Aperiodic Order and Singular Spectra

Dissertation zur Erlangung des akademischen Grades eines
Doktors der Mathematik (Dr. math.)

vorgelegt von

Philipp Nicolai Gohlke

Fakultät für Mathematik
Universität Bielefeld

August 2021

1. Berichterstatter: Prof. Dr. Michael Baake	Universität Bielefeld, Germany
2. Berichterstatter: Prof. Dr. David Damanik	Rice University, Texas, USA
3. Berichterstatter: Prof. Dr. Daniel Lenz	FS-Universität Jena, Germany

Datum der mündlichen Prüfung: 13.12.2021

Contents

Abgrenzung des eigenen Beitrags gemäß §10(2) der Promotionsordnung

Most of the results presented in this thesis have been published in joint work with several co-authors. As a guiding principle, whenever I cite individual results from these publications to which I did not contribute, I explicitly mention the name of the author(s) that I think it should be attributed to.

(1) The results on ergodicity in Section 4.1 are joint work with T. Spindeler and have been published in [GS20]. Here, I provide a slightly different perspective, introducing the ergodic measures as orbit averages of invariant measures, but the core of the arguments remains the same.

(2) The results on topological entropy in Section 4.2.1 are taken from the single-authored paper [Goh20], with a slightly polished presentation.

(3) The results on measure-theoretic entropy in Section 4.2.2 are joint work with A. Mitchell, D. Rust and T. Samuel [GMRS21]. Here, I use a slightly different formalism via conditional entropies.

(4) The results on diffraction in Section 4.3 are original. However, they are meant to be part of an upcoming joint publication with N. Mañibo and D. Rust [GMR].

(5) The results on almost minimal substitutions in Section 5.2 are based on joint work with B. Eichinger [EG21]. I have taken the liberty to discuss a smaller family of substitutions to avoid technicalities.

(6) The results on mixed potentials in Section 5.3 are joint work with J. Chaika, D. Damanik and J. Fillman and will appear in [CDFG].

(7) The results on super-polynomial scaling of g-measures in Section 6.2 and parts of the material in Section 6.1 are essentially taken from joint work with M. Baake, M. Coons and J. Evans [BCEG21]. This includes a result (Proposition 6.1.24) that was announced in [BCEG21] but is derived for the first time in this work. The main result of Section 6.2 (Theorem 6.2.2) is a variant of [BCEG21, Thm. 1.4] that holds for a larger family of examples, but offers a slightly different conclusion.

(8) The results on the multifractal analysis of the Thue–Morse measure in Section 6.3 are joint work with M. Baake, M. Kesseböhmer and T. Schindler [BGKS19]. The material in Section 6.3 is meant to be complementary to [BGKS19], filling in some details, while leaving out others.

Acknowledgements

The past few years have been a beautiful, sometimes intense, journey through several topics in mathematics, but also through the physical world. It would not have been possible without the help of many people, and this work would not be complete without expressing my gratitude to at least a few of them (and to apologize to those that I have left out or forgotten).

First and foremost, I am grateful to my supervisors Michael Baake and David Damanik. To Michael Baake for constantly nourishing my interest in mathematics since my early days as a physics student, for his gentle but skillful guidance over many years, and for generously sharing his insights. To David Damanik for being an amazing host at Rice University, for many illuminating ideas and discussions, and for always being approachable despite a busy calendar.

Many thanks to Britta Heitbreder, Stephan Merkes and Ligia Pesquera Leismer for taking care of all organizational matters. Without your constant support, writing a thesis in two different continents would have been much more of a headache.

I would like to thank Frederic Alberti, Neil Mañibo, Dan Rust and Timo Spindeler for letting me in on this group of young researchers in Bielefeld that was vivid with discussion, sharing ideas and writing publications.

I also thank my further collaborators Michael Baake, Jon Chaika, Michael Coons, David Damanik, Benjamin Eichinger, James Evans, Jake Fillman, Marc Kesseböhmer, Andrew Mitchell, Tony Samuel and Tanja Schindler for sharing their ideas and the sometimes painstaking labour of turning them into a paper.

I am grateful to Benjamin Eichinger for sharing not only a flat in Houston, but also a passion for mathematics and music.

I thank past and present members of the working groups, both in Bielefeld and Houston, for the welcoming atmosphere and stimulating discussions, including Dirk Frettlöh, Franz Gähler, Ethan Gwaltney, Christian Huck, Tobias Jakobi, Hyunkyu Jun, Milivoje Lukić, Daniel Luz, Selim Sukhtaiev, Venta Terauds, Giorgio Young and Peter Zeiner.

I thank the German Research Foundation (DFG) for their financial support within the CRC 1283.

Finally, I owe thanks to many people that have supported me throughout this time with their encouragement and friendship.

To all the amazing people in the Haspel, for giving me a home when I came back to Bielefeld and for letting me be part of this wonderful project. You have made this pandemic, if not pleasant, at least bearable.

To Hayal, for walking (and hiking) a part of this way with me.

To my parents, for giving me roots and letting me fly.

And to Elisa, for all your persistent love and support.

Introduction

A central motive of this thesis is the study of configurations with aperiodic order. Here, a *configuration* may refer to a bi-infinite sequence of symbols, a tiling of the plane or a countable subset of Euclidean space, modelling, for example, the positions of atoms in a material. Although higher-dimensional examples might be visually more appealing (and physically more relevant), we stick to the one-dimensional case. This is because the one-dimensional theory is better understood and generalizations to the higher-dimensional situation are often more involved or simply not available. A configuration x is *periodic* if it coincides with a translated version of itself. That is, $x = x + t$, for some translation vector $t \neq 0$. In this case, every finite pattern repeats at regular distances, given by multiples of t. In the *aperiodic* case, we can get an intuition for the amount of (dis)order in a configuration by asking how close we are to the periodic situation. Does a finite pattern repeat at all? If so, how much does the distance vary between two consecutive appearances of the same pattern? Does the pattern appear with a well-defined frequency? How many different patterns of a predetermined size can we find? How does this number change as we vary the size of the pattern?

All of these questions require the observer to "scan" the configuration, that is, to obtain information from arbitrary positions. A myopic observer may instead prefer to shift the configuration itself. This naturally places us into the mathematical framework of dynamical systems, requiring a small change of perspective. Now, the (symbolic) configuration x is regarded as a "point", and the shift S by a unit length maps x to a new point Sx. Iterating this procedure gives rise to a sequence of points $x, Sx, S^2x, \ldots$ (the forward orbit of x), which are naturally embedded into a larger space of configurations X that is invariant under S. We call the pair (X, S) a *dynamical system*, which can be equipped with additional mathematical structure like a topology, a metric, or a measure. With this, we have a full-fledged and well-developed theory at our disposal. Not only can the questions formulated in the last paragraph be rephrased in the language of dynamical systems; we also obtain a variety of more refined measurements of disorder.

In fact, these measurements of disorder reflect different ways to look at dynamical systems. Some are purely combinatorial in nature, others draw on the topological structure or refer to an associated measure. Of particular importance is *ergodicity*, a measure-theoretic mixing property of a dynamical system. A more information-theoretic perspective yields the concept of *entropy*. This measures to what degree it is possible to predict the next step of a point by measurements that were performed in the past. Yet another approach deals with a unitary operator, introduced by Koopman [Koo31], which maps a function f on X to the function $f \circ S$. The spectrum of this operator is known as the *dynamical spectrum* of (X, S). Curiously, the dynamical spectrum is, under appropriate assumptions, accessible by scattering experiments on the individual configurations, which yields the so-called *diffraction*

spectrum [BL04]. Indeed, diffraction spectra have played an important role in stimulating the interest in aperiodic order. This is due to the discovery of quasicrystals by Dan Shechtman via electron diffraction experiments on an alloy [SBGC84]. The outcome of these experiments, a diffraction image of sharp (Bragg) peaks and a tenfold rotation symmetry, stirred some commotion in the crystallographic community, ranging from skepticism to upright denial. This is because a tenfold rotation symmetry is known to be irreconcilable with periodic structures in 3-dimensional space, challenging the at that time widely accepted belief that Bragg peaks indicate a periodic structure. The existence of quasicrystals was eventually confirmed, earning Shechtman the Nobel Prize in Chemistry in 2011. For more on the intriguing history of quasicrystals, compare [BG13].

Another phenomenological way to access the degree of (dis)order of a material is to measure its electronic transport properties. In the framework of quantum mechanics, a valence electron in a material is described by a wave function that reflects the probability of observing it in certain regions if a measurement is performed. The time evolution of this wave function is driven by the self-adjoint *Schrödinger operator*, modelling the influence of the underlying material. If the material is sufficiently disordered, electrons are "trapped" and electronic transport is inhibited—a phenomenon that became widely known as *Anderson localization*. On the other hand, periodic materials allow for a free transport of valence electrons, provided their energy matches a set of allowed values, known as the electronic band structure. In the mathematical model, this band structure is given by the spectrum of the corresponding Schrödinger operator. Configurations with an intermediate degree of disorder have been shown to exhibit interesting and anomalous behaviour, like giving rise to a spectrum that forms a Cantor set of Lebesgue measure 0 [DFb].

As a set, the spectrum of an operator does not offer sufficient insight into its dynamical properties. It is therefore customary to consider *spectral measures*, which carry more refined spectral information. Such a measure is *singular* if it gives full weight to a set of vanishing Lebesgue measure. In the most extreme case, it is supported on a single point. If single points carry no weight, we speak of a (singular) continuous measure. The implications of singularity vary, depending on the context. If the spectral measure is related to the dynamical spectrum, it points towards an ordered system. If the spectral measure is related to a Schrödinger operator, singularity points towards disorder.

The zoo of the different concepts of disorder is not consistent, in the sense that the same configuration might be classified as ordered if viewed from one angle and as disordered if viewed from another angle. There are some deep and interesting connections between the notions of disorder, but in general, their interdependencies remain obscure and call for the study of specific (families of) examples. It is a non-trivial task to obtain meaningful mathematical models of configurations with aperiodic order that, at the same time, offer some control over its properties. An important tool to generate many such examples in a constructive manner is the concept of a *substitution*. Starting from a finite pattern, it brakes down the pattern into smaller constituents, each of which is replaced by a larger pattern, according to a fixed set of replacement rules. Iterating this procedure eventually leads to an infinite configuration that inherits some hierarchical structure from the production process. Despite being aperiodic in

non-trivial cases, the configurations that arise from a substitution can be deemed to be quite ordered, according to several characterizations. One way to increase the disorder in the construction is to apply the local replacement rules *randomly*; we speak of a *random substitution* in this case. The interest in random substitutions (in different disciplines) has been sporadic over the last decades, and they have gained renewed attention only recently.

Outline of the thesis

After collecting some of the mathematical prerequisites in Chapter 1, we devote Chapter 2 to a short and biased exposition of the theory of dynamical systems, with an emphasis on symbolic dynamics.

Since substitutions provide the foundation for many of the models discussed in this thesis, we take our time discussing their basic properties in Chapter 3. Random substitutions are introduced as natural extensions. The recent literature features random substitutions in two common disguises. Only one of them makes reference to the probabilities assigned to a specific production rule, whereas the other is more combinatorial in nature. We explore the connections and differences in some detail.

Chapter 4 contains the main results on random substitutions. First, we show that iterating a random substitution naturally leads to an invariant distribution, generalizing observations from [BSS18] and [MT-JU18]. Under the classic assumption of *primitivity*, the orbit average of this distribution is shown to define an ergodic measure on the substitution dynamical system. We prove that both the topological and the measure-theoretic entropy of this system can be computed (with explicit bounds on the precision) by following the production process of the random substitution. In many cases, this leads to an explicit formula for the entropy. Regarding diffraction, we take up a key splitting into expectation and variance, that was introduced by Godrèche and Luck to study the random Fibonacci substitution [GL89]. We show for a large class of random substitution that this produces a splitting of the diffraction measure into a pure point part (consisting of Bragg peaks) and an absolutely continuous part (observable as a diffuse background). While the result on the variance part is rather general, the result on the expectation part is formulated under more restrictive (and not yet optimal) conditions. Given the evasive nature of the *Pisot substitution conjecture*—an analogue of this result in the realm of substitutions—this does not come as a surprise.

Schrödinger operators that are associated to dynamical systems are the main objects of interest in Chapter 5. In the first part of this chapter, we consider dynamical systems arising from substitutions. We show that dropping a standard assumption on the nature of the substitution (primitivity) allows for new spectral features to arise that have not been observed for primitive substitutions. Inter alia, we show that, in an appropriately chosen situation, the corresponding Schrödinger operator admits an eigenvalue that is an accumulation point of the spectrum. The corresponding eigenfunction represents a localized electronic state. In the second part of the chapter, we are concerned with an important property of a symbolic dynamical system, called *Boshernitzan's condition*, which is known to lead to a Cantor spectrum of Lebesgue measure 0. We prove that this property is stable under periodic perturbations.

In Chapter 6, we focus on the Thue–Morse measure, the diffraction measure associated to a classic configuration of aperiodic order [Mah27]. This paradigmatic example of a singular continuous measure has been studied extensively, including a *multifractal analysis* of its scaling behaviour in the mathematical physics literature [GL90]. We revisit this analysis and put it in the context of the modern thermodynamic formalism. In this framework, the Thue–Morse measure exhibits a pathology that requires one to extend the existing methods. This technical advancement could be of independent interest.

We close with an outlook on progress and open questions.

1. Preliminaries

It is the aim of this chapter to collect some of the basic mathematical tools and concepts that will find applications as we progress towards our main results. At first, to avoid ambiguities, we fix some notational conventions in Section 1.1.

In Section 1.2, we present *Lebesgue's decomposition theorem*, an important device to categorize the singularity of a measure and associated objects. In particular, this provides a key splitting for the spectrum of Schrödinger operators in Chapter 5. Then, we discuss Radon measures as complex-valued functionals and their Fourier transformability—a problem that will naturally occur in the context of diffraction.

The classic *Perron–Frobenius theorem* on the spectral properties of irreducible non-negative matrices is recalled in Section 1.3. This will be an invaluable tool to analyze how certain characteristics of a pattern change under the *substitutions* that we introduce in Chapter 3.

A random version of substitutions is the main object of interest in Chapter 4. This naturally requires some background on Markov chains, which is provided in Section 1.4. In the form of a random walk on the torus, we will also encounter Markov chains as we discuss g-measures in Chapter 6.

1.1. Basic conventions

Let us agree to call a real number x (strictly) *positive* if $x > 0$ and *non-negative* if $x \geqslant 0$. The natural numbers $\mathbb{N} = \{1, 2, 3, \ldots\}$ are assumed to be positive, and we write $\mathbb{N}_0 = \mathbb{N} \cup \{0\}$ for the non-negative integers. Likewise, we set $\mathbb{R}_+ = \{x \in \mathbb{R} : x > 0\}$ and $\mathbb{R}_{\geqslant 0} = \{x \in \mathbb{R} : x \geqslant 0\}$. The set inclusion $\subset$ is understood to include equality, and we write $\varnothing$ for the empty set. We write $A \sqcup B$ for the disjoint union of sets A, B with $A \cap B = \varnothing$. If the set X is understood and $A \subset X$, we denote the complement of A by $A^C = X \setminus A$. Manipulations of sets are to be understood elementwise. For example, given $A, B \subset \mathbb{R}^d$, we use the Minkowski sum

$$A + B \,=\, \{a + b : a \in A, b \in B\}.$$

A singleton set $\{x\}$ is often identified with x notationally. We refer to countable subsets of $\mathbb{R}^d$ as *point sets*.

The cardinality of a finite set S is denoted by $\#S$. We also adopt the notations $\#S = 0$ if $S = \varnothing$ and $\#S = \infty$ if S is an infinite set. A set $\Lambda \subset \mathbb{R}^d$ is called *locally finite* if, for every compact subset $K \subset \mathbb{R}^d$, we have $\#(\Lambda \cap K) < \infty$.

In several places, we will use the intuitive bra-ket notation on the vector space $\mathbb{R}^d$. In this case, we denote $x \in \mathbb{R}^d$ by $|x\rangle$ if it is understood as a column vector and by $\langle x|$ for the corresponding row vector. Both can be understood as elements of $\mathbb{R}^d$, reflecting the

self-duality of this vector space. Concatenation of these objects is understood as matrix multiplication. In this formalism, the scalar product of $x, y \in \mathbb{R}^d$ is denoted by $\langle x, y\rangle = \langle x|y\rangle$, where one of the vertical bars is dropped, in line with standard conventions. Likewise, the expression $|y\rangle\langle x|$ naturally defines a (projection) matrix on $\mathbb{R}^d$.

For $x \in \mathbb{C}$, we write the complex conjugate of x as $\overline{x}$. We equip $\mathbb{C}^d$ with the scalar product $\langle x, y\rangle = \sum_{j=1}^{d} \overline{x_j} y_j$, which induces the standard metric on $\mathbb{C}^d$. Likewise, we equip $\mathbb{R}^d$ with the Euclidean metric and regard it as a topological space.

Given a topological space X, we write $C(X)$ for the vector space of continuous functions $f\colon X \to \mathbb{C}$. We also set $C_c(\mathbb{R}^d)$ for the functions in $C(\mathbb{R}^d)$ with compact support. Every topological space X is tacitly equipped with the corresponding Borel σ-algebra $\mathcal{B}$, and we write $\mathcal{M}(X)$ for the set of Borel measures on X, as well as $\mathcal{M}^1(X)$ for the simplex of probability measures in $\mathcal{M}(X)$. We denote by $L^p(X, \mu)$ the Banach space of equivalence classes of measurable functions $f\colon X \to \mathbb{C}$, such that $|f|^p$ is integrable. For every μ-integrable function $f\colon X \to \mathbb{C}$, we set $\mu(f) = \int_X f \,\mathrm{d}\mu$. For $p = 2$, the scalar product

$$\langle g, f\rangle \;=\; \int_X \overline{g(x)} f(x) \,\mathrm{d}\mu(x),$$

for all $f, g \in L^2(X, \mu)$, turns $L^2(X, \mu)$ into a Hilbert space. If $X = \mathbb{Z}$, with the discrete topology, and μ is the counting measure on $\mathbb{Z}$, we write $\ell^2(\mathbb{Z})$ instead of $L^2(\mathbb{Z}, \mu)$.

For a measurable function $g\colon X \to \mathbb{R}_{\geqslant 0}$ and a regular Borel measure $\mu \in \mathcal{M}(X)$, we let $g\mu$ be the measure defined by

$$(g\mu)(A) \;=\; \int_X \mathbb{1}_A(x) g(x) \,\mathrm{d}\mu(x),$$

for every Borel set A, where $\mathbb{1}_A(x) = 1$ if $x \in A$ and $\mathbb{1}_A(x) = 0$ otherwise. If X is $\mathbb{R}^d$ (or a compact subset of R^d), we denote by λ_{L} the standard Lebesgue measure on X. In this case, we sometimes identify g and $g\lambda_{\mathrm{L}}$.

In many places, we will be concerned with iterations of a map $f\colon S \to S$ on a set S. Let us write f^n for the n-fold concatenation $f \circ \ldots \circ f$ if $n \in \mathbb{N}$, and $f^0 := \mathrm{Id}\colon x \mapsto x$. If f is invertible, we also write f^{-n} for $(f^{-1})^n$, for all $n \in \mathbb{N}$.

Finally, we introduce some probabilistic notation. Given a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, we call a measurable map $Y\colon \Omega \to \mathbb{R}$ a *random variable*. If $(X, \mathcal{B})$ is a general Borel space, we call a measurable map $Y\colon \Omega \to X$ an X-valued random variable. We adopt the usual convention to write $\{Y \in C\}$ for the set $Y^{-1}(C) = \{\omega \in \Omega : Y(\omega) \in C\}$, provided $C \in \mathcal{B}$. The expectation value of a random variable, Y with respect to $\mathbb{P}$, is generically denoted by $\mathbb{E}[Y]$, and we write $\mathbb{V}[Y] = \mathbb{E}\big[Y^2\big] - \mathbb{E}[Y]^2$ for the corresponding variance.

1.2. Measures and functionals

1.2.1. Lebesgue decomposition

Two measures μ, ν on a measurable space $(X, \mathcal{F})$ are called (mutually) *singular*, denoted by $\mu \perp \nu$, if there exists an $A \in \mathcal{F}$ such that $\mu(A) = 0$ and $\nu(X \setminus A) = 0$. We say that μ is

absolutely continuous with respect to ν, denoted by $\mu \ll \nu$, if $\mu(A) = 0$ whenever $\nu(A) = 0$ for $A \in \mathcal{F}$. The measures μ and ν are called *equivalent*, denoted $\mu \sim \nu$, if $\mu \ll \nu$ and $\nu \ll \mu$. For the following general decomposition theorem, compare for example [Hal74, Sec. 32].

Fact 1.2.1. *Let μ, ν be σ-finite measures on some measurable space $(X, \mathcal{F})$. Then, there is a unique pair of σ-finite measures μ_1, μ_2 such that $\mu = \mu_1 + \mu_2$, $\mu_1 \perp \nu$ and $\mu_2 \ll \nu$.* ∎

A more refined statement is possible if we restrict to regular Borel measures on $\mathbb{R}^d$. Here, we will take the standard Lebesgue measure λ_{L} as a reference measure and say that a measure μ is *absolutely continuous* or *singular* if the corresponding attribute holds relative to Lebesgue measure.

Let $\mu \in \mathcal{M}(\mathbb{R}^d)$ be a regular Borel measure. We call $P_\mu = \{x \in \mathbb{R}^d : \mu(\{x\}) > 0\}$ the set of *pure points* of μ. We say that μ is *continuous* if $P_\mu = \varnothing$, and we say that μ is *pure point* if $\mu(A) = \sum_{x \in A \cap P_\mu} \mu(\{x\})$ for every Borel set A. We call a measure *singular continuous* if it is both singular and continuous. The following classic result is known as the *Lebesgue decomposition theorem*; compare [BG13, Thm. 8.3].

Fact 1.2.2. *Every regular Borel measure μ on $\mathbb{R}^d$ has a unique (Lebesgue) decomposition*

$$\mu \;=\; \mu_{\mathrm{pp}} + \mu_{\mathrm{sc}} + \mu_{\mathrm{ac}},$$

where μ_{pp} is a pure point measure, μ_{sc} is singular continuous, and μ_{ac} is absolutely continuous (with respect to Lebesgue measure). ∎

If only one of the components is present in the Lebesgue decomposition of μ, we say that μ is of *pure type*. In this case, we refer to the attribute $\bullet \in \{\mathrm{pp}, \mathrm{sc}, \mathrm{ac}\}$ such that $\mu = \mu_\bullet$ as the *spectral type* of μ. We also write $\mu_{\mathrm{c}} = \mu_{\mathrm{sc}} + \mu_{\mathrm{ac}}$ for the continuous part of μ and $\mu_{\mathrm{s}} = \mu_{\mathrm{pp}} + \mu_{\mathrm{sc}}$ for the singular part of μ.

