$\partial_{\mathscr{F}^1} = \begin{bmatrix} 1-x & 0 & 0 \\ 0 & 1-x & 0 \\ 0 & 0 & 1-x \\ 1 & -1 & 0 \\ 0 & 1 & -1 \\ -1 & 0 & 1 \end{bmatrix},$$

$$\partial^*_{\mathscr{F}^1} = \begin{bmatrix} 1-x^{-1} & 0 & 0 & 1 & 0 & -1 \\ 0 & 1-x^{-1} & 0 & -1 & 1 & 0 \\ 0 & 0 & 1-x^{-1} & 0 & -1 & 1 \end{bmatrix},$$

and

$$\Delta_{\mathscr{F}^1} = \begin{bmatrix} 4-x-x^{-1} & -1 & -1 \\ -1 & 4-x-x^{-1} & -1 \\ -1 & -1 & 4-x-x^{-1} \end{bmatrix}.$$

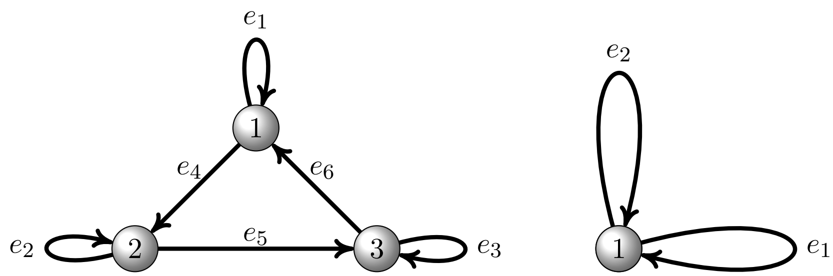


Figure 4.4: $\mathbb{Z}$ and $\mathbb{Z}\times\mathbb{Z}/3\mathbb{Z}$ quotients of the three-ladder graph

The three-ladder graph also admits an action of a group $\mathbb{Z}\times\mathbb{Z}/3\mathbb{Z}$: let $\mathscr{F}_V^2=\{(0,1)\}$ and $\mathscr{F}_{E^o}^2=\{e_{0,1},e_{0,4}\}$ (see Figure 4.4). Elements of $\mathbb{Z}\times\mathbb{Z}/3\mathbb{Z}$ can be written as (n,k), $n\in\mathbb{Z}$ and $k\in\mathbb{Z}/3\mathbb{Z}$.

Let $x_1=x^{(1,0)}$, $x_2=x^{(0,1)}$. Then $x^{(n,k)}=x_1^n x_2^k$, and

$$\partial_{\mathscr{F}^2}=\begin{bmatrix}1-x_1\\1-x_2\end{bmatrix},\quad \partial^*_{\mathscr{F}^2}=\begin{bmatrix}1-x_1^{-1} & 1-x_2^{-1}\end{bmatrix},\text{ and } \Delta_{\mathscr{F}}=4-x_1-x_1^{-1}-x_2-x_2^{-1}.$$

Example 4.3.8 (Triangular graph, Figure 4.5)**.** We consider the triangular graph $G=(V,E)$ where $V=\mathbb{Z}\times B$ with $B=\{1,2\}$ and $E^o=\{e_{k,p},p=1,2,3,k\in\mathbb{Z}\}$ where for $k\in\mathbb{Z}$,

$$e_{k,1}=\{(k,2),(k+1,2)\},\ e_{k,2}=\{(k,2),(k,1)\},\ e_{k,3}=\{(k,1),(k+1,2)\}.$$

Let $\mathscr{F}_V=\{(0,1),(0,2)\}$ and $\mathscr{F}_{E^o}=\{e_{0,1},e_{0,2},e_{0,3}\}$ (see Figure 4.5). Denote by x the horizontal translation by $(1,0)$. Then,

$$\partial_{\mathscr{F}}=\begin{pmatrix}1-x^{-1} & 1 & x^{-1}\\ 0 & -1 & -1\end{pmatrix},\quad \Delta_{\mathscr{F}}=\begin{pmatrix}4-x-x^{-1} & -1-x^{-1}\\ -1-x & 2\end{pmatrix}.$$

Example 4.3.9 (Kagome lattice, Figure 4.6)**.**

$$\partial_{\mathscr{F}}=\begin{pmatrix}1 & -1 & 0 & 1 & -y^{-1} & 0\\ -1 & 0 & -x & -x & 0 & -y^{-1}\\ 0 & 1 & 1 & 0 & 1 & 0\end{pmatrix},$$

$$\Delta_{\mathscr{F}}=\begin{pmatrix}4 & -1-x^{-1} & -1-y^{-1}\\ -1-x & 4 & -x-y^{-1}\\ -1-y & -x^{-1}-y & 4\end{pmatrix}$$

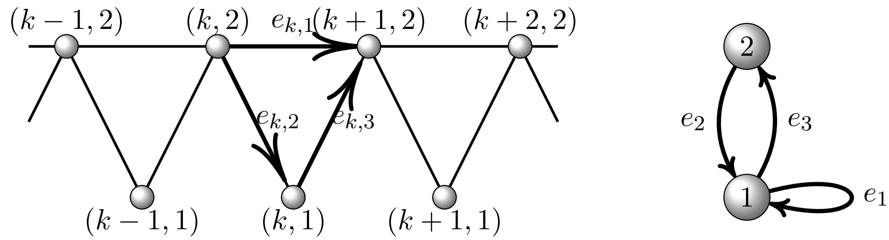


Figure 4.5: A triangular graph with $\mathbb{Z}$ action and the corresponding quotient graph

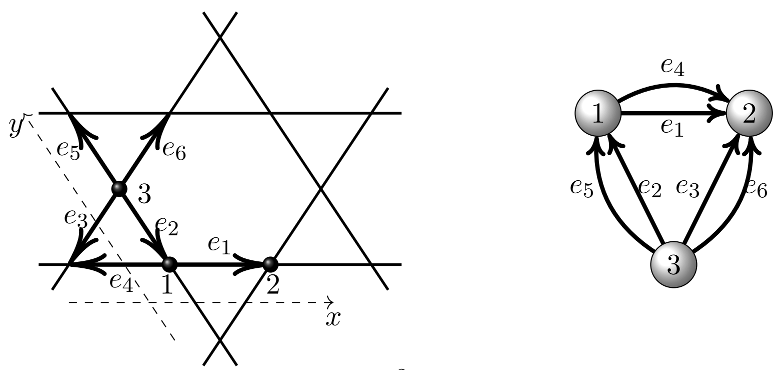


Figure 4.6: A Kagome lattice with $\mathbb{Z}^2$ action and the corresponding quotient graph

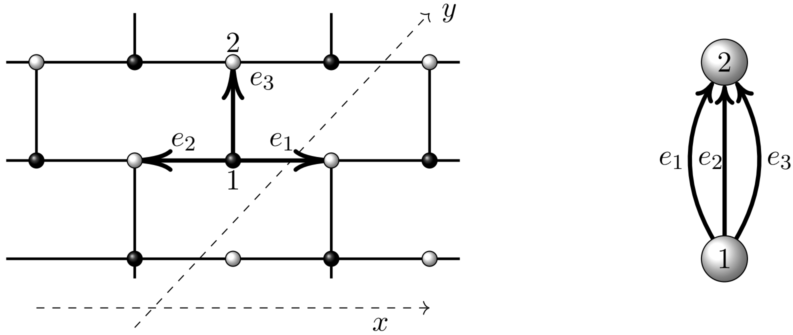


Figure 4.7: A hexagonal lattice with $\mathbb{Z}^2$ action and the corresponding quotient graph

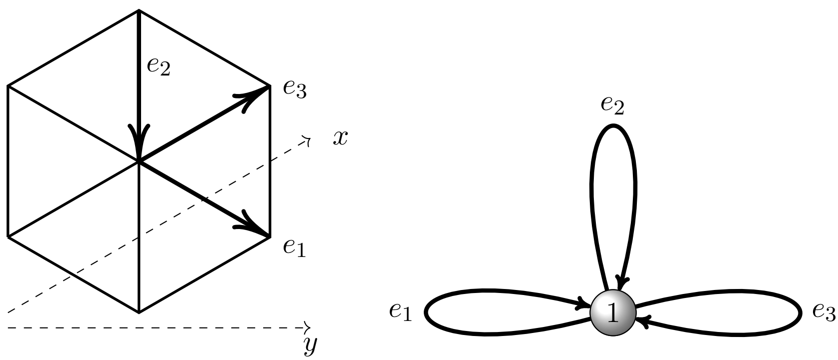


Figure 4.8: A triangular lattice with $\mathbb{Z}^2$ action and the corresponding quotient graph

Example 4.3.10 (Hexagonal lattice, Figure 4.7)**.**

$$\partial_{\mathscr{F}} = \begin{pmatrix} 1 & 1 & 1 \\ -x^{-1}y & -1 & -y \end{pmatrix}, \quad \Delta_{\mathscr{F}} = \begin{pmatrix} 3 & -1-y^{-1}-xy^{-1} \\ -1-y-x^{-1}y & 3 \end{pmatrix}$$

Example 4.3.11 (Triangular lattice, Figure 4.8)**.**

$$\partial_{\mathscr{F}} = \begin{pmatrix} 1 & 1 & 1 \\ -x^{-1}y & -1 & -y \end{pmatrix}, \quad \Delta_{\mathscr{F}} = \begin{pmatrix} 3 & -1-y^{-1}-xy^{-1} \\ -1-y-x^{-1}y & 3 \end{pmatrix}$$

4.4 Explicit expressions of correlation kernels

In this section we show how to explicitly compute the correlation kernel using Theorem 4.2.1. Our method applies to infinite connected graphs with a finitely generated abelian symmetry group Γ. The graphs are assumed to be locally finite, i.e., for any $v \in V$, $\deg v < \infty$.

4.4.1 Duality and Fourier transform

Suppose Γ is a countable finitely generated abelian group. Then Γ is a finite product of infinite and finite cyclic groups: for some $d \geqslant 1$, and $k_1, \ldots, k_m \in \mathbb{N}$, we have

$$\Gamma = \mathbb{Z}^d \times \mathbb{Z}_{k_1} \times \ldots \mathbb{Z}_{k_m}.$$

Let us denote by $\Gamma_0 = \mathbb{Z}_{k_1} \times \ldots \mathbb{Z}_{k_m}$ the finite torsion part of Γ (note that the representation of the torsion part is not unique). The elements of Γ will be written as $\boldsymbol{n} = (n, \ell)$, where $n = (n_1, \ldots, n_d) \in \mathbb{Z}^d$, $\ell = (\ell_1, \ldots, \ell_m) \in \Gamma_0$, with $0 \leqslant l_j < k_j$ for all $j = 1, \ldots, m$. The dual group $\widehat{\Gamma}$ is also easy to describe:

$$\widehat{\Gamma} = \mathbb{T}^d \times \mathbb{Z}_{k_1} \times \ldots \mathbb{Z}_{k_m},$$

where $\mathbb{T}^d = [0,1)^d$. We denote elements of $\widehat{\Gamma}$ by $\boldsymbol{\theta} = (\theta, \varphi)$, $\theta = (\theta_1, \ldots, \theta_d) \in \mathbb{T}^d$, and $\varphi = (\varphi_1, \ldots, \varphi_m)$ with $\varphi_j \in \mathbb{T}$ are such that $k_j \varphi_j = 0$. The corresponding character χ is then

$$\chi_{\boldsymbol{\theta}}(\boldsymbol{n}) = \chi_{(\theta,\varphi)}(n, \ell) = \exp\Big(2\pi i \big(\langle n, \theta\rangle + \langle \ell, \varphi\rangle\big)\Big),$$

$$\langle n, \theta\rangle = \sum_{t=1}^{d} n_t \theta_t, \quad \langle \ell, \varphi\rangle = \sum_{s=1}^{m} \ell_s \varphi_s.$$

We equip $\widehat{\Gamma}$ with the normalized Haar measure $\boldsymbol{\lambda} = \lambda \times \rho$ which is a product of the Lebesgue measure $\lambda = d\theta$ on $\mathbb{T}^d$ and the uniform probability measure ρ on $\widehat{\Gamma}_0 \cong \Gamma_0$.

Consider an unoriented graph $G = (V, E)$, and suppose Γ acts on G by graph automorphisms freely and such that the quotient $G/\Gamma = G_0 = (V_0, E_0)$ is finite. Then, any $v \in V$ can be presented as $v = \mathbf{n} \cdot v_0$ where $v_0 \in V_0$ and $\mathbf{n} \in \Gamma$. Fix an arbitrary orientation E_0^0 for edges in E_0, and let $\overline{E_0} = E_0^0 \sqcup -E_0^0$.

As earlier, consider the Hilbert space $\ell^2(V)$ of square-summable functions on the vertices V equipped with the scalar product

$$\langle f_1, f_2\rangle = \sum_{v \in V} f_1(v)\overline{f_2(v)} = \sum_{v_0 \in V_0} \sum_{\boldsymbol{n} \in \Gamma} f_1(\boldsymbol{n} v_0)\overline{f_2(\boldsymbol{n} v_0)},$$

and the Hilbert space $\ell^2_-(\overline{E})$ of square-summable antisymmetric functions on the oriented edges $\overline{E}$ equipped with a scalar product

$$\langle g_1, g_2\rangle = \frac{1}{2} \sum_{e \in \overline{E}} g_1(e)\overline{g_2(e)}.$$

Definition 4.4.1. A Fourier transform of $f \in \ell^2(V)$ is a function $\mathbb{F}f\colon \widehat{\Gamma} \to \mathbb{C}^{|V_0|}$ given by

$$\mathbb{F}f(\boldsymbol{\theta}, v_0) = \sum_{\boldsymbol{n}\in\Gamma} f(\boldsymbol{n}v_0)\chi_{\boldsymbol{n}}(\boldsymbol{\theta}) = \sum_{(n,l)\in\Gamma} f\big((n,l)v_0\big)e^{2\pi i(\langle n,\theta\rangle+\langle l,\varphi\rangle)}.$$

Note that for $f \in \ell^2(V)$ one has $\mathbb{F}f \in L^2(\widehat{\Gamma}, \mathbb{C}^{|V_0|})$. The space $L^2(\widehat{\Gamma}, \mathbb{C}^{|V_0|})$ is equipped with the inner product

$$\begin{aligned}\langle F_1, F_2\rangle &= \sum_{v_0\in V_0} \int_{\widehat{\Gamma}} F_1(\boldsymbol{\theta}, v_0)\overline{F_2(\boldsymbol{\theta}, v_0)}\boldsymbol{\lambda}(d\boldsymbol{\theta}) \\ &= \sum_{v_0\in V_0,\varphi\in\Gamma_0} \frac{1}{|\Gamma_0|}\int_{\mathbb{T}^d} F_1((\theta,\varphi), v_0)\overline{F_2((\theta,\varphi), v_0)}\lambda(d\theta),\end{aligned}$$

where $\boldsymbol{\lambda}(d\boldsymbol{\theta})$ is the normalized Haar measure on $\widehat{\Gamma}$ and $\lambda(d\theta)$ is the normalised Lebesgue measure on $\mathbb{T}^d$.

Definition 4.4.2. The inverse Fourier transform $\mathbb{F}^{-1}$ of $\widehat{f} \in L^2(\widehat{\Gamma}, \mathbb{C}^{|V_0|})$ is given by

$$\begin{aligned}\mathbb{F}^{-1}\widehat{f}(\boldsymbol{n}, v_0) &= \int_{\widehat{\Gamma}} \widehat{f}(\boldsymbol{\theta}, v_0)\chi_{\boldsymbol{n}}^{-1}(\boldsymbol{\theta})\boldsymbol{\lambda}(d\boldsymbol{\theta}) \\ &= \frac{1}{|\Gamma_0|}\sum_{\varphi\in\Gamma_0}\int_{\mathbb{T}^d} \widehat{f}\big((\theta,\varphi),\ v_0\big)e^{-2\pi i(\langle n,\theta\rangle+\langle l,\varphi\rangle)}\lambda(d\theta).\end{aligned}$$

Note that since for each $v \in V$, there are unique $v_0 \in V_0, \mathbf{n} \in \Gamma$ such that $v = \mathbf{n}v_0$, we can view $\mathbb{F}^{-1}\widehat{f}$ as a function on V:

$$\mathbb{F}^{-1}\widehat{f}(\mathbf{n}v_0) = \mathbb{F}^{-1}\widehat{f}(\mathbf{n}, v_0).$$

Suppose $f \in \mathbb{Z}\Gamma$, $f = \sum_{\mathbf{n}\in\Gamma} f_{\mathbf{n}} \cdot \boldsymbol{x}^{\mathbf{n}}$, we define the Fourier transform of the Laurent polynomial f as

$$\widehat{f}(\boldsymbol{\theta}) = \sum_{\mathbf{n}\in\Gamma} f_{\mathbf{n}}\chi_{\boldsymbol{\theta}}(\mathbf{n}),$$

and similarly for matrices: if $Q = (Q_{jk})$ with $Q_{jk} \in \mathbb{Z}\Gamma$, then

$$\widehat{Q}(\boldsymbol{\theta}) = (\widehat{Q}_{jk}(\boldsymbol{\theta})).$$

Theorem 4.4.3. Let $G = (V, E)$ be a locally finite graph, and Γ is a finitely generated abelian group acting freely on G by graph automorphisms, and such that $G_0 = (V_0, E_0) = G/\Gamma$ is a finite graph. Consider the following $|E_0| \times |E_0|$ matrix valued function on $\widehat{\Gamma}$

$$K(\boldsymbol{\theta}) = \lim_{\epsilon \to 0} D(\boldsymbol{\theta})\Big(D^*(\boldsymbol{\theta})D(\boldsymbol{\theta}) + \epsilon I\Big)^{-1} D^*(\boldsymbol{\theta}),$$

where $D(\boldsymbol{\theta}), D^*(\boldsymbol{\theta})$ are the Fourier transforms of the matrices ∂, ∂^*, respectively (as it was explained before, the dependence on the choice of $\mathscr{F}$ only influences the ordering of the basis vectors of these matrices).

The kernel $\mathbb{K}(e_1, e_2)$, $e_1, e_2 \in E$, of the determinantal USF measure $\mathbb{P}_W$ can be computed as follows. There exist unique $\mathbf{n}_i \in \Gamma$, $e_i^0 \in E_0$, such that $e_i = \mathbf{n}_i e_i^0$, $i = 1, 2$. Then

$$\mathbb{K}(e_1, e_2) = \int_{\widehat{\Gamma}} \chi_{\boldsymbol{\theta}}(\mathbf{n}_1 - \mathbf{n}_2) K(\boldsymbol{\theta})(e_1^0, e_2^0) \boldsymbol{\lambda}(d\boldsymbol{\theta}),$$

where $K(\boldsymbol{\theta})(e_1^0, e_2^0)$ is the (e_1^0, e_2^0)-element of $K(\boldsymbol{\theta})$.

Remark 4.4.4. Consider a family of operators

$$K_\epsilon(\boldsymbol{\theta}) = D(\boldsymbol{\theta})\Big(D^*(\boldsymbol{\theta})D(\boldsymbol{\theta}) + \epsilon I\Big)^{-1} D^*(\boldsymbol{\theta}).$$

By construction, $\lim_{\epsilon \to 0} K_\epsilon(\boldsymbol{\theta}) = K(\boldsymbol{\theta})$. Then, [4]

$$\int_{\widehat{\Gamma}} \langle K_\epsilon(\boldsymbol{\theta}) \mathbb{F}\mathbb{1}_{e_i}, \mathbb{F}\mathbb{1}_{e_j} \rangle d\boldsymbol{\lambda}(d\boldsymbol{\theta}) = H(e_i, e_j),$$

where $H(e_i, e_j)$ is a Green's function of a random walk on G conditioned to be killed at any $v \in V$ with probability ϵ. Moreover, the correlation kernel $\mathbb{K}_\epsilon = (e_1, e_2) = \int_{\widehat{\Gamma}} \chi_{\boldsymbol{\theta}}(\mathbf{n}_1 - \mathbf{n}_2) K_\epsilon(\boldsymbol{\theta})(e_1^0, e_2^0) \boldsymbol{\lambda}(d\boldsymbol{\theta})$ corresponds to a spanning forest determinantal point process on G with an added cemetery state [4].

Remark 4.4.5. By analogy with $d : \ell^2(V) \to \ell^2_-(E)$ and $d^* : \ell^2_-(E) \to \ell^2(V)$, it is natural to view $D(\boldsymbol{\theta}), D^*(\boldsymbol{\theta})$ as linear operators: $D(\boldsymbol{\theta}) : \mathbb{C}^{|V_0|} \to \mathbb{C}^{|E_0|}$ and $D^*(\boldsymbol{\theta}) : \mathbb{C}^{|E_0|} \to \mathbb{C}^{|V_0|}$.

Before we turn to the proof of Theorem 4.4.3, let us give the two examples demonstrating the application of the theorem.

Example 4.4.6 (Square lattice)**.** Consider the lattice $\mathbb{Z}^d$ with the natural shift action of $\Gamma = \mathbb{Z}^d$. As in Example 4.3.5, one readily checks that the quotient graph G/Γ is a d-bouquet graph consisting of a single vertex and d loops. Since $\mathbb{T}^d = \widehat{\mathbb{Z}^d}$, we thus have

$$\partial = \begin{pmatrix} 1 - x_1 \\ \vdots \\ 1 - x_d \end{pmatrix}, \; D(\boldsymbol{\theta}) = \begin{pmatrix} 1 - e^{2\pi i \theta_1} \\ \vdots \\ 1 - e^{2\pi i \theta_d} \end{pmatrix},$$

and

$$\partial^* = \left(1 - x_1^{-1}, 1 - x_2^{-1}, \dots, 1 - x_d^{-1}\right), \; D^*(\boldsymbol{\theta}) = (1 - e^{-2\pi i\theta_1}, \dots, 1 - e^{-2\pi i\theta_d}),$$

where $\boldsymbol{\theta} = (\theta_1, \dots, \theta_d) \in \mathbb{T}^d$. Therefore $D(\boldsymbol{\theta})^* D(\boldsymbol{\theta}) = 2d - 2\sum_{m=1}^d \cos(2\pi\theta_m)$, and hence

$$\begin{aligned} K_\epsilon(\boldsymbol{\theta}) &= D(\boldsymbol{\theta})(D(\boldsymbol{\theta})^* D(\boldsymbol{\theta}) + \epsilon I)^{-1} D(\boldsymbol{\theta}) \\ &= \left(\frac{(1 - e^{2\pi i\theta_j})(1 - e^{-2\pi i\theta_k})}{2d + \epsilon - 2\sum_{m=1}^d \cos(2\pi\theta_m)} \right)_{j,k=1}^d, \end{aligned}$$

and

$$K(\boldsymbol{\theta}) = \lim_{\epsilon \to 0} K_\epsilon(\boldsymbol{\theta}) = \left(\frac{(1 - e^{2\pi i\theta_j})(1 - e^{-2\pi i\theta_k})}{2d - 2\sum_{m=1}^d \cos(2\pi\theta_m)} \right)_{j,k=1}^d.$$

Consider now two edges $e_1 = (\mathbf{n}_1, \mathbf{n}_1 + q_j)$ and $e_2 = (\mathbf{n}_2, \mathbf{n}_2 + q_k)$, where $\mathbf{n}_1, \mathbf{n}_2 \in \mathbb{Z}^d$ and q_j, q_k are the j-the and k-the basis vectors of $\mathbb{Z}^d$, respectively. Then

$$\mathbb{K}(e_1, e_2) = \int_{\mathbb{T}^d} e^{2\pi i\langle \mathbf{n}_1 - \mathbf{n}_2, \boldsymbol{\theta}\rangle} \frac{(1 - e^{2\pi i\theta_j})(1 - e^{-2\pi i\theta_k})}{2d - 2\sum_{m=1}^d \cos(2\pi\theta_m)} \, d\boldsymbol{\theta} =$$

$$\int_{\mathbb{T}^d} \frac{e^{2\pi i\langle \mathbf{n}_1 - \mathbf{n}_2, \boldsymbol{\theta}\rangle} - e^{2\pi i\langle \mathbf{n}_1 + q_j - \mathbf{n}_2, \boldsymbol{\theta}\rangle} - e^{2\pi i\langle \mathbf{n}_1 - (\mathbf{n}_2 + q_k), \boldsymbol{\theta}\rangle} + e^{2\pi i\langle \mathbf{n}_1 + q_j - (\mathbf{n}_2 + q_k), \boldsymbol{\theta}\rangle}}{2d - 2\sum_{m=1}^d \cos(2\pi\theta_m)} \, d\boldsymbol{\theta}$$

If we let $x = \mathbf{n}_1$, $y = \mathbf{n}_1 + q_j$, $z = \mathbf{n}_2$, $w = \mathbf{n}_2 + q_k$, then we immediately see that we have obtained exactly the same expression as in the Example 4.1.3:

$$\mathbb{K}(e_1, e_2) = \frac{1}{2d}\Big[g(z - x) - g(z - y) - g(w - x) + g(w - y) \Big].$$

Note however, that the method does not have to take into account that for $d = 2$ the simple random walk is recurrent, and hence the Green's function has to be redefined appropriately.

Example 4.4.7 (Three-ladder graph, $\mathbb{Z}$-symmetry, Figure 4.4)**.** Consider an unoriented graph G with a vertex set $V = \{(k, i), k \in \mathbb{Z}, i \in \{1, 2, 3\}\}$ and the set of edges

$$E = \{((k, i), (k \pm 1, i))\} \cup \{((k, 1), (k, 2)), ((k, 2), (k, 3)), ((k, 3), (k, 1))\}.$$

The graph G is called a three-ladder graph (see Figure 4.3).

Define an automorphism $x\colon V \to V$ acting by $x(k, i) = (k + 1, i)$. Clearly, the automorphism x implies the action of $\mathbb{Z}$ on G: consider a quotient graph

$G/\mathbb{Z}$ (see Figure 4.4). We write down the corresponding $D(\theta)$ and $D^*(\theta)$ matrices:

$$D(\theta) = \begin{bmatrix} 1-e^{2\pi i\theta} & 0 & 0 \\ 0 & 1-e^{2\pi i\theta} & 0 \\ 0 & 0 & 1-e^{2\pi i\theta} \\ 1 & -1 & 0 \\ 0 & 1 & -1 \\ -1 & 0 & 1 \end{bmatrix},$$

$$D^*(\theta) = \begin{bmatrix} 1-e^{-2\pi i\theta} & 0 & 0 & 1 & 0 & -1 \\ 0 & 1-e^{-2\pi i\theta} & 0 & -1 & 1 & 0 \\ 0 & 0 & 1-e^{--2\pi i\theta} & 0 & -1 & 1 \end{bmatrix}$$

We conclude that

$$D^*(\theta)D(\theta) = \begin{bmatrix} 4-e^{2\pi i\theta}-e^{-2\pi i\theta} & -1 & -1 \\ -1 & 4-e^{2\pi i\theta}-e^{-2\pi i\theta} & -1 \\ -1 & -1 & 4-e^{2\pi i\theta}-e^{-2\pi i\theta} \end{bmatrix}$$

and since $(D^*(\theta)D(\theta))$ is formally invertible,

$$\frac{(e^{2\pi i\theta}-1)^2(e^{4\pi i\theta}-5e^{2\pi i\theta}+1)}{e^{2\pi i\theta}}(D^*(\theta)D(\theta))^{-1} =$$

$$\begin{bmatrix} -e^{4\pi i\theta}+3e^{2\pi i\theta}-1 & e^{2\pi i\theta} & e^{2\pi i\theta} \\ e^{2\pi i\theta} & -e^{4\pi i\theta}+3e^{2\pi i\theta}-1 & e^{2\pi i\theta} \\ e^{2\pi i\theta} & e^{2\pi i\theta} & -e^{4\pi i\theta}+3e^{2\pi i\theta}-1 \end{bmatrix}$$

we do not need to take the limit with respect to ϵ. We finish the calculation by multiplying $K(\theta) = D(\theta)(D^*(\theta)D(\theta))^{-1}D^*(\theta)$:

$$K(\theta) = \frac{1}{5-2\cos(\theta)}\cdot$$

$$\begin{bmatrix} 3-2\cos(2\pi\theta) & 1 & 1 & e^{2\pi i\theta}-1 & 0 & 1-e^{2\pi i\theta} \\ 1 & 3-2\cos(2\pi\theta) & 1 & 1-e^{2\pi i\theta} & e^{2\pi i\theta}-1 & 0 \\ 1 & 1 & 3-2\cos(2\pi\theta) & 0 & 1-e^{2\pi i\theta} & e^{2\pi i\theta}-1 \\ 1-e^{-2\pi i\theta} & e^{-2\pi i\theta}-1 & 0 & 2 & -1 & -1 \\ 0 & 1-e^{-2\pi i\theta} & e^{-2\pi i\theta}-1 & -1 & 2 & -1 \\ e^{-2\pi i\theta}-1 & 0 & 1-e^{-2\pi i\theta} & -1 & -1 & 2 \end{bmatrix}$$

Take $k_1, k_2 \in \Gamma = \mathbb{Z}$. The correlation kernel of two edges $e_{k_1,i}, e_{k_2,j}$ is given by the i,j-th entry of the matrix $\mathbb{K}(k_1, k_1)$ given by the formula

$$\mathbb{K}(k_1, k_2) = \int_0^1 K(\theta)e^{2\pi i(k_1-k_2)\theta}d\theta.$$

Example 4.4.8 (Three-ladder graph, $\mathbb{Z}\times\mathbb{Z}/3\mathbb{Z}$-symmetry, Figure 4.4)**.** Define an automorphism $y\colon V\to V$ acting by the formula

$$y(k,i)=\begin{cases}(k,i+1), i=1,2\\(k,1), i=3\end{cases}$$

Clearly, y is a cyclic permutation on a set of three elements and it is independent of x. Therefore, x and y generate a $\mathbb{Z}\times\mathbb{Z}/3\mathbb{Z}$-action on G. The quotient graph is shown on Figure 4.4. The corresponding $D(\boldsymbol{\theta})$ and $D^*(\boldsymbol{\theta})$ are:

$$D(\boldsymbol{\theta})=\begin{bmatrix}1-e^{2\pi i\theta_1}\\1-e^{2\pi i\theta_2}\end{bmatrix},\quad D^*(\boldsymbol{\theta})=\begin{bmatrix}1-e^{-2\pi i\theta_1} & 1-e^{-2\pi i\theta_2}\end{bmatrix}$$

Then, $(D^*(\boldsymbol{\theta})D(\boldsymbol{\theta}))^{-1}=\frac{1}{4-2\cos(2\pi\theta_1)-2\cos(2\pi\theta_2)}$ and $K(\boldsymbol{\theta})$ is given by

$$\frac{1}{4-2\cos(2\pi\theta_1)-2\cos(2\pi\theta_2)}\begin{bmatrix}2-2\cos(2\pi\theta_1) & (1-e^{2\pi i\theta_1})(1-e^{-2\pi i\theta_2})\\(1-e^{-2\pi i\theta_1})(1-e^{2\pi i\theta_2}) & 2-2\cos(2\pi\theta_2)\end{bmatrix}$$

Take two edges $e_1,e_2\in E$ such that $e_1=\boldsymbol{n}e_i^0$ and $e_2=\boldsymbol{m}e_j^0$, $\boldsymbol{n},\boldsymbol{m}\in\mathbb{Z}\times\mathbb{Z}/3\mathbb{Z}$ and $e_i^0,e_j^0\in E_0$, then their correlation is given by the (i,j)-th element of the matrix

$$\begin{aligned}\mathbb{K}(\boldsymbol{n},\boldsymbol{m})=&\frac{1}{3}\int_0^1 K(\theta,0)e^{2\pi i(n_1-m_1)\theta}d\theta+\frac{1}{3}\int_0^1 K(\theta,1/3)e^{2\pi i(n_1-m_1)\theta}e^{2/3\pi i(n_2-m_2)}d\theta\\&+\frac{1}{3}\int_0^1 K(\theta,-1/3)e^{2\pi i(n_1-m_1)\theta}e^{-2/3\pi i(n_2-m_2)}d\theta\end{aligned}$$

Proof of Theorem 4.4.3. Applying Theorem 4.2.1 to the coboundary operator $d:\ell^2(V)\to\ell^2_-(\overline{E})$ and its adjoint – the divergence operator – $d^*:\ell^2_-(\overline{E})\to\ell^2(V)$ (c.f., (4.6)):

$$df(e)=f(o(e))-f(t(e)),\quad d^*\theta(v)=\sum_{o(e)=v}\theta(e),$$

we conclude that the operators $P_\epsilon=d(d^*d+\epsilon I)^{-1}d^*$ converge, as $\epsilon\to 0$, to the orthogonal projection operator P onto the closure of $\operatorname{Im} d$. For two edges $e_1,e_2\in\overline{E}$

$$\begin{aligned}\mathbb{K}(e_1,e_2)&=\lim_{\epsilon\to0}\left\langle d(d^*d+\epsilon I)^{-1}d^*\mathbb{1}_{e_1},\mathbb{1}_{e_2}\right\rangle_{\ell^2_-(\overline{E})}\\&=\lim_{\epsilon\to0}\left\langle (d^*d+\epsilon I)^{-1}d^*\mathbb{1}_{e_1},d^*\mathbb{1}_{e_2}\right\rangle_{\ell^2(V)}\\&=\lim_{\epsilon\to0}\left\langle \mathbb{F}[(d^*d+\epsilon I)^{-1}d^*\mathbb{1}_{e_1}],\mathbb{F}[d^*\mathbb{1}_{e_2}]\right\rangle_{L^2(\widehat{\Gamma},\mathbb{C}^{|V_0|})}.\end{aligned}\tag{4.12}$$

Now we show how to compute these expressions in the Fourier domain. In order to do it, we use several Lemma's whose proofs we provide in the end of this Subsection.

Lemma 4.4.9. For an edge $e \in E^0$, $\mathbb{F}(d^*\mathbb{1}_e)(\boldsymbol{\theta}, \cdot)$ is a vector of length $|V_0|$ with entries given by

$$\mathbb{F}(d^*\mathbb{1}_e)(\boldsymbol{\theta}, v_0) = \sum_{\mathbf{n}\in\Gamma:\mathbf{n}v_0=o(e)} \chi_{\mathbf{n}}(\boldsymbol{\theta}) - \sum_{\mathbf{n}\in\Gamma:\mathbf{n}v_0=t(e)} \chi_{\mathbf{n}}(\boldsymbol{\theta}).$$

Lemma 4.4.10. For every $f \in \ell^2(V)$ one has

$$\mathbb{F}[(d^*d + \epsilon I)^{-1}f] = (M(\boldsymbol{\theta}) + \epsilon I)^{-1}\mathbb{F}f,$$

where for each $\boldsymbol{\theta} \in \widehat{\Gamma}$, $M(\boldsymbol{\theta})$ is a $|V_0| \times |V_0|$ matrix with entries

$$M(\boldsymbol{\theta})_{v_j,v_k} = \begin{cases} \deg v_j, & v_j = v_k, \\ -\sum\limits_{\in\Gamma:\boldsymbol{v}_k\sim v_j} \chi_{\boldsymbol{\theta}}(-), & v_j \neq v_k. \end{cases}$$

We introduced the $|E_0|\times|V_0|$-incidence matrix $\partial_{\mathscr{F}} = (\partial_{\mathscr{F}_{E^o}}(e,v))_{e\in\mathscr{F}_{E^o},v\in\mathscr{F}_V}$ and its adjoint $\partial^*_{\mathscr{F}} = (\partial^*_{\mathscr{F}}(e,v))_{e\in\mathscr{F}_{E^o},v\in\mathscr{F}_V}$ as

$$\partial_{\mathscr{F}}(e,v) = \mathbb{1}_v(o(e)) - \sum_{\mathbf{n}\in\Gamma:\ \mathbf{n}v=t(e)} \boldsymbol{x}^{\mathbf{n}}, \tag{4.13}$$

$$\partial^*_{\mathscr{F}}(v,e) = \mathbb{1}_v(o(e)) - \sum_{\mathbf{n}\in\Gamma:\ \mathbf{n}v=t(e)} \boldsymbol{x}^{-\mathbf{n}}. \tag{4.14}$$

From now on we will drop the index $\mathscr{F}$. Let $D(\boldsymbol{\theta}) = \widehat{\partial}$ and $D^*(\boldsymbol{\theta}) = \widehat{\partial^*}$ be the corresponding Fourier transforms of matrices ∂, ∂^*, i.e., for $v \in \mathscr{F}_V, e \in \mathscr{F}_{E^o}$,

$$D(\boldsymbol{\theta})(e,v) = \mathbb{1}_v(o(e)) - \sum_{\mathbf{n}\in\Gamma:\ \mathbf{n}v=t(e)} \chi_{\boldsymbol{\theta}}(\mathbf{n}),$$

$$D^*(\boldsymbol{\theta})(v,e) = \mathbb{1}_v(o(e)) - \sum_{\mathbf{n}\in\Gamma:\ \mathbf{n}v=t(e)} \chi_{\boldsymbol{\theta}}(-\mathbf{n}).$$

Then $D^*(\boldsymbol{\theta}) = (D(\boldsymbol{\theta}))^*$. Indeed,

$$\begin{aligned} \widehat{\partial(e,v)} &= \mathbb{1}_v(o(e)) - \sum_{\mathbf{n}\in\Gamma:\ \mathbf{n}v=t(e)} \chi_{\boldsymbol{\theta}}(\mathbf{n}) = \mathbb{1}_v(o(e)) - \sum_{\mathbf{n}\in\Gamma:\ \mathbf{n}v=t(e)} \overline{\chi_{\boldsymbol{\theta}}(-\mathbf{n})} \\ &= \overline{\mathbb{1}_v(o(e)) - \sum_{\mathbf{n}\in\Gamma:\ \mathbf{n}v=t(e)} \chi_{\boldsymbol{\theta}}(-\mathbf{n})} = \overline{\widehat{\partial}^*(v,e)}. \end{aligned}$$

Furthermore, suppose $v_0 \in \mathscr{F}_V$ and $e \in E^o$. Then there exist unique $e_0 \in \mathscr{F}_{E^o}$ and $\mathbf{n}_0 \in \Gamma$ such that $e = \mathbf{n}_0(e_0)$. Therefore,

$$\begin{aligned}
\mathbb{F}(d^*\mathbb{1}_e)(\boldsymbol{\theta}, v_0) &= \sum_{\mathbf{n}\in\Gamma:\mathbf{n}v_0=o(e)} \chi_{\boldsymbol{\theta}}(\mathbf{n}) - \sum_{\mathbf{n}\in\Gamma:\mathbf{n}v_0=t(e)} \chi_{\boldsymbol{\theta}}(\mathbf{n}) \\
&= \sum_{\mathbf{n}\in\Gamma:\mathbf{n}v_0=o(\mathbf{n}_0(e_0))} \chi_{\boldsymbol{\theta}}(\mathbf{n}) - \sum_{\mathbf{n}\in\Gamma:\mathbf{n}v_0=t(\mathbf{n}_0(e_0))} \chi_{\boldsymbol{\theta}}(\mathbf{n}) \\
&= \chi_{\boldsymbol{\theta}}(\mathbf{n}_0)\Big(\mathbb{1}_{v_0}(o(e_0)) - \sum_{\in\Gamma:\boldsymbol{v}_0=t(e_0)} \mathbb{1}_{\boldsymbol{\theta}}()\Big) = \chi_{\boldsymbol{\theta}}(\mathbf{n}_0)D(\boldsymbol{\theta})(e_0, v_0).
\end{aligned} \tag{4.15}$$

For the $|V_0| \times |V_0|$ matrix $M(\boldsymbol{\theta})$ of Lemma 4.4.10 we have

Lemma 4.4.11.

$$M(\boldsymbol{\theta}) = D^*(\boldsymbol{\theta})D(\boldsymbol{\theta}). \tag{4.16}$$

Now we are ready to derive the expression for the kernel K. Suppose $e_1, e_2 \in E^0$, and hence there are unique $\mathbf{n}_j \in \Gamma$, $e_j^0 \in \mathscr{F}_{E^o}$, $j = 1, 2$, such that $e_j = \mathbf{n}_j e_j^0$. Then using the auxiliary Lemmas, and the corresponding expressions (4.15), (4.16), we can continue (4.12) as follows: $K(e_1, e_2) =$

$$\begin{aligned}
&\lim_{\epsilon\to 0}\Big\langle \mathbb{F}[(d^*d+\epsilon I)^{-1}d^*\mathbb{1}_{e_1}], \mathbb{F}[d^*\mathbb{1}_{e_2}]\Big\rangle_{L^2(\widehat{\Gamma},\mathbb{C}^{|V_0|})} \\
&= \lim_{\epsilon\to 0}\Big\langle (M+\epsilon I)^{-1}\mathbb{F}[d^*\mathbb{1}_{e_1}], \mathbb{F}[d^*\mathbb{1}_{e_2}]\Big\rangle_{L^2(\widehat{\Gamma},\mathbb{C}^{|V_0|})} \\
&= \lim_{\epsilon\to 0}\sum_{v_0\in\mathscr{F}_V}\int_{\widehat{\Gamma}}\Big[\sum_{v_1\in\mathscr{F}_V}(M(\boldsymbol{\theta})+\epsilon I)^{-1}_{v_0,v_1}\mathbb{F}d^*\mathbb{1}_{e_1}(\boldsymbol{\theta}, v_1)\Big]\overline{\Big[\mathbb{F}d^*\mathbb{1}_{e_2}(\boldsymbol{\theta}, v_0)\Big]}\boldsymbol{\lambda}(d\boldsymbol{\theta}) \\
&= \lim_{\epsilon\to 0}\int_{\widehat{\Gamma}}\sum_{v_0\in\mathscr{F}_V}\left[\sum_{v_1\in\mathscr{F}_V}(M(\boldsymbol{\theta})+\epsilon I)^{-1}_{v_0,v_1}\chi_{\boldsymbol{\theta}}(\mathbf{n}_1)D(\boldsymbol{\theta})(e_1^0, v_1)\right]\cdot \\
&\quad\cdot\overline{\big[\chi_{\boldsymbol{\theta}}(\mathbf{n}_2)D(\boldsymbol{\theta})(e_2^0, v_0)\big]}\boldsymbol{\lambda}(d\boldsymbol{\theta}) \\
&= \lim_{\epsilon\to 0}\int_{\widehat{\Gamma}}\chi_{\boldsymbol{\theta}}(\mathbf{n}_1-\mathbf{n}_2)\cdot \\
&\quad\cdot\left[\sum_{v_1\in\mathscr{F}_V}\sum_{v_0\in\mathscr{F}_V}D(\boldsymbol{\theta})(e_1^0, v_1)(D^*(\boldsymbol{\theta})D(\boldsymbol{\theta})+\epsilon I)^{-1}_{v_0,v_1}D^*(\boldsymbol{\theta})(v_0, e_2^0)\right]\boldsymbol{\lambda}(d\boldsymbol{\theta}).
\end{aligned}$$

It is now easy to see that the expression in the square brackets is nothing else but the (e_1^0, e_2^0)-entry of the matrix $\mathbb{K}_\epsilon = D(\boldsymbol{\theta})(D^*(\boldsymbol{\theta})D(\boldsymbol{\theta})+\epsilon I)^{-1}D^*(\boldsymbol{\theta})$. Applying Theorem 4.2.1 again, but now in the finite dimensional situation (c.f., (4.7)), we conclude that the limit

$$K(\boldsymbol{\theta}) = \lim_{\epsilon\to 0}D^*(\boldsymbol{\theta})\Big(D^*(\boldsymbol{\theta})D(\boldsymbol{\theta})+\epsilon I\Big)^{-1}D(\boldsymbol{\theta})$$

exists, and in fact, coincides with $D(\boldsymbol{\theta})(D(\boldsymbol{\theta})^*D(\boldsymbol{\theta}))^{-1}D^*(\boldsymbol{\theta})$ since the matrix $D^*(\boldsymbol{\theta})D(\boldsymbol{\theta})$ is invertible on $W = \mathsf{Im}\ D \subset \mathbb{C}^{|E_0|}$. Moreover, by Proposition 4.2.2, for each $\boldsymbol{\theta}$, the norm of $(D^*(\boldsymbol{\theta})D(\boldsymbol{\theta})+\epsilon I)^{-1}$ is bounded by 2, uniformly in ϵ. Hence, by the Lebesgue dominated convergence theorem

$$\begin{aligned}\mathbb{K}(e_1,e_2) &= \lim_{\epsilon\to 0}\int_{\widehat{\Gamma}}\chi_{\boldsymbol{\theta}}(\mathbf{n}_1-\mathbf{n}_2)\mathbb{K}_\epsilon(\boldsymbol{\theta})(e_1^0,e_2^0)\boldsymbol{\lambda}(d\boldsymbol{\theta})\\ &= \int_{\widehat{\Gamma}}\chi_{\boldsymbol{\theta}}(\mathbf{n}_1-\mathbf{n}_2)\Big[\lim_{\epsilon\to 0}\mathbb{K}_\epsilon(\boldsymbol{\theta})(e_1^0,e_2^0)\Big]\boldsymbol{\lambda}(d\boldsymbol{\theta})\\ &= \int_{\widehat{\Gamma}}\chi_{\boldsymbol{\theta}}(\mathbf{n}_1-\mathbf{n}_2)\mathbb{K}(\boldsymbol{\theta})(e_1^0,e_2^0)\boldsymbol{\lambda}(d\boldsymbol{\theta})\end{aligned}$$

This finishes the proof of the Theorem. □

Proof of Lemma 4.4.9. One has

$$d^*\mathbb{1}_e(v) = \begin{cases}1, & v=o(e),\\ -1, & v=t(e),\\ 0, & v\neq o(e),t(e).\end{cases}$$

Therefore,

$$\mathbb{F}(d^*\mathbb{1}_e)(\boldsymbol{\theta},v_0) = \sum_{\mathbf{n}\in\Gamma}(d^*\mathbb{1}_e)(\mathbf{n}v_0)\chi_{\boldsymbol{\theta}}(\mathbf{n}) = \sum_{\mathbf{n}\in\Gamma:\mathbf{n}v_0=o(e)}\chi_{\boldsymbol{\theta}}(\mathbf{n}) - \sum_{\mathbf{n}\in\Gamma:\mathbf{n}v_0=t(e)}\chi_{\boldsymbol{\theta}}(\mathbf{n}).$$

□

Proof of Lemma 4.4.10. The Laplacian $\mathcal{L}=d^*d\colon \ell^2(V)\to\ell^2(V)$ is given by

$$(\mathcal{L}f)(v) = \sum_{y\sim v}\Big(f(v)-f(y)\Big) = \deg v\cdot f(v) - \sum_{y\sim v}f(y),$$

where the sum is taken over all vertices $y\in V$ that are connected to v by an edge. Let us now identify the operator $\widehat{\mathcal{L}}\colon L^2(\widehat{\Gamma},\mathbb{C}^{|V_0|}) \to L^2(\widehat{\Gamma},\mathbb{C}^{|V_0|})$ such that for all $f\in\ell^2(V)$, one has

$$\mathbb{F}(\mathcal{L}f) = \widehat{\mathcal{L}}(\mathbb{F}f).$$

A simple calculation shows that $\mathbb{F}(\mathcal{L}f)(\boldsymbol{\theta}, v_0) =$

$$\begin{aligned}
&= \sum_{\boldsymbol{n}\in\Gamma} (\mathcal{L}f)(\boldsymbol{n}v_0)\chi_{\boldsymbol{\theta}}(\boldsymbol{n}) = \sum_{\boldsymbol{n}\in\Gamma} \Big[\deg(\boldsymbol{n}v_0)\cdot f(\boldsymbol{n}v_0) - \sum_{y\sim \boldsymbol{n}v_0} f(y)\Big]\chi_{\boldsymbol{\theta}}(\boldsymbol{n}) \\
&= \deg(v_0)\mathbb{F}f(\boldsymbol{\theta}, v_0) - \sum_{\boldsymbol{n}\in\Gamma} \chi_{\boldsymbol{\theta}}(\boldsymbol{n}) \sum_{\substack{\mathbf{n}'\in\Gamma, v'\in\mathscr{F}_V \\ \mathbf{n}'v'\sim\mathbf{n}v_0}} f(\boldsymbol{n}'v') \\
&= \deg v_0 \mathbb{F}f(\boldsymbol{\theta}, v_0) - \sum_{\boldsymbol{n}\in\Gamma} \chi_{\boldsymbol{\theta}}(\boldsymbol{n}')\chi_{\boldsymbol{\theta}}(\mathbf{n}-\mathbf{n}') \sum_{\substack{\mathbf{n}'\in\Gamma, v'\in\mathscr{F}_V \\ (\mathbf{n}'-\mathbf{n})v'\sim v_0}} f(\boldsymbol{n}'v') \\
&= \deg(v_0)\mathbb{F}f(\boldsymbol{\theta}, v_0) - \sum_{v'\in\mathscr{F}_V}\sum_{\mathbf{n}'\in\Gamma} \chi_{\boldsymbol{\theta}}(\boldsymbol{n}')f(\boldsymbol{n}'v')\Big(\sum_{\mathbf{n}\in\Gamma:\mathbf{n}'v'\sim\mathbf{n}v_0} \chi_{\boldsymbol{\theta}}(\mathbf{n}-\mathbf{n}')\Big) \\
&= \deg(v_0)\mathbb{F}f(\boldsymbol{\theta}, v_0) - \sum_{v'\in\mathscr{F}_V}\sum_{\mathbf{n}'\in\Gamma} \chi_{\boldsymbol{\theta}}(\boldsymbol{n}')f(\boldsymbol{n}'v')\Big(\sum_{\mathbf{n}\in\Gamma:(\mathbf{n}'-\mathbf{n})v'\sim v_0} \chi_{\boldsymbol{\theta}}(\mathbf{n}-\mathbf{n}')\Big) \\
&= \deg(v_0)\mathbb{F}f(\boldsymbol{\theta}, v_0) - \sum_{v'\in\mathscr{F}_V}\sum_{\mathbf{n}'\in\Gamma} \chi_{\boldsymbol{\theta}}(\boldsymbol{n}')f(\boldsymbol{n}'v')\Big(\sum_{\in\Gamma:\boldsymbol{v}'\sim v_0} \chi_{\boldsymbol{\theta}}(-)\Big) \\
&= \deg(v_0)\mathbb{F}f(\boldsymbol{\theta}, v_0) - \sum_{v'\in\mathscr{F}_V} \mathbb{F}f(\boldsymbol{\theta}, v')\Big(\sum_{\in\Gamma:\boldsymbol{v}'\sim v_0} \chi_{\boldsymbol{\theta}}(-)\Big).
\end{aligned}$$

Hence, for every $\boldsymbol{\theta} \in \widehat{\Gamma}$, one has

$$\mathbb{F}(\mathcal{L} + \epsilon I)f = (M(\boldsymbol{\theta}) + \epsilon I)\mathbb{F}f,$$

where for $V = \{v_1, \dots, v_m\}$,

$$M(\boldsymbol{\theta})_{v_j,v_k} = \begin{cases} \deg v_j, & v_j = v_k, \\ -\sum_{\in\Gamma:\boldsymbol{v}_k\sim v_j} \chi_{\boldsymbol{\theta}}(-), & v_j \neq v_k. \end{cases}$$

Note that for all $\epsilon > 0$, the matrix $M(\boldsymbol{\theta})+\epsilon I$ is diagonally dominant: $|(M(\boldsymbol{\theta})+\epsilon I)_{v,v}| = \deg v + \epsilon > \sum_{v'\neq v} |M(\boldsymbol{\theta})_{v,v'}|$, and hence is invertible.

If $g = (\mathcal{L} + \epsilon I)^{-1}f$, then $\mathbb{F}f = \mathbb{F}(\mathcal{L} + \epsilon I)g = (M(\boldsymbol{\theta}) + \epsilon I)\mathbb{F}g$, and hence, $\mathbb{F}(\mathcal{L} + \epsilon I)^{-1}f = \mathbb{F}g = (M(\boldsymbol{\theta}) + \epsilon I)^{-1}\mathbb{F}f$.

□

Proof of Lemma 4.4.11. For the matrix $D(\boldsymbol{\theta})D^*(\boldsymbol{\theta})$, we have

$$
\begin{aligned}
D(\boldsymbol{\theta})D^*(\boldsymbol{\theta})(v_1, v_2) &= \sum_{e\in\mathscr{F}_{E^o}} D(\boldsymbol{\theta})(v_1, e)D^*(\boldsymbol{\theta})(e, v_2)\\
&= \sum_{e\in\mathscr{F}_{E^o}} \left(\mathbb{1}[v_1 = o(e)] - \sum_{\boldsymbol{n}\in\Gamma} \chi_{\boldsymbol{\theta}}(\boldsymbol{n})\mathbb{1}[\mathbf{n}v_1 = t(e)]\right)\cdot\\
&\quad\cdot\left(\mathbb{1}[v_2 = o(e)] - \sum_{\boldsymbol{n}\in\Gamma} \chi_{\boldsymbol{\theta}}(-\boldsymbol{n})\mathbb{1}[\mathbf{n}v_2 = t(e)]\right)\\
&= \sum_{e\in\mathscr{F}_{E^o}} \Big(\mathbb{1}[v_1 = o(e)]\mathbb{1}[v_2 = o(e)]\\
&\quad + \sum_{\mathbf{n},\in\Gamma} \chi_{\boldsymbol{\theta}}(\boldsymbol{n})\mathbb{1}[\mathbf{n}v_1 = t(e)]\chi_{\boldsymbol{\theta}}(-\boldsymbol{m})\mathbb{1}[\boldsymbol{v}_2 = t(e)]\\
&\quad - \sum_{\boldsymbol{n}\in\Gamma} \chi_{\boldsymbol{\theta}}(\boldsymbol{n})\mathbb{1}[\mathbf{n}v_1 = t(e)]\mathbb{1}[v_2 = o(e)]\\
&\quad - \sum_{\boldsymbol{n}\in\Gamma} \chi_{\boldsymbol{\theta}}(-\boldsymbol{n})\mathbb{1}[\mathbf{n}v_2 = t(e)]\mathbb{1}[v_1 = o(e)]\Big)\\
&= |\{e\in\mathscr{F}_{E^o} : o(e) = v_1\}|\mathbb{1}[v_1 = v_2]\\
&\quad + \sum_{e\in\mathscr{F}_{E^o}} \sum_{\mathbf{n},\in\Gamma} \chi_{\boldsymbol{\theta}}(\boldsymbol{n}-)\mathbb{1}[\mathbf{n}v_1 = \boldsymbol{v}_2 = t(e)]\\
&\quad - \sum_{e\in\mathscr{F}_{E^o}} \sum_{\boldsymbol{n}\in\Gamma} \chi_{\boldsymbol{\theta}}(\boldsymbol{n})\mathbb{1}[\mathbf{n}v_1 = t(e), v_2 = o(e)]\\
&\quad - \sum_{e\in\mathscr{F}_{E^o}} \sum_{\boldsymbol{n}\in\Gamma} \chi_{\boldsymbol{\theta}}(-\boldsymbol{n})\mathbb{1}[\mathbf{n}v_2 = t(e), v_1 = o(e)]\\
&= |\{e\in\mathscr{F}_{E^o} : o(e) = v_1\}|\cdot\mathbb{1}[v_1 = v_2]\\
&\quad + |\{e\in E_0^0 : t(e) = v_1\}|\cdot\mathbb{1}[v_1 = v_2]\\
&\quad - \sum_{e\in\mathscr{F}_{E^o}} \sum_{\boldsymbol{n}\in\Gamma} \chi_{\boldsymbol{\theta}}(\boldsymbol{n})\mathbb{1}[\mathbf{n}v_1 = t(e), v_2 = o(e)]\\
&\quad + \sum_{e\in\mathscr{F}_{E^o}} \sum_{\boldsymbol{n}\in\Gamma} \chi_{\boldsymbol{\theta}}(-\boldsymbol{n})\mathbb{1}[\mathbf{n}v_2 = t(e), v_1 = o(e)].
\end{aligned}
$$

It is clear that the term in brackets is zero if $v_1 = v_2$. For $v_1, v_2 \in V_0$, $v_1 \neq v_2$, if $\mathbf{n} \in \Gamma$ is such that $\mathbf{n}v_2 \sim v_1$, i.e., v_1 and $\mathbf{n}v_2$ are connected by an edge, then either there is an edge $e \in \mathscr{F}_{E^o}$ such that $v_1 = o(e)$ and $\mathbf{n}v_2 = t(e)$, or $v_1 = t(e)$ and $\mathbf{n}v_2 = o(e)$, which is equivalent to $-\mathbf{n}v_1 = t(e')$, $v_2 = o(e')$ for some e'. Hence the term in the brackets is given by $\sum_{\mathbf{n}\in\Gamma:\mathbf{n}v_2\sim v_1} \chi_{\boldsymbol{\theta}}(-\mathbf{n})$.

Thus we can conclude that

$$D(\boldsymbol{\theta})D^*(\boldsymbol{\theta})(v_1,v_2) = \deg(v_1)\mathbb{1}[v_1 = v_2] - \sum_{\mathbf{n}\in\Gamma:\mathbf{n}v_2\sim v_1} \chi_{\boldsymbol{\theta}}(-\mathbf{n}) = M(\boldsymbol{\theta})(v_1,v_2).$$

□

4.4.2 Example: calculation of correlations for the three-ladder graph

Consider an unoriented graph G with a vertex set $V = \{v_{1,i}, v_{2,i}, v_{3,i}\}_{i\in\mathbb{Z}}$ and the set of edges

$$E = \{(v_{k,i}, v_{k,i\pm1})\}_{k=\{1,2,3\},i\in\mathbb{Z}} \cup \{(v_{1,i}, v_{2,i}), (v_{2,i}, v_{3,i}), (v_{3,i}, v_{1,i})\}_{i\in\mathbb{Z}}.$$

We call G a 3-ladder graph (see Figure 4.3). This graph has two symmetries groups: $\mathbb{Z}$ and $\mathbb{Z}\times\mathbb{Z}_3$. Above we showed how to calculate the corresponding determinantal kernels: for $\mathbb{Z}$ reduction we obtain

$$K_{\mathbb{Z}}(\theta) = \frac{1}{5-2\cos(\theta)}.$$

$$\begin{bmatrix} 3-2\cos(2\pi\theta) & 1 & 1 & e^{2\pi i\theta}-1 & 0 & 1-e^{2\pi i\theta} \\ 1 & 3-2\cos(2\pi\theta) & 1 & 1-e^{2\pi i\theta} & e^{2\pi i\theta}-1 & 0 \\ 1 & 1 & 3-2\cos(2\pi\theta) & 0 & 1-e^{2\pi i\theta} & e^{2\pi i\theta}-1 \\ 1-e^{-2\pi i\theta} & e^{-2\pi i\theta}-1 & 0 & 2 & -1 & -1 \\ 0 & 1-e^{-2\pi i\theta} & e^{-2\pi i\theta}-1 & -1 & 2 & -1 \\ e^{-2\pi i\theta}-1 & 0 & 1-e^{-2\pi i\theta} & -1 & -1 & 2 \end{bmatrix}$$

Take $k_1, k_2 \in \Gamma = \mathbb{Z}$. The correlation kernel of two edges $e_{k_1,i}, e_{k_2,j}$ is given by the i,j-th entry of the matrix $\mathbb{K}(k_1,k_1)$ given by the formula

$$\mathbb{K}_{\mathbb{Z}}(k_1,k_2) = \int_0^1 K(\theta)e^{2\pi i(k_1-k_2)\theta}d\theta.$$

For $\mathbb{Z}\times\mathbb{Z}_3$ reduction we obtain that $K_{\mathbb{Z}\times\mathbb{Z}_3}(\boldsymbol{\theta})$ is given by

$$\frac{1}{4-2\cos(2\pi\theta_1)-2\cos(2\pi\theta_2)}\begin{bmatrix} 2-2\cos(2\pi\theta_1) & (1-e^{2\pi i\theta_1})(1-e^{-2\pi i\theta_2}) \\ (1-e^{-2\pi i\theta_1})(1-e^{2\pi i\theta_2}) & 2-2\cos(2\pi\theta_2) \end{bmatrix}$$

Take two edges $e_1, e_2 \in E$ such that $e_1 = \boldsymbol{n}e_i^0$ and $e_2 = \boldsymbol{m}e_j^0$, $\boldsymbol{n},\boldsymbol{m} \in \mathbb{Z}\times\mathbb{Z}/3\mathbb{Z}$ and $e_i^0, e_j^0 \in E_0$, then their correlation is given by the (i,j)-th element of the matrix

$$\begin{aligned}\mathbb{K}_{\mathbb{Z}\times\mathbb{Z}_3}(\boldsymbol{n},\boldsymbol{m}) =& \frac{1}{3}\int_0^1 K(\theta,0)e^{2\pi i(n_1-m_1)\theta}d\theta \\ &+ \frac{1}{3}\int_0^1 K(\theta,1/3)e^{2\pi i(n_1-m_1)\theta}e^{2/3\pi i(n_2-m_2)}d\theta \\ &+ \frac{1}{3}\int_0^1 K(\theta,-1/3)e^{2\pi i(n_1-m_1)\theta}e^{-2/3\pi i(n_2-m_2)}d\theta\end{aligned}$$

In this subsection we calculate the correlations of pairs of edges using two different correlation kernels and show that the results coincide.