1.2.2. Radon measures

Every (positive) locally finite measure $\mu \in \mathcal{M}(\mathbb{R}^d)$ defines a linear functional on the vector space $C_c(\mathbb{R}^d)$ by

$$\mu(g) \;:=\; \int_{\mathbb{R}^d} g(x)\, \mathrm{d}\mu(x), \tag{1.1}$$

for all $g \in C_c(\mathbb{R}^d)$. This functional is *positive* in the sense that $\mu(g) \geqslant 0$, whenever $g \geqslant 0$. In fact, the positive linear functional on $C_c(\mathbb{R}^d)$ are in one-to-one correspondence, via (1.1), to the locally finite, regular Borel measures on $\mathbb{R}^d$. This is due to the Riesz–Markov representation theorem [Rud87, Ch. 2].

Example 1.2.3. Given $x \in \mathbb{R}^d$, we define the *Dirac measure* δ_x via $\delta_x(A) = 1$, precisely if $x \in A$, and $\delta_x(A) = 0$ otherwise, for every Borel set A. As a linear functional, this takes the form $\delta_x(g) = g(x)$, for all $g \in C_c(R^d)$. ◊

We push the correspondence between measures and linear functionals one step further by *defining* measures as linear functionals, even if they are not positive.

Definition 1.2.4. A *Radon measure* on $\mathbb{R}^d$ is a (complex) linear functional μ on $C_c(\mathbb{R}^d)$ such that, for every compact $K \subset \mathbb{R}^d$, there is some constant c_K with $|\mu(g)| \leqslant c_K ||g||_\infty$, for all $g \in C(K)$.

We call a Radon measure μ *real*, if $\mu(f) \in \mathbb{R}$ for every real-valued $f \in C_c(\mathbb{R}^d)$. It is called *positive* if $\mu(g) \geqslant 0$ for all $g \geqslant 0$. Recall that every positive Radon measure may be regarded simply as a measure. For a complex-valued, continuous function f on $\mathbb{R}^d$, we define $f\mu$ via $(f\mu)(g) = \mu(fg)$, for all $g \in C_c(\mathbb{R}^d)$. In particular, if $c \in \mathbb{C}$, we obtain $(c\mu)(g) = c\mu(g)$, for all $g \in C_c(\mathbb{R}^d)$. Given a Radon measure μ, we further define,

$$\widetilde{\mu}\colon g \mapsto \overline{\mu(\widetilde{g})}, \quad \widetilde{g}\colon x \mapsto \overline{g(-x)},$$

for all $g \in \mathrm{C}_c(\mathbb{R}^d)$. This will be important in the context of autocorrelation measures.

Example 1.2.5. The main motivation for the introduction of Radon measures is to obtain a framework for *weighted Dirac combs*. For a locally finite point set $D \subset \mathbb{R}^d$, the Dirac comb

$$\delta_D := \sum_{x \in D} \delta_x$$

forms a locally finite measure. Given a family of complex weights $(w_x)_{x \in D}$, we obtain a Radon measure via

$$\omega = \sum_{x \in D} w_x \delta_x, \quad \omega\colon f \mapsto \sum_{x \in D} w_x f(x).$$

It is straightforward to verify that $\widetilde{\omega} = \sum_{x \in D} \overline{w_x} \delta_{-x}$. ◊

For every Radon measure μ, there exists a smallest *positive* Radon measure $|\mu|$ such that $|\mu(g)| \leqslant |\mu|(g)$ for all $g \in C_c(\mathbb{R}^d)$ with $g \geqslant 0$. The Radon measure μ is said to be *finite* if $|\mu|$ defines a finite measure on $\mathbb{R}^d$. We call μ *translation bounded* if $\sup_{x \in \mathbb{R}^d} |\mu|(x + K) < \infty$ for every compact $K \subset \mathbb{R}^d$.

Definition 1.2.6. The *convolution* of two finite Radon measures μ and ν on $\mathbb{R}^d$ is defined as

$$(\mu * \nu)(g) = \int_{\mathbb{R}^d \times \mathbb{R}^d} g(x + y) \,\mathrm{d}\mu(x) \,\mathrm{d}\nu(y).$$

We call a Radon measure μ on $\mathbb{R}^d$ *positive definite* if $\mu(g * \widetilde{g}) \geqslant 0$ for all $g \in C_c(\mathbb{R}^d)$. If μ is finite, it is straightforward to verify that $\mu * \widetilde{\mu}$ is always positive definite.

1.2.3. Fourier transform

We are interested in the Fourier transform of Radon measures. Naively, we would like to define $\widehat{\mu}(g) = \mu(\widehat{g})$. Unfortunately, the space $C_c(\mathbb{R}^d)$ is not invariant under the Fourier transform and hence, the expression $\mu(\widehat{g})$ is generally not well-defined. We therefore resort to a different space of functions as an intermediate step.

Let $\mathcal{S}(\mathbb{R}^d)$ be the Schwartz space on $\mathbb{R}^d$. The Fourier transform of a function $\phi \in \mathcal{S}(\mathbb{R}^d)$ is defined as

$$\widehat{\phi}(k) = \int_{\mathbb{R}^d} \mathrm{e}^{-2\pi \mathrm{i} k x} \phi(x) \,\mathrm{d}x,$$

for all $k \in \mathbb{R}^d$, using kx as a shorthand for $\langle k, x\rangle$. The Fourier transform is a homeomorphism on $\mathcal{S}(\mathbb{R}^d)$. In particular, $\widehat{\phi} \in \mathcal{S}(\mathbb{R}^d)$ for all $\phi \in \mathcal{S}(\mathbb{R}^d)$. The corresponding *dual space* $\mathcal{S}'(\mathbb{R}^d)$ is the space of continuous linear functionals on $\mathcal{S}(\mathbb{R}^d)$. An element $T \in \mathcal{S}'(\mathbb{R}^d)$ is called a *tempered distribution*. The Fourier transform $\widehat{T}$ of $T \in \mathcal{S}'(\mathbb{R}^d)$, is defined as

$$\widehat{T}(\phi) \,=\, T\big(\widehat{\phi}\,\big),$$

for all $\phi \in \mathcal{S}(\mathbb{R}^d)$. We call a linear functional $T_\mu \colon \mathcal{S}(\mathbb{R}^d) \cup C_c(\mathbb{R}^d) \to \mathbb{C}$ a *tempered measure*, if the restriction T of T_μ to $\mathcal{S}(\mathbb{R}^d)$ is a tempered distribution, and the restriction μ of T_μ to $C_c(\mathbb{R}^d)$ defines a Radon measure. The Fourier transform $\widehat{T}$ may or may not have an extension to a tempered measure $\widehat{T}_\mu$. If it exists, such an extension is unique because $\mathcal{S}(\mathbb{R}^d) \cap C_c(\mathbb{R}^d)$ lies dense in $C_c(\mathbb{R}^d)$. In this case, we write $\widehat{\mu}$ for the restriction of $\widehat{T}_\mu$ to $C_c(\mathbb{R}^d)$, and we call $\widehat{\mu}$ the *Fourier transform* of μ. The following criterion for the Fourier transformability of measures, based on a theorem by Bochner–Schwartz, will be of great avail in the context of diffraction; compare [BG13, Prop. 8.6].

Proposition 1.2.7. *If a Radon measure μ on $\mathbb{R}^d$ is positive definite, its Fourier transform $\widehat{\mu}$ exists, and is a translation bounded (positive) measure on $\mathbb{R}^d$.* ∎

The Fourier transform can be made more explicit if μ is a finite Radon measure. In this case, $\widehat{\mu}$ exists and takes the form $\widehat{\mu} = f_{\widehat{\mu}}\lambda_{\mathrm{L}}$, with

$$f_{\widehat{\mu}}(k) \,=\, \int_{\mathbb{R}^d} \mathrm{e}^{-2\pi \mathrm{i} kx}\, \mathrm{d}\mu(x),$$

defining a bounded, uniformly continuous function on $\mathbb{R}^d$, see [BG13, Ch. 8.6]. With some abuse of notation, we write $\widehat{\mu}(k)$ for $f_{\widehat{\mu}}(k)$. If μ is supported on the unit interval, we call $\widehat{\mu}(k)$, with $k \in \mathbb{Z}$, the *Fourier–Stieltjes coefficients* of μ, see [BG13, Ch. 8.7], and [Rud62] for general background.

1.3. Non-negative matrices

We write $\mathrm{Mat}(d, \mathbb{R})$ for the ring of $d \times d$-matrices with entries in $\mathbb{R}$. This notation extends to the matrix rings $\mathrm{Mat}(d, \mathbb{Z})$ and $\mathrm{Mat}(d, \mathbb{C})$ with obvious meaning. The 1-norm of $v \in \mathbb{R}^d$ is given by $||v||_1 = \sum_{i=1}^d |v_i|$. It will sometimes be convenient to index a (square) matrix by a finite set $\mathcal{A}$ instead of a subset of the natural numbers. In this case, we write $\mathrm{Mat}(\mathcal{A}, \mathbb{R})$ for the ring of matrices $(M_{ab})_{a,b\in\mathcal{A}}$, with $M_{ab} \in \mathbb{R}$, for all $a, b \in \mathcal{A}$.

A matrix $M \in \mathrm{Mat}(d, \mathbb{R})$ is called *positive*, written $M > 0$, if $M_{ij} > 0$ for all $1 \leqslant i, j \leqslant d$. Likewise, a *non-negative* matrix $M \geqslant 0$ is such that all entries of M are non-negative. Given two matrices $M, N \in \mathrm{Mat}(d, \mathbb{R})$, we set $M > N$ if $M - N > 0$ and $M \geqslant N$ if $M - N \geqslant 0$.

Definition 1.3.1. A non-negative matrix $M \in \mathrm{Mat}(d, \mathbb{R})$ is called *irreducible* if for each pair of indices (i, j) there exists a power $q \in \mathbb{N}$ such that $(M^q)_{ij} > 0$. The matrix $M \geqslant 0$ is called *primitive* if there exists a $q \in \mathbb{N}$ such that $M^q > 0$.

Remark 1.3.2. If a matrix M is irreducible, there exists a number $p \in \mathbb{N}$ such that M^p decays into a direct sum of primitive matrices (possibly after a permutation of indices). We call the smallest such number p the *period* of M. Hence an irreducible matrix is primitive if and only if it is *aperiodic*, that is, if its period is $p = 1$. ◊

Arguably the most important tool for the study of non-negative matrices is the Perron–Frobenius (PF) theorem. It goes back to work of Perron, who studied positive matrices, and subsequent generalizations to non-negative matrices by Frobenius. The PF theorem is rather a family of results than a single assertion; for the formulation below compare [BG13, Thm. 2.2, Thm. 2.3]. A more exhaustive treatment, including infinite-dimensional generalizations, can be found in [Sen81] and [Scha99]. Recall that we call a non-zero vector L a *left eigenvector* of M with respect to an eigenvalue λ if $\langle L|M = \lambda\langle L|$.

Theorem 1.3.3 (Perron–Frobenius)**.** *Let $M \in \mathrm{Mat}(d, \mathbb{R})$ be a non-negative irreducible matrix. Then, the spectral radius $\lambda = \rho(M)$ is a simple eigenvalue of M, and there exist corresponding left and right eigenvectors L and R with strictly positive entries and normalized as $\langle L|R\rangle = ||R||_1 = 1$. If M is also primitive, λ is strictly larger in modulus than every other eigenvalue of M, and*

$$\lim_{n\to\infty} \frac{1}{\lambda^n} M^n \;=\; |R\rangle\langle L|,$$

with geometric rate of convergence. ■

In the situation of Theorem 1.3.3, we call λ the *PF eigenvalue* of M, whereas L and R are called the (properly normalized) left and right *PF eigenvectors* of M, respectively.

1.4. Markov chains

We approach the subject of Markov chains in several steps, dealing first with finite, then countable and finally uncountable state spaces. Our focus will shift as we pass through these cases, motivated by the different applications that they find throughout this thesis.

1.4.1. Finite state space

Let S be a finite set, comprising the states of the system. A *Markov matrix* is a non-negative matrix $M \in \mathrm{Mat}(S, \mathbb{R})$ with $\sum_{s\in S} M_{rs} = 1$ for all $r \in S$. It follows that $\lambda = 1$ is an eigenvalue of maximal modulus for M. We call a probability vector π on S an *equilibrium vector* if $\langle\pi| = \langle\pi|M$.

We interpret M_{rs} as the probability of passing from r to s in one time step. We say that s is accessible from r, denoted by $r \rightsquigarrow s$, if there is some $k \in \mathbb{N}_0$ such that $(M^k)_{rs} > 0$. If $r \rightsquigarrow s$ and $s \rightsquigarrow r$, we say that r and s *communicate*, denoted by $r \leftrightarrow s$. Communication is an equivalence relation, the corresponding equivalence classes are called *communication classes*; compare [Bre20, Ch. 2] for background.

Recall that a state s is said to be *recurrent* if the probability of a finite return time to itself is equal to 1. Instead of making this precise, we give an equivalent definition. We emphasize that this works only because S is assumed to be finite.

Definition 1.4.1. We call a state $s \in S$ *recurrent* if $s \rightsquigarrow r$ implies $r \rightsquigarrow s$ for all $r \in S$. It is called *transient* otherwise.

Let $\mathcal{R} \subset S$ be the set of recurrent states. Since recurrence is a class property, $\mathcal{R}$ decays into a finite collection of communication classes $\mathcal{R}_1, \dots, \mathcal{R}_m$, called *recurrence classes*. By construction, the restriction M_j of M to a recurrence class $\mathcal{R}_j$ is irreducible. Let $p_j \in \mathbb{N}$ denote the period of the irreducible matrix M_j. The least common multiple p of all the integers $\{p_j\}_{1 \leqslant j \leqslant m}$ is called the *period* of M. The restriction of M^p to $\mathcal{R}_j$ decays into p_j primitive components, for all $1 \leqslant j \leqslant m$.

Remark 1.4.2. The sets of recurrent and transient states remain the same for every power of the Markov matrix M, however the recurrence classes may become smaller. Hence, what we have gained by passing from M to M^p is that we can assume that the restriction of M^p to each of its recurrence classes is primitive. ◊

Up to replacing M by some of its powers, we can assume that it has period $p = 1$. In this case, we say that M is *aperiodic*. For each recurrence class $\mathcal{R}_j$, there is a unique equilibrium vector π^j of M that is supported on $\mathcal{R}_j$. The long term behaviour of M is captured by the following result; compare [Bre20, Ch. 6]. Here, we denote by e_s the unit vector given by $(e_s)_s = 1$ and $(e_s)_r = 0$ if $r \neq s$.

Fact 1.4.3. *Let M be an aperiodic Markov matrix on the state space S with recurrence classes $\mathcal{R}_1, \dots, \mathcal{R}_m$ and corresponding equilibrium vectors $\pi^1, \dots, \pi^m$. For every $s \in S$, there exists a unique probability vector $c_s = (c_{s,j})_{1 \leqslant j \leqslant m}$ such that*

$$\lim_{n \to \infty} \langle e_s | M^n = \sum_{j=1}^{m} c_{s,j} \langle \pi^j |,$$

holds for all $s \in S$. ■

We call $c_{s,j}$ the *absorption probability* of s by the class $\mathcal{R}_j$. The vector c_s is a unit vector whenever s is itself a recurrent state.

1.4.2. Countable state space

Assume now that S is a countable set. A Markov matrix $M = (M_{rs})_{r,s \in S}$ on S is a non-negative matrix on the *infinite* index set S with the property that $\sum_{s \in S} M_{rs} = 1$ for all $r \in S$. A homogeneous Markov chain on S with Markov matrix M is a sequence of S-valued random variables $(X_n)_{n \in \mathbb{N}_0}$ on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$ such that

$$\mathbb{P}[X_{n+1} = s_{n+1} | X_n = s_n, \dots, X_0 = s_0] = \mathbb{P}[X_{n+1} = s_{n+1} | X_n = s_n] = M_{s_n s_{n+1}}, \tag{1.2}$$

for all $n \in \mathbb{N}$ and $s_0, \dots, s_{n+1} \in S$. The *initial distribution* $\mathbb{P} \circ X_0^{-1}$, together with M, uniquely determines the distribution of the sequence-valued random variable $(X_n)_{n \in \mathbb{N}_0}$ on $S^{\mathbb{N}_0}$. Conversely, given an initial distribution vector π_0 on S and a Markov matrix M, a corresponding Markov process with the required properties always exists. This is due to

Kolmogorov's extension theorem; compare for example [Kle14, Thm. 14.36]. The probability vector π_n, corresponding to the measure $\mathbb{P} \circ X_n^{-1}$ on S, can be calculated from

$$\langle \pi_n | \;=\; \langle \pi_0 | M^n,$$

for all $n \in \mathbb{N}$. We will mostly be concerned with the case that every row of M has only finitely many non-zero entries. In this case, every component of $\langle \pi_n |$ can be calculated from a finite sum.

1.4.3. General state space

Let X be a complete metric space with Borel σ-algebra $\mathcal{B}$. In this case, we replace the concept of a Markov matrix by that of a Markov kernel.

Definition 1.4.4. A *Markov kernel* on $(X, \mathcal{B})$ is a map $\kappa\colon (X, \mathcal{B}) \to [0, 1]$ such that

(1) $x \mapsto \kappa(x, A)$ is a measurable function, for all $A \in \mathcal{B}$.

(2) $A \mapsto \kappa(x, A)$ is a probability measure, for every $x \in X$.

For a general state space, events like $\{X_n = x\}$ typically form a null set. We therefore have to be careful in generalizing the Markov property from (1.2). For an appropriate notion of conditional probabilities, see Appendix A and [Kle14, Ch. 8]. In the following, we use the notation $\sigma(X_0, \dots, X_n)$ for the smallest σ-algebra generated by the random variables $X_0, \dots, X_n$.

Given a Markov kernel κ on $(X, \mathcal{B})$ and an initial probability distribution μ_0, there exists an X-valued, homogeneous Markov chain $(X_n)_{n\in\mathbb{N}_0}$ on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$ such that $\mathbb{P} \circ X_0^{-1} = \mu_0$, and

$$\mathbb{P}[X_{n+1} \in A | \sigma(X_0, \dots, X_n)] \;=\; \kappa(X_n, A), \tag{1.3}$$

holds $\mathbb{P}$-almost surely, for all $A \in \mathcal{B}$ and $n \in \mathbb{N}_0$. This follows from a famous result due to Ionescu-Tulcea; see for example [Kle14, Thm. 14.32].

Together with the initial distribution μ_0, the relation (1.3) fixes the finite-dimensional distributions $\mathbb{P} \circ (X_0, \dots, X_n)^{-1}$ unambiguously. In fact, the distribution of the random sequence $(X_n)_{n\in\mathbb{N}_0}$ is uniquely determined by μ_0 and the transition kernel κ. A relation between the distributions $\mu_n = \mathbb{P} \circ X_n^{-1}$ can be found by integrating (1.3). Indeed, we have $\mu_{n+1} = \kappa^* \mu_n$, for all $n \in \mathbb{N}_0$, where

$$(\kappa^* \mu)(A) \;=\; \int_{x\in X} \kappa(x, A)\, \mathrm{d}\mu_n(x),$$

for all $A \in \mathcal{B}$ and Borel probability measures μ on X. The Markov operator $\kappa^*\colon \mu \mapsto \kappa^* \mu$ is called the *dual* of the transition kernel κ.

2. Dynamical systems

Throughout substantial parts of this thesis, the structures that we investigate can be modelled by bi-infinite sequences. Equipping the space of sequences with an index-shift places us into the realm of *symbolic dynamical systems*, which we introduce in Section 2.1 as our key players.

The field of dynamical systems offers an impressive variety of concepts that assess the degree of (dis)order of the system, most of them invariant under appropriate isomorphisms. For the exposition in this chapter, we cluster them into three different groups.
The mixing properties presented in Section 2.2 characterize how well the underlying space is exhausted by following the orbit of single points (or small neighbourhoods). Some of these notions make reference to an associated measure while others only rely on the topological structure.
In Section 2.3, we adopt a point of view inspired by information theory and introduce the notions of topological and metric entropy. The connection between the two concept emerges via a variational principle.
An operator-theoretic approach yields a spectral characterization of the underlying dynamical system. In Section 2.4, we develop the formalism for both unitary and self-adjoint operators. The unitary Koopman operator gives rise to the dynamical spectrum which is closely related to the (sometimes more easily computable) diffraction spectrum. The self-adjoint Schrödinger operator will be the main object of interest in Chapter 5.

2.1. Symbolic dynamics

2.1.1. From dynamical systems to symbolic sequences

For the sake of this thesis, a (topological) dynamical systems (X, T) consists of a compact space X, equipped with a continuous map $T\colon X \to X$. The map T may or may not be invertible, but for most of the examples that we will encounter, T is surjective. The dynamics on the base space X is induced by iterating the map T. That is, given $x \in X$, we should think of $T^n x$ as the position of x at time $n \in \mathbb{N}_0$. If T is invertible, we even take $n \in \mathbb{Z}$ and interpret negative integers as pertaining to the past. Following this idea, the *orbit* of $x \in X$ under T is defined as,

$$\mathrm{orb}_T(x) \,=\, \{T^n x : n \in \mathbb{I}\},$$

where, here and in the following, $\mathbb{I} = \mathbb{Z}$ if T is invertible and $\mathbb{I} = \mathbb{N}_0$ otherwise. We simply write $\mathrm{orb}(x)$ if the map T is understood. In many situations (X, T) is equipped with a measure μ, which we tacitly assume to be defined on the Borel σ-algebra that is induced by the topology on X. We say that μ is *T-invariant*, if $\mu = \mu \circ T^{-1}$. We denote by $\mathcal{M}_T(X)$ the space of T-invariant (Borel) measures on X, and by $\mathcal{M}_T^1(X)$ the convex subset of probability

measures in $\mathcal{M}_T(X)$. Given $\mu \in \mathcal{M}_T^1(X)$, we call (X, T, μ) a *measure-preserving dynamical system*.

From a general dynamical system, we obtain a symbolic coding by coarse-graining the space X. This procedure is justified by the fact that, in practical situations, only a finite resolution is available. More precisely, we consider a *partition* of X into Borel sets, given by $\alpha = \{X_a\}_{a\in\mathcal{A}}$, where the *alphabet* $\mathcal{A}$ constitutes a finite index set. Being a partition means that $X = \cup_{a\in\mathcal{A}} X_a$ and that, for $a, b \in \mathcal{A}$ with $a \neq b$, the sets X_a and X_b are disjoint. We follow the dynamics of a point $x \in X$ by recording in a string of symbols the succession of sets that are visited by the orbit of x. That is, we obtain a coding map

$$\pi_\alpha \colon x \mapsto (a_n)_{n\in\mathbb{I}},$$

where a_n is the uniquely determined letter such that $T^n x \in X_{a_n}$, for all $n \in \mathbb{I}$. If the partition α is chosen wisely and if the map T is chaotic enough, we can recover the point x from the sequence $\pi_\alpha(x)$ and narrow down the position of x to a small region by just considering a finite number of symbols from the sequence $\pi_\alpha(x)$. Questions about the dynamics on X are hence naturally translated to questions about the obtainable successions of symbols. This point of view is particularly relevant for the discussion of metric entropy in Section 2.3.