1-1 correlation:

$$\mathbb{K}_{\mathbb{Z}}(0,n) = \int_0^1 \frac{3-2\cos(2\pi\theta)}{5-2\cos(2\pi\theta)}e^{2\pi in\theta}d\theta$$

$$\begin{aligned}\mathbb{K}_{\mathbb{Z}\times\mathbb{Z}/3\mathbb{Z}}((0,0),(n,0)) &= \int_0^1 \left(\frac{e^{2\pi in\theta}}{3} + \frac{2(2-2\cos(2\pi\theta))}{3(5-2\cos(2\pi\theta))}e^{2\pi in\theta}\right)d\theta \\ &= \int_0^1 \frac{5-2\cos(2\pi\theta)+4-4\cos(2\pi\theta)}{3(5-2\cos(2\pi\theta))}e^{2\pi in\theta}d\theta \\ &= \int_0^1 \frac{3-2\cos(2\pi\theta)}{5-2\cos(2\pi\theta)}e^{2\pi in\theta}d\theta\end{aligned}$$

1-2 correlation:

$$\mathbb{K}_{\mathbb{Z}}(0,n) = \int_0^1 \frac{1}{5-2\cos(2\pi\theta)}e^{2\pi in\theta}d\theta$$

$$\begin{aligned}&\mathbb{K}_{\mathbb{Z}\times\mathbb{Z}/3\mathbb{Z}}((0,0),(n,1)) = \\ &\qquad\int_0^1 \Big(\frac{1}{3}e^{2\pi in\theta} + \frac{2-2\cos(2\pi\theta)}{3(5-2\cos(2\pi\theta))}e^{2\pi in\theta}e^{2\pi i/3} \\ &\qquad+ \frac{2-2\cos(2\pi\theta)}{3(5-2\cos(2\pi\theta))}e^{2\pi in\theta}e^{-2\pi i/3}\Big) \\ &\qquad= \int_0^1 \frac{1}{5-2\cos(2\pi\theta)}e^{2\pi in\theta}d\theta\end{aligned}$$

1-4 correlation:

$$\mathbb{K}_{\mathbb{Z}}(0,n) = \int_0^1 \frac{e^{-2\pi i\theta}-1}{5-2\cos(2\pi\theta)}e^{2\pi in\theta}d\theta$$

$$
\begin{aligned}
&\mathbb{K}_{\mathbb{Z}\times\mathbb{Z}/3\mathbb{Z}}((0,0),(n,1)) = \\
&\qquad \int_0^1 \frac{1}{3(5-2\cos(2\pi\theta))}(1-e^{-2\pi i\theta})(1-e^{-2\pi i/3})e^{2\pi in\theta}e^{2\pi i/3}+ \\
&\qquad + \int_0^1 \frac{1}{3(5-2\cos(2\pi\theta))}(1-e^{-2\pi i\theta})(1-e^{2\pi i/3})e^{2\pi in\theta}e^{-2\pi i/3} \\
&\qquad = \int_0^1 \frac{e^{-2\pi i\theta}-1}{5-2\cos(2\pi\theta)}e^{2\pi in\theta}d\theta
\end{aligned}
$$

4-5 correlation:

$$
\mathbb{K}_{\mathbb{Z}}(0,n) = \int_0^1 \frac{-1}{5-2\cos(2\pi\theta)}e^{2\pi in\theta}d\theta
$$

$$
\begin{aligned}
\mathbb{K}_{\mathbb{Z}\times\mathbb{Z}/3\mathbb{Z}}((0,0),(n,1)) &= \int_0^1 \left(\frac{3e^{2\pi in\theta}e^{2\pi i/3}}{3(5-2\cos(2\pi\theta))} + \frac{3e^{2\pi in\theta}e^{-2\pi i/3}}{3(5-2\cos(2\pi\theta))}\right)d\theta \\
&= \int_0^1 \frac{-1}{5-2\cos(2\pi\theta)}e^{2\pi in\theta}d\theta
\end{aligned}
$$

4-4 correlation:

$$
\mathbb{K}_{\mathbb{Z}}(0,n) = \int_0^1 \frac{2}{5-2\cos(2\pi\theta)}e^{2\pi in\theta}d\theta
$$

$$
\begin{aligned}
\mathbb{K}_{\mathbb{Z}}((0,0),(n,0)) &= \int_0^1 \left(\frac{3e^{2\pi in\theta}}{3(5-2\cos(2\pi\theta))} + \frac{3e^{2\pi in\theta}}{3(5-2\cos(2\pi\theta))}\right) \\
&= \int_0^1 \frac{2}{5-2\cos(2\pi\theta)}e^{2\pi in\theta}d\theta
\end{aligned}
$$

Remark 4.4.12. The computation of correlation kernels for ladder-like graphs has been presented in [57] (in particular, for the ladder and the 3-ladder graphs, that we also discussed in the present work). In the work mentioned above, it is demonstrated how to compute the correlation kernels using two approaches: the counting approach and the electrical network approach. However, the approach presented in the current work is more universal and, moreover, appeals to simpler mathematical techniques.

Chapter 5

Rates of convergence in Central Limit Theorem for ergodic toral automorphisms[1]

5.1 Introduction and main result

Let $\mathbb{T}^d$ be a d-dimensional torus. Consider the standard projection $\pi\colon \mathbb{R}^d \to \mathbb{T}^d$ given by $\pi(x_1, \dots, x_d) = (x_1 \mod 1, \dots, x_d \mod 1)$ and a matrix $S \in \mathrm{GL}(d, \mathbb{Z})$ such that $\det S = \pm 1$. The *toral automorphism* $T_S\colon \mathbb{T}^d \to \mathbb{T}^d$ associated to the matrix S is given by $\pi \circ S = T_S \circ \pi$. Alternatively, one can simply write $T_S(x) = Sx \mod 1$. The toral automorphism T_S is ergodic if and only if the associated matrix S has no eigenvalues which are roots of unity. The transformation T_S preserves the normalised Lebesgue measure m on $\mathbb{T}^d$.

The eigendirections of a matrix S described above induce a decomposition of $\mathbb{R}^d = E^s_S \oplus E^n_S \oplus E^u_S$ where E^s_S is the eigenspace of S corresponding to the eigenvalues with modulus smaller than 1 (stable directions), E^n_S is the eigenspace of S corresponding to the eigenvalues with modulus 1 (neutral directions), and E^u_S is the eigenspace of S corresponding to the eigenvalues with modulus larger than 1 (unstable directions). An important subclass of ergodic toral automorphisms is formed by the *hyperbolic toral automorphisms* for which $E^n_S = \{0\}$ (see Figure 5.1). In other words, the matrix S associated to a hyperbolic toral automorphism has no eigenvalues of unit absolute value. In summary, we consider the following classes of toral auto-

[1]This chapter is based on: E. Arzhakova, D. Terhesiu, Rates of convergence in Central Limit Theorem for ergodic toral automorphisms, in progress

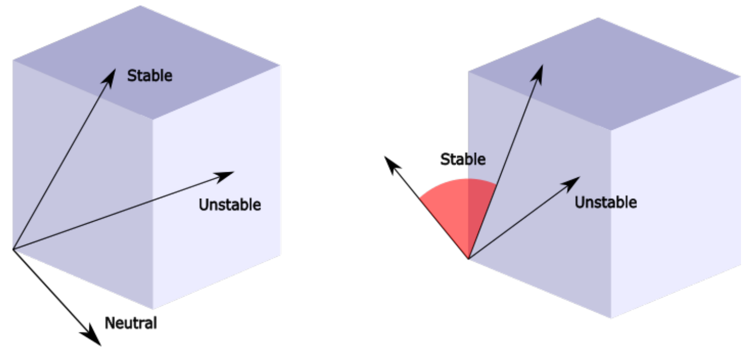


Figure 5.1: Ergodic toral automorphism (left) and a hyperbolic toral automorphism (right). The difference is in the fact that the hyperbolic toral automorphism does not have neutral eigendirections.

morphisms (TA): Toral automorphisms $\supset$ ergodic TA $\supset$ hyperbolic TA.

Example 5.1.1 (Ergodic non-hyperbolic toral automorphism.)**.** The following matrix induces an example of an ergodic non-hyperbolic toral automorphism:

$$S = \begin{pmatrix} 0 & 0 & 0 & -1 \\ 1 & 0 & 0 & 2 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 2 \end{pmatrix}$$

The characteristic polynomial $\chi_S(\lambda) = \lambda^4 - 2\lambda^3 - 2\lambda + 1$ has two real eigenvalues:

$$\lambda_+ = \frac{1}{2}(1 + 3^{1/2} + 12^{1/4}) > 1, \quad \lambda_- = \frac{1}{2}(1 + 3^{1/2} - 12^{1/4}) < 1;$$

and two complex eigenvalues:

$$\lambda_0^{(1)} = \frac{1}{2}(1 - 3^{1/2} + 12^{1/4}i), \quad \lambda_0^{(2)} = \frac{1}{2}(1 - 3^{1/2} - 12^{1/4}i),$$

whose absolute values are equal to 1. Neither of the eigenvalues is a root of unity, therefore, the induced toral automorphism is ergodic. However, it is not hyperbolic due to the presence of two eigenvalues of unit absolute value.

Up to now, several probabilistic aspects of ergodic toral automorphisms have been studied (with respect to the invariant measure m) and we start by recalling some of the landmark results. We remark that hyperbolic toral automorphisms are much easier to study because the dynamical system $(\mathbb{T}^d, T_S)$ has a Markov partition (roughly, every element of the partition gets mapped to a union of partition elements) and as a consequence, the study of $(\mathbb{T}^d, T_S, m)$ can be reduced to that of two-sided finite Markov shifts for which a well developed theory exists (we refer to Section 5.2 for further details).

Due to the presence of neutral directions the study of non-hyperbolic toral automorphisms is much harder; in particular, up to trivial examples, ergodic non-hyperbolic toral automorphisms do not have Markov partitions (see, for instance, [10, 62, 63] and references therein). Resorting to the construction of some clever measurable partition and building on a previous result of Katznelson [54], Lind [63] proved exponential decay of correlation for the general class of θ-Hölder functions $v, w \in C^\theta$ on $\mathbb{T}^d$, that is, $|\int_{\mathbb{T}^d} v\, w \circ T_S^n\, dm - \int_{\mathbb{T}^d} v\, dm \int_{\mathbb{T}^d} w\, dm| \leqslant C\rho^n \|v\|_\theta |w\|_\theta$ for some uniform constant C and some $\rho \in (0, 1)$. Hereafter, we denote the class of θ-Hölder functions by C^θ.

Exploiting the partitions introduced in [63], Le Borgne [10] constructed appropriate filtrations to show that under mild assumptions on the Fourier coefficients on functions v on $\mathbb{T}^d$, the Gordin method [40] of martingale differences can be applied to obtain the Central Limit Theorem (CLT) along with its refinements: Weak Invariance Principle (WIP), that is, convergence to Brownian motion, and Strong Invariance Principle (SIP), which is a strong version of the law of the iterated logarithm. For a rough overview of the martingale difference for dynamical systems we refer to Section 5.2. Below we recall the above mentioned terminology along with the result in [10].

Denote the n-th ergodic sum of $v : \mathbb{T}^d \to \mathbb{R}, v \in L^2(m)$ by $S_n v = \sum_{k=0}^{n-1} v \circ T_S^k$. Given a centered function on $\mathbb{T}^d$ (that is, $\int_{\mathbb{T}^d} v\, dm = 0$), (v, T_S) satisfies the CLT with non-zero variance if there exists $\sigma > 0$ such that $\frac{1}{\sqrt{n}} S_n v$ converges in distribution to a Gaussian random variable $Z \sim \mathcal{N}(0, \sigma^2)$ defined on a probability space $(\Omega, \mathscr{F}, \mathbb{P})$ with mean zero and variance σ^2. This means that as $n \to \infty$,

$$\sup_{\alpha \in \mathbb{R}} \left| m\left(\frac{S_n f}{\sqrt{n}} < \alpha\right) - \mathbb{P}\left(Z < \alpha\right) \right| \to 0. \tag{5.1}$$

Recall that given $v : \mathbb{T}^d \to \mathbb{R}$, (v, T_S) satisfies WIP if $W_n(t) = \{\frac{1}{\sqrt{n}} \sum_{k=0}^{[nt]} v \circ T_S^k,\ t \in [0, 1]\}$ converges in the space $(D[0, 1], \mathbb{R})$ (the space of functions which have left-hand limits and are continuous from the right on $(0, 1)$) to a Brownian motion with variance σ^2. Further, (v, T_S) satisfies the SIP if (enlarging $\mathbb{T}^d$ if necessary) there exists a sequence of independent identically distributed (iid) Gaussian random variables Y_k on $(\mathbb{T}^d, T_S, m)$ with mean zero and variance σ^2 such that

$$\sup_{1\leqslant M\leqslant n} \Big| \sum_{k=0}^{M-1} v\circ T_S^k - \sum_{k=0}^{M-1} Y_k \Big| = o(n^{1/2}(\log\log n)^{1/2}) \text{ almost surely as } n\to\infty. \tag{5.2}$$

With these specified, we recall

Theorem 5.1.2. [10] Let v be a centered function on $\mathbb{T}^d$ and $v \in L^2(m)$. Assume that for every $b > 0$ the Fourier coefficients $\widehat{v}(n)$ of v satisfy

$$\sum_{|n|>b} |\widehat{v}(n)|^2 \leqslant R\log^{-\theta}(b)$$

for some $R > 0, \theta > 2$. Assume that v is not a coboundary, i.e., there exists no $h \in L^2(m)$ such that $v = h - h\circ T_S$.

Then (v, T_S) satisfies: i) the CLT with non-zero variance

$$\sigma^2 = \sum_{k\in\mathbb{Z}} \mathbb{E}_m[v\cdot v\circ T_S^{|k|}];$$

ii) WIP and iii) SIP with rate as in (5.2).

The role of the assumption that v is not coboundary is to ensure that $\sigma > 0$.

Since the works of [10, 63] the results have been improved in two directions. Exponential mixing of all orders of Hölder functions was proved via different methods by Dolgopyat [26] and Pène [90]:

Theorem 5.1.3. [26, 90] Let $v_i \in C^\theta(\mathbb{T}^d)$ for some $\theta \in (0,1)$. Then, there exists $\rho \in (0,1)$ such that for any $n_0, \dots n_s \in \mathbb{N}$,

$$\Big| \int_{\mathbb{T}^d} \prod_{i=1}^s v_i\circ T_S^{n_i}\, dm - \prod_{i=1}^s \left(\int_{\mathbb{T}^d} v_i\, dm \right) \Big| \leqslant C\rho^{\min_{i\neq j}|n_i-n_j|} \prod_{i=1}^s \|v_i\|_{C^\theta},$$

for some uniform constant C.

Combining Theorem 5.1.3 with Theorem 5.1.2, Gorodnik and Spatzier [44] show that Theorem 5.1.2 holds for the whole class of Hölder functions, with no restriction on the Fourier coefficients. In fact, [44, Theorem 6.2] is phrased for the much larger class of ergodic automorphisms on compact nilmanifolds (not just on $\mathbb{T}^d$), of which particularities do not constitute the subject of this work.

In a different direction, a few works obtain rates of convergence in the CLT and the SIP of Theorem 5.1.2. First, we mention that the work of Le Borgne and Pène [15] gives optimal Berry-Essen error rates in CLT for ergodic toral automorphisms on $\mathbb{T}^3$. With the notation used in (5.1), we recall

Theorem 5.1.4. [15, A consequence of Theorem 2.] Consider an ergodic toral automorphism $(\mathbb{T}^3, T_S, m)$. Suppose that $v : \mathbb{T}^3 \to \mathbb{R}$ satisfies the assumptions of Theorem 5.1.2, in particular the same conditions on the decay of Fourier coefficients of v. Then

$$\sup_{\alpha\in\mathbb{R}} \left| m\left(\frac{S_n v}{\sqrt{n}} < \alpha\right) - \mathbb{P}\left(Z < \alpha\right) \right| = O\left(\frac{1}{\sqrt{n}}\right).$$

We also mention the work of Dedecker, Merlevède and Pène [23] who build on the technical part of the proof of Theorem 5.1.2 to enlarge the class of functions (increasing the range of θ from $\theta > 2$ to $\theta > 1$) and also improve the rate in (5.2) from $o(n^{1/2}(\log\log n)^{1/2})$ to $O(n^{1/4}(\log n))$. This estimate is not implied by and does not imply error rates in the CLT.

It is not clear to us how to generalise Theorem 5.1.4 to general $(\mathbb{T}^d, T_S, m)$, $d > 3$ and also to the entire class of Hölder functions. To our knowledge, error rates in CLT for the general class of ergodic toral automorphisms seem to be absent from the up to date literature. Our main focus is to provide promising results in this direction.

To state our main result, we introduce further terminology. Let

$$\Phi_{\sigma^2}(h) = \frac{1}{\sigma\sqrt{2\pi}} \int_{\mathbb{R}} e^{-\frac{\sigma^2}{2}u} h(u)\, du$$

be the expectation of the function $h : \mathbb{R} \to \mathbb{R}$ with respect to the one-dimensional centered distribution $\mathcal{N}(0,\sigma^2)$. Let $\mathcal{W}$ be the class of Lipschitz functions on $\mathbb{R}$. Consider a system $(\mathbb{T}^d, T_S, m)$, a function $v : \mathbb{T}^d \to \mathbb{R}$, and let $S_n v$ be its ergodic sum. Given $Z \sim \mathcal{N}(0,\sigma^2)$ on $(\Omega, \mathscr{F}, \mathbb{P})$, the Wasserstein distance $d_{\mathcal{W}}$ between $\frac{S_n v}{\sqrt{n}}$ and Z is given by

$$d_{\mathcal{W}}\left(\frac{S_n v}{\sqrt{n}}, Z\right) = \sup_{h\in\mathcal{W}} \left| m\left(h\left(\frac{S_n v}{\sqrt{n}}\right)\right) - \Phi_{\sigma^2}(h) \right|. \tag{5.3}$$

Replacing the class $\mathcal{W}$ with the the class of the step functions $\mathbb{K} = \{1_{[-\infty,x]} : x \in \mathbb{R}\}$ gives the Kolmogorov distance, which is the one used in the Berry-Esseen result of Theorem 5.1.4. We recall that results in the Wasserstein distance are weaker since

$$d_{\mathcal{W}}\left(\frac{S_n v}{\sqrt{n}}, Z\right) \leqslant C\left(d_{\mathbb{K}}\left(\frac{S_n v}{\sqrt{n}}), Z\right)\right)^{1/2}, \tag{5.4}$$

for some uniform C.

With these specified, we state the main result of this work

Theorem 5.1.5. Consider $(\mathbb{T}^d, T_S, m)$ and suppose the stable and unstable eigenspaces of S are such that $\dim(E^s_S) = \dim(E^u_S) = 1$. Let $v\colon \mathbb{T}^d \to \mathbb{R}$ be a centered Hölder function such that v is not an L^2-coboundary. Then, (v, T_S) satisfies CLT with non-zero variance $\sigma^2 = \sum_{k\in\mathbb{Z}} \mathbb{E}_m[v \cdot v \circ T_S^{|k|}]$. Moreover,

$$d_{\mathcal{W}}\left(\frac{S_n v}{\sqrt{n}}, Z\right) = O\left(\frac{\log n}{\sqrt{n}}\right).$$

We remark that a similar method of proof works in the situation where the stable and unstable eigendirections of the associated matrix S have different dimensions and also for the multivariate Hölder observable $v : \mathbb{T}^d \to \mathbb{R}^q$, $q \geqslant 1$. For simplicity, in this chapter we omit these generalisations. Given the relation between the Wasserstein and Kolmogorov distances in (5.4), it seems that the current result is far from an optimal Berry-Esseen bound. However, this is not the case for multidimensional observables: in this case the best one can do is to use [91, Theorem 1.1] to improve the result in the Wasserstein distance in Theorem 5.1.5 to $O(n^{-1/2})$, therefore getting rid of $\log n$. In this sense, the present results are very promising.

We emphasise that in the present chapter we provide a new proof of the CLT and that the result of Theorem 5.1.5 is new for $(\mathbb{T}^d, T_S, m)$, $d > 3$ and the entire class of Hölder functions on $\mathbb{T}^d$. Much more importantly, we believe that our proof extends to the cases of random ergodic toral automorphisms, and, eventually, non-linear ergodic toral automorphisms, where previous methods simply break down. This is the subject of work in progress. At this stage we mention that our method of proof relies on the use of the CLT results for dynamical systems via the Stein method obtained by Hella et al. [49] and a careful check of their assumptions using that the eigenfunctions of T_S have components that are Diophantine irrationals.

We conclude this introduction remarking that due to the exponential mixing result for Hölder functions [63], the WIP in Theorem 5.1.2 is a direct consequence of the CLT. This is because the other required condition for WIP, namely, the tightness, can be checked as in the proof of [82, Theorem 1.4]. The SIP in Theorem 5.1.2 is more delicate and we do not address this here.

5.2 A brief survey of the methods of proof of CLT for dynamical systems

In this Section we discuss several methods to prove CLT in the framework of dynamical systems and, in particular, in the framework of hyperbolic toral automorphisms. Where appropriate, we explain why these methods break down for non-hyperbolic ergodic toral automorphisms. The application of some methods (the analogue of the characteristic function for independent random variables and the martingale difference) are illustrated using the simple example of the doubling map.

5.2.1 Gordin's homoclinic points method [41].

Suppose (X, d) is a metric space and $T : X \to X$ is a homeomorphism. Two points $x, y \in X$ are called *homoclinic* if $d(T^n x, T^n y) \to 0$ as $n \to \pm\infty$. The notion goes back to the works of Poincare, and the homoclinic equivalence relation plays an important role in various areas of the theory of dynamical systems.

One of the most convenient settings to study homoclinic structures is that of group automorphisms $T \in \mathrm{Aut}(X)$ of a compact abelian group X. It is sufficient to identify points x that are homoclinic to 0, i.e., $d(T^n x, 0) \to 0$ as $n \to \pm\infty$. Such points form a group of homoclinic points $\Delta(T, X)$.

Gordin also introduced a notion of the homoclinic transformation: an invertible map $R : X \to X$ is called homoclinic to T, if the operators $U_T f(x) = f(Tx)$, $U_R f(x) = f(Rx)$ (called the *Koopman operators* of T and R, respectively), satisfy $U_T^{-n} U_R U_T^n \to \mathrm{Id}$, where Id is the identity operator. Clearly, if $x_0 \in \Delta(X, T)$, then $Rx = x + x_0$ is homoclinic to T. The so-called Gordin group $\mathsf{Gor}(X, T)$ is formed by all invertible non-singular transformations R which are homoclinic to T. For hyperbolic toral automorphisms, the groups $\Delta(\mathbb{T}^d, T_S)$ and $\mathsf{Gor}(\mathbb{T}^d, T_S)$ are isomorphic, i.e., any homoclinic transformation arises from a homoclinic point.

Using Stein's method, Gordin established CLT for functions which are coboundaries with respect to the homoclinic transformations. Here we recall a simplified version of Gordin's CLT for homoclinic points [41]:

Theorem 5.2.1. Suppose T is a group automorphism of a compact abelian group X preserving the Haar measure λ, and $\bar{x} \in \Delta(X,T)$ is a homoclinic point. Suppose a function $f \in L^\infty(X,\lambda)$ satisfies

$$f(x) = F(x+\bar{x}) - F(x)$$

for some $F \in L^2(X,\lambda)$, and moreover,

$$\sum_{n\in\mathbb{Z}} ||f(\cdot + T^n\bar{x}) - f(\cdot)||_{L^\infty} < \infty.$$

Then the sequence

$$Z_n(x) = \frac{1}{\sqrt{n}} \sum_{k=0}^{n-1} f(T^k x)$$

converges to the Gaussian distribution $\mathcal{N}(0,\sigma^2)$, and $\sigma^2 \geqslant 0$ is given by the following absolutely converging series

$$\sigma^2 = \sum_{k=-\infty}^{\infty} \langle F(x), f(T^k x + \bar{x}) - f(T^k x)\rangle_{L^2}.$$

In case of hyperbolic toral automorphisms, the above result immediately applies to a large class of sufficiently regular functions f on $\mathbb{T}^d$.

Unfortunately, non-hyperbolic ergodic toral automorphisms have trivial groups of homoclinic points: $\Delta(\mathbb{T}^d, T_S) = \{0\}$ (see [66] and Theorem 4.1 of [67]), and hence, the above result cannot be applied.

In [42] Gordin extended the homoclinic point approach to non-hyperbolic ergodic toral automorphims using the martingale difference method.

An interesting observation is that the homoclinic point method is also applicable to $\mathbb{Z}^d$-actions [43], and in fact, in striking contrast with $\mathbb{Z}$-actions discussed in this chapter, the method works for some non-expansive algebraic dynamical systems as well, e.g., those which arise naturally in connection to spanning trees.

5.2.2 Characteristic functions method for dynamical systems

We first recall the method in the i.i.d. set up. Consider a sequence of i.i.d. random variables $\{Y_j\}_{j\geqslant 0}$ on $(\Omega, \mathscr{F}, \mathbb{P})$ with mean 0 and positive variance. One easy proof of the CLT goes via the Levy's continuity theorem. Let $\chi(t) = \mathbb{E}_{\mathbb{P}}[e^{itY_j}]$ be the characteristic function of Y_j, set $S_n = \sum_{j=0}^{n-1} Y_j$ and note that $\mathbb{E}_{\mathbb{P}}[e^{it\frac{S_n}{\sqrt{n}}}] = \chi(t/\sqrt{n})^n$. Recall that the characteristic function of $Z \sim \mathcal{N}(0,\sigma^2)$ is $\exp(-\sigma^2 t^2/2)$. Also, since Y_j, $j \geqslant 1$, has finite second moment and zero mean, one has that $1-\chi(t) = \sigma^2 t^2/2(1+o(1))$ as $t \to 0$. Thus, as $n \to \infty$,

$$\chi(t/\sqrt{n})^n = \Big(1 - \frac{\sigma^2 t^2}{2n} + o(\sigma^2 t^2/2)\Big)^n \to \exp(-\sigma^2 t^2/2), \qquad t \in \mathbb{R}.$$

By the Levy's continuity theorem we conclude that the sequence $\frac{1}{\sqrt{n}} S_n$ converges in distribution to $\mathcal{N}(0,\sigma^2)$.

In the framework of measure preserving dynamical systems (X, T, μ), given $f \colon X \to \mathbb{R}$, we are interested in the convergence in distribution of the normalised ergodic sums $\frac{1}{\sqrt{n}} S_n f = \frac{1}{\sqrt{n}} \sum_{k=0}^{n-1} f \circ T^k$ to $Z \sim \mathcal{N}(0,\sigma^2)$, for some $\sigma > 0$. For simplicity, we assume that $\int f \, d\mu = 0$. Using the Levy's continuity theorem, one can rephrase this problem in terms of convergence of the corresponding characteristic functions $\chi_{\frac{1}{\sqrt{n}} S_n f}(t)$ to the characteristic function of a limiting random variable. Originating from the work of Nagaev [84], a standard method to prove this convergence is by using spectral properties of transfer operators. Starting with the work of Aaronson and Denker [2], this method became classic for establishing various probabilistic results, including stable laws and local limit theorems. We briefly recall the main elements of the method.

The transfer operator $\mathcal{L} : L^1(\mu) \to L^1(\mu)$ for (X, T, μ), is defined by the equality $\int_X \mathcal{L}u \cdot v \, d\mu = \int_X u \cdot v \circ T \, d\mu$, where $u, v \in L^1(\mu)$ and $v \in L^\infty(\mu)$. The basic idea of the Nagaev method is that since

$$\int_X e^{itS_n f} u d\mu = \int_X \mathcal{L}(t)^n u d\mu, \quad \text{ where } \mathcal{L}(t)u = \mathcal{L}(e^{itf}u), u \in L^1.$$

the study of characteristic functions for dynamical systems (possibly with heavy dependencies) can be reduced to the study of properties of the perturbed transfer operator $\mathcal{L}(t)$. Note that $\mathcal{L}(0) = \mathcal{L}$. The necessary requirement for this method is to find a good function space on which the family of operators $\{\mathcal{L}(t)\}_{t\geqslant 0}$ satisfies good spectral properties: see, for instance, the survey of Gouëzel [45]. In short, this comes down to finding a Banach space $\mathcal{B}$ with norm $\|\cdot\|$ on which

i) $\mathcal{L}$ has a decomposition of the form
$$\mathcal{L}(0)^n u = \int u\, d\mu + Q(0)^n u,$$
where $Q(0)$ is an operator on $\mathcal{B}$ such that $\|Q(0)^n\| \leqslant \theta^n$, for some $\theta \in (0,1)$.

ii) the family $\{\mathcal{L}(t)\}_{t\geqslant 0}$ satisfies 'good' continuity properties. For the CLT a sufficient (but not necessary, see [45] and references therein) condition is that for $t \in B_\delta(0)$,
$$\|\mathcal{L}(t) - \mathcal{L}(0)\| \leqslant Ct^2$$
for some uniform constant C.

Items (i) and (ii) can be easily established for the simple example of the doubling map $T\colon [0,1] \to [0,1]$ given by the formula $Tx = 2x \mod 1$ (see Figure 5.2). However, apart from simple examples of unit interval maps, establishing (i) and (ii) is highly non-trivial. We refer to the list of references in [45] for some non-trivial examples.

For $Tx = 2x \mod 1$, and letting T^{-1} denote the left inverse branch, the existence of the Markov partition $\mathcal{P} = \{(T^{-(j+1)}1, T^{-j}1]\}_{j\geqslant 0}$ together with the expansion of the map is the key. Using the pointwise formula for the transfer operator as in [45], one establishes (i) and (ii) in the Banach space of piecewise C^2 functions; piecewise C^2 means C^2 on the elements of $\mathcal{P}$. Item (i) holds with $\theta = 1/T' = 1/2$ and item (ii) holds for any observable

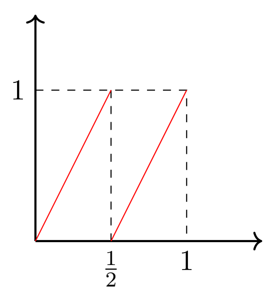


Figure 5.2: The doubling map $Tx = 2x \mod 1$.

with finite second moment.

Nagaev method for hyperbolic toral automorphisms. Nowadays, the hyperbolic toral automorphisms are known to have the spectral gap property (item i) above) in several anisotropic Banach spaces of distributions (i.e., generalised functions): see the survey of Liverani [68]. As invertible transformations, they cannot have spectral gaps in usual Banach spaces embedded in L^∞ [68]. The role of the Banach spaces in [68] is to allow a different treatment of the expanding and contracting directions. In such Banach spaces the existence of the Markov partition for hyperbolic toral automorphisms is not required, though the absence of the neutral direction is crucial.

A more traditional treatment of hyperbolic toral automorphisms exploits the existence of Markov partitions and the isomorphism with the two-sided Markov shift, where classical methods apply. We recall that as in [14] a standard way of treating two-sided Markov shifts is to collapse the stable (contracting) or unstable (expanding) directions. For the one-sided Markov shift there are several Banach spaces known to provide spectral gap: see for instance the work [2] for a brief overview. The lift of limit theorems from one-sided shifts to two-sided shifts is also classic since the work of Bowen [14].

Good Banach spaces for to ergodic toral automorphisms do not exist due to the presence of neutral direction (which in turn, does not not allow one to establish the existence of Markov partitions).

5.2.3 Martingale difference approach

Unlike the homoclinic method and the characteristic functions method, the martingale difference method can be applied in the context of non-hyperbolic ergodic toral automorphisms. The main current stochastic results on the ergodic toral automorphisms, including CLT, are obtained using the martingale difference method. In the following subsection we recall the main ingredients of the method, illustrate them with the simple example of the doubling map, and state the known limit results on the ergodic toral automorphisms.

Consider a probability space $(X, \mathcal{B}, \mu)$. A sequence $Y_n \colon X \to \mathbb{R}$ of random variables is a *martingale difference sequence* if

1. there exists a non-decreasing sequence of σ-algebras (filtration)
$$\mathcal{F}_1 \subset \mathcal{F}_2 \subset \ldots \subset \mathcal{B}$$
such that Y_n is measurable with respect to $\mathcal{F}_n$;
2. the conditional expectations satisfy $\mathbb{E}[Y_{n+1}|\mathcal{F}_n] = 0$ for $n > 0$.

A sequence $Y_n \colon X \to \mathbb{R}$ of random variables is a *reverse martingale difference sequence* if

1. there exists a non-increasing sequence of σ-algebras
$$\mathcal{B} \supset \mathcal{F}_1 \supset \mathcal{F}_2 \supset \ldots$$
such that Y_n is measurable with respect to $\mathcal{F}_n$;
2. the conditional expectations satisfy $\mathbb{E}[Y_n|\mathcal{F}_{n+1}] = 0$ for $n > 0$.

Theorem 5.2.2 ([47,69] The (reverse) martingale difference theorem.)**.** Suppose that $\{Y_n\}$ is a martingale difference with respect to $\{\mathcal{F}_i\}$. If the following two conditions hold:

- $\frac{1}{n}\sum_{i=1}^n \mathbb{E}[Y_i^2|\mathcal{F}_{i-1}] \xrightarrow{P} \sigma^2 < \infty$;
- $\frac{1}{n}\sum_{i=1}^n \mathbb{E}[Y_i^2\mathbf{1}_{|Y_i|>\epsilon\sqrt{n}}|\mathcal{F}_{i-1}] \xrightarrow{P} 0$ for every $\epsilon > 0$,

then the sequence $\frac{1}{\sqrt{n}}\sum_{i=1}^n Y_i$ converges in distribution to $\mathcal{N}(0,\sigma^2)$ [40, 47]. Suppose that $\{Y_n\}$ is a reverse martingale difference with respect to $\{\mathcal{F}_i\}$. If the following two conditions hold:

- $\frac{1}{n}\sum_{i=1}^n \mathbb{E}[Y_i^2|\mathcal{F}_{i+1}] \xrightarrow{P} \sigma^2 < \infty$;
- $\frac{1}{n}\sum_{i=1}^n \mathbb{E}[Y_i^2\mathbf{1}_{|Y_i|>\epsilon\sqrt{n}}|\mathcal{F}_{i+1}] \xrightarrow{P} 0$ for every $\epsilon > 0$,

then the sequence $\frac{1}{\sqrt{n}}\sum_{i=1}^n Y_i$ converges in distribution to $\mathcal{N}(0,\sigma^2)$.

The (reverse) martingale difference method can be applied to prove CLT in the framework of a dynamical system $(X, \mathcal{F}, \mu, T)$ and functions $f \in L^2$.

The idea is to find a non-decreasing (or non-increasing) filtration $\{\mathcal{F}_i\}$ and functions $h, g \in L^2$ such that

$$f = h + g - g \circ T,$$

where $\{h \circ T^i\}$ is a (reverse) martingale difference with respect to $\{\mathcal{F}_i\}$ which satisfies the conditions of Theorem 5.2.2 with $\sigma^2 > 0$. Since $\frac{1}{\sqrt{n}} \sum_1^n f \circ T^i = \frac{1}{\sqrt{n}} \sum_1^n h \circ T^i + \frac{1}{\sqrt{n}}(g - g \circ T^{n+1})$ the CLT for f follows from CLT for h.

Let us illustrate the reverse martingale difference method with a simple example of the doubling map. Consider a system $([0,1], \mathcal{B}, \mu, T)$ where $\mathcal{B}$ is the Borel σ-algebra, μ is the Lebesgue measure, and T is the doubling map given by $Tx = 2x \mod 1$ (see Figure 5.2). Denote by U the Koopman operator of T (i.e., $Uf = f \circ T$) and by $\mathcal{L}$ the transfer operator of T.

Consider f to be Lipschitz, not a coboundary, and with zero mean. Denote the space of Lipschitz functions by $\mathcal{W}$ with norm $\|\cdot\|_{\mathcal{W}}$; since $||\mathcal{L}f||_{\mathcal{W}} \leqslant \frac{1}{2}||f||_{\mathcal{W}}$, the function $g = \sum_{i=1}^{\infty} \mathcal{L}^i f$ is well-defined and Lipschitz. Introduce a function $h = f + g - g \circ T$; we claim that $h \circ T^n$ is a reverse martingale difference with respect to a non-increasing filtration $\mathcal{B} \supset T^{-1}\mathcal{B} \supset T^{-2}\mathcal{B} \ldots$. In order to show this, it suffices to check that $\mathbb{E}[h|T^{-1}\mathcal{B}] = \int_{T^{-1}B} h \, d\mu = 0$ for every $B \in \mathcal{B}$. It is easy to verify that

$$\int_{T^{-1}B} h \, d\mu = \int_B h \cdot (\mathbf{1} \circ T) \, d\mu = \int_B \mathcal{L}h \, d\mu.$$

Note that $\mathcal{L}h = \mathcal{L}f + \mathcal{L} \sum_{i=1}^{\infty} \mathcal{L}^i f - \mathcal{L}U \sum_{i=1}^{\infty} \mathcal{L}^i f = 0$ because $\mathcal{L}U = I$. This ensures that $h \circ T^n$ is a reverse martingale difference. It is stationary and ergodic, therefore, the CLT with positive variance holds, see, for instance, [47].

In [10], the properties of distributions of stable leaves [63] were used to construct a filtration that leads to a proof of the CLT and more refined limit properties using the martingale difference method for ergodic toral automorphisms. We recall Theorem 5.1.2 stated in the introduction.

5.3 Stein's method for establishing CLT with rates of convergence

We remind the reader that the aim of the present work is to study the rates of convergence in CLT for the class of ergodic toral automorphisms, namely, to prove Theorem 5.1.5. So far, we have discussed several methods to prove CLT in the context of dynamical systems. However, neither of the methods mentioned above is suitable to obtain optimal rates of convergence in CLT for ergodic toral automorphisms and a wide class of functions. The characteristic functions method requires the presence of the spectral gap of the transfer operator; the spectral gap is present in some Banach spaces for the hyperbolic toral automorphisms but not for non-hyperbolic ergodic toral automorphisms. The homoclinic method relies on the existence of homoclinic points of the automorphism and, therefore, it only works for the family of hyperbolic toral automorphisms but it is not applicable to the whole family of ergodic toral automorphisms. The martingale difference method allows one to prove CLT for the class of ergodic toral automorphisms. As recalled in Section 5.1, variations of the martingale difference method give error rates in CLT for ergodic toral automorphisms $(\mathbb{T}^3, T_S, m)$ and a large class of observables, see Theorem 5.1.4.

In this Section we discuss the Stein method as in [49] and motivate the application of this particular variation of the Stein method to study the rate of convergence in CLT for ergodic toral automorphisms.

5.3.1 Description of the Stein method with rates of convergence

The Stein method as in [49] studies the Wasserstein distance between $W^N = \frac{1}{\sqrt{N}} \sum_{k=0}^{N-1} f \circ T^k$ and $Z \sim \mathcal{N}(0, \sigma^2)$ under certain conditions on f and T. Therefore, the method allows one not only to prove CLT but also to establish the associated rates of convergence in the Wasserstein distance which provides a smooth metric on the space of distributions. The main difference between the Stein method and the characteristic functions or martingale difference methods is that the Stein method relies on the decay of correlations and it does not use spectral properties of the transfer operator such as the spectral gap. The fact that ergodic toral automorphisms enjoy exponential mixing of all orders [26, 90] ensures that some assumptions of the Stein method are trivial whereas other assumptions require serious effort.

We briefly explain the main idea of the Stein method as in [49]. Suppose that $Z \sim \mathcal{N}(0, \sigma^2)$; for each $h \in \mathcal{W}$ (the space of Lipschitz functions with Lipschitz constant 1) there exists a measurable solution A_h of the Stein equation:

$$\sigma^2 A_h'(x) - x A_h(x) = h(x) - \mathbb{E}[h(Z)].$$

The form of the equation can intuitively be explained by the fact that a random variable $W \sim \mathcal{N}(0, \sigma^2)$ if and only if $\sigma^2 \mathbb{E}[A'(W)] - \mathbb{E}[WA(W)] = 0$ for all absolutely continuous[2] functions A for which these expectations exist. The right hand side of the Stein equation allows us to estimate the distance between the distributions of W and Z. We select the Wasserstein distance (see (5.3) for definition) and obtain

$$d_{\mathcal{W}}(W, Z) = \sup_{h \in \mathcal{W}} \Big| \mathbb{E}[h(W)] - \mathbb{E}[h(Z)] \Big| = \sup_{A_h \colon h \in \mathcal{W}} \Big| \sigma^2 \mathbb{E}[A_h'(W)] - \mathbb{E}[W A_h(W)] \Big|.$$

For each $h \in \mathcal{W}$ such that $||h'||_\infty < \infty$ the solution A_h is a bounded function with bounded first and second derivatives. The upper bound on the Wasserstein distance follows from [49, Lemma 3.2]:

$$d_{\mathcal{W}}(W, Z) \leqslant \sup_{A \in \mathcal{A}_{\mathcal{W}}} \Big| \sigma^2 \mathbb{E}[A'(W)] - \mathbb{E}[WA(W)] \Big|, \tag{5.5}$$

where $\mathcal{A}_{\mathcal{W}} = \{A \colon ||A||_\infty \leqslant 2, ||A'||_\infty \leqslant \sqrt{2/\pi}\sigma^{-1}, ||A''||_\infty \leqslant 2\sigma^{-2}\}$.

Fix $N > 0$ and suppose that $W^N = \frac{1}{\sqrt{N}} \sum_{i=0}^{N-1} X_i$ where $\{X_i\}_{i=0}^{N-1}$ is a set of random variables. Our goal is to estimate $d_{\mathcal{W}}(W^N, Z)$ and by equation (5.5) it suffices to bound the quantity $\Big| \sigma^2 \mathbb{E}[A'(W^N)] - \mathbb{E}[W^N A(W^N)] \Big|$ for $A \in \mathcal{A}_{\mathcal{W}}$. To do this, fix $n, K \leqslant N$ and consider $W^n = W^N - \frac{1}{\sqrt{N}} \sum_{i=\max\{0, n-K\}}^{\min\{n-K, N\}} X_i$. Then,

$$\mathbb{E}[W^N A(W^N)] = \frac{1}{\sqrt{N}} \mathbb{E}\Big[\sum_{i=0}^{N-1} X_i \left(A(W^N) - A(W^i) \right) \Big] + \frac{1}{\sqrt{N}} \mathbb{E}\Big[\sum_{i=0}^{N-1} X_i A(W^i) \Big].$$

Denote the first summand of the right-hand side by S_1 and the second summand by S_2. Note that if the random variables X_i are independent then $S_2 = 0$; it seems plausible that if the correlations of X_i decay fast enough and K is big enough then S_2 is close to zero. However, an upper bound on S_2 is an assumption of the method (see Assumption 2 of Theorem 2.4 in [49]) and has to be checked separately for every system.

Since A, A' are absolutely continuous and $||A''||_\infty < 2\sigma^{-2}$ we have that

$$S_1 \approx \frac{1}{N} \sum_{i=0}^{N-1} \sum_{j=\max\{0, n-K\}}^{\min\{n+k, N-1\}} \mathbb{E}[X_i A'(W^N) X_j].$$

[2] A function $A \colon \mathbb{R} \to \mathbb{R}$ is called absolutely continuous if it has a derivative A' almost everywhere, the derivative is locally integrable, and $A(y) = A(x) + \int_x^y A'(t)dt$.

For large K one can approximate $\frac{1}{N}\sum_{i=0}^{N-1}\sum_{j=\max\{0,n-K\}}^{\min\{n+K,N-1\}}\mathbb{E}[X_iX_j] \approx \sigma^2$ and we conclude that if the correlations of random variables X_i decay fast enough (see Assumption 1 of Theorem 2.4 in [49]) then $\mathbb{E}[W^N A(W^N)] \approx \sigma^2\mathbb{E}[A'(W^N)]$ and, hence, $d_{\mathcal{W}}(W^N, Z) \approx 0$.

Let us now formulate the precise statement of Theorem 2.4 of [49]. Consider a probability space (X,μ), a measure-preserving transformation $T\colon X \to X$, and a bounded measurable centered function $f\colon X \to \mathbb{R}$. For brevity we write $\int f = \int_X f\,d\mu$. Fix two integers $0 \leqslant K < N$ and write $W^N = \sum_{j=0}^{N-1}\frac{1}{\sqrt{N}} f\circ T^j$, $W^n = W^N - \frac{1}{\sqrt{N}}\sum_{j=n-K}^{n+K} f\circ T^j$.

Theorem 5.3.1. [49] Suppose that the following conditions are satisfied for a bounded measurable centered function $f\colon X \to \mathbb{R}$:

1. There exist constants $C_2, C_4 > 0$ and a non-increasing function $\rho\colon \mathbb{N}_0 \to \mathbb{R}_+$ with $\rho(0) = 1$ and $\sum_{i=1}^{\infty} i\rho(i) < \infty$ such that for any $k \geqslant 0$ and $0 \leqslant l \leqslant m \leqslant n < N$,

   (a) $\left|\int f\cdot(f\circ T^k)\right| \leqslant C_2\rho(k)$;

   (b) $\left|\int f\cdot(f\circ T^l)\cdot(f\circ T^m)\cdot(f\circ T^n)\right| \leqslant C_4\min\{\rho(l),\rho(n-m)\}$;

   (c) $\left|\int\Big(f\cdot(f\circ T^l)\cdot(f\circ T^m)\cdot(f\circ T^n)\Big) - \int\Big(f\cdot(f\circ T^l)\Big)\int\Big((f\circ T^m)\cdot(f\circ T^n)\Big)\right| \leqslant C_4\rho(m-l)$.

2. There exists a function $\tilde\rho_N\colon \mathbb{N}_0 \to \mathbb{R}_+$ such that for any differentiable $A\colon \mathbb{R}\to\mathbb{R}$ with A' absolutely continuous and $\max_{0\leqslant k\leqslant 2}||A^{(k)}||_\infty < 1$, and for any $0 \leqslant n < N$,
$$\left|\int(A(W^n)\cdot f\circ T^n)\right| \leqslant \tilde\rho_N(K);$$

3. f is not L^2 - coboundary.

Then, $0 < \sigma^2 = \int(f^2) + 2\sum_{n=1}^{\infty}\int(f\cdot f\circ T^n) < \infty$ and if $Z \sim \mathcal{N}(0,\sigma^2)$ then
$$d_{\mathcal{W}}(W^N, Z) \leqslant C_\#\left(\frac{K+1}{\sqrt{N}} + \sum_{i=K+1}^{\infty}\rho(i)\right) + C'_\#\sqrt{N}\tilde\rho_N(K)$$
where $0 < C_\#, C'_\# < \infty$ do not depend on N, K.

Assumption 1 follows immediately from good enough decay on multiple mixing. We recall this is the case for the doubling map $Tx = 2x \mod 1$. Assumption 2 states conditions on the decay of correlations of $f \circ T^n$ and $A(W^n)$ where A is a bounded function with bounded and absolutely continuous first derivative and bounded second derivative. As we discussed above, Assumption 2 is a natural consequence of the idea of the Stein method to estimate the quantity $d_{\mathcal{W}}(W^N, Z)$ by $\left|\sigma^2\mathbb{E}[A'(W^N)] - \mathbb{E}[W^N A(W^N)]\right|$. This Assumption can be difficult to verify as its proof normally requires several technical steps (for the steps we refer the reader to Section 7.1 of [49]). In the following example we show how to verify Assumption 2 for the doubling map.

Example 5.3.2 (Assumption 2 for the doubling map)**.** Following Section 7.2 of [49] we reason that the Assumption 2 holds in a simple case of the doubling map and a Lipschitz function f with zero mean such that $|f(x) - f(y)| \leqslant L|x-y|$. Fix $0 \leqslant n < N$ and recall that $W^n = W^N - \sum_{j=n-K}^{n+K} f \circ T^j$. It is natural to present W^n as the sum of two parts:

$$W^n = W^n_- + W^n_+ = \sum_{j=0}^{n-K-1} f \circ T^j + \sum_{j=n+K+1}^{N-1} f \circ T^j.$$

To simplify the summand W^n_- we introduce a partition of the unit interval $\{\xi_q = ((q-1)2^{-n}, q2^{-n})\}_{q \in \{1,\dots 2^n\}}$ such that W^n_- is almost equal to a constant c_q on each atom ξ_q of the partition. This allows us to bound the desired quantity $\left|\int (f \circ T^n) \cdot A(W^n_- + W^n_+)\right|$ by a simpler quantity $\left|\sum_q \int_{\xi_q} (f \circ T^n) \cdot A(c_q + W^n_+)\right|$:

$$\left|\int (f \circ T^n) \cdot A(W^n)\right| \leqslant \left|\sum_q \int_{\xi_q} (f \circ T^n) \cdot A(c_q + W^n_+)\right| + \frac{L||A'||_\infty ||f||_\infty}{\sqrt{N} 2^K}.$$

To estimate the quantity $\left|\sum_q \int_{\xi_q} (f \circ T^n) \cdot A(c_q + W^n_+)\right|$ one notices that

$$\begin{aligned}\left|\sum_q \int_{\xi_q} (f \circ T^n) \cdot A(c_q + W^n_+)\right| &= \left|\sum_q \int_{\xi_q} \left(f A(c_q + \tilde{W}^n_+ \circ T^{K+1})\right) \circ T^n\right| \\ &= \left|2^{-n} \sum_q \int_{\mathbb{T}^d} f A(c_q + \tilde{W}^n_+) \circ T^{K+1}\right| \leqslant \frac{L||A||_\infty}{2^K}\end{aligned}$$

where $\tilde{W}^n_+ = W^n_+ \circ T^{n+K+1}$ and the last equality follows from the fact that $||\mathcal{L}^K f||_\theta \leqslant 2^{-K}||f||_\theta$. Since $\max_{0 \leqslant k \leqslant 2} ||A^{(k)}||_\infty < 1$ we have

$$\left|\int (f \circ T^n) \cdot A(W^n)\right| \leqslant \frac{L||A||_\infty}{2^K} + \frac{L||A'||_\infty ||f||_\infty}{\sqrt{N} 2^K} \leqslant \frac{L}{2^K} + \frac{L||v||_\infty}{\sqrt{N} 2^K} = \tilde{\rho}_N(K).$$

We recall that assumption 3 on f not being coboundary is a typical assumption to ensure $\sigma > 0$. The significance of 5.3.1 for proving rates of convergence as in 5.1.5 is emphasised in the following corollary:

Corollary 5.3.3. [49] Suppose that for any $N > 2$ the assumptions of Theorem 5.3.1 hold for f with $\rho(i) = \lambda^i, \tilde{\rho}_N(i) = C'\lambda^i N^a$, and $K = [\log N^b/\log\lambda]$ with fixed $0 < \lambda < 1, 1 < a+1 \leqslant b$. Then, f satisfies CLT and there exists a constant C that does not depend on N such that

$$d_{\mathcal{W}}(W^N, Z) \leqslant C\frac{\log N}{\sqrt{N}}.$$

5.4 Proof of CLT with rates of convergence for ergodic toral automorphisms

In this Section we prove Theorem 5.1.5. The method of proof is to apply Theorem 5.3.1 in the setting of Theorem 5.1.5. We start with a discussion of the three Assumptions of Theorem 5.3.1. Throughout, the integration is with respect to the normalised Lebesgue measure m. The proof below is written making one more simplifying assumption (not appearing as such in 5.1.5) namely that v is Lipschitz. We stress that this is just for the ease of the computation, as to not keep track of Hölder exponents. Else, the proof below can be easily adapted to work for Hölder functions.

Assumption 1: There exist constants $C_2, C_4 > 0$ and a non-increasing function $\rho\colon \mathbb{N}_0 \to \mathbb{R}_+$ with $\rho(0) = 1$ and $\sum_{i=1}^{\infty} i\rho(i) < \infty$ such that for any $k \geqslant 0$ and $0 \leqslant l \leqslant m \leqslant n < N$,

1. $\left|\int v\cdot(v\circ T_S^k)\right| \leqslant C_2\rho(k)$;
2. $\left|\int v\cdot(v\circ T_S^l)\cdot(v\circ T_S^m)\cdot(v\circ T_S^n)\right| \leqslant C_4\min\{\rho(l),\rho(n-m)\}$;
3. $\left|\int\Big(v\cdot(v\circ T_S^l)\cdot(v\circ T_S^m)\cdot(v\circ T_S^n)\Big) - \int\Big(v\cdot(v\circ T_S^l)\Big)\int\Big((v\circ T_S^m)\cdot(v\circ T_S^n)\Big)\right| \leqslant C_4\rho(m-l)$.

This assumption states conditions on decay of correlations of variables of the form $v\circ T_S^k$ of orders 2 and 4. The condition $\sum_{i=1}^{\infty} i\rho(i) < \infty$ implies that the speed of decay is at least $\rho(k) = k^{-2-\epsilon}$. In particular, it is easy to verify that Assumption 1 holds for systems with exponential mixing of all orders (hence, for Hölder functions and ergodic toral automorphisms [26, 90]) since the latter condition implies much stronger mixing properties than Assumption 1 requires.

Recall that the map T_S is exponentially mixing for Hölder functions: see Theorem 5.1.3. Thus, assumptions 1a and 1b of Assumption 1 are immediate. Assumption 1c follows from the exponential mixing of the second order.

Assumption 2: There exists a function $\tilde{\rho}_N : \mathbb{N}_0 \to \mathbb{R}_+$ such that for any differentiable $A\colon \mathbb{R} \to \mathbb{R}$ with A' absolutely continuous and $\max_{0\leqslant k\leqslant 2} ||A^{(k)}||_\infty < 1$, and for any $0 \leqslant n < N$,

$$\left| \int (A(W^n) \cdot v \circ T_A^n) \right| \leqslant \tilde{\rho}_N(K).$$

Assumption 3: This is also an assumption in v 5.1.5.

In short, Assumption 1 holds in the setting of Theorem 5.1.5 and Assumption 3 is an assumption of Theorem 5.1.5. Therefore, in order to apply Theorem 5.3.1 we are left to verify that Assumption 2 holds for (v, T_S); in other words, we are left to prove the following proposition:

Proposition 5.4.1. Let $(\mathbb{T}^d, T_S, m)$ and $v\colon \mathbb{T}^d \to \mathbb{R}$ satisfy the assumptions of 5.1.5. Fix $0 \leqslant K < N$ and for $0 \leqslant n < N$ denote

$$W^n = \frac{1}{\sqrt{N}} \left(\sum_{i=0}^{n-K-1} v \circ T_S^i + \sum_{i=n+K+1}^{N-1} v \circ T_S^i \right).$$

Then, for any $A\colon \mathbb{R} \to \mathbb{R}$ with A', A'' defined almost everywhere such that $\max_{k=0,1,2} ||A^{(k)}||_\infty \leqslant 1$ there exist constants $|\theta| < 1$ and C which do not depend on N, K, n such that

$$\left| \int (v \circ T_S^n) \cdot A(W^n) \right| \leqslant C \cdot \theta^K \sqrt{N} \tag{5.6}$$

holds for all $0 \leqslant n < N$.

Proof. The matrix S associated to the ergodic toral automorphism $T_S : \mathbb{T}^d \to \mathbb{T}^d$ has a characteristic polynomial p_S which is irreducible and of Salem type.

In other words, it has a single root $\lambda \in \mathbb{R}$ outside the unit disc, a single root λ^{-1} inside the unit disc, and other $d-2$ roots lie on the unit circle. Because of the ergodicity of T_S no root of p_S is a root of unity. Since the automorphism T_S is ergodic with respect to the normalised Lebesgue measure m, it is also exponentially mixing for pairs of Hölder observables [44, 63].

For every $x \in \mathbb{T}^d$ the tangent space T_x is isomorphic to $\mathbb{R}^d$ and admits eigenspace decomposition: $T_x = E^u_x \oplus E^n_x \oplus E^s_x$ where E^u_x is a 1-dimensional unstable eigenspace corresponding to the largest eigenvalue λ, E^s_x is a 1-dimensional stable eigenspace corresponding to the eigenvalue $\pm\lambda^{-1}$, and E^n_x is a $(d-2)$-dimensional neutral eigenspace corresponding to eigenvalues having unit absolute values. Since p_S is irreducible, it follows that all its roots are distinct, and the roots on the unit circle come in pairs of complex conjugates. Hence, the action induced by T_S on E^n_x is an isometry. The action induced by T_S on E^u_x is expanding and the action induced by T_S on E^s_x is contracting. Therefore, even though v is Lipschitz, the function W^n is not Lipschitz continuous uniformly in N; for instance, it grows rapidly in the unstable direction. Thus, we cannot directly apply the results of [44, 63] on decay of correlations for Lipschitz functions to prove Proposition 5.4.1.

Let us present W^n as a sum of two terms, $W^n = W^n_- + W^n_+$, where

$$W^n_- = \frac{1}{\sqrt{N}} \sum_{i=0}^{n-K-1} v \circ T^i_S \quad \text{and} \quad W^n_+ = \frac{1}{\sqrt{N}} \sum_{i=n+K+1}^{N-1} v \circ T^i_S.$$

Lemma 5.4.2. There exist a finite partition $\{\xi_q\}_{q\in Q}$ of $\mathbb{T}^d$ and a set of numbers $\{c_q\}_{q\in Q}$ such that

$$\int_{\mathbb{T}^d} A(W^n)\cdot(v\circ T^n_S)\, dm \leqslant \Big| \sum_q \mu(\xi_q) \int_{\xi_q} A(c_q + W^n_+)\cdot(v\circ T^n_S)\, d\nu_q \Big| + C_1 \lambda^{-K/3(d-2)}\sqrt{N},$$

where $\nu_q(\cdot) = m(\xi_d)^{-1} m(\cdot \cap \xi_q)$ and C_1 does not depend on N, K, n.

Proof. Introduce a partition $\{\xi_q\}_{q\in Q}$ of $\mathbb{T}^d$ into approximately $\lambda^{n+\frac{d-1}{3(d-2)}K}$ parallelepipeds of equal size with sides parallel to eigendirections. The length of the side parallel to the unstable direction is λ^{-n}, the length of the side parallel to the stable direction is $\lambda^{-K/3(d-2)}$, and the lengths of all sides parallel to neutral directions is $\lambda^{-K/3(d-2)}$. Note that the action of T_S is expanding in the unstable direction with coefficient λ, contracting in the stable direction with coefficient λ^{-1}, and is an isometry in neutral eigenspace; the action of T_S on the atoms of the partition and on the partition described above is shown in Figure 5.3.