2.1.2. Symbolic notation

In what follows, we set up the basic notation for combinatorics on words. This is mostly in line with the conventions in [BG13, Ch. 4] to which we refer for further details; see also [LM95]. We start with a finite set $\mathcal{A}$, called *alphabet*. Elements of $\mathcal{A}$ are called *letters*. To exclude trivial cases, we assume that $\#\mathcal{A} \geqslant 2$. A *word* is an element of $\mathcal{A}^+ = \bigcup_{n\in\mathbb{N}} \mathcal{A}^n$. If $u \in \mathcal{A}^n$, we say that u is of *length* n, and write $|u| = n$ in this case. Given $1 \leqslant j \leqslant n$ and $u \in \mathcal{A}^n$, we denote by u_j the j-th entry of the n-tuple u. In line with standard conventions, we write $u_1u_2\cdots u_n$ instead of $(u_1, u_2, \ldots, u_n)$ for a word u in $\mathcal{A}^n$. The *concatenation* of two words $u = u_1\cdots u_n \in \mathcal{A}^n$ and $v = v_1\cdots v_m \in \mathcal{A}^m$, is given by $uv = u_1\cdots u_n v_1\cdots v_m \in \mathcal{A}^{n+m}$. Thus, the set of words $\mathcal{A}^+$ forms a (free) semigroup under concatenation. It can be extended to a monoid $\mathcal{A}^* = \mathcal{A}^+ \cup \{\varepsilon\}$ by adding the empty word ε, satisfying $\varepsilon v = v\varepsilon = v$ for all $v \in \mathcal{A}^+$.

We call a word v a *prefix* of another word u if there is a word $v' \in \mathcal{A}^*$ such that $u = vv'$. Similarly, v is a *suffix* of u, if there is $v' \in \mathcal{A}^*$ such that $u = v'v$. Given $u \in \mathcal{A}^n$ and $1 \leqslant j \leqslant k \leqslant n$, we call

$$u_{[j,k]} := u_j \cdots u_k$$

a *subword* of u. By convention, $u_{[j,j]} = u_j$ for all $1 \leqslant j \leqslant |u|$. For two words u, v with $|v| \leqslant |u|$, we set

$$|u|_v = \#\{1 \leqslant j \leqslant |u| - |v| + 1 : u_{[j,j+|v|-1]} = v\}$$

to be the number of times that v occurs in u as a subword. Note that different occurrences of v in u may overlap. We write $v \triangleleft u$ if v occurs as a subword of u. In the special case that $v = a \in \mathcal{A}$ is a letter, we obtain $|u|_a = \#\{1 \leqslant j \leqslant |u| : u_j = a\}$. The *Abelianisation* of a word u is given by

$$\Phi(u) = (|u|_a)_{a\in\mathcal{A}}.$$

Intuitively, the Abelianisation erases the specific order of letters in a word as it amounts to counting their absolute numbers.

For $n \in \mathbb{N}$ and $u \in \mathcal{A}^+$, we write $u^n = u \cdots u$ for the n-fold concatenation of the word u with itself. Formally, we set $u^0 = \varepsilon$. In Chapter 5, we will also be concerned with rational powers of a word u. For $u = u_1 \cdots u_m$, and $1 \leqslant k < m$, we write $u^{k/m} = u_1 \cdots u_k$, for the prefix of u of length k. Accordingly, given $s = n + k/m$, for $n \in \mathbb{N}$ and $1 \leqslant k < m$, we set

$$u^s \;=\; u^n u_1 \cdots u_k,$$

for all $u \in \mathcal{A}^m$.

2.1.3. From symbolic sequences to dynamics

In order to get back to dynamical systems, we consider the spaces of (bi-)infinite sequences of symbols, $\mathcal{A}^{\mathbb{N}}$ and $\mathcal{A}^{\mathbb{Z}}$. Let $\mathbb{I}$ stand for $\mathbb{N}$ or $\mathbb{Z}$ in the following. We equip $\mathcal{A}^{\mathbb{I}}$ with the product topology inherited from the discrete topology on $\mathcal{A}$. Thus, $\mathcal{A}^{\mathbb{I}}$ is a compact space due to Tychonoff's theorem [Cec37, Tyc30]. It is noteworthy that the topology on $\mathcal{A}^{\mathbb{I}}$ is metrizable. Somewhat arbitrarily, we fix a metric on $\mathcal{A}^{\mathbb{I}}$ by

$$d(x, y) \;=\; \sum_{n \in \mathbb{I}} \frac{\delta_{x_n, y_n}}{2^n},$$

that is, two sequences $x, y \in \mathcal{A}^{\mathbb{I}}$ are close precisely if they agree on a large window around the origin. For $x = (x_n)_{n \in \mathbb{I}} \in \mathcal{A}^{\mathbb{I}}$, and $j, k \in \mathbb{I}$ satisfying $j \leqslant k$ we set $x_{[j,k]} = x_j \cdots x_k$. The topology on $\mathcal{A}^{\mathbb{I}}$ is generated by cylinder sets of the form

$$[v]_j \;=\; \{x \in \mathcal{A}^{\mathbb{I}} : x_{[j, j+|v|-1]} = v\},$$

with $v \in \mathcal{A}^+$ and $j \in \mathbb{I}$. We write $[v]$ for the cylinder set starting at the origin, that is, we set $[v] = [v]_1$ if $\mathbb{I} = \mathbb{N}$ and $[v] = [v]_0$ if $\mathbb{I} = \mathbb{Z}$. These sets are both closed and open and hence termed *clopen*. In the bilateral case, we also make use of the notation

$$[u.v] \;=\; \{x \in \mathcal{A}^{\mathbb{Z}} : x_{[-|u|, |v|-1]} = uv\},$$

for $u, v \in \mathcal{A}^+$. We define the (left) shift action $S \colon \mathcal{A}^{\mathbb{I}} \to \mathcal{A}^{\mathbb{I}}$, via the prescription $(Sx)_n = x_{n+1}$ for all $n \in \mathbb{I}$. We call the compact dynamical system $(\mathcal{A}^{\mathbb{I}}, S)$ the *full shift* on $\mathcal{A}$. The map S is invertible on $\mathcal{A}^{\mathbb{I}}$ precisely if $\mathbb{I} = \mathbb{Z}$. A bi-infinite sequence $x = (x_n)_{n \in \mathbb{Z}} \in \mathcal{A}^{\mathbb{Z}}$ is often represented in the form

$$x \;=\; \cdots x_{-2} x_{-1} . x_0 x_1 x_2 \cdots,$$

where the lower dot indicates the position between the entries of index -1 and 0. If x is S-periodic, it can be written in the form

$$x \;=\; v^{\mathbb{Z}} = \cdots vvv.vvv \cdots,$$

for some $v \in \mathcal{A}^+$. We say that x is *eventually periodic to the right* if there is some $k \in \mathbb{Z}$ such that

$$(S^k x)_{[0,\infty)} \;=\; v^{\mathbb{N}},$$

for some $v \in \mathcal{A}^+$, with obvious meaning. Likewise, x is called *eventually periodic to the left* if $(S^k x)_{(-\infty,-1]} = v^{-\mathbb{N}}$, for some $k \in \mathbb{Z}$ and $v \in \mathcal{A}^+$. If $v \in \mathcal{A}$ is a letter, we use the term *eventually constant* instead of eventually periodic.

Definition 2.1.1. A *subshift* of $(\mathcal{A}^{\mathbb{I}}, S)$ is a dynamical system $(\mathbb{X}, S)$, given by an S-invariant closed subset $\mathbb{X}$ of $A^{\mathbb{I}}$, equipped with the restriction of S to $\mathbb{X}$. The *language* of $\mathbb{X}$ is defined as

$$\mathcal{L}(\mathbb{X}) \,=\, \{v \in \mathcal{A}^+ : [v] \cap \mathbb{X} \neq \varnothing\},$$

words in $\mathcal{L}(\mathbb{X})$ are called *legal*, and we set $\mathcal{L}_n(\mathbb{X}) = \mathcal{L}(\mathbb{X}) \cap \mathcal{A}^n$, for all $n \in \mathbb{N}$. We call $(\mathbb{X}, S)$ a *one-sided* (two-sided) subshift if $\mathbb{I} = \mathbb{N}$ ($\mathbb{I} = \mathbb{Z}$).

Remark 2.1.2. In Definition 2.1.1, we identify the restriction of S to $\mathbb{X}$ with S, by slight abuse of notation. Also, the term *subshift* sometimes refers to just $\mathbb{X}$ instead of $(\mathbb{X}, S)$. If the subshift is understood, we frequently write $[v]_j$ in place of $[v]_j \cap \mathbb{X}$, for $v \in \mathcal{A}^+$ and $j \in \mathbb{I}$. Since $\mathbb{X}$ is S-invariant by definition, the legal words are precisely those that appear as a subword of some element in $\mathbb{X}$. ◊

Let us present some particularly prominent examples of subshifts [LM95]. In Chapter 3, we will focus on another paradigmatic family of subshifts, arising from *substitutions*.

Example 2.1.3. Given $v \in \mathcal{A}^+$, the periodic orbit $\mathbb{X} = \mathrm{orb}_S(v^{\mathbb{Z}})$ is a finite set, hence closed, and clearly S-invariant. That is, $(\mathbb{X}, S)$ is a subshift. The cardinality of $\mathbb{X}$ is bounded by $|v|$ but might be smaller if v is itself an integer power of a shorter word. ◊

Example 2.1.4. Let $\#\mathcal{A} = n$ and $A \in \mathrm{Mat}(n, \mathbb{R})$ a matrix with entries in $\{0, 1\}$, indexed by the alphabet $\mathcal{A}$. We call

$$\mathbb{X}_A \,=\, \{x \in \mathcal{A}^{\mathbb{I}} : A_{x_i x_{i+1}} = 1 \text{ for all } i \in \mathbb{I}\}$$

a *Markov subshift* with *transition matrix* A (unless $\mathbb{X}_A = \varnothing$). By construction, $(\mathbb{X}_A, S)$ is a subshift. ◊

Example 2.1.5. Let $\mathcal{F} \subset \mathcal{A}^+$ be a finite collection of *forbidden words*. The subshift given by

$$\mathbb{X}_{\mathcal{F}} \,=\, \{x \in \mathcal{A}^{\mathbb{I}} : x_{[j,k]} \notin \mathcal{F} \text{ for all } j, k \in \mathbb{I}, \text{ with } j \leqslant k\}$$

is called a *subshift of finite type* (SFT), provided it is non-empty. ◊

Every Markov subshift $\mathbb{X}_A$ is an SFT, with $\mathcal{F} = \{ab \in \mathcal{A}^2 : A_{ab} = 0\}$. Conversely, every SFT is *conjugate* to a Markov subshift [LM95] in the following sense.

Definition 2.1.6. Let (X_1, T_1) and (X_2, T_2) be compact dynamical systems. We say that (X_2, T_2) is a (topological) *factor* of (X_1, T_1) if there exists a continuous, surjective map $\pi\colon X_1 \to X_2$ such that $\pi \circ T_1 = T_2 \circ \pi$. In this case, we call π a *factor map*. If π is even a homeomorphism, we say that (X_1, T_1) and (X_2, T_2) are (topologically) conjugate and call π a topological conjugation.

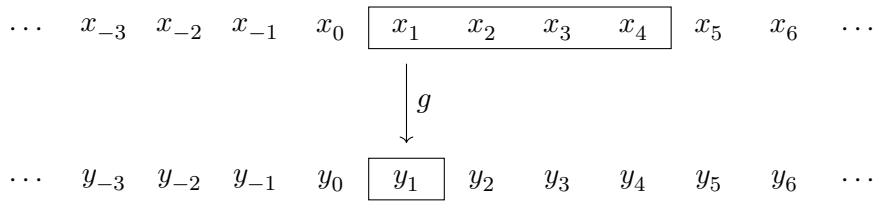

Figure 2.1.: Sliding block code ϕ_g acting on x, yielding the sequence $y = \phi_g(x)$.

Topological conjugacy is an important concept in the study of dynamical system because it preserves the dynamical properties, due to the defining commutation relation. In particular, all of the topological characteristics of dynamical systems that we encounter in the following sections are invariant under conjugacy. If (X_1, T_1) and (X_2, T_2) are both subshifts, factor maps and conjugations take a particular form. For simplicity, we assume that both subshifts are defined over the same alphabet (possibly by taking the union of the original alphabets).

Definition 2.1.7. Let $(\mathbb{X}_1, S)$ and $(\mathbb{X}_2, S)$ be subshifts of $(\mathcal{A}^{\mathbb{I}}, S)$. Given $n \in \mathbb{N}$ and a function $g\colon \mathcal{L}_n(\mathbb{X}_1) \to \mathcal{L}_1(\mathbb{X}_2)$, we define the (basic) *sliding block code* $\phi_g\colon \mathbb{X}_1 \to \mathbb{X}_2$ via

$$\phi_g\colon x \mapsto \phi_g(x), \quad (\phi_g(x))_j \,=\, g(x_{[j,j+n-1]}),$$

for all $j \in \mathbb{I}$. A *general* sliding block code is of the form $\phi_g \circ S^k$, for some $k \in \mathbb{I}$.

An illustration of a basic sliding block code for some function $g\colon \mathcal{L}_4(\mathbb{X}_1) \to \mathcal{L}_1(\mathbb{X}_2)$ is shown in Figure 2.1. It is easy to verify that every sliding block code is continuous and commutes with the shift action and is hence a factor map. On the other hand, given a factor map $\pi\colon \mathbb{X}_1 \to \mathbb{X}_2$ between subshifts, we can deduce that π is a sliding block code from the requirement that $\pi^{-1}([a])$ is clopen for all $a \in \mathcal{L}_1(\mathbb{X}_2)$ and $\pi \circ S = S \circ \pi$. Hence, sliding block codes are *precisely* the factor maps between subshifts. This is the content of the Curtis–Lyndon–Hedlund theorem; compare [LM95, Thm. 6.2.9].

There is a natural analogue of topological factor maps and topological conjugations to measure-preserving dynamical systems, which again offers a natural tool for their classification.

Definition 2.1.8. Let (X_1, T_1, μ_1) and (X_2, T_2, μ_2) be measure-preserving dynamical systems. We call (X_2, T_2, μ_2) a (measure-theoretic) *factor* of (X_1, T_1, μ_1) if there is a measurable map $\pi\colon X_1 \to X_2$ such that $\mu_2 = \mu_1 \circ \pi^{-1}$ and $\pi \circ T_1 = T_2 \circ \pi$ holds almost surely. In this case, we call π a factor map. If π is a bijection (up to null sets), we say that (X_1, T_1, μ_1) and (X_2, T_2, μ_2) are *metrically ismorphic* and call π a *metric isomorphism*.

Every topological conjugation π between (X_1, T_1) and (X_2, T_2) also constitutes a metric isomorphism between (X_1, T_1, μ) and $(X_2, T_2, \mu \circ \pi^{-1})$, for every $\mu \in \mathcal{M}_T^1(X_1)$. However, not every metric ismorphism arises in this manner; in general, it corresponds to a more coarse-grained perspective on the underlying spaces.

All of the measure-theoretic characteristics that we present in the following are invariant under metric isomorphisms. Within a class of dynamical systems, we call a characteristic a

complete invariant if it is the same for two dynamical systems *precisely* if these systems are isomorphic. We invite the reader to think of a complete invariant as a fingerprint of dynamical systems.

2.2. Mixing properties

2.2.1. Topological characteristics

Definition 2.2.1. A dynamical system (X, T) is said to be *topologically transitive* if for every pair of non-empty open sets $U, V \in X$, there exists an integer $n \in \mathbb{N}$ such that

$$T^n(U) \cap V \neq \varnothing. \tag{2.1}$$

In this situation, (X, T) is *topologically mixing* if there exists an $n_0 \in \mathbb{N}$, depending on U and V, such that (2.1) holds for all $n \geqslant n_0$.

Remark 2.2.2. If $(X, T) = (\mathbb{X}, S)$ is a subshift, both topological transitivity and topological mixing can be expressed as a combinatorial condition. Indeed, $(\mathbb{X}, S)$ is topologically transitive if and only if, for all $u, v \in \mathcal{L}(\mathbb{X})$, there is a word w such that $uwv \in \mathcal{L}(\mathbb{X})$. If $|w|$ can take any integer value greater than some n_0, the system $(\mathbb{X}, S)$ is even topologically mixing. We refer to [LM95, Ch. 6] for details. ◊

Example 2.2.3. Let $(\mathbb{X}_A, S)$ be a Markov subshift with transition matrix A. Let $\widehat{A}$ be the restriction of A to the index set given by $\mathcal{L}(\mathbb{X}_A)$, that is, we set $\widehat{A}_{ab} = A_{ab}$ for all $a, b \in \mathcal{L}(\mathbb{X}_A)$. Then, $(\mathbb{X}_A, S)$ is topologically transitive if and only if $\widehat{A}$ is irreducible and $(\mathbb{X}_A, S)$ is topologically mixing if and only if $\widehat{A}$ is primitive. This is straightforward to verify from Remark 2.2.2. ◊

For compact and metrizable dynamical systems, there is a useful characterization of topological transitivity that follows from a more general result by Silverman [Sil92]; see also [AC12].

Fact 2.2.4. *Let X be a compact metrizable space. If (X, T) is topologically transitive, there exists a point $x \in X$ such that $\mathrm{orb}_T(x)$ is dense in X. The converse holds if X has no isolated points.* ■

Points with a dense orbit in X are sometimes called *hypercyclic*. In particularly convenient situations, *all* points in X are hypercyclic.

Definition 2.2.5. A dynamical system (X, T) on a compact metrizable space X is called *minimal* if, for every $x \in X$, the orbit $\mathrm{orb}_T(x)$ is dense in X.

Another equivalent definition of minimality often found in the literature is that $X, \varnothing$ are the only T-invariant subsets of X. Periodic orbits are particularly simple examples of minimal systems; we will encounter more interesting examples in Chapter 3.

2.2.2. Measure-theoretic characteristics

For the remainder of this section, we assume that a dynamical system (X, T) is equipped with a T-invariant Borel measure μ. A useful property of such a measure with far-reaching consequences is *ergodicity*.

Definition 2.2.6. Given a dynamical system (X, T) and $\mu \in \mathcal{M}_T(X)$, we say that the measure μ is *ergodic* if, for every measurable set A with $A = T^{-1}A$, we have $\mu(A) = 0$ or $\mu(A^C) = 0$.

We will encounter dynamical systems with an *infinite* (but σ-finite) ergodic measure in Chapter 5. For background on infinite ergodic theory, we refer to the book by Aaronson [Aar97]. In the following, we stick to the more common case that μ is a probability measure. Under this assumption, there are many useful characterizations of ergodicity in the literature; compare for example [Pet83, Wal82]. A particularly useful property of ergodic probability measures is that they allow us to replace orbit averages for typical points by integration over the underlying space. This is the content of Birkhoff's ergodic theorem [Bir31]. In the following form, it is taken from [Pet83, Thm. 4.4].

Theorem 2.2.7 (Birkhoff)**.** *Let (X, T, μ) be a measure-preserving dynamical system. The measure μ is ergodic if and only if, for all $f \in L^1(X, \mu)$,*

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} f(T^k x) \;=\; \int_X f \,\mathrm{d}\mu$$

holds for μ-almost every $x \in X$. ∎

The ergodic probability measures on a dynamical system (X, T) are precisely the extremal points of the convex set of T-invariant Borel probability measures $\mathcal{M}_T^1(X)$. In particular, there always exists at least one ergodic measure.

Definition 2.2.8. We call a dynamical system (X, T) *uniquely ergodic* if there exists precisely one ergodic probability measure on (X, T). If, in addition, (X, T) is minimal, we call it *strictly ergodic*.

Although we may think of unique ergodicity as the analogue of minimality in the measure-theoretic category, the two notions are independent in general. A condition by Boshernitzan establishes a connection between the two properties for subshifts.

Definition 2.2.9. Let $(\mathbb{X}, S)$ be a subshift. For an invariant measure μ on $(\mathbb{X}, S)$, we set $\underline{\mu}(n) = \min\{\mu[v] : v \in \mathcal{L}_n(\mathbb{X})\}$. We say that $(\mathbb{X}, S)$ satisfies *Boshernitzan's condition* if there exists an invariant measure μ such that

$$\limsup_{n\to\infty} n\underline{\mu}(n) \;>\; 0.$$

If, in addition, $(\mathbb{X}, S)$ is minimal, we call it a *Boshernitzan subshift*.

If $(\mathbb{X}, S)$ is minimal, Boshernitzan's condition implies that it is also uniquely ergodic [Bos92]. It has also become an invaluable condition in the analysis of Schrödinger operators defined on symbolic dynamical systems; we will come back to this in Chapter 5.

There is yet another formulation of ergodicity, which underlines its interpretation as a mixing property. A measure $\mu \in \mathcal{M}_T^1(X)$ is ergodic on (X, T) precisely if

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} \mu(T^{-k}A \cap B) \;=\; \mu(A)\mu(B),$$

for all measurable $A, B \in X$. That is, ergodicity means convergence of $\mu(T^{-k}A \cap B)$ to $\mu(A)\mu(B)$ in a Cesàro sense. Stronger notions of convergence can be related to stronger versions of mixing.

Definition 2.2.10. Let $\mu \in M_T^1(X)$. We say that μ is *weakly mixing* on (X, T) if

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} |\mu(T^{-k}A \cap B) - \mu(A)\mu(B)| \;=\; 0,$$

for all measurable $A, B \in X$. We say that μ is *strongly mixing* on (X, T) if

$$\lim_{n\to\infty} \mu(T^{-n}A \cap B) \;=\; \mu(A)\mu(B), \tag{2.2}$$

for all measurable $A, B \in X$.

From these definitions, it is immediate that strong mixing implies weak mixing, which in turn implies ergodicity. Hence, within this hierarchy, ergodicity is in fact the weakest notion of mixing.

Example 2.2.11. Consider the full shift $(\mathcal{A}^{\mathbb{Z}}, S)$, equipped with the Bernoulli measure μ, defined via

$$\mu([v_1 \cdots v_n]) \;=\; (\#\mathcal{A})^{-n},$$

for all $v_1 \cdots v_n \in \mathcal{A}^n$ and $n \in \mathbb{N}$. It is easy to verify the defining relation for strong mixing in (2.2) if A, B are cylinder sets. This is in fact enough to conclude strong mixing because the cylinder sets (determining the first position) form a semi-algebra of sets that generate the Borel σ-algebra; compare the discussion in [Pet83, Ch. 2.5]. ♢

2.3. Entropy and complexity

2.3.1. Measure-theoretic entropy

Picture a situation where the orbit of a point is observed with finite resolution, modelled by some coarse-graining of the underlying space. The entropy of the dynamical system measures the average amount of information about the initial point that we can gain per step, provided we have a complete understanding of the underlying dynamics.

We start from a measure-preserving dynamical system (X, T, μ). As in Section 2.1, we model the coarse-graining by some finite measurable partition $\alpha = \{X_a\}_{a\in\mathcal{A}}$. To avoid technicalities, we further assume that $\mu(X_a) > 0$ for all $a \in \mathcal{A}$. A small value of $\mu(X_a)$ corresponds to a small a-priori probability of finding a point x in X_a and hence to a larger gain of information if $x \in X_a$ is revealed to hold true. More formally, we define the *information function* as $I_\alpha(x) = -\log(\mu(X_a))$, whenever $x \in X_a$. We obtain the corresponding entropy by averaging over all $x \in X$.