On each atom ξ_q we introduce an induced measure $\nu_q(\cdot) = m(\xi_d)^{-1}m(\cdot \cap \xi_q)$ which is the normalised Lebesgue measure conditioned to ξ_q. Then, W_-^n is nearly constant on each ξ_q; in other words, the variation $W_-^n|_{\xi_q}$ is proportional to $N^{\frac{1}{2}}\lambda^{-K/3(d-2)}$. Indeed, take

$$c_q = \int W_-^n \, d\nu_q = m(\xi_q)^{-1} \int_{\xi_q} W_-^n \, dm.$$

Then

$$\begin{aligned}
\sup_{x\in\xi_q} |W_-^n(x) - c_q| &\leq \sup_{x,y\in\xi_q} |W_-^n(x) - W_-^n(y)| \\
&\leq \frac{1}{\sqrt{N}} \sum_{j=0}^{n-K-1} \sup_{x,y\in\xi_q} |v^j(x) - v^j(y)| \\
&\leq \frac{L}{\sqrt{N}} \sum_{j=0}^{n-K-1} \operatorname{diam}(T_S^j \xi_q),
\end{aligned}$$

where $\operatorname{diam}(T_S^j\xi_q)$ stands for $\sup_{x,y\in T_S^j\xi_q} |x-y|$. We can bound the diameter of the parallelepiped ξ_q by the sum of its sides: $\operatorname{diam}(\xi_q) \leqslant \lambda^{-n} + \lambda^{-K/3(d-2)} + \lambda^{-K/3(d-2)}(d-2)$. The automorphism T_S acts on the distances in the unstable direction by multiplying it by λ, in the stable direction – by multiplying it by λ^{-1}, and T_S is an isometry in the neutral directions. We conclude that

$$\begin{aligned}
\sup_{x\in\xi_q} |W_-^n(x) - c_q| &\leq \frac{L}{\sqrt{N}} \sum_{j=0}^{n-K-1} \left(\lambda^j\lambda^{-n} + \lambda^{-j}\lambda^{-K/3(d-2)} + (d-2)\lambda^{-K/3(d-2)}\right) \\
&\leq \frac{L}{\sqrt{N}} \left(\frac{\lambda^{n-K}-1}{\lambda-1}\lambda^{-n} + \frac{\lambda}{\lambda-1}\lambda^{-K/3(d-2)}\right) \\
&+ \frac{L}{\sqrt{N}} \left((n-K)(d-2)\lambda^{-K/3(d-2)}\right) \\
&\leq \frac{L}{\sqrt{N}}\lambda^{-K/3(d-2)} \left(\frac{1}{\lambda-1} + N(d-2)\right).
\end{aligned}$$

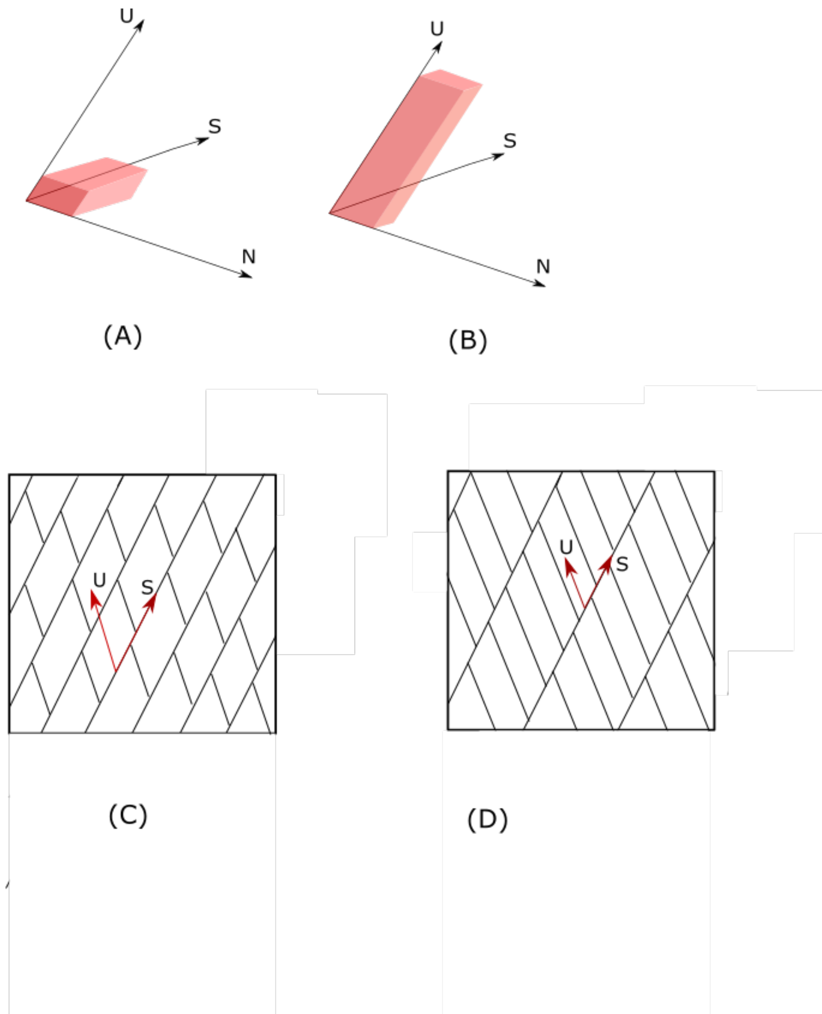


Figure 5.3: (A): an atom of the partition $\{\xi_q\}_{q\in Q}$ is a parallelepiped ξ_q whose sides are parallel to eigendirections. (B): the image of ξ_q under the action of T_S. (C): a projection on stable and unstable directions of the partition $\{\xi_q\}_{q\in Q}$ of $\mathbb{T}^d$. (D): the image of the projection on stable and unstable directions of the partition $\{\xi_q\}_{q\in Q}$ of $\mathbb{T}^d$ under the action of T_S.

By the mean value theorem we conclude that

$$\begin{aligned}
\max_{q\in Q}\sup_{x\in\xi_q}|A(W^n(x)) &- A(c_q+W_+^n(x))| \\
&\le \|A'\|_\infty \sup_{q\in Q, x\in\xi_q}|W^n(x)-(c_q+W_n^+(x))| \\
&\le \frac{L\|A'\|_\infty\lambda^{-K/3(d-2)}}{\sqrt{N}}\left(\frac{1}{\lambda-1}+N(d-2)\right)
\end{aligned}$$

and

$$\begin{aligned}
\textstyle\int_{\mathbb{T}^d} & A(W^n)\cdot v\circ T_S^n\,dm \le \left|\sum_q m(\xi_q)\int_{\xi_q}A(W^n)\cdot v\circ T_S^n\,d\nu_q\right| \\
\le & \Big|\sum_q m(\xi_q)\int_{\xi_q}A(c_1+W_+^n)\cdot v\circ T_S^n\,d\nu_q + \\
+ & \sum_q m(\xi_q)\int_{\xi_q}\frac{L\|A'\|_\infty\lambda^{-K/3(d-2)}}{\sqrt{N}}\left(\frac{1}{\lambda-1}+N(d-2)\right)\cdot v\circ T_S^n\,d\nu_q\Big| \\
\le & \Big|\sum_q m(\xi_q)\int_{\xi_q}A(c_q+W_+^n)\cdot v\circ T_S^n\,d\nu_q\Big| \\
+ & \frac{L\|A'\|_\infty\|f\|_\infty\lambda^{-K/3(d-2)}}{\sqrt{N}}\left(\frac{1}{\lambda-1}+N(d-2)\right).
\end{aligned}$$

The proof is complete. □

Let us rewrite $W_+^n = \widetilde{W}_+^n\circ T_S^{n+K+1}$ where $\widetilde{W}_+^n = \frac{1}{\sqrt{N}}\sum_{i=0}^{N-n-K-2} v\circ T_S^i$. Since $T_S^n:\xi_q\to T_S^n(\xi_q)$ is a diffeomorphism on $\{\xi_q\}$ and, therefore, $T_S^n(\xi_q)$ has no self-intersections, $m_q := (T_S^n)_*\nu_q = m(\xi_q)m(\cdot\cap T_S^n(\xi_q))$ is Lebesgue measure conditioned to $T_S^n(\xi_q)$. The sets $T_S^n(\xi_q)$ form a partition of $\mathbb{T}^d$ consisting of skinny parallelepipeds of approximate unstable side length 1, stable side length $\lambda^{-n-K/3(d-2)}$, and neutral sides length $\lambda^{-K/3(d-2)}$. Then,

$$\begin{aligned}
\sum_{q\in Q} m(\xi_q) & \int_{\xi_q}A(c_q+W_+^n)\cdot v\circ T_S^n\,d\nu_q \\
&= \sum_{q\in Q}m(\xi_q)\int_{\xi_q}\left(v\cdot A(c_q+\widetilde{W}_+^n)\circ T_S^{K+1}\right)\circ T_S^n\,d\nu_q \\
&= \sum_{q\in Q}m(\xi_q)\int_{T_S^n(\xi_q)}v\cdot A(c_q+\widetilde{W}_+^n)\circ T_S^{K+1}\,dm_q \\
&= \int_{\mathbb{T}^d}v\cdot A(c(x)+\widetilde{W}_+^n)\circ T_S^{K+1}\,dm.
\end{aligned}$$

where $c(x) = c_q$ if $x \in T_S^n(\xi_q)$. The function $A(c_q + \widetilde{W}_+^n)$ is discontinuous at $\partial T_S^n(\xi_q)$ and not uniformly Lipschitz inside each $T_S^n(\xi_q)$, so we cannot directly apply the exponential decay of correlations result to finish the proof. Instead, our next step is to prove the correlation bounds using measure disintegration and the Koksma inequality.

Lemma 5.4.3. There exists a constant C_2 which does not depend on N, K, n such that

$$\int_{\mathbb{T}^d} v \cdot A(c(x) + \widetilde{W}_+^n) \circ T_S^{K+1}\, dm \leqslant C_2 \lambda^{-K/3(d-2)}.$$

Proof. Divide each atom $T_S^n(\xi_q)$ into approximately $\lambda^{K/3(d-2)}$ parallelepipeds $\{\zeta_{q'}\}_{q' \in Q'}$ of unstable side length $\lambda^{-K/3(d-2)}$. Then, each $\zeta_{q'}$ has unstable side length $\lambda^{-K/3(d-2)}$, stable side length $\lambda^{-n-K/3(d-2)}$, and neutral sides length $\lambda^{-K/3(d-2)}$; the set $\{\zeta_{q'}\}_{q' \in Q'}$ is a partition of $\mathbb{T}^d$. On each $\zeta_{q'}$, v varies no more than $Ld\lambda^{-K/3(d-2)}$, so for $m_{q'} = \frac{m(\cdot \cap \zeta_{q'})}{m(\zeta_{q'})}$ and $v_{q'} := \int v\, dm_{q'}$,

$$\begin{aligned} \Big| \int_{\mathbb{T}^d} v \cdot A(c(x) + \widetilde{W}_+^n(x)) \circ T_S^{K+1}\, dm - \\ - \sum_{q' \in Q'} m(\zeta_{q'}) v_{q'} \int A(c(x) + \widetilde{W}_+^n(x)) \circ T_S^{K+1}\, dm_{q'} \Big| \\ \leq Ld \|A\|_\infty \lambda^{-K/3(d-2)}. \end{aligned}$$

Let P be a $(d-1)$-dimensional section in the direction $E^c \oplus E^u$ passing through 0 in $\mathbb{T}^d$. Let $\{\ell_y\}_{y \in P}$ be a partition of $\mathbb{T}^d$ into arcs in the stable direction such that each ℓ_y intersects P in its endpoints (one of which is y) and doesn't intersect P in its interior. Each ℓ_y is an interval on $\mathbb{T}^d$ whose length depends on y; however, locally the lengths are the same and they change discretely in a finite number of discontinuity points. The number of discontinuity points is bounded from above and the bound depends only on d. It follows that the leaves ℓ_y of the same length form a finite number of parallelepipeds that partition the torus (see Figure 5.4) and the lengths of ℓ_y are bounded from above and from below.

The stable foliation induces a disintegration of the Lebesgue measure m: for each measurable $v : \mathbb{T}^d \to \mathbb{R}$,

$$\int_{\mathbb{T}^d} v\, dm = \int_P \int_{\ell_y} v\, d\nu_{\ell_y}\, dm_P.$$

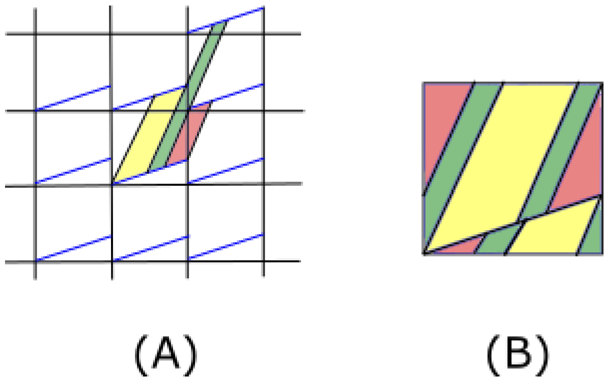


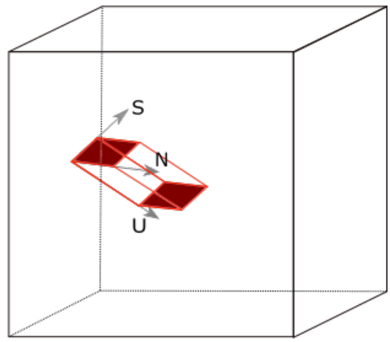
Figure 5.4: (A): Two-dimensional projection of $\mathbb{T}^d$ is a factor of $\mathbb{R}^2$ by $\mathbb{Z}^2$. The projection of P is shown by blue lines. The fibers of the stable foliation $\{\ell_y\}_{y\in P}$ are shown in yellow, green, and red; the boxes of different colours depict the regions of $\mathbb{T}^d$ with different fiber length. (B): The boxes where the fibers $\{\ell_y\}_{y\in P}$ have different length induce a partition of $\mathbb{T}^d$. The number of boxes is bounded from above and the bound depends only on d.

Note that since ℓ_y are intervals, the measure ν_{ℓ_y} is a scaled copy of the 1-dimensional Lebesgue measure for each y.

For $q' \in Q'$ consider $Z_{q'} := T_S^{K+1}(\zeta_{q'})$: it is a long skinny parallelepiped that wraps about $\lambda^{K+1-K/3(d-2)}$ around the torus in the unstable direction, and has width $\lambda^{-K/3(d-2)}$ in the neutral directions, and has width equal to $\lambda^{-n-K-1-K/3(d-2)}$ in the stable direction (see Figure 5.5). The parallelepiped $Z_{q'}$ intersects ℓ_y in intervals $\{I_j^y\}_{j=1,\dots,j_y}$ of length $\lambda^{-n-1-K(1+1/3(d-2))}$. Since the intervals I_j^y have length $\lambda^{-n-K-1-K/3(d-2)}$, the function v restricted to I_j^y varies very little. Hence, if we replace $\nu_y|_{I_j^y}$ by a single Dirac mass $\nu_y(I_j^y)\delta_{x_j^y}$ for some $x_j^y \in I_j^y$, we make only an error of order $\lambda^{-n-1-K(1+1/3(d-2))}$. Such errors are naturally absorbed in the other exponential errors we identified above.

Since $Z_{q'}$ consists of roughly $\lambda^{K+1-K/3(d-2)}$ pieces of unit (unstable) length, (stable) width $\lambda^{-n-K-1-K/3(d-2)}$, and projecting along the stable direction to $(d-1)$-dimensional parallelepipeds of $(d-1)$-dimensional area equal to $\left(\lambda^{-K/3(d-2)}\right)^{d-2}$ in P (which are roughly distributed uniformly over P), $Z_{q'}$ intersect ℓ_y roughly

$$j_y \sim \operatorname{len}(\ell_y)\lambda^{K+1-K/3(d-2)}\left(\lambda^{-K/3(d-2)}\right)^{d-2} = \operatorname{len}(\ell_y)\lambda^{\frac{K}{3}(2-\frac{1}{d-2})}$$



(A)

Figure 5.5: (A): an atom $\zeta_{q'}$. S, N, U stand for stable, neutral, and unstable eigendirections, respectively. (B): the parallelepiped $Z_{q'} := T_S^{K+1}(\zeta_{q'})$ is a skinny parallelepiped which wraps around $\mathbb{T}^d$ multiple times in the unstable direction.

times. This is the number of intervals I_j^y.

Since all eigenspaces are in irrational algebraic directions, consecutive intersections of parallel translations of such eigenspaces are obtained by a rotation over an irrational algebraic number. The components $\{v_i^u\}_{i=1,\dots,d}$ of the unit vector $\vec{v^u} \in E^u$ are rationally independent because T_S is ergodic. Also $\{v_i^u\}$ are algebraic numbers, and hence, by the Siegel-Roth theorem [96, 101], they are Diophantine of order ϵ for any $\epsilon > 0$. [3] It follows that the points x_j^y are roughly uniformly distributed on ℓ_y in the sense that the *discrepancy* (see [28, 60])

$$\mathfrak{D}_R^*(\{x_n\}_n) = \sup_k \sup_{[a,b)\subset\ell} \left| \frac{1}{R}\#\{k+1 \le n \le k+R : x_n \in [a,b)\} - (b-a) \right|$$

behaves as that of the consecutive points in the orbit of a rotation over a Diophantine number:

$$\mathfrak{D}_{j_y}^*(\{\alpha n\}) \le C_\epsilon j_y^{-\frac{1}{1+\nu}+\epsilon} \le C\lambda^{-\frac{K}{3}(2-\frac{1}{d-2})},$$

see [60, Theorem 3.2-3.4]. Recall that f and therefore $\widetilde{W}_+^n$ and also $A(c+\widetilde{W}_-^n)$ is Lipschitz (uniformly in N, n and K) in the stable direction. Now we can

[3] A number α is called Diophantine if for every $\epsilon > 0$ there is a constant C such that $|\alpha - \frac{p}{q}| \ge Cq^{-2-\epsilon}$ for all $p, q \in \mathbb{Z}$, $q \neq 0$.

estimate $\int_{\ell_y} g\,d\nu_{\ell_y}$ using the Koksma inequality:

$$\left|\sum_{j=1}^{N} g(x_j^y) - \int_{\ell_y} g(x)\,dx\right| \le \mathrm{Var}(g)\,\mathfrak{D}^*_{j_y}((x_j^y)).$$

Taking $g(x) = A(c(x) + \widetilde{W}_+^n(x))$, we obtain

$$\begin{aligned}\Big|\frac{1}{j_y}\sum_{j=1}^{j_y} A(c(x_j^y) + \widetilde{W}_+^n(x_j^y)) \quad &- \quad \int_{\ell_y} A(c(x) + \widetilde{W}_+^n)\,d\nu_\ell\Big| \\ &\le \quad \mathrm{Var}(A(c(\cdot) + \widetilde{W}_+^n(\cdot))|_{\ell_y})\mathfrak{D}^*_{j_y} \\ &\ll \quad L\mathrm{Var}(A)\lambda^{-\frac{K}{3}(2-1/(d-2))}.\end{aligned}$$

Using the estimate above and the measure disintegration, we can estimate the integral over $Z_{q'}$:

$$\begin{aligned}&\int \quad A(c(x) + \widetilde{W}_+^n(x)) \circ T_S^{K+1}\,dm_{q'} \\ \approx \quad &\int_P\int_{\ell_y} A(c + \widetilde{W}_-^n(x))d\nu_\ell\,dm_S + O(L\|A'\|_\infty\lambda^{-\frac{K}{3}(2-\frac{1}{d-2})}) \\ = \quad &\int_{\mathbb{T}^d} A(c + \widetilde{W}_-^n)dm + O(L\|A'\|_\infty\lambda^{-\frac{K}{3}(2-\frac{1}{d-2})}).\end{aligned}$$

Finally, summing over all $\zeta_{q'}$ with weights $v_{q'}$, we find

$$\begin{aligned}&\int_{\mathbb{T}^2} v\cdot A(c(x) + \widetilde{W}_+^n)\circ T_S^{K+1}\,dm \\ &= \sum_{q'\in Q'} m(\zeta_{q'})\Big(v_{q'}\int A(c(x) + \widetilde{W}_+^n)\circ T_S^{K+1}\,dm_{q'} + O(Ld\|A\|_\infty\lambda^{-K/3(d-2)})\Big) \\ &\le \sum_{q'\in Q'} m(\zeta_{q'})\left(v_{q'}\int_{\mathbb{T}^d} A(c + \widetilde{W}_-^n)\,dm + O(L\|A'\|_\infty\lambda^{-\frac{K}{3}(2-\frac{1}{d-2})})\right) \\ &+ O(Ld\|A\|_\infty\lambda^{-K/3(d-2)}) \\ &\le \left(\sum_{q'\in Q'} m(\zeta_{q'})v_{q'}\int_{\mathbb{T}^d} A(c + \widetilde{W}_-^n)\,dm\right) \\ &+ O\left(L\|A'\|_\infty\lambda^{-\frac{K}{3}(2-\frac{1}{d-2})} + Ld\|A\|_\infty\lambda^{-K/3(d-2)}\right) \\ &= \int v\,dm\int A(c + \widetilde{W}_-^n)\,dm + O\left(L\|A'\|_\infty\lambda^{-\frac{K}{3}(2-\frac{1}{d-2})} + Ld\|A\|_\infty\lambda^{-K/3(d-2)}\right) \\ &= O\left(L\|A'\|_\infty\lambda^{-\frac{K}{3}(2-\frac{1}{d-2})} + Ld\|A\|_\infty\lambda^{-K/3(d-2)}\right)\end{aligned}$$

where the last equality follows from the assumption that $\int v\,dm = 0$. $\square$

Lemma 5.4.2 and Lemma 5.4.3 prove the Assumption 2 of Theorem 5.3.1 with $\tilde{\rho}_N(k) = \tilde{C}\theta^k\sqrt{N}$. Theorem 5.3.1 implies that

$$d_{\mathcal{W}}(W^N, Z) \leqslant C_{\#}\left(\frac{K+1}{\sqrt{N}} + \sum_{i=K+1}^{\infty}\rho(i)\right) + C'_{\#}\sqrt{N}\tilde{\rho}_N(K) \tag{5.7}$$

where $\rho(k) = \theta^k$ and $\tilde{\rho}_N(k) = \tilde{C}\theta^k\sqrt{N}$ for some $\theta < 1$. Plugging ρ and $\tilde{\rho}_N$ into 5.7 and choosing $K(N) = \frac{3\log N}{2\log\theta}$ (see Corollary 5.3.3) yields

$$d_{\mathcal{W}}(W^N, Z) \leqslant C_{\#}\left(\frac{3\log N + 1}{2\log\theta\sqrt{N}} + \frac{\theta}{N^{3/2}(1-\theta)}\right) + \frac{C'_{\#}}{\sqrt{N}} \leqslant C\frac{\log N}{\sqrt{N}}$$

which concludes the proof of Theorem 5.1.5.

$\square$

Chapter 6

Connectivity of real isoperiodic sets on a torus with 3 poles[1]

6.1 Introduction

A Riemann surface is a connected manifold of complex dimension one that is equipped with a complex structure, i.e., with an atlas of open charts $\{U_i\}$ and a collection of homeomorphisms to the open disk $f_i\colon U_i \to D \subset \mathbb{C}$; the transition functions g_{ij} between the charts U_i and U_j are given by the equality $f_i = g_{ij} \circ f_j$. The transition maps are required to be holomorphic, i.e., differentiable at any point of their domain. Any open set in $\mathbb{C}$ is naturally a Riemann surface; some of the examples include the unit disk $\mathbb{D} = \{z\colon |z| < 1\}$ and the upper half-plane $\mathbb{H} = \{z \in \mathbb{C}\colon \mathrm{Im} z > 0\}$. The simplest example of a compact Riemann surface is the sphere $\widehat{\mathbb{C}} = \mathbb{C} \cup \{\infty\}$ with charts $U_1 = \mathbb{C}$ and $U_2 = \widehat{\mathbb{C}} - \{0\}$ and homeomorpisms $f_1 = z$ and $f_2 = 1/z$. Then, both transition maps $g_{1,2}$ and $g_{2,1}$ are given by a holomorphic function $z \mapsto 1/z$. By the classification theorem, any orientable compact surface X is homeomorphic to either a sphere $\widehat{\mathbb{C}}$ or a g-holed torus with $g \geqslant 1$ [27]. The number g is called the genus of the surface. For some applications it is important to consider pointed Riemann surfaces, i.e., the data of the Riemann surface with a finite number of points on it.

It is convenient to consider Riemann surfaces with fixed genus g and fixed number of marked points as elements of some space. There are many choices of such spaces, the most notable ones include the Teichmuller space and

[1]This chapter is based on: E. Arzhakova, G. Calsamiglia, B. Deroin, Isoperiodic moduli spaces of meromorphic forms, in progress

the moduli space.

- The Teichmuller space. Fix a reference surface – an oriented closed surface S of genus g with $n \geqslant 0$ ordered distinct marked points. The Teichmuller space $T(S)$ associated to S is a space of equivalence classes of pairs (X, f) where X is a Riemann surface with n ordered distinct marked points and $f\colon S \to X$ is a diffeomorphism between the surfaces that maps the ordered marked points of S to the ordered marked points of X. The equivalence in $T(S)$ is described as follows: two pairs (X_1, f_1) and (X_2, f_2) are equivalent if $f_1 \circ f_2^{-1}\colon X_2 \to X_1$ is isotopic to a holomorphic diffeomorphism. The Teichmuller space is connected and has a canonical complex manifold structure [3, 8].

  Example 6.1.1 (Teichmuller space of a torus)**.** Consider the reference surface to be a torus $\mathbb{T} = \mathbb{R}^2/\mathbb{Z}^2$. Any complex structure on a torus can be realised by a Riemann surface of a form $\mathbb{C}/(\mathbb{Z} + \tau\mathbb{Z})$ where $\tau \in \mathbb{C}$ is such that $\mathrm{Im}\tau > 0$. Note that such complex numbers form an upper half-plane $\mathbb{H} = \{z \in \mathbb{C}\colon \mathrm{Im}z > 0\}$. The map $\mathbb{H} \to T(\mathbb{T})$ given by $\tau \mapsto \mathbb{C}/\langle 1, \tau\rangle$ is a bijection and, therefore, $T(\mathbb{T}) = \mathbb{H}$.

- The moduli space. The mapping class group $\mathrm{Mod}(S)$ of a surface S is the group of isotopy classes of homeomorphisms of S that fix each marked point. In other words, $\mathrm{Mod}(S) = \mathrm{Homeo}(S)/\mathrm{Homeo}_0(S)$ where $\mathrm{Homeo}_0(S)$ are the homeomorphisms isotopic to identity. The moduli space $\mathcal{M}_S$ is given by the quotient $T(S)/\mathrm{Mod}(S)$. In fact, the moduli space depends only on the genus g and number of marked points n of the surface S, therefore, it is usually denoted as $\mathcal{M}_{g,n}$. The moduli space has an orbifold structure, it is typically not a manifold.

  Example 6.1.2 (Moduli space of a torus)**.** The mapping class group of a torus $\mathrm{Mod}(\mathbb{T})$ is isomorphic to $SL(2, \mathbb{Z})$. It follows that $\mathcal{M}_{1,0} = \mathbb{H}/SL(2, \mathbb{Z})$.

Differential forms can be defined on the Riemann surfaces, in particular, the space of the 1-forms is the dual vector space to the tangent space of a surface. In a local coordinate z given by the complex structure a differential 1-form can be written as $\omega = f(z)\,dz$.

- A 1-form ω on a surface X is called *a holomorphic differential* if $f(z)$ is holomorphic, i.e., a complex differentiable function. Denote by N the set of zeroes of ω; $N = \{x \in X\colon \omega(x) = 0\}$. Then, $X\backslash N$ inherits a flat metric and in the neighbourhood of a zero this metric admits a conical singularity of angle $2\pi(k+1)$. In other words, in the neighborhood of a zero the 1-form ω is locally given by $z^k\, dz, k \geqslant 1$. In this case, we say that k is the order of the zero. The flat metric defined locally by a zero of order k is a ramified covering over a flat disk of order $k+1$ that is ramified at zero.

- A 1-form ω on a surface X is called *a meromorphic differential* if $f(z)$ is a meromorphic function. i.e., it is holomorphic everywhere except in a discrete set of points that are called the poles of the function. Locally at the pole $\omega = z^{-k}\, dz$ and k is the order of the pole. The singularities of a pair (X, ω) consist of poles and zeroes; let us define the degree of a singularity to be k if it is a zero of order k, and $-k$ if it is a pole of order k. Denote by n_i the degree of the i-th singularity of X; then, as a consequence of the Riemann-Roch theorem, we obtain that $\sum n_i = 2g-2$ [102].
  Select a pole on X and denote it by p. Choose a short closed curve γ_p going around p, i.e., a curve that has no other singularity in its interior and does not wind around genus. The complex number $\mathrm{res}_p(\omega) = \frac{1}{2\pi i}\int_{\gamma_p} \omega$ is called the residue of the form ω at p and it does not depend on the choice of γ_p. Let (X, ω) be a Riemann surface with poles $p_1, \dots, p_n$. Then, by the residue theorem, $\sum_1^n \mathrm{res}_p(\omega) = 0$ [102].

Consider a closed curve c on a Riemann surface X and introduce a 1-form ν_c such that for every closed 1-form α, $\int_c \alpha = -\int \alpha \wedge \nu_c = (\alpha, \star\nu_c)$ where $\star$ is the Hodge star. Then, we define an intersection of two closed curves a and b on X as $a \cdot b = \int \nu_a \wedge \nu_b$. The intersection form is an anti-symmetric form with its image contained in $\mathbb{Z}$ and the intersection of a and b depends only on the homology classes of a and b. Denote a surface of genus g with n poles as $\Sigma_{g,n}$. The intersection form enables us to select a basis of the fundamental group $\pi_1(\Sigma_{g,n})$ given by $\{a_1, b_1, \dots, a_g, b_g; \pi_1, \dots, \pi_n\}$ such that $a_i \cdot b_j = \delta_{i,j}, a_i \cdot \pi_j = b_i \cdot \pi_j = \pi \cdot \pi_j = 0$.

In the follow-up we are interested in comparing integrals of 1-forms over a basis of $H_1(X, \mathbb{Z})$. Therefore, it is natural to seek some space of Riemann surfaces that identifies the curves in $H_1(X, \mathbb{Z})$ for different X. The most convenient space for this goal is the Torelli space and it is defined as follows. Denote by Σ_{g,n^*} a surface obtained by making punctures at each marked point of $\Sigma_{g,n}$, i.e., $\Sigma_{g,n^*} = \Sigma_{g,n} \backslash N$ where N is the set of n marked points. The subgroup $\mathcal{I}_{g,n}$ of $\mathrm{Mod}(\Sigma_{g,n})$ that acts trivially on $H_1(\Sigma_{g,n^*})$ is called the

Torelli group of Σ_{g,n^*}. The Torelli space $\mathcal{S}_{g,n}$ is given by $T_{g,n}/\mathcal{I}_{g,n}$ and the pullback of the 1-form bundle $\Omega\mathcal{M}_{g,n}$ over the moduli space by the cover $\mathcal{S}_{g,n^*} \to \mathcal{M}_{g,n}$ is denoted by $\Omega\mathcal{S}_{g,n^*}$. A point in $\Omega\mathcal{S}_{g,n^*}$ is therefore described by a fourtuple $(X, N, [f], \omega)$ where $[f]$ is the equivalence class of the homotopical collapse map f under the action of the Torelli group.

Definition 6.1.3. The map $\mathrm{Per}_{g.n}\colon \Omega\mathcal{S}_{g,n} \to \mathrm{Hom}(H_1(\Sigma_{g,n^*}, \mathbb{C})$ is defined by the formula

$$\mathrm{Per}_{g,n}(X, N, [f], \omega) = \Big\{p\colon \gamma \to p(\gamma) = \int_{f_*\gamma} \omega\Big\}.$$

The map p is called the period map of ω and it provides coordinates on the space $\Omega\mathcal{S}_{g,n}$. Notably, the period coordinates do not allow to recover the differential even on infinitesimal level [17]. Indeed, the isoperiodic deformations define a foliation of $\Omega\mathcal{S}_{g,n}$ which is called the isoperiodic foliation. Some of the first results on the isoperiodic foliation of holomorphic differentials over the moduli space include that the isoperiodic leaves are Euclidean spaces with complete metric [81]. This work also includes the study of the isoperiodic sets in the holomorphic case with $g = 2$. Then, in a fundamental work of Calsamiglia, Deroin, and Francaviglia [17] it was shown that the leaves of the isoperiodic foliation of holomorphic differentials are connected for $g \geqslant 2$ and primitive degree at least three. The method used in [17] involves the degeneration of the Riemann surface into a nodal curve which allows to simplify the problem to surfaces of lower genera.

The method proposed in [17] cannot be applied to the meromorphic case because meromorphic differentials can have real periods (i.e., the image $\mathrm{Im}p$ is contained in $\mathbb{R}$). It is not possible to degenerate a meromorphic form with real periods into a union of forms that includes holomorphic parts because holomorphic forms do not admit real periods (a consequence of Riemann's bilinear relations, [9]). Therefore, the case of real periods of meromorphic forms requires new tools in order to prove connectivity of the isoperiodic sets. In [18] the non-emptiness and connectivity of the isoperiodic leaves in $\Omega\mathcal{S}_{g,n}$ is shown for meromorphic forms with 2 poles and $g \geqslant 1$ of degree at least 3. However, the method of proof relies on the combinatorial properties of having exactly two poles. Our result is the extension of the study of connectivity of the isoperiodic leaves to a higher number of poles in the case of real period. Denote by $\Sigma_{1,3}$ a surface of genus 1 with 3 marked points:

Theorem 6.1.4. Denote by Π the peripheral module of $\Sigma_{1,3}$, i.e., a module Π such that $H_1(\Sigma_{1,3}, \mathbb{Z}) = H_1(\Sigma_1, \mathbb{Z}) + \Pi$; let $p \in \mathrm{Hom}(H_1(\Sigma_{1,3^*}, \mathbb{Z}), \mathbb{C})$ be a period map. If the image of p in $\mathbb{C}$ is real, then the level $\mathrm{Per}^{-1}(p)$ is connected in $\Omega\mathcal{S}_{1,3}$ if the image of p is not contained in the $\mathbb{Q}$-vector space generated by $p(\Pi)$.

This result is a new step towards proving the connectivity of the isoperiodic sets in general case. The strategy of treating the cases of higher genera and larger number of poles often relies on the degeneration into simpler bits and applying induction. Therefore, the result of Theorem 6.1.4 is meant to serve as a base of induction for our work in progress that studies the connectivity of the leaves of the isoperiodic foliations with real periods. We emphasise that the benefit of the geometrical method used in the present work is that it can be applied to the surfaces of any (small) genus and number of poles unlike the methods used in [18, 59]. However, we believe that the complexity of the proof is growing very fast with larger genus and larger number of poles. To demonstrate the universality of the method for low g, n, we supplement the proof of Theorem 6.1.4 with Appendix which contains the proof of a similar statement for $\Sigma_{1,2}$.

The strategy of the present work relies on applying a local surgery called the Schiffer variation [103]. The Schiffer variation is an isoperiodic surgery of the surface which provides a tool to connect two different forms in $\Omega\mathcal{S}_{1,3}$ by an isoperiodic path. It is performed by selecting two twins leaving or entering the zero in the same direction with an angle 2π between them and making two cuts of the same length along them. Then the sides of the two cuts are re-glued: the left side of the first cut is glued to the right side of the second cut, and vice verca. In this manner the zero can be isoperiodically moved to a different position on the surface.

The main directions of our further study is proving the connectivity of the isoperiodic sets both in real and complex cases for larger genus and number of poles using induction. Another interesting direction of research is to check the ergodicity of the isoperiodic foliation following [17, 18, 48]. An interesting approach of proving ergodicity of a real isoperiodic foliation of forms with a single double pole is proposed in [59] and involves using the cut diagrams.

6.2 Rigid forms

In this Section we introduce am important subclass in $\Omega\mathcal{S}_{1,3}$ of rigid forms, i.e., forms with one zero. We show that any form in $\Omega\mathcal{S}_{1,3}$ can be connected to a rigid form; therefore, in order to prove Theorem 6.1.4 it suffices to prove the isoperiodic connectedness of the rigid forms. Therefore, it is natural to study possible topological types of the rigid forms.

Definition 6.2.1 (A rigid form)**.** A rigid form $(X, N, [f], \omega) \in \Omega\mathcal{S}_{g,n}$ is a form with a single zero. By the Riemann-Roch theorem, the multiplicity of the single zero is equal to $2g - 2 + n$ and the angle at the zero is $2\pi(2g - 1 + n)$.

Example 6.2.2. A rigid form in $\Omega\mathcal{S}_{1,3}$ has a single zero of order 3 with an angle 8π around it.

The importance of the subclass of the rigid forms in the context of Theorem 6.1.4 is explained in the following lemma:

Lemma 6.2.3. Any form in $\Omega\mathcal{S}_{1,3}$ with real period map is isoperiodically connected to a rigid form.

Proof. A generic form in $\Omega\mathcal{S}_{1,3}$ has three simple zeroes and the distances between zeroes do not coincide. Let us select two zeroes with the shortest saddle connection between them. This saddle connection has a twin of the same length emerging at an angle 2π from one of the two zeroes. If this twin ends at a regular point, performing Schiffer variation along the saddle connection and its twin yields a double zero which is not a node. In the same manner we can connect the double zero to the remaining simple zero, thus, obtaining a rigid form. If the twin described above ends in the same zero, i.e., forms a loop, an infinitesimal perturbation of the form will ensure that it ends at a regular point. Then, the previous argument applies.

If the twin ends at the second zero then the double zero formed by the corresponding Schiffer variation is a node. Note that since the cycle formed by two twins integrates to zero, the node has zero residue. We can show by contradiction that it is a non-separating node. First, it cannot separate the surface into a component with three poles and holomorphic component since holomorphic component does not allow real periods. It also cannot separate one pole from the other two by the residue theorem. We conclude that it is a non-separating node. In this case, we apply the degeneration

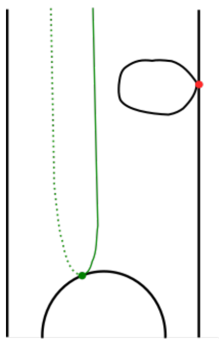


Figure 6.1: A decomposition of the surface $\Sigma_{1,3}$ with a double zero into a sphere with 2 poles and a torus with 2 poles and a double zero.

technique in order to connect such form to a rigid form.

Let us perform a Schiffer variation along two twins that leave the remaining simple zero in the positive imaginary direction: it results in a decomposition of the initial surface into a sphere with 2 poles and no zeroes and a torus with 2 poles and a double zero which is a node (see Figure 6.1). The toral component can be isoperiodically perturbed into a torus with 2 simple poles and 2 zeroes. In [18] it is shown that a torus with 2 simple poles and 2 zeroes can be isoperiodically deformed into a rigid form with 2 simple poles. In the end, we glue the two parts by selecting a geodesic in the positive imaginary direction emerging from an arbitrary point on the sphere, and a geodesic on the torus leaving the double zero in the positive imaginary direction. We glue the surfaces along these geodesics obtaining a torus with three poles and a single zero of the third order. $\square$

Lemma 6.2.3 implies that to prove Theorem 6.1.4 it suffices to consider only the rigid forms. Therefore, it is natural to study further the structure and types of the rigid forms in $\Omega\mathcal{S}_{1,3}$. We start the discussion with studying the separatrices in real directions that pass through the zero of the rigid form in $\Omega\mathcal{S}_{1,3}$.

Select a regular point z_0 on $\Sigma_{1,3}$ and consider an integral $f(z) = \int_{z_0}^{z} \mathrm{Im}\omega$. Since the periods of ω are real the imaginary part of ω does not have monodromy, i.e., f is a real-valued univalent function $f\colon X \to \mathbb{R}$. The levels of f are the leaves of the real foliation on X. Since residues around each pole are real, one can view poles as semi-infinite annuli where all real leaves are closed and compact manifolds of dimension 1 and no leaf is minimal even

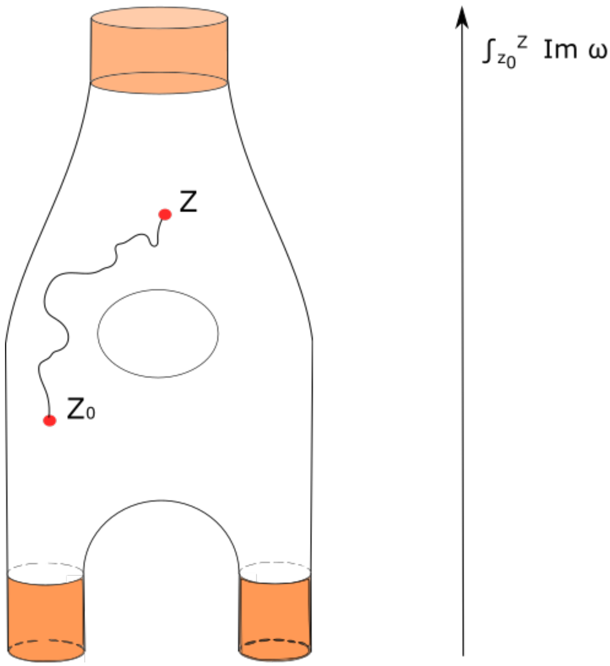


Figure 6.2: Integral $\int_{z_0}^{z} \mathrm{Im}(\omega)$ of the imaginary part of a 1-form ω defines a function $f(z)$ on the torus X. The levels of the function are the leaves of the real foliation of ω. These leaves form closed compact 1-dimensional manifolds in the neighbourhood of the poles.

locally (see Figure 6.2). We conclude that any saddle connection leaving a zero in real direction cannot escape to a pole because it cannot transversally cross the leaves of the real foliation. Instead, each saddle connection leaving a zero in real direction has to end in some zero.

It follows for a rigid form in $\Omega\mathcal{S}_{1,3}$ that all 4 separatrices that leave the zero along real directions come back to the zero along the real directions. Moreover, the outgoing and ingoing real separatrices alternate in order.

6.2.1 Octopodes and butterflies

The outgoing separatrices in real directions have to return to the zero; in this subsection we investigate in which order the separatrices return. The order defines the topological type of the rigid form. There are two ways to graphically depict a topological type of a rigid form (see Figure 6.4):

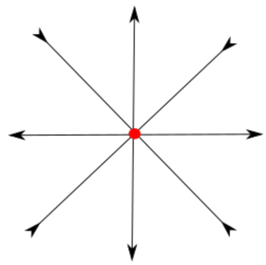


Figure 6.3: The structure of a zero of a rigid form in $\Omega\mathcal{S}_{1,3}$: the angle around the zero is 8π and in real directions there are 4 outgoing and 4 ingoing separatrices that alternate in order.

- **Radial diagram.** The radial diagram features a zero in its center and the saddle connections leaving the zero and entering the zero. The order in which the in-going and out-going separatrices are connected defines the topological type of the rigid form.

- **Circle diagram.** Both sets of in-going and out-going separatrices in the neighborhood of the zero are presented as two sets of points on a circle. Each out-going point is bijectively connected to an in-going point. The order in which they are connected defines the topological type of the rigid form.

The correspondence between the radial diagrams and the circle diagrams is easy to establish: for the convenience of the reader, we demonstrate it on Figure 6.4. In the Figures hereafter we will be using circle diagrams. It turns out that the order in which the outgoing separatrices enter the zero is not arbitrary, but restricted by the topology of $\Sigma_{1,3}$ to few options as we see in the following Lemma.

Lemma 6.2.4. Up to the change of orientation, there are only 2 possible combinatorial types of rigid forms in $\Sigma_{1,3}$.

Proof. Each separatrix that leaves the zero of the third order must eventually come back: there are 4 separatrices leaving the zero and 4 entering it. Let us number both sets counter-clockwise from 1 to 4. We need to understand in which order the separatrices that leave the zero come back: it is natural to think about this correspondence as of possible permutations on 4 elements.

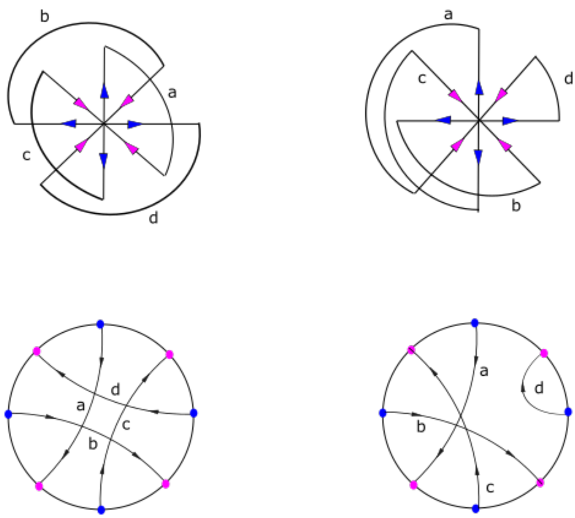


Figure 6.4: Correspondence between the radial (top) and the circle (bottom) diagrams. Left: radial and circle diagrams of the butterflies form, right: radial and circle diagrams of the octopus form.

The geometrical type of the zero does not depend on the rotation of the chosen numbering: therefore, one does not need to consider all possible 24 permutations separately. Instead, it suffices to consider only 10 conjugacy classes by two cyclic permutations (1234) and (1432):

1. constant permutation $(1)(2)(3)(4)$ forms a conjugacy class of 1 element;
2. transpositions of 2 neighboring elements $(12)(3)(4) \to (23)(4)(1) \to (34)(1)(2) \to (41)(2)(3)$ form a class of 4 elements;
3. transpositions of 2 non-neighboring elements $(13)(2)(4) \to (24)(3)(1)$ form a class of 2 elements;
4. 2 non-intersecting transpositions of two pairs of neighboring elements $(12)(34) \to (23)(14)$ form a class of 2 elements;
5. 2 non-intersecting transpositions of two pairs of non-neighboring elements $(13)(24)$ forms a class of 1 element;
6. three-cycles oriented counterclockwise $(123)(4) \to (234)(1) \to (341)(2) \to (412)(3)$ form a class of 4 elements;

7. three-cycles oriented clockwise $(132)(4) \to (243)(1) \to (314)(2) \to (421)(3)$ form a class of 4 elements;

8. four-cycle oriented counterclockwise (1234) forms a class of 1 element;

9. four-cycle oriented clockwise (1432) forms a class of 1 element;

10. four-cycles with transposition $(1243) \to (2314) \to (3421) \to (4132)$ form a class of 4 elements.

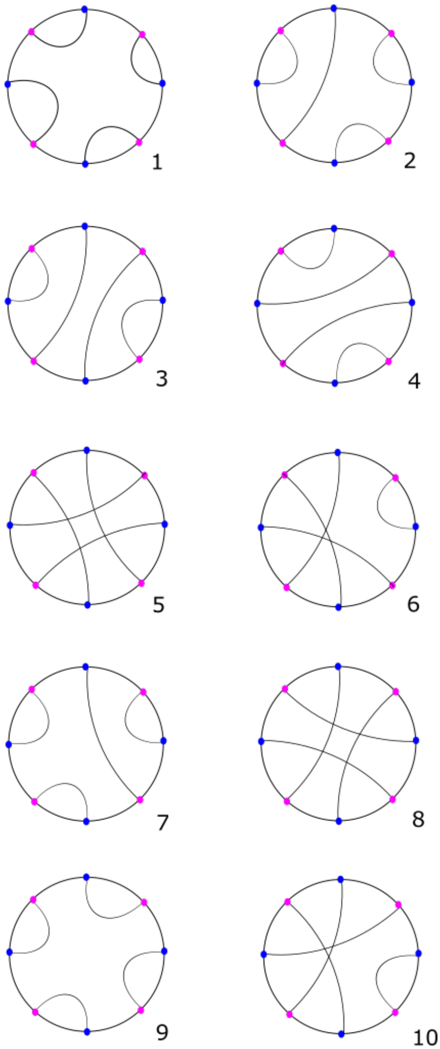


Figure 6.5: Possible combinatorics at a zero of order 3. Figures 1-4, 7,9 correspond to $g = 0, n = 4$, Figures 5,6,8,10 correspond to $g = 1, n = 3$.

If we cut the surface along the paths following the directions of separatrices on the left hand side and on the right hand side we will obtain the decomposition of the surface into several semi-infinite cylinders. The considered surface $\Sigma_{1,3}$ can only be decomposed into 3 semi-infinite cylinders because it has three poles. The Figure 6.5 shows that there are only 2 possible combinatorial types where the surface decomposes into 3 semi-infinite cylinders: the first corresponding to cases $(5), (8)$ and the second corresponding to the cases $(6), (10)$. The rest of diagrams represent spheres with 5 punctures. $\square$

Let us discuss the two topological types of rigid forms more closely. In order to do it we agree on the following conventions and notations:

The surface $\Sigma_{1,3}$ has three poles whose real residues sum up to zero, so it either has 1 or 2 poles with positive residues. The number of poles with positive residue defines the orientation of the associated form; without loss of generality we assume that all rigid forms have 1 positive pole and 2 negative poles. Then, changing the signs of all residues corresponds merely to switching the orientation of the surface.

Denote the positive pole by s_+ and the two negative poles by $s_-^{\leq}$ and $s_-^{\geq}$ where $s_-^{\geq}$ stands for the negative pole with a residue equal or larger in absolute value than the residue of $s_-^{\leq}$. Denote the short closed curves going around the poles as π_+, $\pi_-^{\leq}$ and $\pi_-^{\geq}$, respectively, and note that $p(\pi_+) + p(\pi_-^{\leq}) + p(\pi_-^{\geq}) = 0$ by the residue theorem.

Let us define the two types of the rigid forms:

- **Butterflies** We say that a rigid form associated with marking (a, b, c, d) is in $\Omega\mathcal{S}_{1,3}$ and is of type "butterflies" if
  - $a \cdot b = b \cdot c = c \cdot d = d \cdot a = 1,\ a \cdot c = b \cdot d = 0$;
  - $a + c = -\pi_-^{\geq},\ b + d = -\pi_-^{\leq},\ a + b + c + d = \pi_+$;
  - $p(a), p(b), p(c), p(d) > 0$.

  An example of radial and circle diagrams of butterflies is provided on Figure 6.4. For convenience, we denote by $B(a, c \,|\, b, d)$ a butterflies form that satisfies the three conditions listed above.

- **Octopus** We say that a rigid form associated with marking of curves (a, b, c, d) is in $\Omega\mathcal{S}_{1,3}$ and is of type "octopus" if

- $a \cdot b = b \cdot c = c \cdot a = 1$, $a \cdot d = b \cdot d = c \cdot d = 0$;
- $a + b + c = -\pi_-^{\geq}$, $d = -\pi_-^{\leq}$ or
  $a + b + c = -\pi_-^{\leq}$, $d = -\pi_-^{\geq}$;
- $a + b + c + d = \pi_+$;
- $p(a), p(b), p(c), p(d) > 0$.

An example of radial and circle diagrams of octopus is provided on Figure 6.4. Note that the connection d is distinguished from the other three connections: it is the connection that starts and finishes at adjacent points. We call the connection d *the head of the octopus*. The head of the octopus generates a closed curve that goes around a pole. By the length considerations, this pole cannot be positive, so it is one of the negative poles (see Figure 6.6). Therefore, there are two distinct cases: $a+b+c = -\pi_-^{\geq}, d = -\pi_-^{\leq}$ and $a+b+c = -\pi_-^{\leq}, d = -\pi_-^{\geq}$. We call a form associated to the first case *the small head octopus* (SHO) and a form associated to the second case *the large head octopus* (LHO). For convenience, we denote by $O(a, b, c \,|\, d)$ an octopus form that satisfies the three conditions listed above. When the distinction between the large head octopodes and the small head octopodes is important, we use $LHO(a, b, c \,|\, d)$ and $SHO(a, b, c \,|\, d)$, respectively.

Note that any octopus admits two different orientations and it has to do with the signs of the residues at the three poles (one positive and two negative, or two positive and one negative). However, there is an equivalence between the "left-handed" and "right-handed" octopodes achieved by switching the signs of the residues; hence, there is no necessity to distinguish between these two cases.

6.2.2 Schiffer variation on circle diagram.

To perform the Schiffer variation at a zero on the circle diagram we select two neighbouring (i.e., with angle 2π in between) separatrices of the same direction and of different lengths. Denote the point corresponding to the shorter separatrix as "Short" and the point corresponding to the shorter separatrix as "Long". Denote the shorter separatrix by S and the longer separatrix by L. There is a unique separatrix between them that goes in an opposite direction - denote it by C and the corresponding point by "Central". Perform the Schiffer variation along L and S along the whole length of S: it

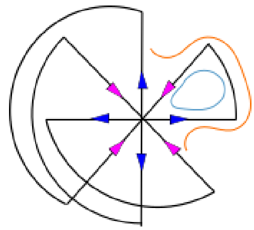


Figure 6.6: The blue curve inside the head of the octopus goes around one of the poles. The blue curve is shorter than the orange curve whose fragment is shown on the Figure. Since $\pi_+ = \pi_-^{\leq} + \pi_-^{\geq}$, the blue curve cannot go around n_+ by the length consideration.

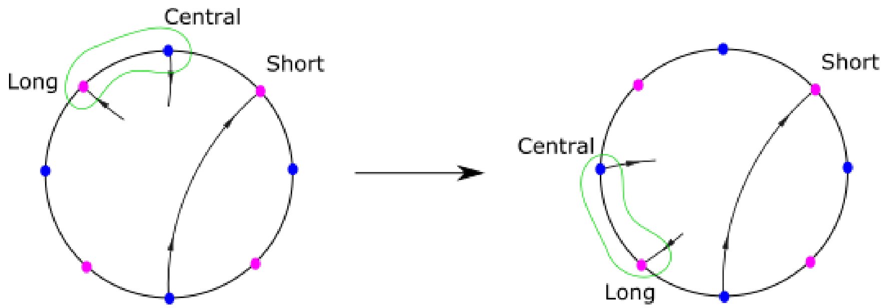


Figure 6.7: The Schiffer variation on a circle diagram: the group (L, C) of the endpoints of the long and the central connections are "sliding" all the way to the opposite endpoint of the short connection.

yields a new rigid form where the zero is in the same position but the length and order of the saddle connections is changed.

The lengths of the saddle connections will be $p(L') = p(L) - p(S), p(S') = p(S), p(C') = p(C + S)$. The positions of the points "Long" and "Central" shift to the opposite end of the short connection S (see Figure 6.7).

Remark 6.2.5. Note that if the two points "Long" and "Central" are the only two points inside the arch of S then the order of the points does not change.

6.2.3 Connecting different types of rigid forms in $\Omega\mathcal{S}_{1,3}$.

According to Lemma 6.2.4, each rigid form in $\Omega\mathcal{S}_{1,3}$ is either butterflies or octopus. In this subsection we show how to isoperiodically connect any rigid form to a large head octopus under mild conditions. The first step is to connect a small head octopus to butterflies and the second step is to connect the butterflies to the large head octopus.

Lemma 6.2.6. If the period map of a small head octopus form p is not contained in the $\mathbb{Q}$-vector space generated by $p(\Pi)$, it can be connected to a butterflies form.

Proof. Consider an octopus $SHO(a, b, c \mid d)$. If $p(d) > p(c)$ (i.e., if the opposite arm is shorter than the head), then if we perform the Schiffer variation along d and c we will reach butterflies in one step. If $p(d) \leqslant p(c)$ we treat several cases:

- $p(a)$ and $p(b)$ are *not rationally dependent*. Without loss of generality, assume $p(a) > p(b)$. There are two possible types of Schiffer variations along a and b: one changes the order of the marking (namely, $O(a-b, b, c+b \mid d)$), and the other does not (namely, $O(b, c+b, a-b \mid d)$). By performing Schiffer variation along a and b that does not change the order of the marking we can make $p(a)$ and $p(b)$ arbitrarily small while $p(c)$ grows (similar to Euclidean algorithm). Then, we perform the Schiffer variation along a and b that changes the order of arms; depending on their periods, one of them becomes the arm of the octopus which is opposite to the head. The period of the opposite arm is now smaller than the period of the head, therefore, it can be connected to butterflies.

- If $p(a)$ and $p(b)$ are *rationally dependent*. All four periods $p(a), p(b), p(c)$ and $p(d)$ cannot be rationally dependent because it contradicts the assumption on the image of the period map not being contained in the $\mathbb{Q}$-vector space generated by $p(\Pi)$. If $p(c)$ is not rationally dependent with $p(a)$ we perform a Schiffer variation along a and c which does not change the order of the marking. Now the two arms that are not opposite of the head are not rationally dependent: we proceed as in the previous cases. If $p(c)$ is rationally dependent with $p(a)$, then $p(d)$ is not rationally dependent with any of them. We perform a Schiffer variation along c and d which results in an octopus with not all arms rationally dependent. We proceed as in previous cases.

□

Lemma 6.2.7. If the image of p of the butterflies form is not contained in the

$\mathbb{Q}$-vector space generated by $p(\Pi)$, the butterflies form can be connected to a large head octopus form.

Proof. Label the butterflies as $B(a, c \,|\, b, d)$ and assume wlog that $p(a+c) \geqslant p(b+d)$. With an appropriate choice of direction, performing a Schiffer variation along neighbouring saddle connections yields an octopus. In this manner one can reach at most 4 different octopodes in one Schiffer variation (see Figure 6.8):

- using saddle connections a and d: if $p(a) < p(d)$, we can reach an octopus $O(d-a, b, a \,|\, a+c)$ with head $(a+c)$,and if $p(d) < p(a)$, we can reach an octopus $O(c, a-d, d \,|\, b+d)$ with head $(b+d)$;
- using saddle connections a and b: if $p(a) < p(b)$, we can reach an octopus $O(d, b-a, a \,|\, a+c)$ with head $(a+c)$, and if $p(b) < p(a)$, we can reach an octopus $O(a-b, c, b \,|\, b+d)$ with head $(b+d)$;
- using saddle connections b and c: if $p(b) < p(c)$, we can reach an octopus $O(a, c-b, b \,|\, b+d)$ with head $(b+d)$, and if $p(c) < p(b)$, we can reach an octopus $O(b-c, d, c \,|\, a+c)$ with head $(a+c)$;
- using saddle connections c and d: if $p(c) < p(d)$, we can reach an octopus $O(b, d-c, c \,|\, a+c)$ with head $(a+c)$, and if $p(d) < p(c)$, we can reach an octopus $O(c-d, a, d \,|\, b+d)$ with head $(b+d)$.

There are two types of the head that we can obtain in this manner: $a+c$ and $b+d$, which corresponds to the large head octopodes and the small head octopodes. If we are able to reach an octopus with the head $a+c$, i.e., a large head octopus, we have confirmed the statement of the lemma. If we are not able to reach a large head octopus, then by the list above it follows that $p(b), p(d) \leqslant p(a), p(c)$. Note that it is not possible that all four pairs $(p(a), p(b))\,, (p(a), p(d))\,, (p(c), p(b))\,, (p(c), p(d))$ are rationally dependent because it contradicts the assumption of the lemma. Without loss of generality, assume that the pair $(p(a), p(b))$ is rationally independent.

By assumption, $p(b) < p(a)$; using Schiffer variations along a and b several times we can go to another couple of butterflies $B(a - q \times b, c + q \times b \,|\, b, d)$ where $q \in \mathbb{Z}^+$ is chosen such that $0 < p(a - q \times b) < p(b)$. Using the fact that $p(a-q\times b) < p(b)$, we use a corresponding Schiffer variation to go to an octopus $O(d, (q+1)\times b - a, a - q \times b \,|\, a+c)$ which is a large head octopus. $\square$

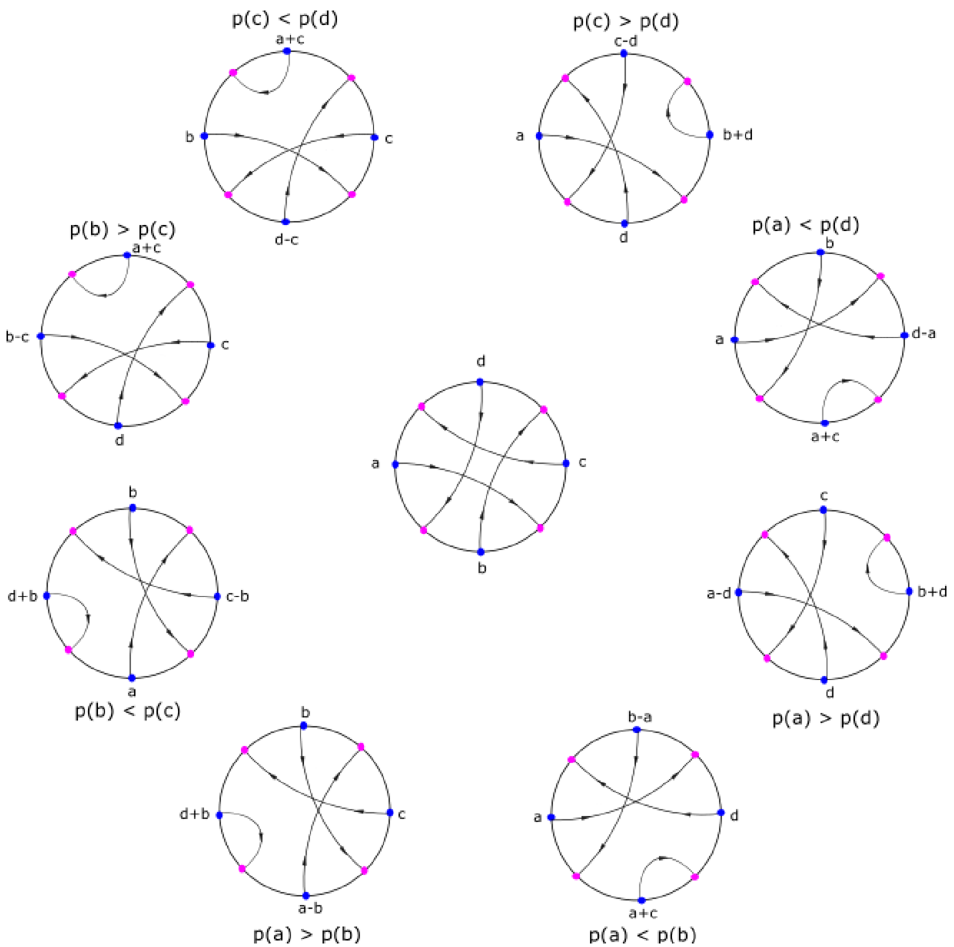


Figure 6.8: Possible octopodes that can be reached from a butterflies form.

Corollary 6.2.8. To prove Theorem 6.1.4 it suffices to show that any two LHO with same real periods are isoperiodically connected if the image of their period map p is not contained in $\mathbb{Q}p(\Pi)$.