Definition 2.3.1. Let $\alpha = \{X_a\}_{a\in\mathcal{A}}$ be a finite, measurable partition of X and μ a probability measure on X. The *entropy* of the partition α is defined as

$$H_\mu(\alpha) \,=\, \int_X I_a \,\mathrm{d}\mu \,=\, -\sum_{a\in\mathcal{A}} \mu(X_a)\log\mu(X_a).$$

Let us denote by $\mathcal{P}$ the collection of all finite measurable partitions of X such that every element of the partition has positive μ-measure. Given $\alpha, \beta \in \mathcal{P}$, we call

$$\alpha \vee \beta \,=\, \{A \cap B : A \in \alpha, B \in \beta\}$$

the *common refinement* of α and β. More generally, we say that β is a *refinement* of α, denoted by $\alpha \leqslant \beta$ if, for every $B \in \beta$, there is some $A \in \alpha$ such that $B \subset A$. We say that α and β are independent if $\mu(A \cap B) = \mu(A)\mu(B)$ for all $A \in \alpha$ and $B \in \beta$. The definition of entropy is chosen in such a way that $H_\mu(\alpha \vee \beta) = H_\mu(\alpha) + H_\mu(\beta)$, whenever α and β are independent.

In several applications, we are interested in the *additional* gain of information that we obtain from a partition $\alpha \in \mathcal{P}$, provided that the chosen element of a second partition $\beta \in \mathcal{P}$ is already known. Given $B \in \beta$, we define the conditional measure μ_B via

$$\mu_B(A) \,=\, \mu(A|B) \,=\, \frac{\mu(A \cap B)}{\mu(B)},$$

for all measurable $A \subset X$. The *conditional information function* is defined as the function $I_{\alpha|\beta}$, satisfying $I_{\alpha|\beta}(x) = -\log(\mu_B(A))$, whenever $x \in A \cap B$, for $A \in \alpha$ and $B \in \beta$. Likewise, the *conditional entropy* is defined as

$$H_\mu(\alpha|\beta) \,=\, \int_X I_{\alpha|\beta} \,\mathrm{d}\mu \,=\, \sum_{B\in\beta} \mu(B) H_{\mu_B}(\alpha).$$

Let us state a few properties of (conditional) entropy that will become useful in Chapter 4. The proofs are textbook material, compare for example [Kel98, Pet83].

Fact 2.3.2. *The (conditional) entropy satisfies the following properties.*

(1) *$H_\mu(\alpha) \leqslant H_\mu(\alpha \vee \beta)$ for all $\alpha, \beta \in \mathcal{P}$, with equality if and only if $\beta \leqslant \alpha$ up to null sets.*

(2) *$H_\mu(\alpha \vee \beta) = H_\mu(\beta) + H_\mu(\alpha|\beta)$ for all $\alpha, \beta \in \mathcal{P}$.*

(3) *$H_\mu(\alpha|\beta) \leqslant H_\mu(\alpha)$ for all $\alpha, \beta \in \mathcal{P}$, with equality if and only if α and β are independent.*

■

Remark 2.3.3. We can also associate a partition α to a random variable V on (X, μ) that can take values in a finite set S. More precisely, we set $\alpha(V) = \{V^{-1}(s) : s \in S\}$. In this case, we define $H_\mu(V) = H_\mu(\alpha(V))$ with slight abuse of notation. Given two such random variables V and W on (X, μ), we similarly define $H_\mu(V|W) = H_\mu(\alpha(V)|\alpha(W))$. This point of view will become relevant in Section 4.2. ◊

Let us introduce some reference to the *dynamics* to our discussion of entropy. Given $\alpha \in \mathcal{P}$, we collect the information that we obtain from observing the succession of sets in α that are visited by the orbit of some unknown point $x \in X$. For $n \in \mathbb{N}$, the condition $T^n x \in A$ can alternatively be formulated as $x \in T^{-n}A$. That is, observing the orbit at time n corresponds to choosing one of the sets in $T^{-n}\alpha$. Hence, the accumulated (average) information that we have obtained after n steps is given by the entropy of the partition

$$\alpha_n := \alpha \vee T^{-1}\alpha \vee \cdots \vee T^{-(n-1)}\alpha.$$

In general, $H_\mu(\alpha_n)$ will diverge as $n \to \infty$. However, the entropy *per step* is often a useful characteristic. This is well-defined due to a result by Fekete [Fek23] because $(H_\mu(\alpha_n))_{n\in\mathbb{N}}$ is a subadditive sequence by virtue of Fact 2.3.2.

Definition 2.3.4. Given a measure-preserving dynamical system (X, T, μ) and a partition $\alpha \in \mathcal{P}$, we define

$$h_\mu(T, \alpha) = \lim_{n\to\infty} \frac{1}{n} H_\mu(\alpha_n).$$

The (measure-theoretic) *entropy* of μ on (X, T) is given by

$$h_\mu(T) = \sup_{\alpha\in\mathcal{P}} h_\mu(T, \alpha).$$

Measure-theoretic entropy also appears under the terms *metric entropy* and *Kolmogorov–Sinai entropy* in the literature. Let $\mathbb{I} = \mathbb{Z}$ if T is invertible and $\mathbb{I} = \mathbb{N}_0$ otherwise. We say that α is a *μ-generator* with respect to T, if the smallest σ-algebra containing all of the sets in $T^{-n}\alpha$, for all $n \in \mathbb{I}$, is the Borel σ-algebra on X, up to μ-null sets. The following result goes back to Kolmogorov [Kol59] and Sinai [Sina59], in this form it is taken from [Kel98, Thm. 3.2.18].

Theorem 2.3.5 (Kolmogorov–Sinai)**.** *Let α be a μ-generator with respect to T. Then, $h_\mu(T) = h_\mu(T, \alpha)$.* ■

Example 2.3.6. Let $(\mathbb{X}, S)$ be a subshift over some alphabet $\mathcal{A}$. We note that the partition $\alpha = \{[a] : a \in \mathcal{A}\}$, together with the shift action S, generates the collection of cylinder sets and hence the complete Borel σ-algebra. That is, the entropy of $\mu \in \mathcal{M}^1_S(\mathbb{X})$ is given by

$$h_\mu(S) = -\lim_{n\to\infty} \frac{1}{n} \sum_{v\in\mathcal{L}_n(\mathbb{X})} \mu([v]) \log(\mu([v])).$$

A trivial upper bound for $h_\mu(S)$ is given by $\log(\#\mathcal{A})$. ◊

If the measure μ is not only T-invariant but even ergodic, it is possible to obtain the entropy $h_\mu(T)$ from individual points almost surely. This became known as the Shannon–MacMillan–Breiman theorem. We formulate it here in the special case that (X, T) is a subshift which is most relevant for us; compare [Kel98, Thm. 3.2.7] for the more general result.

Theorem 2.3.7 (Shannon–MacMillan–Breiman)**.** *Let $(\mathbb{X}, S)$ be a subshift, equipped with an ergodic measure $\mu \in \mathcal{M}_S^1(\mathbb{X})$. Then,*

$$h_\mu(S) \;=\; \lim_{n\to\infty} -\frac{1}{n}\log(\mu([x_1 \cdots x_n])),$$

holds for μ-almost every $x \in \mathbb{X}$. ∎

2.3.2. Topological entropy

Topological entropy was introduced by Adler, Konheim and McAndrew as an analogue of measure-theoretic entropy [AKM65]. Indeed, the construction closely follows the ideas presented in the last section; we refer to [Wal82, Ch. 7] for background. Instead of a measurable partition, we start from a finite open cover $\mathcal{U}$ of the compact space X. We define $N(\mathcal{U})$ as the cardinality of the smallest open subcover of $\mathcal{U}$. The *entropy* of $\mathcal{U}$ is given by $H(\mathcal{U}) = \log N(\mathcal{U})$. The common refinement of two finite open covers $\mathcal{U}, \mathcal{V}$ is given by

$$\mathcal{U} \vee \mathcal{V} \;=\; \{A \cap B : A \in \mathcal{U}, B \in \mathcal{V}, A \cap B \neq \varnothing\}.$$

Since T is continuous, each of the collections $T^{-n}(\mathcal{U})$ is a finite open cover of X, and the same holds for

$$\mathcal{U}_n \;=\; \mathcal{U} \vee T^{-1}\mathcal{U} \vee \cdots \vee T^{-(n-1)}\mathcal{U},$$

for all $n \in \mathbb{N}$. Again, $H(\mathcal{U}_n)_{n\in\mathbb{N}}$ is a subadditive sequence and hence, the limit in the following definition is well-defined.

Definition 2.3.8. Given a dynamical system (X, T) and a finite open cover $\mathcal{U}$ of X, we define

$$h_{\mathrm{top}}(T,\mathcal{U}) \;=\; \lim_{n\to\infty} \frac{1}{n} H(\mathcal{U}_n).$$

The *topological entropy* on (X, T) is given by $h_{\mathrm{top}}(T) \;=\; \sup_{\mathcal{U}} h_{\mathrm{top}}(T,\mathcal{U})$.

The concept of topological entropy can be extended to non-compact metric spaces X using the concept of (n, ε)-separated sets (or alternatively (n, ε)-spanning sets), going back to ideas of Bowen [Bow73]. For compact metric spaces, all three definitions coincide. A relation between topological entropy and measure-theoretic entropy was established by Dinaburg [Din70], Goodman [Goo71] and Goodwyn [Goo72]. This takes the form of a *variational principle* [Wal82, Thm. 8.6].

Theorem 2.3.9. *Let (X, T) be a dynamical system on a compact metric space X. Then,* $h_{\mathrm{top}}(T) = \sup\{h_\mu(T) : \mu \in \mathcal{M}_T^1(X)\}$. ∎

We call μ a *measure of maximal entropy* if $h_{\mathrm{top}}(T) = h_\mu(T)$. In Chapter 6, we will encounter a related, but slightly more general variational principle.

Example 2.3.10. Let $(\mathbb{X}_A, S)$ be a Markov subshift with irreducible transition matrix A. Let λ be the PF eigenvalue of A and L, R the left and right PF eigenvectors, respectively. The *Parry measure* μ on $(\mathbb{X}_A, S)$ is defined via

$$\mu([x_1 \cdots x_n]) \;=\; \frac{L_{x_1} R_{x_n}}{\lambda^{n-1}}$$

for all $x_1 \cdots x_n \in \mathcal{L}_n(\mathbb{X}_A)$ and $n \geqslant 2$. It can be shown that μ is the unique measure of maximal entropy on $(\mathbb{X}_A, S)$ [Par64], see also [Mañ87]. ◊

For a subshift $(\mathbb{X}, S)$, the topological entropy can in fact be calculated as a combinatorial quantity. Let us define the *complexity function* corresponding to $\mathbb{X}$ by

$$c\colon \mathbb{N} \to \mathbb{N}, \quad c(n) = \#\mathcal{L}_n(\mathbb{X}).$$

Then, the topological entropy of $(\mathbb{X}, S)$ is given by

$$h_{\text{top}}(S) = \lim_{n\to\infty} \frac{\log(c(n))}{n};$$

compare [Pet83, Ex. 3.6]. If the topological entropy of $(\mathbb{X}, S)$ is equal to 0, it is customary to study the growth behaviour of the complexity function instead.

2.4. Spectral properties

2.4.1. Spectral theory for normal operators

In this section, we give a brief exposition of the spectral theorem; see for example [Con90] for a more thorough introduction to the topic. Let N be a normal, bounded linear operator on a separable Hilbert space $\mathcal{H}$. The *spectrum* of N is defined as

$$\sigma(N) = \{z \in \mathrm{C} : (N - z) \text{ is not invertible}\}.$$

The spectral theorem asserts that N can be represented by a multiplication operator on an appropriate function space; compare [Pet83, Ch. 1.4], [Con90, Thm. 10.1] and [EFHN15, Thm. 18.2].

Theorem 2.4.1 (Spectral theorem)**.** *There is a sequence of finite Borel measures $(\nu_n)_{n\in I}$ on $\mathbb{C}$, with I finite or countable, such that $\nu_1 \gg \nu_2 \gg \cdots$, and such that N is unitarily equivalent to the operator $M = \oplus_{n\in I} M_n$ on $\oplus_{n\in I} L^2(\mathbb{C}, \nu_n)$, where*

$$M_n\colon L^2(\mathbb{C}, \nu_n) \to L^2(\mathbb{C}, \nu_n), \quad (M_n f)(z) = zf(z),$$

for all $n \in I$. Each of the measures ν_n, with $n \in I$, is unique up to equivalence. ■

The spectrum of each of the operators M_n is given by the topological support of ν_n [Con90, Ex. 2.5]. Since the measures ν_n are nested by absolute continuity (and since unitary equivalence preserves the spectrum), we observe

$$\sigma(N) = \operatorname{supp}(\nu_1).$$

We call ν_1 a *measure of maximal spectral type*. Identifying the complex plane $\mathbb{C}$ with $\mathbb{R}^2$, we can apply the Lebesgue decomposition to ν_1, yielding

$$\nu_1 = \nu_{1,\text{pp}} + \nu_{1,\text{sc}} + \nu_{1,\text{ac}}.$$

This provides refined spectral information about N. For example, a number $\lambda \in \mathbb{C}$ is an eigenvalue of N if and only if $\nu_{1,\mathrm{pp}}(\{\lambda\}) > 0$. This motivates us to define, for $\bullet \in \{\mathrm{pp}, \mathrm{sc}, \mathrm{ac}\}$,

$$\sigma_\bullet(N) \;=\; \operatorname{supp}(\nu_{1,\bullet}).$$

Note that $\sigma(N) = \sigma_{\mathrm{pp}}(N) \cup \sigma_{\mathrm{sc}}(N) \cup \sigma_{\mathrm{ac}}(N)$, but in general this union is not disjoint.

In some cases, we are interested in the iterates of N on a single element $h \in \mathcal{H}$. Assume that the unitary equivalence between N and M relates h to $(f_n)_{n\in I}$. Then, for all $n \in \mathbb{N}$, we obtain

$$\langle h, N^n h\rangle_{\mathcal{H}} \;=\; \sum_{m\in I} \int_{\mathbb{C}} z^n |f_m(z)|^2 \,\mathrm{d}\nu_m(z) \;=\; \int_{\mathbb{C}} z^n \,\mathrm{d}\nu_h(z),$$

where $\nu_h = \sum_{m\in I} |f_m|^2 \nu_m$. We call ν_h the *spectral measure* of h, and note $\nu_h \ll \nu_1$.

The Schrödinger operators that we shall consider in Chapter 5 are bounded, *self-adjoint* operators on a separable Hilbert space and are hence covered by the results in this section. In this case, the spectrum is a subset of the real line. For a discussion of spectral theory with a view towards Schrödinger operators, see [DFa, Ch. 1].

2.4.2. The dynamical spectrum

Of particular importance in the study of a measure-preserving dynamical system (X, T, μ) is the *Koopman operator*

$$U_T \colon L^2(X,\mu) \to L^2(X,\mu), \quad f \mapsto f \circ T.$$

Let us further assume that X is a compact metric space. In this case, the Hilbert space $L^2(X,\mu)$ is separable, as follows from [Bil95, Thm. 19.2]. The operator U_T is unitary, hence both normal and bounded, and the results of the preceding section are applicable. We call $\sigma(U_T)$ the *dynamical spectrum* of (X, T). This set is contained in the complex unit circle $\mathbb{S}^1$ because U_T is an isometry. Given a measure of maximal spectral type ν of U_T, we relate attributes of the equivalence class of ν to the dynamical spectrum. For example, we say that the dynamical spectrum is pure point if $\nu = \nu_{\mathrm{pp}}$.

The measure-theoretic mixing properties of the system (X, T, μ) discussed earlier can in fact be rephrased as spectral properties of the Koopman operator; compare [Que10, Prop. 3.1] and [Que10, Prop. 3.10] for the proof of the following result. Recall that we assume X to be a compact topological space and $\mu \in \mathcal{M}^1_T(X)$ a T-invariant probability measure on X.

Fact 2.4.2. *The measure-preserving dynamical system (X, T, μ) is ergodic if and only if $\lambda = 1$ is a simple eigenvalue of U_T, that is, the constant functions in $L^2(X,\mu)$ are the only functions with $f = f \circ T$. The system (X, T, μ) is weakly mixing if and only if $\lambda = 1$ is a simple eigenvalue of U_T and U_T has no further eigenvalues.* ∎

It is also possible to rephrase strong mixing in terms of properties of the measure of maximal spectral type ν [Que10, Prop. 3.9]. Further, it is worth pointing out that, due to a theorem by von Neumann, every *ergodic* and invertible dynamical system (X, T, μ) with pure point dynamical spectrum is isomorphic to a rotation on some compact abelian group, equipped

with the Haar measure [PF02, Thm. 1.5.4]. For this class of systems, the dynamical spectrum is a complete invariant. As a rough guiding principle, the more ordered a system (X, T, μ), the more singular is the dynamical spectrum. We refer to [Que10, Ch. 3] for a more detailed discussion.

2.4.3. Diffraction as a spectral measure

If the dynamical system is a subshift $(\mathbb{X}, S, \mu)$, a useful family of spectral characteristics are *diffraction measures*. These can be constructed as particular spectral measures. We start by choosing an arbitrary weight vector $\tau \in \mathbb{C}^{\mathcal{A}}$. This induces a function $\pi_\tau \in L^2(\mathbb{X}, \mu)$ via $\pi_\tau(x) = \tau_{x_0}$, for all $x \in \mathbb{X}$. The corresponding spectral measure $\nu_\tau = \nu_{\pi_\tau}$ satisfies by definition

$$\gamma(n) := \langle \pi_\tau, U_S^n \pi_\tau \rangle_{L^2(\mathbb{X},\mu)} = \int_{\mathbb{S}^1} z^n \, \mathrm{d}\nu_\tau(z), \tag{2.3}$$

for all $n \in \mathbb{Z}$. We call $\gamma \colon \mathbb{Z} \to \mathbb{C}$ the *autocorrelation sequence* of $(\mathbb{X}, S, \mu)$. We remark that there is a *unique* finite Borel measure ν_τ satisfying (2.3) for all $n \in \mathbb{Z}$; see [BG13, Prop. 8.7] and [PF02, Thm. 1.5.7]. Hence, the spectral measure ν_τ is in fact independent of the choice of the measures $\nu_1 \gg \nu_2 \gg \cdots$ from the spectral theorem. Identifying $\mathbb{S}^1$ with the torus $\mathbb{T}$ via the map $t \mapsto \mathrm{e}^{2\pi \mathrm{i} t}$, we may regard ν_τ as a measure on $\mathbb{T}$, yielding

$$\gamma(n) = \int_{\mathbb{T}} \mathrm{e}^{2\pi \mathrm{i} n t} \, \mathrm{d}\nu_\tau(t).$$

That is, the sequence $(\gamma(n))_{n\in\mathbb{Z}}$ arises from the (inverse) Fourier transform of the finite measure ν_τ on $\mathbb{T}$. We call the measure ν_τ the *diffraction* of $(\mathbb{X}, S, \mu)$ with respect to τ. Since the diffraction ν_τ is absolutely continuous with respect to the measure of maximal spectral type ν, every component that is present in the Lebesgue decomposition of ν_τ is also present in the Lebesgue decomposition of ν. Often, it is even possible to obtain the measure of maximal spectral type from (generalized) diffraction measures [BLvE15].

Let us further assume that μ is ergodic on $(\mathbb{X}, S)$. In this case, we obtain by Birkhoff's ergodic theorem

$$\gamma(n) = \int_{\mathbb{X}} \overline{\pi_\tau(y)} \pi_\tau(S^n y) \, \mathrm{d}\mu(y) = \lim_{m\to\infty} \frac{1}{m} \sum_{j=0}^{m-1} \overline{\tau}_{x_j} \tau_{x_{j+n}}, \tag{2.4}$$

for μ-almost every $x \in \mathbb{X}$. This gives a justification to call γ the autocorrelation sequence.

2.4.4. Diffraction for point sets and measures

In line with [BG13, Ch. 9], we discuss a slightly different notion of diffraction, related to point sets and associated measures. Thinking of a sequence $x \in \mathcal{A}^{\mathbb{Z}}$ as a way to store a succession of symbols, it is natural to define the *symbolic point sets*

$$\Lambda_a^s(x) = \{n \in \mathbb{Z} : x_n = a\}, \tag{2.5}$$

for all $a \in \mathcal{A}$. By construction, $\mathbb{Z} = \bigcup_{a\in\mathcal{A}} \Lambda_a^s(x)$ holds for all $x \in \mathcal{A}^{\mathbb{Z}}$. This corresponds to placing the symbols in x on the real line with equal spacing. More generally, we choose a

length vector $L \in \mathbb{R}_+^{\mathcal{A}}$ and leave a gap of length L_a after a letter a before placing the next symbol. Thus, we obtain a sequence of positions $(\ell_n(x))_{n\in\mathbb{Z}}$, defined recursively via $\ell_0(x) = 0$ and $\ell_{n+1}(x) = \ell_n(x) + L_{x_n}$ for all $n \in \mathbb{Z}$. In this case, we obtain

$$\Lambda_a(x) \,=\, \{\ell_n(x) : x_n = a\},$$

which implicitly depends on L. We recover the symbolic point sets in (2.5) if we set $L_a = 1$ for all $a \in \mathcal{A}$.

Regarding the letters in $\mathcal{A}$ as place holders for different types of atoms, we assign distinct *scattering amplitudes* according to some $\tau \in \mathbb{C}^{\mathcal{A}}$. The atomic structure is then modelled by

$$\omega(x) \,=\, \sum_{a\in\mathcal{A}} \tau_a \delta_{\Lambda_a(x)}, \quad \delta_{\Lambda_a(x)} \,=\, \sum_{r\in\Lambda_a(x)} \delta_r.$$

The weighted Dirac comb $\omega(x)$ naturally defines a translation bounded (complex) Radon measure on $\mathbb{R}$. If $x \in \mathcal{A}^{\mathbb{N}}$ is a one-sided sequence, the above quantities are defined analogously, and $\omega(x)$ is a Radon measure on $\mathbb{R}_+$. In the following, we fix an *averaging sequence* $(B_n)_{n\in\mathbb{N}}$ by $B_n = [-n, n]$ if $x \in \mathcal{A}^{\mathbb{Z}}$ and $B_n = [0, n]$ if $x \in \mathcal{A}^{\mathbb{N}}$. Further, we denote by $\omega_n = \omega|_{B_n}$ the restriction of the Radon measure ω to B_n.

Definition 2.4.3. Given $\omega = \omega(x)$, an *autocorrelation measure* of x (with respect to the averaging sequence $(B_n)_{n\in\mathbb{N}}$) is a vague accumulation point $\gamma(x)$ of the sequence $(\gamma_n)_{n\in\mathbb{N}}$, where

$$\gamma_n \,=\, \frac{\omega_n * \widetilde{\omega_n}}{|B_n|}.$$

Since all the measures $\gamma_n(x)$ are positive-definite, so is $\gamma(x)$. Hence, by the theorem of Bochner–Schwartz [ReS80, Thm. IX.10], the Radon measure $\gamma(x)$ has a well-defined Fourier transform $\widehat{\gamma}(x) = \widehat{\gamma(x)}$, as a positive tempered measure; see Proposition 1.2.7. We call $\widehat{\gamma}(x)$ a *diffraction* of x.