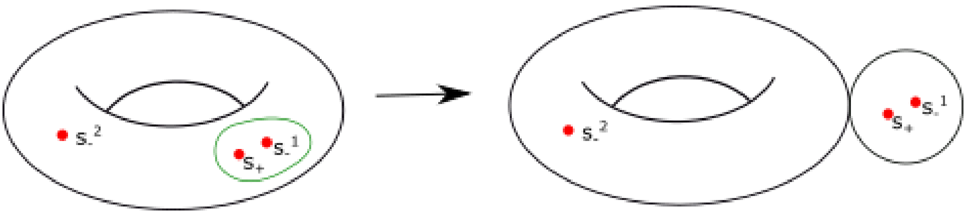


Figure 6.9: Arm module $M_{s_-^1}$ and the corresponding decomposition of $\Sigma_{1,3}$.

6.2.4 Arm modules

In order to study the octopodes and the isoperiodic connections between them it is useful to introduce the arm modules. The idea is to degenerate the surface $\Sigma_{1,3}$ into a nodal stable curve. This degeneration induces a decomposition of the group $H_1(\Sigma_{1,3},\mathbb{Z})$. Select one of the negative poles s_-^1 of $\Sigma_{1,3}$ and consider a curve that goes around it and the positive pole, does not wind around genus and does not have zeroes and the third pole in its interior. Using Schiffer variation we can degenerate this curve to a node, thus, decomposing $\Sigma_{1,3}$ into the union of Σ_{1,s_-^1} and Σ_{0,s_+,s_-^2} (see Figure 6.9).

Definition 6.2.9. An arm module $M_{s_-^1}$ associated to a negative pole s_-^1 is a rank 3 submodule of $H_1(\Sigma_{1,3},\mathbb{Z})$ such that

$$H_1(\Sigma_{1,3},\mathbb{Z}) = M_{s_-^1} + \Pi_{s_-^1}, \tag{6.1}$$

where $M_{s_-^1}$ has rank 3 and $\Pi_{s_-^1} = \{\pi(s_-^1), \pi(s_+)\}$ has rank 2. The two modules intersect: $M_{s_-^1} \cap \Pi_{s_-^1} = \mathbb{Z}\pi(s_-^2)$. The arm module $M_{s_-^1}$ has a basis a, b, c where $p(a), p(b), p(c) > 0$ and $a \cdot b = b \cdot c = c \cdot a = 1$. Moreover, $a+b+c = -\pi_{s_-^2}$.

The map $\mu_{s_-^1} : H_1(\Sigma_{1,3},\mathbb{Z}) \to H_1(\Sigma_{1,2},\mathbb{Z})$ restricted to the arm module $M_{s_-^1}$ is an isomorphism. The form $p_{M_{s_-^1}} := p \circ \mu_{s_-^1}^{-1} \in \mathrm{Hom}(H_1(\Sigma_{1,2},\mathbb{Z}),\mathbb{R})$ is called the period of an arm module. Define the quotient arm modules of $M_{s_-^{\le}}$ and $M_{s_-^{\ge}}$ as $M_{s_-^{\le}}/\mathbb{Z}\pi_-^{\ge}$ and $M_{s_-^{\ge}}/\mathbb{Z}\pi_-^{\le}$, respectively. Without loss of generality, let us consider a quotient arm module $M_{s_-^{\le}}/\mathbb{Z}\pi_-^{\ge}$: the map $q : H_1(\Sigma_{1,2},\mathbb{Z}) \to H_1(\Sigma_1,\mathbb{Z})$ has $\pi_-^{\ge}$ in its kernel, and induces an isomorphism between $M_{s_-^{\le}}/\mathbb{Z}\pi_-^{\ge}$ and $H_1(\Sigma_1,\mathbb{Z})$. The reduction $p_N := p_M \mod \mathbb{Z}p(\pi_-^{\ge}) \in H^1(\Sigma_1, \mathbb{R}/\mathbb{Z}p(\pi_-^{\ge}))$ is well-defined and called the period of the quotient arm module.

The set of arm modules and the set of quotient arm modules associated to a given pole both have a natural structure of affine space directed by

$H^1(\Sigma_1, \mathbb{Z})$; in other words, given two arm modules M, M' (resp. quotient arm modules N, N') associated to the pole $s_-^{\geq}$, there exists a unique $\psi \in N^* \simeq H^1(\Sigma_1, \mathbb{Z})$ such that M' (resp. N') is the image of M (resp. N) by the map $id_{|M} + (\psi \circ q) \cdot \pi_-^{\geq}$ (resp. $id_{|N} + \psi \cdot \pi_-^{\geq}$). We will denote $M' = M + \psi$ and $N' = N + \psi$ in the sequel.

Proposition 6.2.10. To each octopus one can associate an arm module: it is M_s where s is the pole associated to the head. Two marked octopodes with the same head and the same arm module can be connected with a finite number of Schiffer variations.

Proof. By Figure 6.9 it suffices to appeal to the connectedness of the isoperiodic space of meromorphic forms on a torus with two poles [18], and a convenient attaching map (similar to Lemma 6.2.3 and Figure 6.1). □

6.3 Proof of the connectivity of real isoperiodic sets in $\Omega\mathcal{S}_{1,3}$

This Section contains the proof of Theorem 6.1.4. Proving Theorem 6.1.4 is equivalent to proving Corollary 6.2.8, i.e., proving that the large head octopodes are isoperiodically connected under the conditions of Theorem 6.1.4. We do it in several steps:

1. we show that we can connect $LHO(a, b, c \,|\, d)$ with head d and arm module $M = \{a, b, c\}$ to another LHO with arm module $M + a^*$ where $a^* \in H^1(\Sigma_1, \mathbb{Z})$ is the dual of a with respect to he intersection form. In other words, $a^*(a) = 0, a^*(b) = 1, a^*(c) = -1$ (Lemma 6.3.1);
2. using the previous step, we prove that a LHO with an arm module M can be connected to a LHO with arm module $M + \varphi$ if $p(\varphi^*) \notin \mathbb{Q}/\mathbb{Z}$ (Lemma 6.3.2);
3. we conclude the proof of Corollary 6.2.8 by showing that any two LHO can be connected if the image of the period map is not contained in $\mathbb{Q}p(\Pi)$ (Lemma 6.3.3).

Let us start with establishing the first step:

Lemma 6.3.1. Consider $LHO(a, b, c \,|\, d)$ with an arm module M generated by a, b, c. Then there exists a concatenation of a finite number of Schiffer variations that connects this octopus to a LHO with an arm submodule M' such that $M' = M + \psi$ with $\psi = a^*$ where a^* is defined as $\mathrm{Ker}(a^*) = a$.

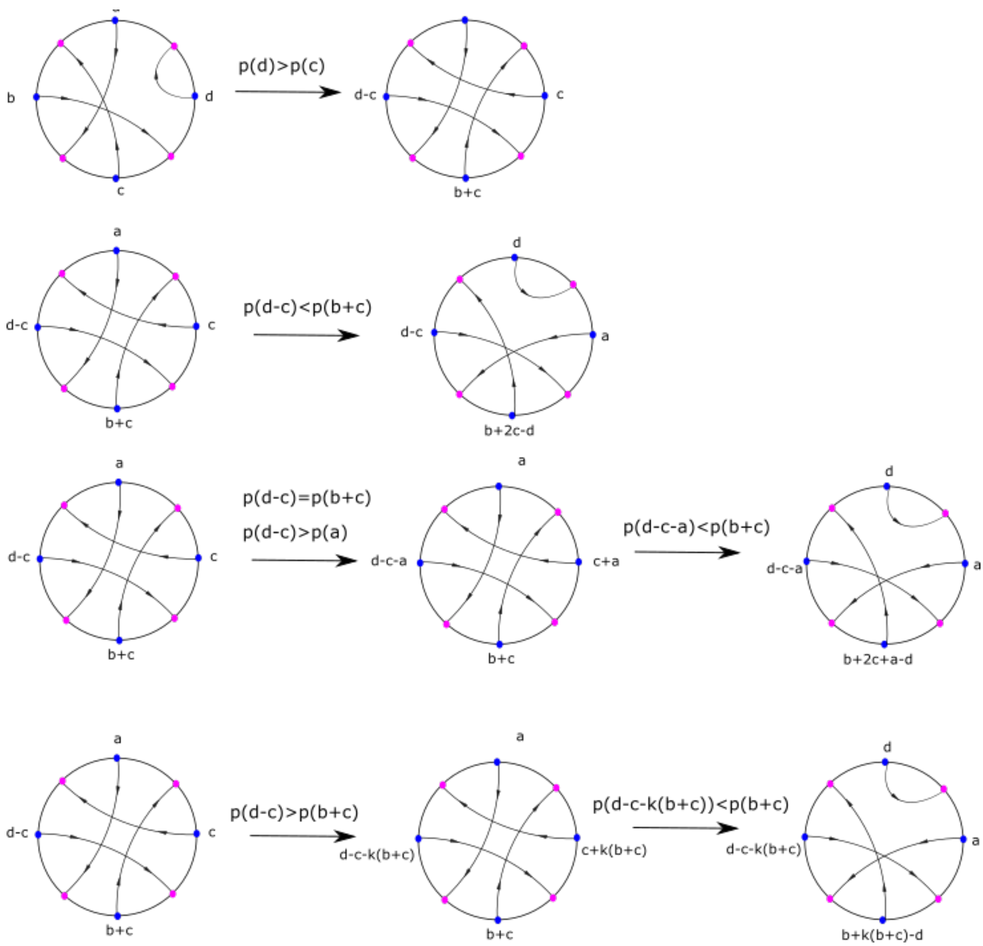


Figure 6.10: Graphical proof of Lemma 6.3.1.

Proof. Since the octopus has a large head, let us perform a Schiffer variation along the head d and the opposite arm c which will result in butterflies with marking $B(a, c|d-c, b+c)$. Consider the cases below - for graphical proof, see Figure 6.10:

1. First, let us consider the case when $p(d) - p(c) < p(b) + p(c)$. It follows that we can perform a Schiffer variation along $d-c$ and $b+c$ which will result in an octopus with a marking $O(b+2c-d, a, d-c \,|\, d)$.

2. Consider the case $p(d) - p(c) = p(b) + p(c)$; since $p(d) - p(c) > p(a)$, perform a Schiffer variation along a and $d-c$ which results in $B(a, b+c \,|\, d-c-a, c+a)$. Now $p(d) - p(c) - p(a) < p(b) + p(c)$ and so we can go to an octopus $O(b+2c+a-d, a, d-c-a \,|\, d)$.

3. Now consider that $p(d) - p(c) > p(b) + p(c)$. Let us perform k Schiffer variations that will result in $B(a, b + c \,|\, d - c - k \times (b + c), c + k \times (b + c))$ where $k \in \mathbb{Z}^+$ is such that $0 < p(d) - p(c) - k(p(b) + p(c)) < p(b) + p(c)$. After this we can perform a Schiffer variation along $d - c - k \times (b + c)$ and $b + c$, which will result in an octopus with a marking $O\left((k+1) \times (b + c) + c - d, a, d - c - k \times (b + c) \,|\, d\right)$.

The octopodes obtained in all three cases have large heads. Observe that in all cases $\psi = a^*$. Let us check it in the first case; M' is generated by $\{a, d-c, b+2c-d\}$. We see that $\psi(a) = 0$, $\psi(b) = 1$, $\psi(c) = -1$ and, therefore, $\psi = a^*$. The other two cases are similar. □

Two arm modules M and M' are connected (denoted by $M' \sim M$) if some LHO with an arm module M and head d is connected to some LHO with an arm module M' and head d. The connectedness of the set of arm modules implies that each octopus with arm module M is connected to each octopus with an arm module M' as we showed in Proposition 6.2.10. In the following lemma's we will be connecting two arm modules which should be understood as finding and connecting two representatives of each module.

Lemma 6.3.2. Let M, M' be arm modules of two large head octopodes and $M' = M + \psi$. If $p_M(\psi^*) \notin \mathbb{Q}/\mathbb{Z}$ then $M' \sim M$.

Proof. We need to construct an arm basis $\{a, b, c\}$ of M which satisfies the following three properties:

1. $a \mod \pi_-^{\geq} = \psi^*$;
2. $a \cdot b = b \cdot c = c \cdot a = 1$;
3. $p(a), p(b), p(c) > 0$;

By definition, $M \mod \pi_-^{\leq} = N$. Taking for simplicity $p(-\pi_-^{\leq}) = 1$ we see that $p_N \colon N \to \mathbb{R}/\mathbb{Z}$. Assume that $p_N(\psi^*) \notin \mathbb{Q}$ and consider the case when ψ^* is primitive.

Select $a_0 = \psi^*$ and $a = a_0 + k_a \pi_-^{\geq}$ such that $p(a) > 0$. By assumption, the number $p(a_0)$ is irrational. Select an element $b_0 \in N$ such that $a \cdot b_0 = 1$ and select k_b such that $p(b_0 + k_b \pi_-^{\geq}) > 0$. Note that $a \cdot (b_0 + k_b \pi_-^{\geq}) = 1$ since $\pi_-^{\geq} \cdot a = 0$.

If $p(a)+p(b_0+k_b\pi_{-}^{\geq})<1$ then we can uniquely choose the last element c such that $p(c)=1-p(a)-p(b_0+k_b\pi_{-}^{\geq})$. If $p(a)+p(b_0+k_b\pi_{-}^{\geq})>1$ introduce a constant $\bar{k}_b$ such that $0<p(b_0+k_b\pi_{-}^{\geq}+\bar{k}_b a_0)<1-p(a)$ (denote $b_0+k_b\pi_{-}^{\geq}+\bar{k}_b a_0$ by b). This is possible since $p(a_0)$ is irrational by assumption. Note that $a\cdot b=1$. Now the reasoning follows the case $p(a)+p(b_0+k_b\pi_{-}^{\geq})<1$.

By Lemma 6.3.1 we can connect the LHO with the arm module a,b,c that we constructed above to a LHO with an arm module M' such that $M'=M+\psi=M+a^*$ which concludes the proof for primitive ψ^*. If ψ^* is not primitive, we can connect M and $M'=M+\psi$ in several similar steps, but for this we require that $p(\psi^*)$ is irrational. □

Lemma 6.3.3. Any two arm modules M and M' associated to the same pole are connected if the image of the period map is not contained in the rational space $\mathbb{Q}\Pi$ generated by the peripheral periods.

Proof. Assume that $M'=M+\psi$. If $p_M(\psi^*)\notin\mathbb{Q}/\mathbb{Z}$, then, by Lemma 6.3.2, M is connected to M', and the proof is complete. If $p_M(\psi^*)\in\mathbb{Q}/\mathbb{Z}$, we construct an auxiliary arm module M'' such that $M\sim M''$ and $M''\sim M'$.
Note that the auxiliary arm module M'' is completely defined by φ where $M''=M'+\varphi$; in this case, $M''=M+\varphi+\psi$. It follows that we need to find φ such that $p_{N'}(\varphi^*)$ is irrational and $p_N(\varphi^*+\psi^*)$ is irrational; then, by the previous Lemma, $M\sim M''\sim M'$.

Fix a basis $\{x,y\}$ of $H_1(\Sigma_1,\mathbb{Z})$: then, in this basis the element $\varphi^*\in H_1(\Sigma_1,\mathbb{Z})$ is given by a pair $(n_x x,n_y y),n_x,n_y\in\mathbb{Z}$. Then, $p_N(\varphi^*)=n_x\alpha+n_y\beta$ where $\alpha=p_N(x)\in\mathbb{R}/\mathbb{Z},\beta=p_N(y)\in\mathbb{R}/\mathbb{Z}$, and by the assumptions of the lemma $\alpha\notin\mathbb{Q}\Pi$ or $\beta\notin\mathbb{Q}\Pi$. By the linearity, $p_{N'}(\varphi^*)=p_N(\varphi)+\psi(\varphi^*)p(d)$. Note that $\psi(\varphi^*)p(d)\in\mathbb{Z}p(d)$; so, $p_{N'}(\varphi^*)\notin(\mathbb{Q}+\mathbb{Q}p(d))/\mathbb{Z}$ is equivalent to $p_N(\varphi^*)\notin(\mathbb{Q}+\mathbb{Q}p(d))/\mathbb{Z}$. Additionally, take $p_N(\psi^*)=\gamma\in\mathbb{R}/\mathbb{Z}$. Then, we need to find $n_x,n_y\in\mathbb{Z}$ such that

$$p_N(\varphi^*+\psi^*)=n_x\alpha+n_y\beta+\gamma\notin(\mathbb{Q}+\mathbb{Q}p(d))/\mathbb{Z}\quad\text{and}$$

$$p_N(\varphi^*)=n_x\alpha+n_y\beta\notin(\mathbb{Q}+\mathbb{Q}p(d))/\mathbb{Z}.$$

Denote the space $(\mathbb{Q}+\mathbb{Q}p(d))/\mathbb{Z}$ by Q. Select arbitrary $(n_x',n_y')\in\mathbb{Z}^2$; if $\alpha n_x+\beta n_y\notin Q$ we fix this selection $(n_x,n_y)=(n_x',n_y')$. If $\alpha n_x'+\beta n_y'\in Q$ then $\alpha(n_x'+1)+\beta n_y'\notin Q$ since either $\alpha\notin Q$ or $\beta\notin Q$, so, wlog, we assume it for α. Then, we fix the selection $(n_x,n_y)=(n_x'+1,n_y')$. If $\alpha n_x+\beta n_y+\gamma\notin Q$ then we found (n_x,n_y) that satisfy the conditions above and we are done. If $\alpha n_x+\beta n_y+\gamma\in Q$ we conclude that $\alpha 2n_x+\beta 2n_y+\gamma\notin Q$ because $\alpha n_x+\beta n_y\notin Q$ by construction. Then, $(2n_x,2n_y)$ satisfy the conditions above and we are done. □

Lemma 6.3.3 sums up the proof of Theorem 6.1.4.

6.4 Appendix: connectivity of real isoperiodic sets in $\Omega\mathcal{S}_{1,2}$

In this Section we show that the level $\mathrm{Per}^{-1}(p)$ is connected in $\Omega\mathcal{S}_{1,2}$ if the image of p is contained in $\mathbb{R}$. A simple argument can be found in Section 3 of [18]; here, we present a lengthier geometrical argument which inspired the geometrical proof of Theorem 6.1.4. Using the notation proposed in the introduction we formulate the following proposition:

Proposition 6.4.1. The level $\mathrm{Per}^{-1}(p)$ is connected in $\Omega\mathcal{S}_{1,2}$ if the image of p is real.

Consider a torus $\Sigma_{1,2}$ equipped with a meromorphic differential ω having two simple poles s_+ and s_-. Let us denote by X^* the torus with punctures at s_+ and s_- and consider a marking of X^*, i. e., a basis $m \in H_1(X^*, \mathbb{Z})$. As X is compact, the sum of the residues around the poles is equal to zero by the residue theorem. The basis m has three components a, b, c such that $a+b+c = \pi_+$, where π_+ is a curve going around a marked "positive" pole s_+ (the choice of a “positive” and a “negative” pole is included in the marking).

Let us equip the group $H_1(X^*, \mathbb{Z})$ with a standard intersection form $(\cdot)$ such that $a \cdot b = b \cdot c = c \cdot a = 1$. Then, it is easy to check that π_+ belongs to the kernel of the intersection form. Moreover, $\mathrm{Ker}(\cdot) = \mathbb{Z}\pi_+$. Fix three real numbers $(\alpha, \beta, \gamma) \in \mathbb{R}^3$ as the real period coordinates and consider $M_{(\alpha,\beta,\gamma)}$ - the Torelli space of marked meromorphic differentials (X^*, ω, a, b, c) with

$$p(a) = \alpha, p(b) = \beta, p(c) = \gamma.$$

Note that

$$\mathrm{res}_{s_+} = \alpha + \beta + \gamma \neq 0,$$

because s_+ is a simple pole of ω (see Figure 6.11). Without loss of generality we can assume that $\mathrm{res}_{s_+} = 1$ and $\mathrm{res}_{s_-} = -1$. The form ω that has 2 poles also has two zeroes by the Riemann-Roch theorem. Assume that $\mathrm{Im}(\int_{z_2}^{z_1} \omega)$ is not zero: let us perform the Schiffer variations until $\mathrm{Im}(\int_{z_2}^{z_1} \omega) = 0$. It can happen that after this operation the two zeroes coincide forming a zero of order two. In this case is easy to see that

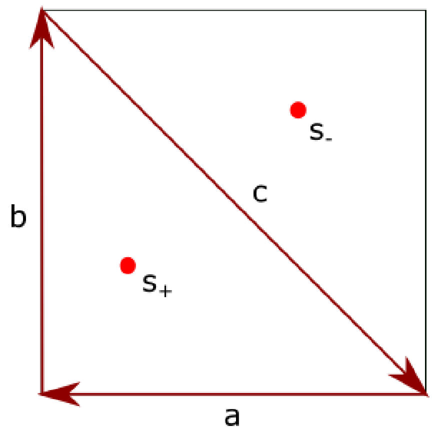


Figure 6.11: A marking on $\Sigma_{1,2}$.

Lemma 6.4.2. A zero of order two of a meromorphic differential with two poles on a torus has a unique topological type.

Proof. The proof is similar to the proof of Lemma 6.2.4, therefore, we provide a short version here. Each separatrix that leaves the double zero in a real direction has to come back along a real direction. Therefore, we have to consider the three different orders (up to rotation) in which the three outgoing separatrices come back (see Figure 6.12). Note that cutting the surface along the left and the right directions along separatrices decomposes the surface $\Sigma_{1,2}$ into a union of semi-infinite cylinders. Since the number of poles is 2, the number of semi-infinite cylinders is aso 2. We see that the types 1 and 2 on Figure 6.12 are therefre no possible; the only possible combinatorics is type 3 on Figure 6.12. □

Let us fix a marking $m = (a, b, c)$ of $H_1(\Sigma_{1,2}, \mathbb{Z})$. Any other marking $m' = (a', b', c')$ differs from $m = (a, b, c)$ by an automorphism of $(H_1(X^*, \mathbb{Z}), \cdot, \pi_+)$, i. e., by a positively oriented automorphism of the first homology group preserving the intersection form and both cycles π_+ and π_-. Therefore, the set of markings is $\mathrm{Aut}(H_1(X^*, \mathbb{Z}), \cdot, \pi_+)$.

Lemma 6.4.3. The family of automorphisms gives rise to a short exact sequence:

$$0 \to \mathbb{Z}^2 \to \mathrm{Aut}(H_1(X^*, \mathbb{Z}), \cdot, \pi_+) \to \mathrm{Aut}(H_1(X^*, \mathbb{Z})/\mathrm{Ker}(\cdot), \cdot) \to 0$$

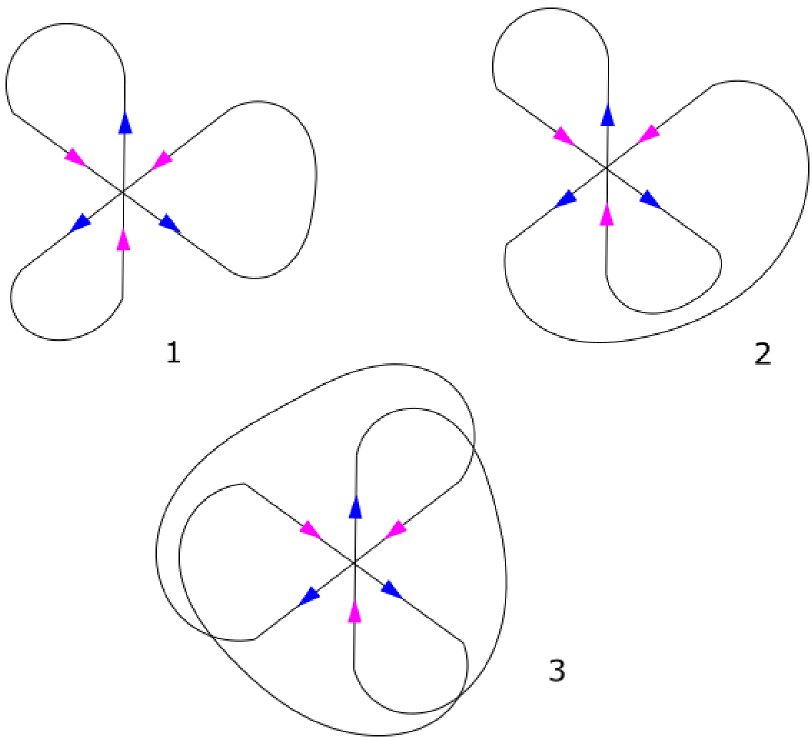


Figure 6.12: Possible combinatorics of a double zero on $\Sigma_{1,2}$. Since the number of poles is 2, tyes 1 and 2 are not possible, and type 3 is the only possible combinatorics.

The group $\mathrm{Ker}(\cdot)$ is a free $\mathbb{Z}$-module of rank 2 over the set of markings and, hence, its action is isomorphic to $\mathbb{Z}^2$. Therefore, $\mathrm{Aut}(H_1(X^*, \mathbb{Z})/\mathrm{Ker}(\cdot), \cdot)$ is isomorphic to the group $\mathrm{SL}(2, \mathbb{Z})$.

Proof. The sum of any triple of periods $\alpha' + \beta' + \gamma'$ is fixed to be equal to 1 because of the normalization. As the intersection form is preserved and its kernel is generated by π_+, for any other marking (a', b', c') it holds that $a' = a + q\pi_+$, and $b' = b + r\pi_+$. Then, c' is determined to be $c - (q + r)\pi_+$ to satisfy the sum condition. Therefore, any element in the kernel of the action of $\mathrm{Ker}(\cdot)$ on the markings is determined by a pair of integers q, r. Vice verca, any $q, r \in \mathbb{Z}$ define an element $f_{(q,r)} \in \mathrm{Ker}^2(\cdot)$ that sends a triple (a, b, c) to a triple $(a + q\pi_+, b + r\pi_+, c - (q + r)\pi_+)$. The defined map is an isomorphism. Therefore, the action of $\mathrm{Ker}(\cdot)$ on the set of markings is isomorphic to $\mathbb{Z}^2$. It follows that a group of automorphisms $\mathrm{Aut}(H_1(X^*, \mathbb{Z})/\mathrm{Ker}(\cdot), \cdot)$ is isomorphic to $\mathrm{SL}(2, \mathbb{Z})$. It acts as a matrix on the first two entries a and b of the marking; the third entry is defined via the sum condition. $\square$

Lemma 6.4.4. Let us call a marking (a', b', c') positive if $p(a), p(b), p(c) > 0$ and their sum is equal to 1. There is a bijection between the set of positive markings and a group $\mathrm{PSL}(2, \mathbb{Z})/\mathbb{Z}_3$. Since the Cayley graph of $\mathrm{PSL}(2, \mathbb{Z})/\mathbb{Z}_3$ with generators $z \to \pm 1$ and $z \to \frac{z}{z+1}$ is connected we conclude that the isoperiodic foliation is also connected.

We start the proof by claiming that there is a bijection between the set of the surfaces with a double zero and the positive markings up to a cyclic permutation. Indeed, every meromorphic differential ω with two simple poles and a double zero on a torus injectively corresponds to a positive marking (without loss of generality, we can assume that the periods are positive for every meromorphic differential).

To prove the second inclusion, let us construct a 1-form ω given a positive marking $(\alpha', \beta', \gamma')$. The marking determines the three loops at the double zero and their ordering. Therefore, it determines π_+ and π_- in the neighborhood of the loops. Take two semi-infinite cylinders C_1 and C_2 both with the circular circumference of length 1. Glue the circular boundary of C_1 to π_+ and the circular boundary of C_2 to π_-. The resulting surface is a torus with two poles and marking $(\alpha', \beta', \gamma')$. We conclude the bijection between the set of positive markings and a factor of $\mathrm{PSL}(2, \mathbb{Z})$ with respect to $\mathbb{Z}_3$.

Fix a marking (a, b, c) in $H_1(X^*/\mathrm{Ker}(\cdot), \cdot)$ and act on it with $\mathrm{PSL}(2, \mathbb{Z})$. We claim that in each fiber of this action there is a unique positive marking. Let

(a', b', c') be an image of (a, b, c) under an action of an element of $\mathrm{PSL}(2,\mathbb{Z})$. Pass to a triple of lengths $(\alpha', \beta', \gamma')$ and take their positive fractional parts, arriving to a triple $([\alpha'], [\beta'], [\gamma'])$ (since $\pi_+ = 1$ is in $\mathrm{Ker}(\cdot)$). It is clear that the sum of the fractional parts is an integer; therefore, it is either 1 or 2. If it is 1, then this marking is positive and it satisfies the sum condition; therefore, it belongs to the image of $\mathrm{PSL}(2,\mathbb{Z})$. If the sum is 2, take $([\alpha'] - 1, [\beta'] - 1, [\gamma'] - 1)$. For $\mathrm{PSL}(2,\mathbb{Z})$ it is equivalent to $(1 - [\alpha'], 1 - [\beta'], 1 - [\gamma'])$, which is then a positive marking satisfying the sum condition; therefore, it belongs to the image of $\mathrm{PSL}(2,\mathbb{Z})$. Vice versa, every positive marking lies in the image of $\mathrm{PSL}(2,\mathbb{Z})$ due to the short exact sequence.

Let us construct a graph $\mathfrak{G}$ representing the surfaces for which $\mathrm{Im}(\int_{z_2}^{z_1} \omega) = 0$. This graph is a retraction of the Torelli space to the subspace given by the condition $\mathrm{Im}(\int_{z_2}^{z_1} \omega) = 0$. The vertices of the graph are the surfaces having a double zero. Two vertices are connected with an edge if one can be transformed into another using the Schiffer variation in one step keeping the condition $\mathrm{Im}(\int_{z_2}^{z_1} \omega) = 0$. Next step is to show that each vertex of the graph belongs to three edges.

Indeed, without loss of generality, assume that $\alpha < \beta < \gamma$. Then, to conserve the positivity of the triple, one can only obtain three other triples with a single Schiffer variation, namely, $(\alpha, \beta - \alpha, \gamma + \alpha)$, $(\alpha, \beta + \alpha, \gamma - \alpha)$, and $(\alpha+\beta, \beta, \gamma-\beta)$. The corresponding markings are $(a, b-a, c+a)$, $(a, b+a, c-a)$, and $(a + b, b, c - b)$. As the transformations are reversible, the graph is not oriented. It follows that the graph $\mathfrak{G}$ is a three-valent graph; moreover, we continue to show that there is a bijection between its set of vertices and $\mathrm{PSL}(2,\mathbb{Z})/\mathbb{Z}_3$. We show that the graph $\mathfrak{G}$ is connected; therefore, the corresponding isoperiodic set is also connected.

To see it, take a standard fundamental domain of $\mathrm{PSL}(2,\mathbb{Z})$ on the upper half plane. Acting on it with $\mathrm{PSL}(2,\mathbb{Z})$ we cover the whole upper half plane with the images of the standard fundamental domain. Consider the boundary of this covering as a graph $\mathfrak{G}'$ ignoring the edges that go to infinity on the upper half-plane (i.e., vertical edges going upwards to infinity and the edges reaching the bottom line).

Take a point $e^{i\pi/3}$ which is a vertex of $\mathfrak{G}'$. It is easy to see that the three vertices connected to $e^{i\pi/3}$ are the images of $e^{i\pi/3}$ under the transformations $z \to \pm 1$ and $z \to \frac{z}{z+1}$. As the graph $\mathfrak{G}'$ is transitive with respect to the group $\mathrm{PSL}(2,\mathbb{Z})$, this holds for every vertex and its three neighbours. However, the

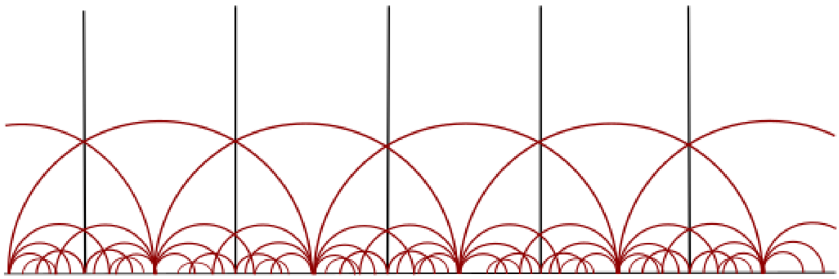


Figure 6.13: Fundamental domains of $\mathrm{PSL}(2,\mathbb{Z})$ covering the upper half plane

same thing holds for the graph $\mathfrak{G}$: as the order in the triple is not relevant, reorder it so that $\alpha < \beta < \gamma$. Then, the allowed moves from the triple (a, b, c) are: $(a, b-a, c+a), (a, b+a, c-a), (a+b, b, c-b)$. If we restrict ourselves to the first two entries, we have $(a, b) \to (a, b+a), (a, b-a), (a+b, b)$, which, in $\mathrm{PSL}(2,\mathbb{Z})$ correspond to $z \to \pm 1$ and $z \to \frac{z}{z+1}$. Therefore, $\mathfrak{G}'$ and $\mathfrak{G}$ are the same graph. As $\mathfrak{G}'$ is connected, so is $\mathfrak{G}$. This concludes the proof of the connectivity of the isoperiodic foliation of $\Sigma_{1,2}$.

Bibliography

[1] Elizaveta Arzhakova, Douglas Lind, Klaus Schmidt, and Evegeny Verbitskiy, *Decimation limits of principal algebraic $\mathbb{Z}^d$ - actions*, Preprint (2021).

[2] Jon Aaronson and Manfred Denker, *Local limit theorems for partial sums of stationary sequences generated by Gibbs–Markov maps*, Stochastics and Dynamics **1** (2001), no. 02, 193–237.

[3] Lars Ahlfors, *The complex analytic structure of the space of closed Riemann surfaces*, Princeton University Press, 2015.

[4] Luca Avena and Alexandre Gaudilliere, *A proof of the transfer-current theorem in absence of reversibility*, Statistics & Probability Letters **142** (2018), 17–22.

[5] Michael Francis Atiyah and Ian Grant Macdonald, *Introduction to commutative algebra*, CRC Press, 2018.

[6] Gerald Beer, *A natural topology for upper semicontinuous functions and a Baire category dual for convergence in measure*, Pacific J. Math. **96** (1981), no. 2, 251–263. MR637972

[7] Jean Berstel and Maurice Mignotte, *Deux propriétés décidables des suites récurrentes linéaires*, Bull. Soc. Math. France **104** (1976), no. 2, 175–184 (French). MR414475

[8] Lipman Bers, *Spaces of Riemann surfaces as bounded domains*, Bulletin of the American Mathematical Society **66** (1960), no. 2, 98–103.

[9] Alexander Bobenko, *Introduction to compact Riemann surfaces*, Springer, 2011.

[10] Stéphane Le Borgne, *Limit theorems for non-hyperbolic automorphisms of the torus*, Israel Journal of Mathematics **109** (1999), no. 1, 61–73.

[11] David W. Boyd, *Uniform approximation to Mahler's measure in several variables*, Canad. Math. Bull. **41** (1998), no. 1, 125–128, DOI 10.4153/CMB-1998-019-6. MR1618904

[12] Itai Benjamini, Russell Lyons, Yuval Peres, and Oded Schramm, *Special invited paper: uniform spanning forests*, Annals of probability (2001), 1–65.

[13] Mike Boyle and Douglas Lind, *Expansive subdynamics*, Trans. Amer. Math. Soc. **349** (1997), no. 1, 55–102, DOI 10.1090/S0002-9947-97-01634-6. MR1355295

[14] Rufus Bowen, *Equilibrium states and the ergodic theory of Anosov diffeomorphisms*, Springer Lecture Notes in Math **470** (1975), 78–104.

[15] Stéphane Le Borgne and Françoise Pène, *Vitesse dans le théorème limite central pour certains systèmes dynamiques quasi-hyperboliques*, Bulletin de la Société Mathématique de France **133** (2005), no. 3, 395–417.

[16] Robert Burton and Robin Pemantle, *Local characteristics, entropy and limit theorems for spanning trees and domino tilings via transfer-impedances*, The Annals of Probability (1993), 1329–1371.

[17] Gabriel Calsamiglia, Bertrand Deroin, and Stefano Francaviglia, *A transfer principle: from periods to isoperiodic foliations*, arXiv preprint arXiv:1511.07635 (2015).

[18] Gabriel Calsamiglia and Bertrand Deroin, *Isoperiodic meromorphic forms: two simple poles*, arXiv:2109.01796 (2021).

[19] Claude Chabauty, *Limite d'ensembles et géométrie des nombres*, Bull. Soc. Math. France **78** (1950), 143–151 (French). MR38983

[20] Keith Conrad, *Galois descent*, `https://kconrad.math.uconn.edu/blurbs/galoistheory/galoisdescent.pdf`.

[21] Yves Cornulier, *On the Chabauty space of locally compact abelian groups*, Algebr. Geom. Topol. **11** (2011), no. 4, 2007–2035, DOI 10.2140/agt.2011.11.2007. MR2826931

[22] Henry Cohn, Richard Kenyon, and James Propp, *A variational principle for domino tilings*, J. Amer. Math. Soc. **14** (2001), no. 2, 297–346.

[23] Jérôme Dedecker, Florence Merlevède, and Françoise Pene, *Rates in the strong invariance principle for ergodic automorphisms of the torus*, Stochastics and Dynamics **14** (2014), no. 02, 1350021.

[24] Deepak Dhar, *Theoretical studies of self-organized criticality*, Physica A: Statistical Mechanics and its Applications **369** (2006), no. 1, 29–70.

[25] Vesselin Dimitrov, *Convergence to the Mahler measure and the distribution of periodic points for algebraic Noetherian $\mathbb{Z}^d$-actions*, arXiv 1611.04664 (2016).

[26] Dmitry Dolgopyat, *Limit theorems for partially hyperbolic systems*, Transactions of the American Mathematical Society **356** (2004), no. 4, 1637–1689.

[27] Simon Donaldson, *Riemann surfaces*, Oxford University Press, 2011.

[28] Michael Drmota and Robert F. Tichy, *Sequences, discrepancies and applications*, Springer, 2006.

[29] Dominique Duval, *Absolute factorization of polynomials: a geometric approach*, SIAM J. Comput. **20** (1991), no. 1, 1–21, DOI 10.1137/0220001. MR1082133

[30] Freeman J. Dyson, *Statistical theory of the energy levels of complex systems. I*, Journal of Mathematical Physics **3** (1962), no. 1, 140–156.

[31] Manfred Einsiedler, Douglas Lind, Richard Miles, and Thomas Ward, *Expansive subdynamics for algebraic $\mathbb{Z}^d$-actions*, Ergodic Theory Dynam. Systems **21** (2001), no. 6, 1695–1729, DOI 10.1017/S014338570100181X. MR1869066

[32] Graham Everest, Alf van der Poorten, Igor Shparlinski, and Thomas Ward, *Recurrence sequences*, Mathematical Surveys and Monographs, vol. 104, American Mathematical Society, Providence, RI, 2003. MR1990179

[33] Gabor Elek and Gabor Tardos, *On roughly transitive amenable graphs and harmonic Dirichlet functions*, Proceedings of the American Mathematical Society (2000), 2479–2485.

[34] Ilya Itenberg and Grigory Mikhalkin, *Geometry in the tropical limit*, Math. Semesterber. **59** (2012), no. 1, 57–73.

[35] Shuhong Gao, *Absolute irreducibility of polynomials via Newton polytopes*, J. Algebra **237** (2001), no. 2, 501–520, DOI 10.1006/jabr.2000.8586. MR1816701

[36] Israel M. Gelfand, Mikhail M. Kapranov, and Andrei V. Zelevinsky, *Discriminants, resultants, and multidimensional determinants*, Mathematics: Theory & Applications, Birkhäuser Boston, Inc., Boston, MA, 1994. MR1264417

[37] Aleksandr O. Gelfond, *Transcendental and algebraic numbers*, Dover, New York, 1960.

[38] Jean Ginibre, *Statistical ensembles of complex, quaternion, and real matrices*, Journal of Mathematical Physics **6** (1965), no. 3, 440–449.

[39] Vadim Gorin, *Lectures on random lozenge tilings*, Cambridge Univ. Press, to appear.

[40] Mikhail Gordin, *The central limit theorem for stationary processes*, Doklady Akademii Nauk, 1969, pp. 739–741.

[41] ______, *A homoclinic version of the central limit theorem*, Journal of Mathematical Sciences **68** (1994), no. 4, 451–458.

[42] ______, *Double Extensions of Dynamical Systems and Construction of Mixing Filtrations. II. Quasihyperbolic Toral Automorphisms*, Journal of Mathematical Sciences **109** (2002), no. 6, 2103–2114.

[43] Mikhail Gordin and Michel Weber, *On the Almost Sure Central Limit Theorem for a Class of Z d-Actions*, Journal of Theoretical Probability **15** (2002), no. 2, 477–501.

[44] Alexander Gorodnik and Ralf Spatzier, *Exponential mixing of nilmanifold automorphisms*, Journal d'Analyse Mathématique **123** (2014), no. 1, 355–396.

[45] Sébastien Gouëzel, *Stable laws for the doubling map*, preprint (2008).

[46] Paul R. Halmos, *On automorphisms of compact groups*, Bull. Amer. Math. Soc. **49** (1943), 619–624, DOI 10.1090/S0002-9904-1943-07995-5. MR0008647

[47] Brian C. Hall, *Quantum theory for mathematicians*, Vol. 267, Springer, 2013.

[48] Ursula Hamenstädt, *Ergodicity of the absolute period foliation*, Israel Journal of Mathematics **225** (2018), no. 2, 661–680.

[49] Olli Hella, Juho Leppänen, and Mikko Stenlund, *Stein's method of normal approximation for dynamical systems*, Stochastics and Dynamics **20** (2020), no. 04, 2050021.

[50] John Hubbard and Ibrahim Pourezza, *The space of closed subgroups of* $\mathbf{R}^2$, Topology **18** (1979), no. 2, 143–146, DOI 10.1016/0040-9383(79)90032-6. MR544155

[51] Ildar A. Ibragimov, *Some limit theorems for stationary processes*, Theory of Probability & Its Applications **7** (1962), no. 4, 349–382.

[52] Anatole Katok, Svetlana Katok, and Klaus Schmidt, *Rigidity of measurable structure for* $\mathbb{Z}^d$*-actions by automorphisms of a torus*, Comment. Math. Helv. **77** (2002), no. 4, 718–745, DOI 10.1007/PL00012439. MR1949111

[53] Piet W. Kasteleyn, *Dimer statistics and phase transitions*, J. Mathematical Phys. **4** (1963), 287–293.

[54] Yitzhak Katzenlson, *Ergodic automorphisms of T n are Bernoulli shifts*, Israel Journal of Mathematics **10** (1971), no. 2, 186–195.

[55] Richard Kenyon, Andrei Okounkov, and Scott Sheffield, *Dimers and amoebae*, Ann. of Math. (2) **163** (2006), no. 3, 1019–1056, DOI 10.4007/annals.2006.163.1019. MR2215138

[56] Richard Kenyon, *An introduction to the dimer model*, arXiv preprint math/0310326 (2003).

[57] Achim Klenke, *The random spanning tree on ladder-like graphs*, arXiv preprint arXiv:1704.00182 (2017).

[58] Andrei N. Kolmogorov, *A new metric invariant of transient dynamical systems and automorphisms in Lebesgue spaces*, Doklady Akademii Nauk, 1958, pp. 861–864.

[59] Igor Krichever, Sergei Lando, and Alexandra Skripchenko, *Real-normalized differentials with a single order 2 pole*, Letters in Mathematical Physics **111** (2021), no. 2, 1–19.

[60] Lauwerens Kuipers and Harald Niederreiter, *Uniform distribution of sequences*, Courier Corporation, 2012.

[61] François Ledrappier, *Un champ markovien peut être d'entropie nulle et mélangeant*, C. R. Acad. Sci. Paris Sér. A-B **287** (1978), no. 7, A561–A563 (French, with English summary). MR512106

[62] V. P. Leonov, *On the central limit theorem for ergodic endomorphisms of compact commutative groups*, Doklady Akademii Nauk, 1960, pp. 258–261.

[63] Douglas Lind, *Dynamical properties of quasihyperbolic toral automorphisms*, Ergodic Theory and Dynamical Systems **2** (1982), no. 1, 49–68.

[64] Douglas Lind, Klaus Schmidt, and Tom Ward, *Mahler measure and entropy for commuting automorphisms of compact groups*, Invent. Math. **101** (1990), no. 3, 593–629. MR1062797

[65] Douglas Lind, Klaus Schmidt, and Evgeny Verbitskiy, *Homoclinic points, atoral polynomials, and periodic points of algebraic $\mathbb{Z}^d$-actions*, Ergodic Theory Dynam. Systems **33** (2013), no. 4, 1060–1081.

[66] Douglas Lind and Klaus Schmidt, *Homoclinic points of algebraic $\mathbb{Z}^d$-actions*, Journal of the American Mathematical Society **12** (1999), no. 4, 953–980.

[67] Elon Lindenstrauss and K Schmidt, *Symbolic representations of nonexpansive group automorphisms*, Israel Journal of Mathematics **149** (2005), no. 1, 227–266.

[68] Carlangelo Liverani, *Statistical properties of uniformly hyperbolic maps and transfer operators' spectrum*, arXiv preprint arXiv:1810.05924 (2018).

[69] R. M. Loynes, *The central limit theorem for backwards martingales*, Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete **13** (1969), no. 1, 1–8.

[70] Johannes Lundqvist, *An explicit calculation of the Ronkin function*, Ann. Fac. Sci. Toulouse Math. (6) **24** (2015), no. 2, 227–250, DOI 10.5802/afst.1447 (English, with English and French summaries). MR3358612

[71] Russell Lyons and Yuval Peres, *Probability on trees and networks*, Vol. 42, Cambridge University Press, 2017.

[72] Russell and Steif Lyons Jeffrey E, *Stationary determinantal processes: phase multiplicity, Bernoullicity, entropy, and domination*, Duke Mathematical Journal **120** (2003), no. 3, 515–575.

[73] Russell Lyons, *Determinantal probability measures*, Publications Mathématiques de l'IHÉS **98** (2003), 167–212.

[74] Diane Maclagan and Bernd Sturmfels, *Introduction to tropical geometry*, Graduate Studies in Mathematics, vol. 161, American Mathematical Society, Providence, RI, 2015. MR3287221

[75] Arturo Magidin and David McKinnon, *Gauss's lemma for number fields*, Amer. Math. Monthly **112** (2005), no. 5, 385–416, DOI 10.2307/30037491. MR2139573

[76] K. Mahler, *On lattice points in n-dimensional star bodies. I. Existence theorems*, Proc. Roy. Soc. London Ser. A **187** (1946), 151–187, DOI 10.1098/rspa.1946.0072. MR17753

[77] ———, *On some inequalities for polynomials in several variables*, J. London Math. Soc. **37** (1962), 341–344, DOI 10.1112/jlms/s1-37.1.341. MR0138593

[78] Vincent Maillot, *Géométrie d'Arakelov des variétés toriques et fibrés en droites intégrables*, Mém. Soc. Math. Fr. (N.S.) **80** (2000), vi+129, DOI 10.24033/msmf.393 (French, with English and French summaries). MR1775582

[79] Odile Macchi, *The coincidence approach to stochastic point processes*, Advances in Applied Probability **7** (1975), no. 1, 83–122.

[80] Howard Masur and John Smillie, *Hausdorff dimension of sets of nonergodic measured foliations*, Annals of Mathematics **134** (1991), no. 3, 455–543.

[81] Curtis T. McMullen, *Moduli spaces of isoperiodic forms on Riemann surfaces*, Duke Mathematical Journal **163** (2014), no. 12, 2271–2323.

[82] Florence Merlevàde and Magda Peligrad, *The functional central limit theorem under the strong mixing condition*, The Annals of Probability **28** (2000), no. 3, 1336–1352.

[83] Grigory Mikhalkin, *Amoebas of algebraic varieties and tropical geometry*, Different faces of geometry, 2004, pp. 257–300.

[84] Sergey V. Nagaev, *Some limit theorems for stationary Markov chains*, Theory of Probability & Its Applications **2** (1957), no. 4, 378–406.

[85] Sergei P. Novikov, *The Hamiltonian formalism and a many-valued analogue of Morse theory*, Uspekhi Matematicheskikh Nauk **37** (1982), no. 5, 3–49.

[86] Andrei Okounkov, *Limit shapes, real and imagined*, Bull. Amer. Math. Soc. (N.S.) **53** (2016), no. 2, 187–216, DOI 10.1090/bull/1512. MR3474306

[87] Richard Kenyon and Andrei Okounkov, *Planar dimers and Harnack curves*, Duke Math. J. **131** (2006), no. 3, 499–524.

[88] Mikael Passare and Hans Rullgård, *Amoebas, Monge-Ampère measures, and triangulations of the Newton polytope*, Duke Math. J. **121** (2004), no. 3, 481–507, DOI 10.1215/S0012-7094-04-12134-7. MR2040284

[89] Robin Pemantle, *Choosing a spanning tree for the integer lattice uniformly*, The Annals of Probability (1991), 1559–1574.

[90] Françoise Pène, *Averaging method for differential equations perturbed by dynamical systems*, ESAIM: Probability and Statistics **6** (2002), 33–88.

[91] ______, *Rate of convergence in the multidimensional central limit theorem for stationary processes. Application to the Knudsen gas and to the Sinai billiard*, The Annals of Applied Probability **15** (2005), no. 4, 2331–2392.

[92] Kevin Purbhoo, *A Nullstellensatz for amoebas*, Duke Math. J. **141** (2008), no. 3, 407–445, DOI 10.1215/00127094-2007-001. MR2387427

[93] Richard Kenyon, James G Propp, and David B Wilson, *Trees and matchings*, arXiv preprint math/9903025 (1999).

[94] R. Tyrrell Rockafellar, *Convex analysis*, Princeton Mathematical Series, No. 28, Princeton University Press, Princeton, N.J., 1970. MR0274683

[95] L. I. Ronkin, *On zeros of almost periodic functions generated by functions holomorphic in a multicircular domain*, Complex analysis in modern mathematics (Russian), FAZIS, Moscow, 2001, pp. 239–251 (Russian). MR1833516

[96] Klaus Friedrich Roth, *Rational approximations to algebraic numbers*, Mathematika **2** (1955), no. 1, 1–20.

[97] Murray Rosenblatt, *A central limit theorem and a strong mixing condition*, Proceedings of the National Academy of Sciences of the United States of America **42** (1956), no. 1, 43.

[98] Klaus Schmidt, *Dynamical systems of algebraic origin*, Progress in Mathematics, vol. 128, Birkhäuser Verlag, Basel, 1995. MR1345152

[99] Klaus Schmidt and Evgeny Verbitskiy, *Abelian sandpiles and the harmonic model*, Communications in Mathematical Physics **292** (2009), no. 3, 721–759.

[100] ———, *New directions in algebraic dynamical systems*, Regular and Chaotic Dynamics **16** (2011), no. 1, 79–89.

[101] Theodor Schneider, *Einführung in die transzendenten Zahlen*, Vol. 81, Springer, 1957.

[102] Martin Schlichenmaier, *An introduction to Riemann surfaces, algebraic curves and moduli spaces*, Springer Science & Business Media, 2010.

[103] Menahem Schiffer and Donald Clayton Spencer, *Functionals of finite Riemann surfaces*, Princeton University Press, 2015.

[104] Yakov G. Sinai, *On the notion of entropy of a dynamical system*, Doklady of Russian Academy of Sciences, 1959, pp. 768–771.

[105] C. J. Smyth, *On measures of polynomials in several variables*, Bull. Austral. Math. Soc. **23** (1981), no. 1, 49–63, DOI 10.1017/S0004972700006894. MR615132

[106] ———, *A Kronecker-type theorem for complex polynomials in several variables*, Canad. Math. Bull. **24** (1981), no. 4, 447–452, DOI 10.4153/CMB-1981-068-8. MR644534

[107] H. N. V. Temperley, *Combinatorics: Proceedings of the British Combinatorial Conference 1973. 202-204* (1974).

[108] Anton Zorich, *Flat surfaces*, arXiv preprint math/0609392 (2006).

Samenvatting

Deze scriptie omschrijft het limietgedrag van dynamische systemen zoals deze voorkomen in verscheidene wiskundige deelgebieden, zoals algebra, meetkunde en kansrekening.

Objecten in de wereld om ons heen bestaan uit een gegeven aantal deeltjes die onderling interageren. Dit wordt op een natuurlijke manier omschreven als een dynamisch systeem. Soms is het aantal deeltjes zodanig dat het redelijk is om aan te nemen dat het er oneindig veel zijn. Zulke oneindige systemen kunnen worden omschreven door ze te benaderen met een reeks dynamische systemen met een eindig veel en toenemend aantal deeltjes. Het is niet noodzakelijk makkelijk om de eindige systemen te omschrijven. Een bekend voorbeeld is het n-lichamen probleem, dat voor $n > 2$ geen gesloten vorm oplossingen heeft. Maar in sommige situaties is het mogelijk om het zogenaamde limietgedrag, dat wil zeggen, eigenschappen van het oneindige systeem te bepalen met behulp van deze reeks van eindige dynamische systemen.

In hoofdstukken twee en drie onderzoeken we het limietgedrag van de coefficienten van gedecimeerde Laurent polynomen. Het is bekend dat de schaallimiet bestaat voor de karakteristieke polynomen van Dimer modelen. Het bewijs van het bestaan van deze limiet hangt af van de eigenschappen van het Dimer model. Daarom kunnen ze niet zondermeer gegeneraliseerd worden naar andere systemen. In hoofdstuk twee bespreken we het bewijs van het bestaan van een *convexe romp* van de geschaalde coefficienten van de gedecimeerde Laurent polynoom. Analoog aan de schaallimiet van het Dimer model wordt deze convexe romp gegeven door de Legendre transformatie van de Ronkin functie van deze polynoom. Hoofdstuk drie herhaalt dit resultaat met een alternatief bewijs en leidt de implicaties af voor gedecimeerde voornaamste algebraische acties.

In hoofdstuk vier bestuderen we het determinante puntproces horend bij uniform opspannende bossen op $\mathbb{Z}^d$-periodieke grafen. Het voornaamste

resultaat in het hoofdstuk is een expliciete uitdrukking van de kansmaat geassocieerd met het uniforme opspannende bos DPP op $\mathbb{Z}^d$-periodieke grafen. Het bestaan van een dergelijke maat is aangetoond in [12] als de limiet van een reeks van maten op eindige grafen. Het is opmerkelijk dat de limietmaat hangt af van de randvoorwaarden en deze maat is daardoor niet noodzakelijk uniek. De grafen die worden bestudeerd in dit hoofdstuk daarentegen hebben wel een unieke limietmaat waarvoor een expliciete uitdrukking is afgeleid, waarbij slechts de fundamentele domeinen van de $\mathbb{Z}^d$-periodieke graaf zijn gebruikt. Het is een open vraag of dezelfde aanpak gebruikt kan worden voor de Cayleygrafen van groepen anders dan $\mathbb{Z}^d$, zoals bijvoorbeeld de Heisenberggroep.

In hoofdstuk vijf bestuderen we de snelheid van convergentie in de Centrale Limiet Stelling voor ergodische torische automorfismen. De resultaten zijn beschikbaar voor een subklasse van ergodische torische automorfismen, namelijk de hyperbolische torische automorfismen. Spectrumeigenschappen van niet-hyperbolische torische automorfismen zijn zwakker dan voor de hyperbolische torische automorfismen. Daardoor kunnen de bekende methodes niet worden toegepast voor de studie van de snelheid van convergentie van algemene ergodische torische automorfismen. In hoofdstuk vijf bewijzen we een centrale limietstelling voor Ergdoische torische automorfismen met behulp van de Steinmethode, zoals in [49]. De twee voordelen van deze methode zijn als volgt: In de eerste plaats geeft het op natuurlijke wijze een grens voor de snelheid van convergentie. Als tweede gebruikt het de mixeigenschappen van automorfismen (ergodische torische automorfismen zijn exponentieel mixend voor Hölder observabelen) in plaats van de spectrumeigenschappen.

Hoofdstuk zes focust op het meetkundige probleem van de studie van isoperiodieke foliaties in de Torelliruimte van Riemannoppervlakken. Voor holomorfe differentialen zijn veel eigenschappen van de isoperiodieke foliaties bekend, zoals bijvoorbeeld ergodiciteit en of de vezels verbonden zijn. De studie van meromorfe differentialen daarentegen is veel gefragmenteerder. De verbondenheid is alleen bekend in het geval van twee polen met een willekeurige genus. In hoofdstuk zes bewijzen we dat elke vezel van de reële isoperiodieke foliaties is verbonden (onder milde voorwaarden) in de Torelliruimte van Riemannoppervlakken van genus één met drie simpele polen. Dit geval is informatief om twee redenen: In de eerste plaats is het moeilijker dan een geval met meer polen, omdat de degeneratie techniek niet kan worden toegepast. De tweede reden is dat het geval van genus één met drie polen kan dienen als een basis voor de inductie van een algemenere uitspraak.

Acknowledgements

This thesis was made possible with the support of many people who I would like to thank in this section.

I would like to thank my promotor and the head of the Probability group, Prof. Dr. Frank den Hollander. Your support and advice through my PhD were invaluable, and I am happy and proud to have been a part of your group. I am grateful to my supervisor, Prof. Dr. Evgeny Verbitskiy, for suggesting interesting projects and for our frequent stimulating discussions. Lastly, I would like to thank my copromotor Dr. Dalia Terhesiu, for opening new mathematical horizons to me and for dazzling me with beautiful mathematical ideas.

Many projects have been done in collaboration with colleagues from outside of Leiden. Here, I thank Prof. Dr. Douglas Lind, Prof. Dr. Klaus Schmidt, and Prof. Dr. Tomoyuki Shirai. It has been a great experience to work together, and I learned a lot from you.

I am deeply grateful to Prof. Dr. Bertrand Deroin and Dr. Gabriel Calsamiglia. Our weekly online Wednesday meetings were to look forward to! I also cherish memories of my visit to Paris in 2021 to finish the project, when we could work on real blackboards instead of virtual ones. Working with you has been truly inspiring, and I deeply value our collaboration and, if I may add, friendship.

My sincere thanks goes to Dr. Alexandra Skripchenko who made me aware of the opening PhD position in Leiden University. As my master thesis advisor, Alexandra passed on her passion for mathematics to me, and has been my role model ever since.

I am grateful to all members of my Promotion Committee, in particular, to Prof. Dr. Hanfeng Li, Prof. Dr. Sergey Shadrin, and Prof. Dr. Beatrice de Tilière, for reviewing my manuscript. Beatrice, thank you very much for the extensive feedback that you provided.

I would like to thank my university peers who made my PhD years colourful: my very first friend in Leiden, Leonardo Morelli, my dear housemate Benthen Zeegers, my brilliant paranymphs Mia Jukic and Amine Hadji, and many more. We have travelled the PhD road together, sharing both fun and deep thoughts. You have shaped my Leiden University.

Last but not the least, I would like to thank those closest to me. Mom and Dad, thank you for your love and support, and for believing in me. Steven, thank you for bright and perpetual sunshine that you bring to my life. Finally, I am grateful to mijn Nederlandse familie: Yvonne, Johan, Mariska, Leonie, and Jasper, for welcoming me as their own.

Curriculum vitae

Elizaveta Maksimovna Arzhakova was born on August 12, 1993 in Moscow, Russia. She graduated from a mathematical class of the Moscow Gymnasium no. 1543 in 2011. She obtained a Bachelor of Science in Mathematics in the Higher School of Economics in Moscow in 2015 where she wrote a graduation thesis *Dynamics on Heisenberg group* under the supervision of Prof. dr. M. Verbitsky.

In 2015 Elizaveta entered the Master programme in Mathematics in the Higher School of Economics, Moscow. In 2016 she conducted a research project on isoperiodic sets of meromorphic differentials in École Normale Supérieure under the supervision of Prof. dr. B. Deroin. Elizaveta obtained a SwissMap fellowship to participate in the Master class in Geometry, Topology and Physics in Université de Genève in 2016-2017. She obtained two Master of Science degrees, both in Mathematics. The first degree was obtained in the Higher School of Economics in 2017 with a thesis *On thermodynamic formalism for quadratic differentials* under supervision of Dr. A. Skripchenko. The second degree was obtained in Université de Genève in 2018 with a thesis *On Pitman Theorem for* $SU(2)$ supervised by Prof. dr. A. Alexeev.

In September 2017 Elizaveta started a PhD research project in Leiden University under the supervision of Prof. dr. E. Verbitskiy and supported by the Huygens fellowship. During her doctoral studies, she served as a teaching assistant in many courses and visited several scientific conferences. In 2019 she supervised a project on domino tilings of the Leiden PRE-University programme. In 2020 she took part in organising the transition into the online education due to Covid-19 pandemic in the Mathematical Institute in Leiden.