Remark 2.4.4. Assume that μ is an ergodic probability measure on a subshift $(\mathbb{X}, S)$. Then, there exists an almost sure autocorrelation γ, that is, for μ-almost every $x \in \mathbb{X}$ and $\omega = \omega(x)$, the limit

$$\gamma(x) \,=\, \lim_{n\to\infty} \frac{\omega_n * \widetilde{\omega_n}}{|B_n|}$$

exists and equals γ. This follows by Birkhoff's ergodic theorem; we provide a proof of a slightly more general result in Lemma 4.3.1. Hence, there is also an almost sure diffraction measure $\widehat{\gamma}$. In the *symbolic setting*, that is, if $L_a = 1$ for all $a \in \mathcal{A}$, a straightforward calculation yields

$$\gamma \,=\, \sum_{n\in\mathbb{Z}} \gamma(n)\delta_n,$$

with $\gamma(n)$ as in (2.4); compare [BG13, Ex. 9.1]. Interpreting ν_τ as a measure on $[0, 1)$, we obtain

$$\widehat{\gamma} \,=\, \delta_{\mathbb{Z}} * \nu_\tau,$$

which establishes a link between the two notions of diffraction measures. This correspondence can be pushed further to more general choices of L by considering the dynamical system that arises from a *suspension* of $(\mathbb{X}, S, \mu)$; compare [BLvE15] for the details of this relation, and [CFS82, Ch. 11] for background on suspension flows. $\Diamond$

2.4.5. D-function of minimal systems

The D-function of a minimal dynamical system (X, T) was advertised as a new topological invariant of dynamical systems in [Ye92]. It was also implicitly used in the work of Dekking [Dek78] and Kamae [Kam72]. As we discuss in this section, the D-function is just a different perspective on some part of the dynamical spectrum.

Throughout this section, assume that (X, T) is a minimal and invertible dynamical system on a compact metric space X. We call $\lambda \in \mathbb{S}^1$ a *topological eigenvalue* of (X, T) if there exists a *continuous* function $f \in \mathcal{C}(X)$ such that

$$f(Tx) \,=\, \lambda f(x),$$

for all $x \in X$. We say that $\lambda = \mathrm{e}^{2\pi \mathrm{i} q}$ with $q \in [0, 1)$ is *rational* if $q \in \mathbb{Q}$. Note that λ is an eigenvalue of U_T, for every invariant measure μ on (X, T) with full topological support. Conversely, a *rational* eigenvalue of $U_T \colon L^2(X, \mu) \to L^2(X, \mu)$ is indeed a topological eigenvalue if (X, T) is a Boshernitzan subshift. The proof of this statement will appear in [CDFG].

We are interested in the mixing behaviour of powers of T on X. Unfortunately, (X, T^n) need not be minimal for general $n \in \mathbb{N}$. But it decays into minimal components in a satisfactory manner, as was shown in [Ye92, Thm. 3.1]; see also [GH55, Thm. 2.24].

Fact 2.4.5. *For every $n \in \mathbb{N}$, there exists a unique number $s(n) \in \mathbb{N}$ that divides n and a finite partition $\{X_0, \cdots, X_{s(n)-1}\}$ of X such that each X_j is T^n-minimal, and we have $TX_j = X_{j+1 \mod s(n)}$, for all $1 \leqslant j \leqslant s(n)$.* ∎

We call the assignment $n \mapsto s(n)$ the *D-function* of (X, T). Using some basic properties of the D-function stated in [Ye92], the following relation to topological eigenvalues is straightforward to show; we provide a proof in [CDFG].

Fact 2.4.6. *For every n, the number $s(n)$ is the largest divisor k of n with the property that $\mathrm{e}^{2\pi \mathrm{i}/k}$ is a topological eigenvalue of (X, T).* ∎

In particular, it follows from Fact 2.4.2 that (X, T^n) is minimal for all $n \in \mathbb{N}$ if (X, T) admits a weakly mixing probability measure of full support. In fact, it suffices to show the absence of rational topological eigenvalues other than 1 to reach the same conclusion.

3. Substitutions and their generalizations

It is the aim of this chapter to introduce substitutions as a device to produce self-similar and non-periodic sequences, using just a finite number of symbols. A stochastic generalization of the substitution concept permits one to produce sequences with positive entropy, while keeping the long-range correlations of substitution sequences that are responsible for the sharp peaks in the diffraction image.

3.1. Substitutions

The material in the first part of this section is standard and we largely follow [BG13] in our notation. As we wish to set the stage for the introduction of random substitutions, we focus on concepts that readily generalize to the random setting. We will make extensive use of the symbolic notation introduced in Section 2.1.2.

Definition 3.1.1. Let $\mathcal{A}$ be a finite alphabet. A *substitution* on $\mathcal{A}$ is a map $\varrho\colon \mathcal{A} \to \mathcal{A}^+$. It extends to an endomorphism $\mathcal{A}^+ \to \mathcal{A}^+$. That is, for $u = u_1 \cdots u_n \in \mathcal{A}^n$ we define

$$\varrho(u_1 \cdots u_n) \;=\; \varrho(u_1) \cdots \varrho(u_n),$$

to be understood as a concatenation of words.

We call a word of the form $\varrho^n(a)$ for some $n \in \mathbb{N}$ and $a \in \mathcal{A}$ a (level-n) *inflation word*. A substitution is said to be of *constant length* $\ell \in \mathbb{N}$ if $|\varrho(a)| = \ell$ for all $a \in \mathcal{A}$.

Example 3.1.2. The *period doubling* substitution $\varrho\colon a \mapsto ab, b \mapsto aa$ is a substitution of constant length 2 on the binary alphabet $\mathcal{A} = \{a, b\}$. The substitution can be iterated on a letter as follows:

$$a \;\mapsto\; ab \;\mapsto\; \varrho(ab) = \varrho(a)\varrho(b) = abaa \;\mapsto\; abaaabab \;\mapsto\; \cdots .$$

It is remarkable that $\varrho^n(a)$ is a prefix of $\varrho^m(a)$ for all $m \geqslant n$. We will come back to this observation later. ◊

The self-similar structure associated to a substitution is best understood in an infinite setting. On a sequence $x = (x_n)_{n\in\mathbb{Z}}$, we define the action of a substitution ϱ as

$$\varrho(x) \;=\; \cdots \varrho(x_{-2})\varrho(x_{-1}).\varrho(x_0)\varrho(x_1)\varrho(x_2) \cdots ,$$

thus extending ϱ to a (continuous) map $\mathcal{A}^{\mathbb{Z}} \to \mathcal{A}^{\mathbb{Z}}$. We are interested in structures that can be generated by repeated application of the substitution ϱ. For finite words, this leads to the

concept of admissible and legal words and gives rise to the concept of a substitution subshift in the setting of bi-infinite sequences. We call a word $u \in \mathcal{A}^+$ *admissible* (under ϱ) if there exists a letter $a \in \mathcal{A}$ and a power $n \in \mathbb{N}$ such that $u \triangleleft \varrho^n(a)$. It is straightforward to verify that the *substitution subshift*

$$\mathbb{X}_\varrho = \{x \in \mathcal{A}^{\mathbb{Z}} : x_{[j,k]} \text{ is admissible for all } j, k \in \mathbb{Z} \text{ with } j \leqslant k\}$$

is indeed closed and S-invariant, as well as ϱ-invariant (provided that it is non-empty). In line with Definition 2.1.1, the language of a substitution is given by

$$\mathcal{L}(\mathbb{X}_\varrho) = \{x_{[j,k]} : x \in \mathbb{X}_\varrho \text{ and } j, k \in \mathbb{Z}, \text{ with } j \leqslant k\}.$$

We often write $\mathcal{L}$ for $\mathcal{L}(\mathbb{X}_\varrho)$ if the substitution is understood, and $\mathcal{L}_n = \mathcal{L} \cap \mathcal{A}^n$. A word $u \in \mathcal{A}^+$ is called *legal* if $u \in \mathcal{L}$. We note that every legal word is admissible, but not necessarily vice versa. For a more detailed account of this distinction, we refer to [MR18]. The following result shows that, for most substitutions ϱ, the substitution subshift and the language are in fact meaningful objects.

Lemma 3.1.3. *Both $\mathbb{X}_\varrho$ and $\mathcal{L}(\mathbb{X}_\varrho)$ are non-empty precisely if there exists a letter $a \in \mathcal{A}$ such that $|\varrho^n(a)| \to \infty$ as $n \to \infty$.*

Proof. If $|\varrho^n(a)|$ remains bounded for all $n \in \mathbb{N}$ and $a \in \mathcal{A}$, there is a maximal length of admissible words, implying that $\mathbb{X}_\varrho = \varnothing$. Conversely, assume that $\lim_{n\to\infty} |\vartheta^n(a)| = \infty$. Let $x^{(n)}$ be the sequence

$$x^{(n)} = \cdots \varrho^n(a)\varrho^n(a).\varrho^n(a)\varrho^n(a)\cdots$$

and set $y^{(n)} = S^{\lfloor |\varrho^n(a)|/2 \rfloor} x^{(n)}$, which amounts to shifting the origin approximately to the middle of the word $\varrho^n(a)$. Since $\mathcal{A}^{\mathbb{Z}}$ is compact, the sequence $(y^{(n)})_{n\in\mathbb{N}}$ has an accumulation point $y \in \mathcal{A}^{\mathbb{Z}}$. By construction, all subwords of y appear in $\varrho^n(a)$ for some $n \in \mathbb{N}$ and hence $y \in \mathbb{X}_\varrho$. □

A similar compactness argument shows that every point $x \in \mathbb{X}_\varrho$ can be de-substituted in the sense that there are $y \in \mathbb{X}_\varrho$ and $0 \leqslant j \leqslant |\varrho(y_0)|$ such that $x = S^j(\varrho(y))$, see for example [EG21, Lemma 2.2]. If the choice for y and j are unique for every $x \in \mathbb{X}_\varrho$, the substitution ϱ is called *recognizable*. It was shown only recently (in this generality) that a substitution ϱ is recognizable precisely if $(\mathbb{X}_\varrho, S)$ is not periodic [BSTY19, Thm. 5.3]. A classic analogue of this result was shown by Mossé under the additional assumption that the substitution ϱ is *primitive* [Mos92]. Primitivity is in fact a central property that is satisfied by many of the classic examples [BG13]. We devote the next section to the definition of primitivity and the exploration of some of its consequences.

3.1.1. Primitive substitutions

The structure of the substitution subshift $(\mathbb{X}_\varrho, S)$ becomes somewhat simpler if we can guarantee that all letters indeed occur in all level-n inflation words for large enough n. Since this property depends only on the Abelianisation of the inflation words, it is convenient to

introduce a device that determines how the Abelianisation of a word u changes as we let ϱ act on u. The matrix $M \in \mathrm{Mat}(\mathcal{A}, \mathbb{Z})$ with entries $M_{ab} = |\varrho(b)|_a$ for $a, b \in \mathcal{A}$, is called the *substitution matrix* associated to ϱ. Indeed, we have

$$|\Phi(\varrho(u))\rangle \;=\; \Bigl(\sum_{b\in\mathcal{A}} |u|_b |\varrho(b)|_a\Bigr)_{a\in\mathcal{A}} \;=\; M|\Phi(u)\rangle,$$

for every $u \in \mathcal{A}^+$.

Definition 3.1.4. A substitution ϱ is called *primitive* precisely if its substitution matrix M is primitive.

Hence, ϱ is primitive iff there exists a power $m \in \mathbb{N}$ such that all entries of M^m are positive. This amounts to requiring that $\varrho^m(a)$ contains every letter of the alphabet for all $a \in \mathcal{A}$. In particular, we have $|\varrho^m(a)| \to \infty$ as $m \to \infty$ for all $a \in \mathcal{A}$ and hence the subshift $\mathbb{X}_\varrho$ is non-empty. Given a primitive substitution matrix M, we write λ for its PF eigenvalue and L, R for the corresponding left and right PF eigenvectors, normalized as $\langle L|R\rangle = \|R\|_1 = 1$; compare Theorem 1.3.3. Primitivity is a central property of substitutions and will be assumed for the bulk of this thesis. In Chapter 5, however, we will be concerned with a family of non-primitive substitutions.

Example 3.1.5. Let $\varrho_F \colon a \mapsto ab, b \mapsto a$, which constitutes the famous *Fibonacci substitution*. The name reflects the property that the length of inflation words $(|\varrho_F^n(b)|)_{n\in\mathbb{N}}$ is precisely the Fibonacci sequence. This is derived in a straightforward manner from the observation that the substitution matrix takes the form

$$M \;=\; \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix}.$$

Since $M^2 > 0$, the Fibonacci substitution is primitive. Indeed, both a and b appear in every level-2 inflation word. The eigenvalues of M are the golden ratio $\tau = (1+\sqrt{5})/2$ and its algebraic conjugate, $\tau' = -1/\tau$. ◊

Example 3.1.6. The substitution $\varrho \colon a \mapsto a, b \mapsto bba$ is a non-primitive substitution. As we iterate ϱ, we obtain

$$b \mapsto bba \mapsto bbabbaa \mapsto bbabbaabbabbaaa$$

and observe that the word a^n occurs at the end of $\varrho^n(b)$ for all $n \in \mathbb{N}$. This is easily proved by induction and implies that there can be arbitrarily large gaps between two occurrences of b in legal words. ◊

Remark 3.1.7. If the substitution ϱ is primitive, there is an alternative route to the definition of the subshift $\mathbb{X}_\varrho$. We first construct an explicit element of $\mathbb{X}_\varrho$ that is invariant under (some power of) ϱ and then obtain the whole subshift as an orbit closure under the shift action. This is the approach adopted in [BG13, Ch. 4]. Take an admitted word u_1u_2 of length 2 and iterate it under the substitution ϱ. By the pigeon hole principle, there are letters $a, b \in \mathcal{A}$ and powers $m < n$ such that b is the last letter of the words $\varrho^m(u_1)$ and $\varrho^n(u_1)$ and a is the

first letter of both $\varrho^m(u_2)$ and $\varrho^n(u_2)$. Hence, both $\varrho^m(u_1).\varrho^m(u_2)$ and $\varrho^n(u_1).\varrho^n(u_2)$ have the word $a.b$ around the origin. With $r = n - m$, we get that b is the last letter of $\varrho^r(b)$ and a is the first letter of $\varrho^r(a)$. By induction, we obtain that $\varrho^{rj}(a)$ is a prefix of $\varrho^{rk}(a)$ for all $j, k \in \mathbb{N}$ with $j \leqslant k$. Similarly, $\varrho^{rj}(b)$ is a suffix $\varrho^{rk}(b)$ for $j \leqslant k$. The sequence

$$x^\star \,=\, \lim_{n\to\infty} \cdots \varrho^{nr}(b)\varrho^{nr}(b).\varrho^{nr}(a)\varrho^{nr}(a)\cdots$$

is therefore a well-defined element of $\mathbb{X}_\varrho$. Note that the periodic continuation of the words $\varrho^{nr}(a)$ and $\varrho^{nr}(b)$ is immaterial for the limiting sequence $x^\star$, since all but the central copies get "shifted out to infinity" in both directions. With this in mind, we also write

$$x^\star \,=\, (\varrho^r)^\infty(b).(\varrho^r)^\infty(a) \,=\, \lim_{n\to\infty} \varrho^{nr}(b).\varrho^{nr}(a),$$

with slight abuse of notation. It is straightforward to verify that $x^\star$ is invariant under ϱ^r. Every power of a primitive substitution in fact gives rise to the same subshift, hence there is no loss of generality in assuming that $r = 1$. ◊

This is an instance of the self-similarity alluded to earlier: If we replace every letter a in $x^\star$ by its level-n inflation word $\varrho^n(a)$ for an arbitrary $n \in \mathbb{N}$, we recover the same sequence. This property is at the heart of powerful renormalisation techniques [BG13, BGM19, Mañ19] and has strong consequences. For instance, it guarantees that every legal word occurs with bounded gaps and with a well-defined frequency in every sequence $x \in \mathbb{X}_\varrho$. In the language of dynamical systems, these observations take the following form; compare [BG13, Thm. 4.3].

Fact 3.1.8. *If ϱ is a primitive substitution, the dynamical system $(\mathbb{X}_\varrho, S)$ is strictly ergodic, that is, minimal and uniquely ergodic.* ■

Recall that minimality means that every point in the space has a dense orbit. In particular, we can choose the explicit point $x^\star$ and obtain

$$\mathbb{X}_\varrho \,=\, \overline{\{S^j x^\star : j \in \mathbb{Z}\}}.$$

Similarly, there is a natural way to construct the unique shift-invariant (and hence ergodic) measure as an orbit average of the Dirac measure $\delta_{x^\star}$, supported on the point $x^\star$. Indeed, the weak limit

$$\mu_\varrho \,=\, \lim_{n\to\infty} \frac{1}{n} \sum_{j=0}^{n-1} \delta_{x^\star} \circ S^{-j}.$$

exists and is shift-invariant by construction. The convergence follows because every weak accumulation point of this sequence is necessarily shift-invariant and the existence of more than one accumulation point would contradict the unique ergodicity of $(\mathbb{X}_\varrho, S)$.

By the Morse–Hedlund theorem [MH38], the subshift $(\mathbb{X}_\varrho, S)$ has a bounded complexity function precisely if it is periodic. If it is non-periodic and ϱ is primitive, its complexity function $\mathrm{c}\colon \mathbb{N} \to \mathbb{N}$ satisfies $\mathrm{c}(n) \in \Theta(n)$, where $\Theta(f(n))$ denotes the class of functions $g\colon \mathbb{N} \to \mathbb{N}$ such that $f(n)/g(n)$ is bounded from above and below by positive constants, for all $n \in \mathbb{N}$. For arbitrary substitutions, the complexity function falls into one of the classes

$\Theta(1), \Theta(n), \Theta(n \log\log n), \Theta(n \log n)$ or $\Theta(n^2)$. Compare [Pan84] for a proof of this statement in the context of one-sided sequences, as well as a neat characterization of the different cases. A presentation of this result in English may be found in [DLR13]. In any case, the entropy is trivial.

Fact 3.1.9. *For every substitution subshift $(\mathbb{X}_\varrho, S)$, the topological entropy is* 0. ∎

Finally, let us expand a bit on the interpretation of the right and left PF eigenvectors of a primitive substitution. Although the following is well-known [Que10, Ch. 5], we provide a short proof, as it illustrates how PF theory translates to properties of primitive substitutions. Furthermore, the general idea of this proof readily generalizes to random substitutions.

Lemma 3.1.10. *Let ϱ be a primitive substitution, with PF eigenvalue λ and R, L the right and left PF eigenvectors, respectively, normalized as $\langle L|R\rangle = \|R\|_1 = 1$. Then,*

$$\begin{aligned} L_a &= \lim_{n\to\infty} \frac{|\varrho^n(a)|}{\lambda^n}, \\ R_a &= \lim_{n\to\infty} \frac{|\varrho^n(b)|_a}{|\varrho^n(b)|} = \mu_\varrho([a]), \end{aligned}$$

for all $a, b \in \mathcal{A}$.

Proof. Since $\Phi(\varrho^n(u)) = M^n \Phi(u)$ for all $n \in \mathbb{N}$ and $u \in \mathcal{A}^+$, the substitution matrix of ϱ^n is given by M^n for all $n \in \mathbb{N}$. By a straightforward application of the PF theorem, we get

$$\frac{|\varrho^n(a)|}{\lambda^n} = \frac{1}{\lambda^n} \sum_{b\in\mathcal{A}} (M^n)_{ba} \xrightarrow{n\to\infty} \sum_{b\in\mathcal{A}} R_b L_a = L_a,$$

for all $a \in \mathcal{A}$. Due to the unique ergodicity of $(\mathbb{X}_\varrho, S, \mu_\varrho)$, the letter frequencies exist uniformly in sequences $x \in \mathbb{X}_\varrho$ and are given by $\mu_\varrho([a])$ for all $a \in \mathcal{A}$. In particular, they can be calculated from large inflation words as

$$\mu_\varrho([a]) = \lim_{n\to\infty} \frac{|\varrho^n(b)|_a}{|\varrho^n(b)|} = \lim_{n\to\infty} \frac{\lambda^{-n}(M^n)_{ab}}{\sum_{c\in\mathcal{A}} \lambda^{-n}(M^n)_{cb}} = \frac{R_a L_b}{\sum_{c\in\mathcal{A}} R_c L_b} = R_a,$$

for an arbitrary $b \in \mathcal{A}$. □

3.1.2. The geometric picture

By Lemma 3.1.10, R_a is the frequency of the letter a in large legal words, whereas L_a controls the relative size of large inflation words. More precisely, we have $|\varrho^n(a)| = L_a \lambda^n + o(\lambda^n)$, where $o(\lambda^n)$ denotes an error term that, divided by λ^n, converges to 0 as $n \to \infty$. We can make this relation exact, removing the error term $o(\lambda^n)$, by changing to the *geometric setting*. Here, we replace every letter $a \in \mathcal{A}$ by an interval I_a of length L_a. A word $u_1 \cdots u_n$ is replaced by a string of intervals $I_{u_1} \cdots I_{u_n}$ that are positioned next to each other. Accordingly, the *geometric length* of the word $u = u_1 \cdots u_n \in \mathcal{A}^+$ is given by

$$L(u) = L_{u_1} + \cdots + L_{u_n} = \sum_{a\in\mathcal{A}} L_a |u|_a = \langle L|\Phi(u)\rangle.$$

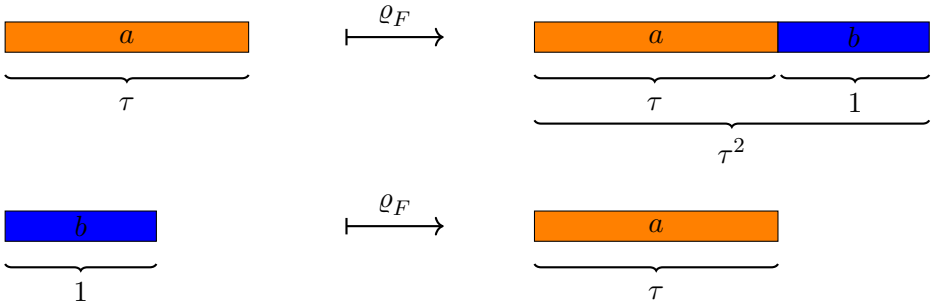

Figure 3.1.: Geometric analogue of the Fibonacci substitution $\varrho_F\colon a \mapsto ab, b \mapsto a$. All lengths have been scaled by a factor $1 + \tau^{-2}$ relative to the entries of L.

Instead of replacing a letter by a word, the substitution now acts by replacing an interval by a string of adjacent intervals, called a *patch*; see Figure 3.1 for an illustration. Since

$$L(\varrho(u)) \,=\, \langle L|\Phi(\varrho(u))\rangle \,=\, \langle L|M|\Phi(u)\rangle \,=\, \lambda L(u),$$

the geometric length of a word changes *exactly* by the factor λ under the substitution—this is a consequence of choosing the lengths of the intervals according to the entries of the left PF eigenvector L. The action of ϱ therefore amounts to an *inflation rule*: We stretch every interval by a factor λ and cut up the resulting interval into smaller intervals according to the substitution. A sequence $x \in \mathbb{X}_\varrho$ is replaced by a bi-infinite string of intervals (tiling) in the geometric picture. Placing a point at the left endpoint of each interval of type I_a we thus obtain a point set $\Lambda_a(x)$ for all $a \in \mathcal{A}$ and we set $\Lambda(x) = \cup_{a\in\mathcal{A}}\Lambda_a(x)$ for the set of all endpoints of intervals. For a more formal definition of these sets, recall Section 2.4.4. By convention, we always have $0 \in \Lambda(x)$ and $\Lambda(x)$ is a *Delone set* for all $x \in \mathbb{X}_\varrho$, that is, the distance between two adjacent points is bounded from above and from below by two universal, positive constants; compare [BG13, Ch. 2] for background. The normalization $\langle L|R\rangle = 1$ ensures that the density of the point set $\Lambda(x)$ is given by 1 for all $x \in \mathbb{X}_\varrho$.

Example 3.1.11. Recall the Fibonacci substitution $\varrho_F\colon a \mapsto ab$, $b \mapsto a$ from Example 3.1.5, with PF eigenvalue $\tau = (1+\sqrt{5})/2$. The right and left PF eigenvectors are $R = \tau^{-2}(\tau, 1)$ and $L = (1 + \tau^{-2})^{-1}\,(\tau, 1)$, respectively. There are precisely two fixed points under ϱ_F^2 in $\mathbb{X}_{\varrho_F}$, one of them is given by

$$x^\star \,=\, \cdots ababa.abaab \cdots .$$

In the geometric setting, this gives rise to a self-similar tiling, compare Figure 3.2. ◊

This procedure can be formalized further and extended to work in higher dimensions, we refer to [BG13, Ch. 5] for more details. An alternative way to produce aperiodic Delone sets is the powerful notion of a *cut and project scheme*; see [BG13, Ch. 7] for a detailed account of this method, including its relation to substitutions. In one dimension, a geometric analogue of $(\mathbb{X}_\varrho, S)$ is most elegantly formalized as a *suspension*. We refer the interested reader to [CFS82, Ch. 11] for general background on suspension flows and to [Mol13] for its application in the context of substitutions. Since we will mostly work in the symbolic setting, we will be content with the more informal description given in this section.

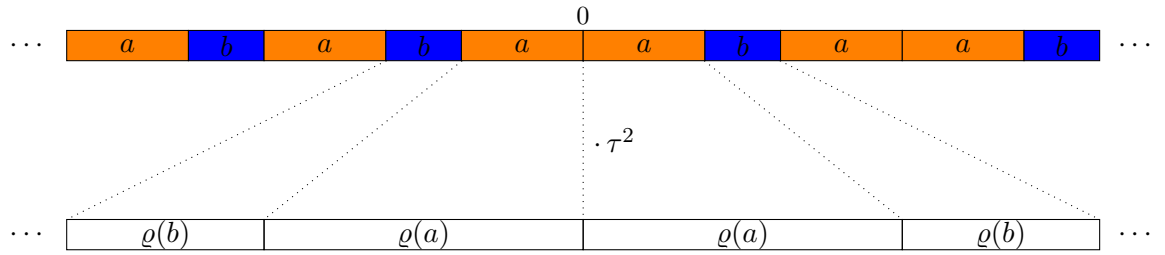

Figure 3.2.: Inflation rule, performed on a fixed point $x^\star$ of $\varrho = \varrho_F^2$, with $\lambda = \tau^2$. Using $\varrho(a) = aba$ and $\varrho(b) = ab$ we obtain the same tiling under inflation.

3.1.3. Almost minimal substitutions

Almost minimal substitutions were introduced by Yuasa in [Yua07] and later generalized to the class of *substitutions of some primitive components* in [HY11]. The main outcome of Yuasa's study is a characterization of the set of (not necessarily finite) ergodic measures on the substitution subshift for such substitutions. In this section, we highlight some of the key differences and similarities between almost minimal and primitive substitutions. For simplicity, we mostly stick to the case of a binary alphabet.

Recall the non-primitive substitution $\varrho\colon a \mapsto a, b \mapsto bba$ from Example 3.1.6. Like for primitive substitutions, this substitution admits a fixed point, of the form

$$x^\star = \cdots aaaaaaaaaa.bbabbaabbabbaaa\cdots = \lim_{n\to\infty} \varrho^n(b).\varrho^n(b),$$

and we again find $\mathbb{X}_\varrho = \overline{\{S^j x^\star : j \in \mathbb{Z}\}}$. In this sense, ϱ is close to a primitive substitution. However, the point $a^{\mathbb{Z}} \in \mathbb{X}_\varrho$ clearly does not have a dense orbit and hence the subshift $(\mathbb{X}_\varrho, S)$ is not minimal. A closer inspection yields that it is *almost minimal* in the following sense.

Definition 3.1.12. A subshift $(\mathbb{X}, S)$ is *almost minimal* if it contains precisely one fixed point $y = Sy$ under the shift map and all points in $\mathbb{X} \setminus \{y\}$ have a dense S-orbit in $\mathbb{X}$. We call a substitution ϱ almost minimal if $(\mathbb{X}_\varrho, S)$ is almost minimal.

Remark 3.1.13. In general, non-primitive substitutions need not be almost minimal. We refer to [MR18] for an overview of some of the topological phenomena that can occur in the absence of primitivity. However, in the special case of a binary alphabet $\mathcal{A} = \{a, b\}$, every (non-empty) substitution subshift $(\mathbb{X}_\varrho, S)$ is either minimal or almost-minimal [EG21]. If the subshift $(\mathbb{X}_\varrho, S)$ is minimal, it is in fact conjugate to a subshift that arises from a primitive substitution [MR18]. ◊

For the remainder of this section, let us fix $\mathcal{A} = \{a, b\}$. We will not be interested in the *trivial* cases, when $\mathbb{X}_\varrho$ is either empty or finite. Up to a renaming of the letters, every non-primitive substitution on $\mathcal{A}$ is either trivial or of the form

$$\varrho\colon a \mapsto a^p,\ b \mapsto u, \tag{3.1}$$

for some $p \in \mathbb{N}$ and $u \in \mathcal{A}^+$. We distinguish several cases, depending on p and $r = |u|_b$. First, note that non-triviality requires that $r > 0$ and $|u|_a > 0$.

- If $r = 1$, we either have $\mathbb{X}_\varrho = \{a^{\mathbb{Z}}\}$ or $\mathbb{X}_\varrho$ is the orbit closure of $x = a^\infty.ba^\infty$. We add this to the list of trivial cases.
- If $r \geqslant 2$ and $u = bvb$ for some $v \in \mathcal{A}^+$, then $(\mathbb{X}_\varrho, S)$ is minimal precisely if $p = 1$ and almost minimal otherwise.
- If $r \geqslant 2$ and $u = av$ or $u = va$ for some $v \in \mathcal{A}^+$, the subshift $(\mathbb{X}_\varrho, S)$ is almost minimal.

The minimal case in the second item was investigated in [dOL02], so we will focus on the cases where ϱ is almost minimal in the following. Returning to our initial example of an almost minimal substitution $\varrho\colon a \mapsto a, b \mapsto bba$, we observe that the letter frequencies of both a and b are not well-defined in the fixed point $x^\star$. Hence, its orbit closure $(\mathbb{X}_\varrho, S)$ cannot be uniquely ergodic. This turns out to be a general feature of almost minimal substitutions of the form in (3.1), as long as $p < r$. If $p \geqslant r$, the occurrences of b in points $x \in \mathbb{X}_\varrho$ become so rare that b has a vanishing frequency. A more detailed analysis reveals that, apart from the trivial ergodic measure $\delta_{a^{\mathbb{Z}}}$, the subshift $(\mathbb{X}_\varrho, S)$ always supports an additional ergodic measure that is finite precisely if $p < r$. This is a special case of [Yua07, Prop. 5.4 and Thm. 5.6]. The precise statement is as follows.

Proposition 3.1.14 ([Yua07])**.** *Let ϱ be a non-trivial almost minimal substitution on $\mathcal{A} = \{a, b\}$ of the form in* (3.1)*. Up to a scaling factor, there exists a unique non-atomic ergodic measure μ on $(\mathbb{X}_\varrho, S)$ which is finite on every clopen set disjoint from $a^{\mathbb{Z}}$. With an appropriate scaling, this measure satisfies*

$$\mu([u]) \,=\, \lim_{n\to\infty} \frac{|\varrho^n(b)|_u}{|\varrho^n(b)|_b},$$

for all $u \in \mathcal{A}^+$. The measure μ is finite if and only if $|\varrho(a)|_a < |\varrho(b)|_b$. ■

Remark 3.1.15. The question whether μ is a finite measure depends on the growth behaviour of $|\varrho^n(b)|$, compared to $|\varrho^n(b)|_b = r^n$ as $n \to \infty$. Taking powers of the substitution matrix, we find that $|\varrho^n(b)| \sim r^n$ if $p < r$, $|\varrho^n(b)| \sim nr^n$ if $p = r$, and $|\varrho^n(b)| \sim p^n$ if $p > r$. Here, we use the notation $f(n) \sim g(n)$ to denote that $f(n)/g(n) \to 1$ as $n \to \infty$. The variety of this behaviour is rich enough to provide examples of almost minimal substitutions on $\mathcal{A} = \{a, b\}$ with complexity functions in each of the classes $\Theta(n), \Theta(n \log\log n), \Theta(n \log n)$ and $\Theta(n^2)$, compare [EG21] and [Pan84] for details. More refined estimates of the complexity function for such examples can be found in [Cas97]. ◊

For the sake of brevity and transparency of the exposition, we restrict to an even smaller class of almost minimal substitutions that already exhibits many of the decisive features that can occur for a binary alphabet. A more exhaustive treatment of *all* almost minimal substitutions on $\mathcal{A} = \{a, b\}$ can be found in [EG21].

Definition 3.1.16. A *simple* almost minimal substitution on $\mathcal{A} = \{a, b\}$ is a substitution of the form

$$\varrho_{p,r,j}\colon a \mapsto a^p,\ b \mapsto b^j a b^{r-j},$$

for some $p, r \geqslant 2$ and $1 \leqslant j \leqslant r-1$. We denote by $\mathsf{Sams} = \{\varrho_{p,r,j} : p, r \geqslant 2, 1 \leqslant j \leqslant r\}$ the collection of all simple almost minimal substitutions.

We readily verify that for a simple almost minimal substitution all admitted words are in fact legal. Hence we drop this distinction in the following. Like with the introductory example, we can construct $\mathbb{X}_\varrho$ from an eventually constant fixed point of ϱ for all $\varrho \in \mathsf{Sams}$. Let $\mathbb{X}_\varrho^{\mathrm{ec}}$ be the set of all eventually constant sequences in $\mathbb{X}_\varrho$. These sequences are still exceptional in the sense that $\mathbb{X}_\varrho^{\mathrm{ec}}$ is a countable subset of the uncountable set $\mathbb{X}_\varrho$. More precisely, we have the following.

Lemma 3.1.17. *For every $\varrho \in \mathsf{Sams}$, there are exactly 4 fixed points of ϱ in $\mathcal{A}^{\mathbb{Z}}$, given by*

$$x_{cd}^\star = \lim_{n\to\infty} \varrho^n(c).\varrho^n(d),$$

for $c, d \in \mathcal{A}$. We have $x_{cd}^\star \in \mathbb{X}_\varrho$ precisely if cd is a legal word. Every eventually constant sequence $y \in \mathbb{X}_\varrho$ is an element of the S-orbit of one of these fixed points.

Proof. Let $x^\star \in \mathcal{A}^{\mathbb{Z}}$ be a fixed point under ϱ and set $c = x_{-1}$ and $d = x_0$. Since $\varrho(b)$ has b as its first and its last letter, the limit $\lim_{n\to\infty} \varrho^n(b)$ is well-defined, both as a left-sided and a right-sided sequence. The same holds for $\lim_{n\to\infty} \varrho^n(a) = a^\infty$. Hence,

$$x^\star = \lim_{n\to\infty} \varrho^n(x^\star) = \lim_{n\to\infty} \varrho^n(\cdots x_{-3}x_{-2})\varrho^n(x_{-1}).\varrho^n(x_0)\varrho^n(x_1x_2\cdots) = \lim_{n\to\infty} \varrho^n(c).\varrho^n(d),$$

since both $|\varrho^n(c)| \to \infty$ and $|\varrho^n(d)| \to \infty$ as $n \to \infty$. As soon as cd is a legal word, all subwords of $x^\star$ are legal and therefore $x^\star \in \mathbb{X}_\varrho$.
The last claim is obvious for $y = a^{\mathbb{Z}} = x_{aa}^\star$. Hence, assume that $y \neq a^{\mathbb{Z}}$ and $y \in \mathbb{X}_\varrho$ is eventually constant to the left. Up to a finite shift, y is of the form $y = a^\infty.y^+$ for some $y^+ \in \mathcal{A}^{\mathbb{N}}$ with $y_0^+ = b$. We can decompose y into level-n inflation words for arbitrary n. Since $\varrho^n(a) = a^{p^n}$ and $\varrho^n(b)$ always starts with the letter b, this decomposition is of the form $\varrho^n(a).\varrho^n(b)$ around the origin. Performing the limit $n \to \infty$, we find that $y = x_{ab}^\star$. If $y \in \mathbb{X}_\varrho$ is eventually constant to the right, we argue analogously. □

Note that aa, ab, ba are always legal for $\varrho = \varrho_{p,r,j} \in \mathsf{Sams}$ and bb is legal precisely if $r \geqslant 3$. The fixed point $x_{bb}^\star$ is the only fixed point that is not eventually constant. However, since the measure μ is non-atomic and $\mathbb{X}_\varrho^{\mathrm{ec}}$ is countable, μ-almost every sequence $x \in \mathbb{X}_\varrho$ has infinitely many occurrences of b to both sides of the origin.

3.1.4. Substitutions on a compact alphabet

In our setting, substitutions on a compact alphabet arise as a book-keeping device in the context of random substitutions. More precisely, we will see in Chapter 4 that taking expectation values naturally semi-conjugates random substitutions to substitutions on a compact alphabet in an appropriate sense. Another application of substitutions on a compact alphabet arises if the fixed point of an almost-minimal substitution is decomposed into so-called *return words*; we refer to [EG21] for details.

Apparently, substitutions on compact alphabets have not received much attention in the literature, a notable exception being [DOP18]. A systematic study of their properties has been initiated recently [MRW21, MRW], and we largely follow their notational conventions in this short exhibition.

In this section, let $\mathcal{A}$ be a compact topological Hausdorff space, called a *compact alphabet*. We can again concatenate letters to words and carry over much of the symbolic notation introduced in Section 2.1.2. Naturally, we equip $\mathcal{A}^n$ with the product topology for $n \in \mathbb{N}$, and $\mathcal{A}^+ = \sqcup_{n\in\mathbb{N}}\mathcal{A}^n$ with the corresponding disjoint union topology.

Definition 3.1.18. A substitution on the compact alphabet $\mathcal{A}$ is defined as a *continuous* map $\varrho\colon \mathcal{A} \to \mathcal{A}^+$.

In contrast to the case of a finite alphabet, the continuity of ϱ is not automatic because small neighbourhoods of letters are no longer singleton sets in general. In particular, if $\mathcal{A}$ is connected, this implies that ϱ is of constant length; see [MRW]. Just as we have seen for substitutions, the action of ϱ naturally extends to a map on $\mathcal{A}^+$ and on $\mathcal{A}^{\mathbb{Z}}$. An adequate definition of primitivity needs a slightly more careful approach; compare [DOP18, MRW].

Definition 3.1.19. A substitution ϱ on a compact alphabet $\mathcal{A}$ is called *primitive* if for every non-empty open subset $U \subset \mathcal{A}$, there exists a number $m \in \mathbb{N}$ such that for all $a \in \mathcal{A}$, the word $\varrho^m(a)$ contains some letter in U.

With this definition, many of the properties of primitive substitutions on a compact alphabet resemble those in the finite alphabet case. We will not use these properties and therefore refer to [MRW21, MRW] for further details.

3.2. Random substitutions

As we discussed in Section 3.1, substitutions give rise to self-similar structures of relatively low complexity. In particular, every substitution subshift has vanishing topological entropy, compare Fact 3.1.9. On the other hand, many classical examples of subshifts in the positive entropy regime like subshifts of finite type or sofic subshifts lack any reasonable notion of long-range order [LM95]. Random substitutions provide examples of structures that combine long-range correlations with a positive topological and measure-theoretic entropy. This was illustrated by Godrèche and Luck in their pioneering work [GL89] at the example of the *random Fibonacci substitution*. Over the decades, random substitutions have reappeared in several disguises and across multiple disciplines. They made their possibly first appearance as *D0L-systems* in the context of formal language theory [RoS76], were introduced as *M-systems* in a seminal paper by Peyrière [Pey80] and have served as models for structures in DNA sequences under the name *expansion-modification systems* [Li89]. In the context of random percolation, they were taken up by Dekking, Meester and van der Wal to investigate the occurrence of phase transitions [DM90,DvdW01]. Another application of random substitution deals with the undecidability of the domino problem on certain surface groups [ABM19].

Besides these somewhat scattered appearances in the literature, random substitutions have been the object of renewed interest in recent years. Koslicki and Denker took up ideas of Peyrière and formalized random substitutions as *substitution Markov chains* in [Kos12, KD16]. This was later extended to the setting of sequence spaces in [MT-JU18]. Building on [GL89], the diffraction measure was analyzed for more families of random substitutions

by Moll [Mol14] and by Baake, Spindeler and Strungaru [BSS18], while entropy calculations for several random substitutions were performed by Nilsson [Nil12, Nil13]. A systematic study of topological properties of random substitution subshifts was initiated by Rust and Spindeler [RS18] and complemented over the last few years in a series of papers [EMM21, GRS19, MMRT20, MRST21, Rus20]. In Chapter 4, we will outline results on the ergodicity of random substitution subshifts [GS20] and provide general tools to calculate (or at least approximate) the topological and measure-theoretic entropy associated with these subshifts [Goh20, GMRS21].

3.2.1. Set-valued substitutions

Recall that a substitution is determined by its action on letters of the alphabet $\mathcal{A}$. The basic idea behind a *random* substitution ϑ is to introduce some freedom which word in $\mathcal{A}^+$ a letter gets mapped to, under the application of ϑ. We will stick with the standard convention that there is only a finite number of choices for each letter. Depending on the context and research interest, these choices are either considered as a set of words, or equipped with a probability assignment. For the sake of definiteness, we introduce both concepts under different names, but will later refer to both as *random substitutions* if the context is clear. Sometimes, we will refer to substitutions in the sense of Definition 3.1.1 as *deterministic substitutions* when we wish to emphasize on the difference to random substitutions.

Let $\mathcal{A}$ be a finite alphabet and $\mathcal{F}(\mathcal{A}^+)$ denote the collection of finite subsets of $\mathcal{A}^+$. In analogy to the concatenation of words, we define a concatenation of sets $A, B \in \mathcal{F}(\mathcal{A}^+)$ as

$$AB \,=\, \{uv \in \mathcal{A}^+ : u \in A, v \in B\},$$

which naturally turns $\mathcal{F}(\mathcal{A}^+)$ into a semigroup. The concatenation of sets satisfies the distributive laws $A(B \cup C) = AB \cup AC$ and $(B \cup C)A = BA \cup CA$ for all $A, B, C \in \mathcal{F}(\mathcal{A}^+)$. Hence, $\mathcal{F}(\mathcal{A}^+)$ forms a *semiring* with $\cup$ as addition and concatenation as multiplication. Given a function f on words and $A \subset \mathcal{A}^+$, we set $f(A) = \{f(u) : u \in A\}$. In particular, for $A \in \mathcal{F}(\mathcal{A}^+)$, $|A|, |A|_a$ and $\Phi(A)$ are finite sets of integers or vectors, respectively. If B is a singleton set, we identify it with its unique element.

Definition 3.2.1. A *set-valued substitution* is a map $\vartheta\colon \mathcal{A} \to \mathcal{F}(\mathcal{A}^+)$. It extends uniquely to a map $\vartheta\colon \mathcal{F}(\mathcal{A}^+) \to \mathcal{F}(\mathcal{A}^+)$ via

$$\vartheta(u) \,=\, \vartheta(u_1) \cdots \vartheta(u_n),$$

for all $u = u_1 \cdots u_n \in \mathcal{A}^+$ and

$$\vartheta(A) \,=\, \bigcup_{u \in A} \vartheta(u),$$

for all $A \in \mathcal{F}(\mathcal{A}^+)$.

In this setting, we call every $u \in \vartheta^n(a)$ for some $n \in \mathbb{N}$ and $a \in \mathcal{A}$ a (level-n) *inflation word*. Just like a substitution preserves the *semigroup* structure on $\mathcal{A}^+$, a set-valued substitution preserves the *semiring* structure on $\mathcal{F}(\mathcal{A}^+)$.

Lemma 3.2.2. *The map $\vartheta\colon \mathcal{F}(\mathcal{A}^+) \to \mathcal{F}(\mathcal{A}^+)$ is a semiring endomorphism on $\mathcal{F}(\mathcal{A}^+)$.*

Proof. The relation $\vartheta(A \cup B) = \vartheta(A) \cup \vartheta(B)$ for all $A, B \in \mathcal{F}(\mathcal{A}^+)$ is immediate from the definition of ϑ. Similarly, it is easy to verify that $\vartheta(uv) = \vartheta(u)\vartheta(v)$ for all $u, v \in \mathcal{A}^+$. This implies for all $A, B \in \mathcal{F}(\mathcal{A}^+)$,

$$\vartheta(AB) = \bigcup_{w \in AB} \vartheta(w) = \bigcup_{u \in A, v \in B} \vartheta(u)\vartheta(v) = \Big(\bigcup_{u \in A} \vartheta(u)\Big)\Big(\bigcup_{v \in B} \vartheta(v)\Big) = \vartheta(A)\vartheta(B),$$

where we have used of the distributive law in the penultimate step. □

The extension to $\mathcal{F}(\mathcal{A}^+)$ allows us to concatenate set-valued substitutions and in particular to take powers of a single set-valued substitution ϑ.

Example 3.2.3. The set-valued Fibonacci substitution is given by $\vartheta\colon a \mapsto \{ab, ba\}$, $b \mapsto \{a\}$. On the word $ab \in \mathcal{A}^+$, the set-valued Fibonacci substitution acts as

$$\vartheta(ab) = \vartheta(a)\vartheta(b) = \{ab, ba\}\{a\} = \{aba, baa\}.$$

The square of ϑ is given by

$$\begin{aligned}\vartheta^2(b) &= \vartheta(\vartheta(b)) = \vartheta(a) = \{ab, ba\},\\ \vartheta^2(a) &= \vartheta(\vartheta(a)) = \vartheta(\{ab, ba\}) = \{aba, baa\} \cup \{aab, aba\} = \{aab, aba, baa\}.\end{aligned}$$

We note that $\vartheta^n(b) = \vartheta^{n-1}(a)$ for all $n \in \mathbb{N}$. ◊

The substitution subshift and the language of a set-valued substitution ϑ are defined in a similar manner as for deterministic substitutions, compare Section 3.1. We call a word $u \in \mathcal{A}^+$ *admissible* (for ϑ) if there exists some $n \in \mathbb{N}, a \in \mathcal{A}$ and $v \in \vartheta^n(a)$ such that u is a subword of v. The subshift $\mathbb{X}_\vartheta$ comprises all sequences $x \in \mathcal{A}^{\mathbb{Z}}$ such that every subword of x is admissible, and the language $\mathcal{L} = \mathcal{L}(\mathbb{X}_\vartheta)$ is the set of all subwords that appear in sequences of $\mathbb{X}_\vartheta$. In general, we cannot expect $(\mathbb{X}_\vartheta, S)$ to be minimal. That is because it contains a lot of substitution subshifts that we get by restricting each set $\vartheta(a)$ to one of its elements.

Definition 3.2.4. Let ϑ be a set-valued substitution on $\mathcal{A}$. We call a substitution ϱ on $\mathcal{A}$ a *marginal* of ϑ if $\varrho(a) \in \vartheta(a)$ for all $a \in \mathcal{A}$.

Whenever ϱ is a marginal of ϑ^n for some $n \in \mathbb{N}$, we clearly have $\mathbb{X}_\varrho \subset \mathbb{X}_\vartheta$. In fact, it is not hard to show that the union of these substitution subshifts lies dense in $\mathbb{X}_\vartheta$ [RS18]. It is therefore intuitive that the behaviour of ϑ depends on the properties of these marginals.

Definition 3.2.5. We call a set-valued substitution ϑ *primitive* if there exists a power $n \in \mathbb{N}$ such that some marginal of ϑ^n is primitive.

Suppose that ϱ is a primitive marginal of ϑ^n for some $n \in \mathbb{N}$. Then, $\varnothing \neq \mathbb{X}_\varrho \subset \mathbb{X}_\vartheta$ and hence, the subshift of a primitive set-valued substitution is always non-empty. Also, $\mathbb{X}_\vartheta = \mathbb{X}_{\vartheta^n}$ for all $n \in \mathbb{N}$ if ϑ is primitive [RS18]. Unfortunately, unlike in the deterministic setting, primitivity is not yet enough for the powerful machinery of PF theory to be applicable to set-valued substitutions. However, the analysis becomes easier if there are some properties that are shared by *all* marginals of ϑ.

Definition 3.2.6. A set-valued substitution ϑ is said to be of *constant length* ℓ, with $\ell \in \mathbb{N}$, if all marginals of ϑ are of constant length ℓ. We call ϑ *compatible* if every marginal of ϑ has the same substitution matrix M. In this case, we call $M_\vartheta = M$ the substitution matrix of ϑ.

Both of these properties are stable under taking powers of the set-valued substitution ϑ. This also shows that a *compatible* set-valued substitution is primitive if and only if its substitution matrix is primitive. Compatibility can alternatively be expressed by demanding that $\Phi(\vartheta(a))$ is a singleton for all $a \in \mathcal{A}$. In that case, $M_{ab} = |\vartheta(b)|_a$ for all $a, b \in \mathcal{A}$. Similarly, ϑ is of constant length $\ell \in \mathbb{N}$ precisely if $|\vartheta(a)| = \ell$ for all $a \in \mathcal{A}$.

Example 3.2.7. We call $\vartheta\colon a \mapsto \{ab, ba\}, b \mapsto \{aa\}$ the set-valued period doubling substitution. Its marginals are given by $\varrho_1\colon a \mapsto ab$, $b \mapsto aa$ and $\varrho_2\colon a \mapsto ba$, $b \mapsto aa$. The first marginal ϱ_1 is the standard period doubling substitution from Example 3.1.2. In fact, both marginals give rise to the same subshift $\mathbb{X}_\varrho$. The set-valued substitution ϑ is clearly of constant length 2 since $|\vartheta(a)| = |\vartheta(b)| = 2$. Also, ϑ is compatible, with uniform Abelianisation vectors $\Phi(\vartheta(a)) = (1, 1)$ and $\Phi(\vartheta(b)) = (2, 0)$. Its substitution matrix

$$M \,=\, \begin{pmatrix} 1 & 2 \\ 1 & 0 \end{pmatrix}$$

is primitive and hence the same applies to ϑ. Somewhat surprisingly, the subshift $\mathbb{X}_\vartheta$ contains a dense set of S-periodic sequences [Rus20]. ◊

It is readily verified that the set-valued Fibonacci substitution from Example 3.2.3 is also both compatible and primitive. Combining compatibility and primitivity has strong consequences. Some of them will be explored in Chapter 4. Here, we give a result on uniform letter frequencies as a foretaste. The proof is basically the same as for the corresponding result on deterministic substitutions [Que10, Prop. 5.9]. We give a sketch for the sake of completeness.

Lemma 3.2.8. *Let ϑ be a primitive, compatible set-valued substitution with substitution matrix M and corresponding right PF eigenvector R. For every $\varepsilon > 0$, there exists a number $n_0 \in \mathbb{N}$ such that for all $u \in \mathcal{L}(\mathbb{X}_\vartheta)$ with $|u| \geqslant n_0$, we have*

$$\left| \frac{|u|_a}{|u|} - R_a \right| \,<\, \varepsilon,$$

for all $a \in \mathcal{A}$.

Proof. Suppose $u \in \mathcal{L}_n$ for some large $n \in \mathbb{N}$. Let $m \in \mathbb{N}$ be the maximal level of an inflation word that is completely contained in u. We can write u as

$$u \,=\, v_0 v_1 \cdots v_{m-1} v_m v'_{m-1} \cdots v'_1 v'_0,$$

where for all j, we have $v_j \in \vartheta^m(w_j) \cup \{\varepsilon\}$ for some $w_j \in \mathcal{L}$ and $v'_j \in \vartheta^j(w'_j) \cup \{\varepsilon\}$ for some $w'_j \in \mathcal{L}$. By the maximality of m, we have $|w_m| \leqslant \ell := \max_{a\in\mathcal{A}} |\vartheta(a)|$. We can further assume that $|w_j|, |w'_j| \leqslant \ell$ for all $0 \leqslant j \leqslant m-1$. Let λ be the PF eigenvalue of M. By the PF theorem, for every $b \in \mathcal{A}$ and $j \in \mathbb{N}$, we have

$$\big| |\vartheta^j(b)|_a - |\vartheta^j(b)| R_a \big| \,\leqslant\, c r^j,$$

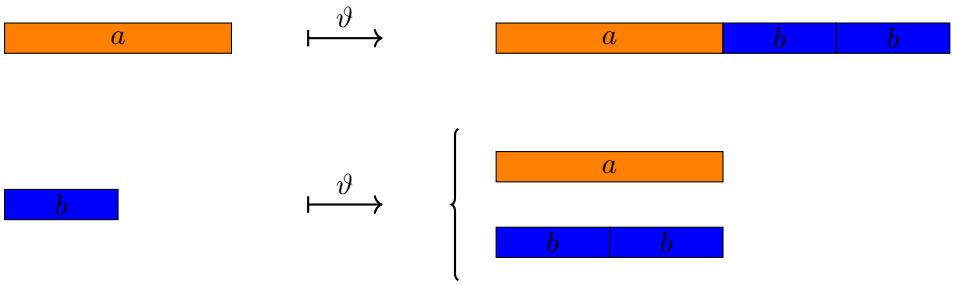

Figure 3.3.: The geometrically compatible set-valued substitution $\vartheta\colon a \mapsto \{abb\}, b \mapsto \{a, bb\}$.

for some $c > 0$ and $0 < r < \lambda$. This implies that

$$||u|_a - |u|R_a| \leqslant 2c\ell \sum_{j=0}^{m} r^j \leqslant Cr^m,$$

for some $C > 0$. Since $|u| \geqslant c'\lambda^m$ for some $c' > 0$, the claim follows. □

It is a natural question to ask, under which conditions we can interpret a set-valued substitution ϑ as a (set-valued) *inflation rule*, compare the discussion in Section 3.1.2. For this, we require that both the inflation factor and the lengths of the intervals are fixed and there is only freedom in which way to cut up the interval corresponding to $\vartheta(a)$ (for some $a \in \mathcal{A}$) into smaller intervals. Set-valued substitutions that are both primitive and compatible meet these requirements. In this case, the inflation factor is given by the PF eigenvalue λ and the interval lengths are encoded in the left PF eigenvector L. Likewise, if ϑ is of constant length ℓ, we can choose ℓ as the inflation factor and take the unit length for each of the intervals I_a, with $a \in \mathcal{A}$. If ϑ meets the minimal requirements to be interpreted as an inflation rule, we call it geometrically compatible.

Definition 3.2.9. A set-valued substitution ϑ is called *geometrically compatible* if there exists a real value $\lambda > 1$ and a vector $L \in \mathbb{R}^{\mathcal{A}}$ with strictly positive entries such that, for every marginal ϱ of ϑ, λ is an eigenvalue of M_ϱ with left eigenvector L. In this situation, we call λ the *inflation factor* of ϑ.

If ϑ is geometrically compatible, then so are all of its powers ϑ^n, for $n \in \mathbb{N}$. The class of geometrically compatible set-valued substitutions is larger than just the union of the primitive compatible class and the constant-length class. We illustrate this with a couple of examples.

Example 3.2.10. The set-valued substitution $\vartheta\colon a \mapsto \{abb\}$, $b \mapsto \{a, bb\}$ is neither compatible nor of constant length. But it is geometrically compatible, since both marginals share the same left eigenvector $L = (2, 1)$ with eigenvalue $\lambda = 2$; compare Figure 3.3.
An example with a non-integer inflation factor is given by $\vartheta\colon a \mapsto \{ab, ca\}$, $b, c \mapsto \{a\}$. This is yet another version of the Fibonacci substitution. We have $\lambda = \tau$ and can choose $L = (\tau, 1, 1)$. Both set-valued substitutions in this example are primitive. ♢

Unlike for primitive deterministic substitutions, the letter frequencies need not be well-defined for geometrically compatible set-valued substitutions. Hence, there is no obvious choice for the normalization of the left eigenvector L and there remains a scaling factor as a free parameter. At least if ϑ is primitive, the letter frequencies exist in an almost sure sense, if we assign probability vectors to the set $\vartheta(a)$, with $a \in \mathcal{A}$. This naturally leads us to the definition of a substitution Markov matrix in the next section.

3.2.2. Substitution Markov matrices

We obtain a random substitution from a set-valued substitution ϑ by choosing a probability vector

$$P_a \colon \vartheta(a) \to (0,1], \quad u \mapsto P_{a,u}$$

with $\sum_{u \in \vartheta(a)} P_{a,u} = 1$, for all $a \in \mathcal{A}$. The value $P_{a,u}$ is to be interpreted as the probability of mapping a to u under the application of ϑ.

Definition 3.2.11. A *substitution Markov matrix* is a Markov matrix P, that is indexed by $\mathcal{A} \times \mathcal{A}^+$ and contains only finitely many non-zero entries. We write $P_a = (P_{a,u})_{u \in \mathcal{A}^+}$ for the probability vector corresponding to $a \in \mathcal{A}$. The set-valued substitution *associated* to P is given by

$$\vartheta \colon a \mapsto \operatorname{supp}(P_a) \,=\, \{u \in \mathcal{A}^+ : P_{a,u} > 0\},$$

for all $a \in \mathcal{A}$. The pair $\vartheta_P = (\vartheta, P)$ is called a *random substitution*.

Remark 3.2.12. Since ϑ is uniquely determined by P, we might as well call P a random substitution. We have chosen to include ϑ in the definition because it is customary to put emphasis on the set of words that can be reached under a random substitution. ◊

Example 3.2.13. The random Fibonacci substitution ϑ_P on $\mathcal{A} = \{a, b\}$ is determined by $P_{a,ab} = p, P_{a,ba} = 1 - p$ and $P_{b,a} = 1$. It has one free parameter $p \in (0,1)$. We represent ϑ_P in a visually more appealing form as

$$\vartheta_P \colon \begin{cases} a \mapsto \begin{cases} ab, & \text{with probability } p, \\ ba & \text{with probability } 1-p, \end{cases} \\ b \mapsto a. \end{cases}$$

This will be the standard format to represent explicit examples of random substitutions in the following. The marginals of ϑ_P are recovered in the limiting cases $p \to 0$ and $p \to 1$. ◊

We extend the action of a random substitution from letters to words by requiring that each letter is mapped *independently*. Hence, the tuple $(v_1, \cdots, v_n)$ of letters is mapped to the tuple of words $(u_1, \cdots, u_n)$ with probability $P_{v_1,u_1} \cdots P_{v_n,u_n}$ under ϑ. At first sight, it might seem tempting to take this as the probability for mapping $v = v_1 \cdots v_n$ to $u = u_1 \cdots u_n$. However, the word $u \in \mathcal{A}^+$ does not necessarily determine such a decomposition into inflation words $(u_1, \cdots, u_n)$ uniquely. We therefore need to proceed with more caution. Given $u, v \in \mathcal{A}^+$, with $v = v_1 \cdots v_n \in \mathcal{L}_n$, we define

$$\mathcal{D}_v(u) \,=\, \big\{(u_1, \cdots, u_n) : u = u_1 \cdots u_n, \text{ and } u_j \in \vartheta(v_j) \text{ for all } 1 \leqslant j \leqslant n\big\}$$

to be the set of all *inflation word decompositions* of u induced by v. Note that $\mathcal{D}_v(u)$ is empty unless $u \in \vartheta(v)$. Finally, we extend P to a Markov matrix, indexed by $\mathcal{A}^+$, via

$$P_{v,u} = \sum_{(u_1,\cdots u_n)\in\mathcal{D}_v(u)} \prod_{j=1}^{n} P_{v_j,u_j},$$

for all $u, v \in \mathcal{A}^+$ with $|v| = n$. By convention, $P_{v,u} = 0$ if $\mathcal{D}_v(u) = \varnothing$ and hence $P_{v,u} > 0$ if and only if $u \in \vartheta(v)$.

Remark 3.2.14. If $|\vartheta(a)|$ is a singleton set for all $a \in \mathcal{A}$, the set $\mathcal{D}_v(u)$ is either empty or a singleton set for all $u, v \in \mathcal{A}^+$ because v uniquely determines the "cutting points" in the inflation word decomposition of u. This holds in particular if ϑ is compatible or of constant length, and thus for many cases of interest. Changing to the geometric picture, we easily verify that the same is true if ϑ is geometrically compatible. On the other hand, consider $\vartheta\colon a \mapsto \{a, ab\}$, $b \mapsto \{ba, a\}$. Here, $\mathcal{D}_{ab}(aba) = \{(a, ba), (ab, a)\}$ has cardinality 2. ◊

Starting from an initial word $u \in \mathcal{A}^+$, we would like to describe the stochastic process of words that we obtain by iterated applications of the random substitutions. This was formalized by Peyrière [Pey80] and later by Denker and Koslicki [KD16]; compare also [GMRS21].

Definition 3.2.15. Let ϑ_P a random substitution on $\mathcal{A}$ and $u \in \mathcal{A}^+$. A corresponding *substitution Markov chain* is a word-valued homogeneous Markov chain $(\vartheta_P^n(u))_{n\in\mathbb{N}_0}$ on some probability space $(\Omega_u, \mathcal{F}_u, \mathbb{P}_u)$ that satisfies $\vartheta_P^0(u) = u$ and

$$\mathbb{P}_u\big[\vartheta_P^{n+1}(u) = w | \vartheta_P^n(u) = v\big] = P_{v,w},$$

for all $v, w \in \mathcal{A}^+$ and $n \in \mathbb{N}_0$.

Remark 3.2.16. Instead of starting from $u \in \mathcal{A}^+$, we could start from an arbitrary distribution of words. The Markov matrix P uniquely determines all the (finite) joint distributions of random words in $(\vartheta_P^n(u))_{n\in\mathbb{N}}$. This will suffice for most of this work. There is still some freedom in the choice of the probability space $(\Omega_u, \mathcal{F}_u, \mathbb{P}_u)$. There is an explicit construction following ideas from the theory of Galton–Watson processes, compare [GS20]. This has the advantage of an additional structure that ensures the measurability of certain *induced substitutions*, constructed from ϑ_P. If the context is clear, we mostly write $\mathbb{P}$ instead of $\mathbb{P}_u$. It is important to note that the construction of P ensures that $\vartheta_P^n(a)\vartheta_P^n(b)$ (as a random variable on $(\Omega_a \times \Omega_b, \mathcal{F}_a \times \mathcal{F}_b, \mathbb{P}_a \otimes \mathbb{P}_b)$) defines the same distribution as $\vartheta_P^n(ab)$ on $(\Omega_{ab}, \mathcal{F}_{ab}, \mathbb{P}_{ab})$, for all $n \in \mathbb{N}$ and $a, b \in \mathcal{A}$. We refer to [GS20] for details. ◊

The distribution of $\vartheta_P^n(u)$ is given by $\mathbb{P}[\vartheta_P^n(u) = v] = P_{uv}^n$ and hence coincides with the distribution of $\vartheta_{P^n}(u)$ for all $n \in \mathbb{N}$. Since ϑ^n is the set-valued substitution associated to P^n, we consistently define $\vartheta_P^n = (\vartheta^n, P^n)$ to be the nth power of ϑ_P for all $n \in \mathbb{N}$.

We carry over some of the characterizations of set-valued substitutions introduced in Section 3.2.1, mutatis mutandis, to random substitutions. That is, we say that a random substitution $\vartheta_P = (\vartheta, P)$ is primitive/compatible/geometrically compatible/of constant length if the corresponding attribute holds for ϑ. Likewise, the substitution subshift and the language

of $\vartheta_P = (\vartheta, P)$ are given by the corresponding objects for ϑ. The advantage of the random setting is that, even if $\Phi(\vartheta(a))$ is a set (and hence $\Phi(\vartheta_P(a))$ a random vector), we can average over all possibilities to get a unique value.

Definition 3.2.17. The *substitution matrix* $M = M_{\vartheta_P} \in \mathrm{Mat}(\#\mathcal{A}, \mathbb{R})$ of a random substitution ϑ_P is defined by $M_{ab} = \mathbb{E}|\vartheta_P(b)|_a$ for all $a, b \in \mathcal{A}$.

This should not be confused with the substitution Markov matrix P. Again, M is related to the Abelianisation of ϑ_P. More precisely,

$$\mathbb{E}\big[\,\big|\Phi\big(\vartheta_P^{n+1}(u)\big)\big\rangle\,\big|\,\Phi\big(\vartheta_P^{n}(u)\big)\,\big] \;=\; M\big|\Phi\big(\vartheta_P^{n}(u)\big)\big\rangle,$$

for all $n \in \mathbb{N}_0$, as follows from a straightforward calculation [GS20, Prop. 3.3]. As a direct consequence, we observe that

$$\mathbb{E}\big[|\vartheta_P^n(b)|_a\big] \;=\; \mathbb{E}\big[\big\langle e_a\big|\Phi\big(\vartheta_P^n(b)\big)\big\rangle\big] \;=\; \langle e_a|M^n|\Phi(b)\rangle \;=\; (M^n)_{ab}$$

for all $a, b \in \mathcal{A}$. Hence, M^n is the substitution matrix for ϑ_P^n for all $n \in \mathbb{N}$. The substitution matrix provides an alternative characterization of primitivity, even if ϑ is not compatible.

Lemma 3.2.18. *A random substitution ϑ_P is primitive if and only if its substitution matrix M is primitive with PF eigenvalue $\lambda > 1$.*

Proof. First, assume that $\vartheta_P = (\vartheta, P)$ is primitive. Let ϱ be a primitive marginal of ϑ^n. Then, M_ϱ^m is strictly positive for some $m \in \mathbb{N}$. Since ϱ^m is a marginal of ϑ^{nm}, this implies that $(M_{\vartheta_P}^{nm})_{ab} = \mathbb{E}[|\vartheta^{nm}(b)|_a] > 0$ for all $a, b \in \mathcal{A}$. Hence, M_{ϑ_P} is primitive. By a similar argument, $\mathbb{E}\big[|\vartheta_P^{nm}(a)|\big] > 1$ for all $a \in \mathcal{A}$, implying that $\lambda > 1$.
Conversely, assume that M is primitive with PF eigenvalue $\lambda > 1$ and PF eigenvectors L, R. Since $M^n = \lambda^n |R\rangle\langle L| + o(\lambda^n)$, we can choose $n \in \mathbb{N}$ large enough to ensure that $\mathbb{E}[|\vartheta_P^n(a)|_b] \geqslant 1$ for all $a, b \in \mathcal{A}$. This also implies $\mathbb{E}[|\vartheta_P^n(a)|] \geqslant \#\mathcal{A}$ for all $a \in \mathcal{A}$. For a fixed $a \in \mathcal{A}$, choose $u^a \in \vartheta^n(a)$ with $|u^a| \geqslant \#\mathcal{A}$. For each $b \in \mathcal{A}$, choose a letter u_j^a and $v_j \in \vartheta^n(u_j^a)$ such that $|v_j|_b \geqslant 1$. Since $|u^a| \geqslant \#\mathcal{A}$, we can choose a different j for each b. Due to this construction, the word

$$v^a \;=\; v_1 \cdots v_{|u^a|} \in \vartheta^n(u_a) \;\subset\; \vartheta^{2n}(a)$$

contains every letter $b \in \mathcal{A}$. We repeat this procedure for every $a \in \mathcal{A}$. Setting $\varrho\colon a \mapsto v^a$, we obtain a marginal of ϱ^{2n} with strictly positive substitution matrix M_ϱ. We conclude that ϑ_P is primitive. □

Unlike for deterministic substitutions, having a primitive substitution matrix M for a random substitution does *not* imply that its PF eigenvalue λ is larger than 1. This can be seen from the example $\vartheta_P = (\vartheta, P)$ with $\vartheta\colon a, b \mapsto \{a, b\}$. Demanding $\lambda > 1$ excludes such trivial cases. Primitivity also has the following convenient consequence.

Lemma 3.2.19. *Let ϑ_P be a primitive random substitution. Then, $\lim_{n\to\infty} |\vartheta_P^n(a)| = \infty$ holds $\mathbb{P}_a$-almost surely for all $a \in \mathcal{A}$.*

Proof. Let $a \in \mathcal{A}$. The limit always exists because $(|\vartheta_P^n(a)|)_{n\in\mathbb{N}}$ is a non-decreasing sequence of random variables. Possibly replacing ϑ_P by one of its powers, we can assume $\mathbb{E}[|\vartheta_P(b)|] > 1$ and hence $\mathbb{P}[|\vartheta_P(b)| \geqslant 2] > 0$ for all $b \in \mathcal{A}$, due to primitivity. Given $N \in \mathbb{N}$, the probability that $|\vartheta_P^n(a)| \leqslant N$ therefore decays exponentially with n. By an application of the Borel–Cantelli lemma, this implies that $\lim_{n\to\infty} |\vartheta_P^n(a)| > N$ holds almost surely. Since N was arbitrary, the claim follows. □

As a last step, let us extend the action of a random substitution to sequence spaces. This is the point of view adopted in [MT-JU18]. Since we leave the realm of countable state spaces, the substitution Markov matrix P is replaced by a *Markov kernel* $P\colon \mathcal{A}^{\mathbb{N}} \times \mathcal{B} \to [0,1]$ on the Borel space $(\mathcal{A}^{\mathbb{N}}, \mathcal{B})$; compare Section 1.4. Given $x \in \mathcal{A}^{\mathbb{N}}$ and $u \in \mathcal{A}^n$, it is defined on the corresponding cylinder set via

$$P(x, [u]) \,=\, \mathbb{P}[\vartheta(x_1 \cdots x_n)_{[1,n]} = u] \,=\, \sum_{v, v_{[1,n]}=u} P_{x_{[1,n]},v}\,.$$

Since the cylinder sets generate the clopen topology, this uniquely determines P. Measurability in the first argument follows from the observation that the above expression is locally constant in x. Given $u, v \in \mathcal{A}^n$, it is natural to define $P([v],[u])$ as the common value $P(x,[u])$ for all $x \in [v]$. More specifically,

$$P([v],[u]) \,=\, \mathbb{P}\big[\vartheta(v)_{[1,n]} = u\big].$$

The dual operator P^* on the space $\mathcal{M}^1(\mathcal{A}^{\mathbb{N}})$ satisfies

$$(P^*\mu)(B) \,=\, \int_{\mathcal{A}^{\mathbb{N}}} P(x,B)\,\mathrm{d}\mu(x),$$

for every Borel set $B \subset \mathcal{A}^{\mathbb{N}}$ and $\mu \in \mathcal{M}(\mathcal{A}^{\mathbb{N}})$. For $u \in \mathcal{A}^n$, this yields

$$(P^*\mu)([u]) \,=\, \sum_{v\in\mathcal{A}^n} \mu([v])\, P([v],[u]).$$

The extension to $\mathcal{A}^{\mathbb{Z}}$ works similarly. To be precise, for $u \in \mathcal{A}^m, v \in \mathcal{A}^n$ and $x \in \mathcal{A}^{\mathbb{Z}}$, we set

$$P(x, [u.v]) \,=\, \mathbb{P}\big[\vartheta(x_{-m} \cdots x_{-1}.x_0 \cdots x_{n-1})_{[-m,n-1]} = u.v\big].$$

For $\mathbb{I} \in \{\mathbb{Z}, \mathbb{N}\}$, the transition kernel P induces a substitution Markov chain on $\mathcal{A}^{\mathbb{I}}$, just like in Definition 3.2.15. Given an initial distribution μ on $\mathcal{A}^{\mathbb{I}}$, $\vartheta_P(\mu)$ has therefore a well-defined distribution $P^*\mu$. With some abuse of notation, we also write $\vartheta_P(\mu)$ for its distribution $P^*\mu$. We summarize this as follows.

Definition 3.2.20. Let ϑ_P be a random substitution on $\mathcal{A}$ and $\mu \in \mathcal{M}^1(\mathcal{A}^{\mathbb{N}})$. We denote by $\vartheta_P(\mu)$ the Borel measure defined by

$$(\vartheta_P(\mu))([u]) \,=\, \sum_{v,|v|=|u|} \mu([v])\, \mathbb{P}\big[\vartheta_P(v)_{[1,n]} = u\big],$$

for all $u \in \mathcal{A}^+$. Similarly, for $\mu \in \mathcal{M}^1(\mathcal{A}^{\mathbb{Z}})$, and $u \in \mathcal{A}^m, u' \in \mathcal{A}^n$,

$$(\vartheta_P(\mu))([u.u']) \,=\, \sum_{v\in\mathcal{A}^m, v'\in\mathcal{A}^n} \mu([v.v'])\, \mathbb{P}\big[\vartheta_P(v.v')_{[-m,n-1]} = u.u'\big]$$

defines a Borel measure $\vartheta_P(\mu) \in \mathcal{M}^1(\mathcal{A}^{\mathbb{Z}})$.

4. Random substitutions

A random substitution $\vartheta_P = (\vartheta, P)$ and its associated set-valued substitution ϑ are intimately related. For notational simplicity, we will mostly refer to both concepts as the "random substitution ϑ" in the following. Hence, the expression $\vartheta(a)$ stands for both a random word $\vartheta_P(a)$ and for the set of words $\operatorname{supp}\vartheta_P(a)$. The actual meaning will be clear from the context. Some operations on $\vartheta(a)$ pertain unequivocally to either a random setting (like probabilities or expectations) or to a set notation (like cardinalities or elements). On the other hand, operations on words can be applied to $\vartheta(a)$ in both interpretations, and the outcome is consistent. To be more precise, let f be a function on $\mathcal{A}^+$. Then, $f(\vartheta_P(a)) = f \circ \vartheta_P(a)$ is a concatenation of a function with a random variable and hence a random variable. Likewise, $f(\operatorname{supp}\vartheta_P(a)) = \operatorname{supp} f(\vartheta_P(a))$ is a well-defined set. Hence, this notational ambiguity should hopefully not lead to confusion.

Let us give a short outline of the main results of this chapter. Generalising the idea of a fixed point, we show in Proposition 4.1.3 that every non-trivial random substitution ϑ admits an invariant distribution (under some power of the random substitution). If the random substitution is primitive, the orbit average of this invariant distribution produces a measure μ_ϑ on the subshift $(\mathbb{X}_\vartheta, S)$ that reflects the frequencies of words in large random inflation words. We prove in Theorem 4.1.13 that the measure μ_ϑ is ergodic on $(\mathbb{X}_\vartheta, S)$.

The entropy of random substitution subshifts is considered in Section 4.2. If the random substitution is primitive and compatible, we show that both the topological and the measure-theoretic entropy can be obtained from the substitution Markov chain. For $v \in \mathcal{A}^+$ and $n \in \mathbb{N}$, recall the notation $H_{\mathbb{P}}(\vartheta^n(v))$ for the entropy of the finite partition generated by the word-valued random variable $\vartheta^n(v)$; see Remark 2.3.3.

Theorem 4.0.1. *Let $(\mathbb{X}_\vartheta, S)$ be the subshift of a primitive compatible random substitution. The topological entropy $s_\vartheta = h_{\mathrm{top}}(S)$ of $(\mathbb{X}_\vartheta, S)$ satisfies*

$$s_\vartheta \;=\; \lim_{n\to\infty} \frac{\log \#\vartheta^n(v)}{|\vartheta^n(v)|},$$

for all $v \in \mathcal{A}^+$. The entropy $h_\vartheta = h_{\mu_\vartheta}(S)$ of the ergodic measure μ_ϑ on $(\mathbb{X}_\vartheta, S)$ is given by

$$h_\vartheta \;=\; \lim_{n\to\infty} \frac{H_{\mathbb{P}}(\vartheta^n(v))}{|\vartheta^n(v)|},$$

for all $v \in \mathcal{A}^+$.

The statement of this result is refined in Theorem 4.2.10 and Proposition 4.2.17. There, we show that an appropriate averaging procedure provides explicit upper and lower bounds in

terms of the family $\{\vartheta^n(a)\}_{a\in\mathcal{A}}$, for every $n \in \mathbb{N}$. We also provide sharp conditions for these bounds to be attained.

The last part of this chapter, Section 4.3, deals with the diffraction of $(\mathbb{X}_\vartheta, S)$, where ϑ is a primitive compatible random substitution. Due to the ergodicity of μ_ϑ, the diffraction measure $\widehat{\gamma} = \widehat{\gamma}(x)$ is the same for μ_ϑ-almost every $x \in \mathbb{X}_\vartheta$. We show that this diffraction measure can be recovered from the substitution Markov chain via

$$\widehat{\gamma} \,=\, \lim_{n\to\infty} \widehat{\gamma}(\vartheta^n(a)),$$

for all $a \in \mathcal{A}$, where $\widehat{\gamma}(\vartheta^n(a))$ is the expected value of the (renormalized) diffraction measure of the random word $\vartheta^n(a)$. Using an idea of Godrèche and Luck [GL89], the approximants can be decomposed into a sum, corresponding to first and second moments, as

$$\widehat{\gamma}(\vartheta^n(a)) \,=\, \widehat{\gamma}^{\mathrm{ex}}(\vartheta^n(a)) + \widehat{\gamma}^{\mathrm{var}}(\vartheta^n(a)).$$

See Definition 4.3.3 for details. The following will be proved in Theorem 4.3.4.

Theorem 4.0.2. *Let ϑ be a primitive, geometrically compatible random substitution. For every $a \in \mathcal{A}$, the vague limits*

$$\widehat{\gamma}^{\mathrm{ex}} \,=\, \lim_{n\to\infty} \widehat{\gamma}^{\mathrm{ex}}(\vartheta^n(a)) \quad \text{and} \quad \widehat{\gamma}^{\mathrm{var}} \,=\, \lim_{n\to\infty} \widehat{\gamma}^{\mathrm{var}}(\vartheta^n(a))$$

both exist, and are independent of $a \in \mathcal{A}$. The measure $\widehat{\gamma}^{\mathrm{var}}$ is absolutely continuous with respect to Lebesgue measure.

An expression for the density of $\widehat{\gamma}^{\mathrm{var}}$ in terms of an infinite series is provided in Proposition 4.3.9. The spectral type of $\widehat{\gamma}^{\mathrm{ex}}$ is more difficult to determine. If ϑ is of constant length ℓ, we call it *scrambling* if, for all $a, b \in \mathcal{A}$, there exists a position $1 \leqslant j \leqslant \ell$ such that the random letters $\vartheta(a)_j$ and $\vartheta(b)_j$ coincide with positive probability.

Theorem 4.0.3. *Let ϑ be a primitive compatible random substitution of constant length. Assume further that some power of ϑ is scrambling. Then, $\widehat{\gamma}^{\mathrm{ex}}$ is a pure point measure.*

Both $\widehat{\gamma}^{\mathrm{ex}}$ and $\widehat{\gamma}^{\mathrm{var}}$ can also be obtained from the ϑ-invariant distribution; we refer to Theorem 4.3.27 for the details. In Section 4.3.3, we show that the assumption of constant length can be relaxed to that of an integer inflation factor λ, provided ϑ is geometrically compatible.

4.1. Invariant and ergodic measures

The subshift $\mathbb{X}_\vartheta$ of a random substitution ϑ is a purely combinatorial object as it can be constructed from the corresponding set-valued substitution. It therefore lacks a probabilistic interpretation. Another apparent deficit of $\mathbb{X}_\vartheta$ is that many properties like the frequencies of letters are in general not well-defined. Even in the compatible case, where letter frequencies *do* exist uniformly, the frequencies of larger words still fail to exist. As a result, there is no natural autocorrelation or diffraction measure associated to $(\mathbb{X}_\vartheta, S)$ as a topological dynamical

system. We address both points by choosing an appropriate measure on $\mathbb{X}_\vartheta$ that reflects the substitution probabilities encoded in the substitution Markov matrix P. With respect to this measure, the word frequencies will be defined almost surely. It is possible to choose this measure to be either invariant under (some power of) ϑ or under S (but not both). We start with the discussion of substitution invariant measures on $\mathbb{X}_\vartheta$.

Recall from Remark 3.1.7 that a primitive substitution ϱ admits a fixed point x^* under ϱ^r for some $r \in \mathbb{N}$. By the strict ergodicity of $(\mathbb{X}_\varrho, S)$, the autocorrelation and diffraction measure are the same for every $x \in \mathbb{X}$ and can hence be calculated from x^*. The point x^* is convenient because it is constructive and allows us to use renormalisation techniques. For a random substitution ϑ, the concept of a fixed *point* x^* is naturally replaced by that of a fixed *measure* ν^*. Just like x^* emerges as a limit of iterating ϱ on $\mathcal{A}^{\mathbb{Z}}$, we obtain ν^* as a limit of iterating ϑ on $\mathcal{M}^1(\mathcal{A}^{\mathbb{Z}})$. In principle, there might be several such limiting measures, but we will we see that they are all convex combinations of measures from a finite set.

For ease of notation, we first work on $\mathcal{A}^{\mathbb{N}}$. The arguments readily generalize to $\mathcal{A}^{\mathbb{Z}}$. Convergence is understood in the weak topology. On sequence spaces, this is equivalent to convergence on every cylinder set [Bil99, Thm. 2.2]. That is, we say that $\lim_{n\to\infty} \vartheta^n(\mu) \to \nu^*$ if for all $u \in \mathcal{A}^+$,

$$\lim_{n\to\infty} \vartheta^n(\mu)([u]) \,=\, \nu^*([u]).$$

Recall that for $u \in \mathcal{A}^m$,

$$\vartheta(\mu)([u]) \,=\, \sum_{v\in\mathcal{A}^m} \mu([v])\, \mathbb{P}\big[\vartheta(v)_{[1,m]} = u\big].$$

This has the structure of a product between a vector and a matrix.

Definition 4.1.1. Let ϑ be a random substitution on $\mathcal{A}$ and $m \in \mathbb{N}$. The *m-prefix matrix* $M^{(m)}$ is a Markov matrix on $\mathcal{A}^m$ with entries

$$M^{(m)}_{u,v} \,=\, \mathbb{P}\big[\vartheta(u)_{[1,m]} = v\big],$$

for all $u, v \in \mathcal{A}^m$. The *m-prefix vector* $\mu^{(m)}$ of a measure $\mu \in \mathcal{M}^1(\mathcal{A}^{\mathbb{N}})$, is a $\mathcal{A}^m$-indexed probability vector, with $\mu^{(m)}_v = \mu([v])$, for all $v \in \mathcal{A}^m$.

With this notation, $\lim_{n\to\infty} \vartheta^n(\mu) = \nu$ is equivalent to

$$\lim_{n\to\infty} \big\langle \mu^{(m)} \big| (M^{(m)})^n \,=\, \big\langle \nu^{(m)} \big|,$$

for all $m \in \mathbb{N}$. Recall the basic notions of Markov chains on a finite state space from Section 1.4.

Definition 4.1.2. Let ϑ be a random substitution with 1-prefix matrix $M^{(1)}$ and let r be the period of $M^{(1)}$. Then, we call r the *period of* ϑ.

Possibly replacing ϑ by ϑ^r, we assume that ϑ is of period 1 in the following. Hence, the restriction of $M^{(1)}$ to each of its recurrence classes $\mathcal{R}_j$ is primitive. Let us denote by $\mathbb{X}^+_\vartheta \subset \mathcal{A}^{\mathbb{N}}$ the image of $\mathbb{X}_\vartheta$ under the projection $x \mapsto x_{[1,\infty)}$.

Proposition 4.1.3. *Let ϑ be a random substitution of period 1 such that $|\vartheta^n(a)| \to \infty$ as $n \to \infty$ holds almost surely for all $a \in \mathcal{A}$. Assume that $\mathcal{R}_1, \dots, \mathcal{R}_s \subset \mathcal{A}$ are the recurrence classes of $M^{(1)}$ and let c_{aj} be the absorption probability from $a \in \mathcal{A}$ to class $\mathcal{R}_j$. Then, for each $1 \leqslant j \leqslant s$, there is a probability measure ν_j on $\mathcal{A}^{\mathbb{N}}$ with the following properties.*

(1) *$\vartheta(\nu_j) = \nu_j$.*

(2) *If $\mu \in \mathcal{M}^1(\mathcal{A}^{\mathbb{N}})$, with $\sum_{a \in R_j} \mu[a] = 1$, we have $\lim_{n\to\infty} \vartheta^n(\mu) = \nu_j$.*

(3) *The topological support of ν_j is contained in $\mathbb{X}_\vartheta^+$ and $\sum_{a\in\mathcal{R}_j} \nu_j[a] = 1$.*

In general, $\lim_{n\to\infty} \vartheta^n(\mu) = \sum_{a\in\mathcal{A}} \mu[a] \sum_{j=1}^{s} c_{aj}\nu_j$, for all $\mu \in \mathcal{M}^1(\mathcal{A}^{\mathbb{N}})$.

Proof. Let $j \in \{1, \dots, s\}$ be fixed and assume that $\sum_{a\in\mathcal{R}_j} \mu[a] = 1$. In words, the 1-prefix vector $\mu^{(1)}$ is supported on $\mathcal{R}_j$. Since the restriction of $M^{(1)}$ to $\mathcal{R}_j$ is primitive, with some equilibrium vector π^j, we find

$$\lim_{n\to\infty} \bigl\langle \mu^{(1)} \bigr| (M^{(1)})^n \;=\; \langle \pi^j|.$$

Now, let $m > 1$ and

$$\mathcal{L}_m(j) \;=\; \{u \in \mathcal{L}_m : \mathbb{P}[\vartheta^n(a)_{[1,m]} = u] > 0 \text{ for some } n \in \mathbb{N},\, a \in \mathcal{R}_j\}.$$

We prove that $\mathcal{L}_m(j)$ is a recurrence class of $M^{(m)}$ with period 1. As a first step, we show that the restriction of $M^{(m)}$ to $\mathcal{L}_m(j)$ is primitive. For each $u \in \mathcal{L}_m(j)$, choose a pair (a_u, n_u) in $\mathcal{R}_j \times \mathbb{N}$ such that $\mathbb{P}[\vartheta^{n_u}(a_u)_{[1,m]} = u] > 0$. Since $M^{(1)}$ is primitive on $\mathcal{R}_j$, there exists a k_0 such that $\mathbb{P}[\vartheta^k(a)_1 = a_u] > 0$ for all $k \geqslant k_0$. Hence, there exists a number $n \in \mathbb{N}$ such that

$$\mathbb{P}[\vartheta^n(a)_{[1,m]} = u] \;>\; 0,$$

for all $u \in \mathcal{L}_m(j)$ and $a \in \mathcal{R}_j$. Since every word in $\mathcal{L}_m(j)$ starts with a letter in $\mathcal{R}_j$, this implies that

$$(M^{(m)})^n_{v,u} = \mathbb{P}[\vartheta^n(v)_{[1,m]} = u] > 0,$$

for all $u, v \in \mathcal{L}_m(j)$, that is, $M^{(m)}$ restricts to a primitive matrix on $\mathcal{L}_m(j)$.

Note that the vector $\mu^{(m)}$ is supported on the set $\mathcal{A}^m(j) = \{u \in A^m : u_1 \in \mathcal{R}_j\} \supset \mathcal{L}_m(j)$. We claim that all words in $\mathcal{A}^m(j) \setminus \mathcal{L}_m(j)$ are transient under $M^{(m)}$. Let $u \in \mathcal{A}^m(j) \setminus \mathcal{L}_m(j)$ with $u_1 = a$. There exists an $n \in \mathbb{N}$ and $w \in \vartheta^n(a)$ such that $|w| \geqslant m$. That is, $w_{[1,m]} \in \mathcal{L}_m(j)$ is a prefix of some word in $\vartheta^n(u)$. That means that u is absorbed with positive probability by the recurrence class $\mathcal{L}_m(j)$ and hence it is transient. Since $\mathcal{L}_m(j)$ is the only recurrence class in $\mathcal{A}^m(j)$, the absorption probability is in fact 1.
Let $\pi^{j,m}$ be the equilibrium vector of $M^{(m)}$ on $\mathcal{L}_m(j)$. By Fact 1.4.3, we obtain

$$\lim_{n\to\infty} \bigl\langle \mu^{(m)} \bigr| (M^{(m)})^n \;=\; \bigl\langle \pi^{j,m} \bigr|,$$

for all $m \in \mathbb{N}$. Defining $\nu_j([v]) = \pi^{j,m}_v$ for all $v \in \mathcal{A}^m, m \in \mathbb{N}$, we get $\lim_{n\to\infty} \vartheta^n(\mu) = \nu_j$. By construction, ϑ is a continuous operator on $\mathcal{M}^1(\mathcal{A}^{\mathbb{N}})$ in the weak topology and hence

$$\vartheta(\nu_j) \;=\; \vartheta\bigl(\lim_{n\to\infty} \vartheta^n(\mu)\bigr) \;=\; \lim_{n\to\infty} \vartheta^{n+1}(\mu) \;=\; \nu_j,$$

with μ as above. Since $\mathcal{L}_m(j) \subset \mathcal{L}$ for all $m \in \mathbb{N}$, it follows immediately that the support of ν_j is contained in $\mathbb{X}_\vartheta^+$. For the general case, note that a word $u \in \mathcal{A}^m$ with $u_1 = a$ is absorbed with probability c_{aj} by $A^m(j)$ and hence by $\mathcal{L}_m(j)$. Another application of Fact 1.4.3 yields

$$\lim_{n\to\infty} \langle \mu^{(m)} | (M^{(m)})^n \;=\; \sum_{a\in\mathcal{A}} \mu[a] \sum_{j=1}^{s} c_{aj} \langle \pi^{j,m} |,$$

and the last claim follows. □

Remark 4.1.4. As an immediate consequence of Proposition 4.1.3, we obtain that $M^{(1)}$ has a unique recurrence class in $\mathcal{A}$ if and only if there is a unique ϑ-invariant measure $\nu \in \mathcal{M}^1(\mathcal{A}^{\mathbb{N}})$ such that $\lim_{n\to\infty} \vartheta^n(\mu) = \nu$ for all $\mu \in \mathcal{M}^1(\mathcal{A}^{\mathbb{N}})$. We note that the arguments in the proof of Proposition 4.1.3 do not rely on ϑ acting independently on neighboring letters. This yields an improvement of [MT-JU18, Thm. 1], where independence was not required, but it was implicitly assumed that $M^{(m)}$ is a primitive matrix on $\mathcal{A}^m$ for all $m \in \mathbb{N}$. ◊

Remark 4.1.5. The discussion of ϑ-invariant distributions on the *two-sided* subshift $(\mathbb{X}_\vartheta, S)$ can be done in an analogous fashion. Here, instead of just considering prefixes, we consider prefixes and suffixes simultaneously. The analogue of $M^{(1)}$ is a matrix M, indexed by $\mathcal{L}_2(\mathbb{X}_\vartheta)$, where the entry $M_{u_1u_2,v_1v_2}$ is the probability that the last letter of $\vartheta(u_1)$ is v_1 and the first letter of $\vartheta(u_2)$ is v_2. That is, we think of $\mathcal{L}_2(\mathbb{X}_\vartheta)$ as legal seeds, placed symmetrically around the origin; compare the discussion in Remark 3.1.7. If M has period 1 (otherwise we take powers of ϑ), the recurrence classes of M are in 1-to-1 correspondence with the ϑ-invariant distributions on $(\mathbb{X}_\vartheta, S)$. ◊

We can find a more explicit form for the invariant measures ν_j if ϑ satisfies some uniform growth assumption.

Definition 4.1.6. We say that a random substitution *grows uniformly* if

$$\lim_{n\to\infty} \min\{|u| : u \in \vartheta^n(a)\} \;=\; \infty,$$

for all $a \in \mathcal{A}$.

Lemma 4.1.7. *Let ϑ be a random substitution satisfying the assumptions in Proposition* 4.1.3 *and assume that ϑ grows uniformly. For $m \in \mathbb{N}$, let $n_m \in \mathbb{N}$ be the minimal number such that $|\vartheta^{n_m}(a)| \geqslant m$ for all $a \in \mathcal{A}$. Then, given $v \in \mathcal{A}^m$, we have*

$$\nu_j([v]) \;=\; \sum_{b\in\mathcal{R}_j} \pi_b^j \, \mathbb{P}\big[\vartheta^n(b)_{[1,m]} = v\big],$$

for all $n \geqslant n_m$, where π^j denotes the equilibrium vector of $M^{(1)}$ restricted to the class $\mathcal{R}_j$.

Proof. Let $m \in \mathbb{N}$ be fixed. Comparing with the proof of Proposition 4.1.3, we see that it suffices to show that the probability vector q on $\mathcal{A}^m$ with

$$q_v \;=\; \sum_{b\in\mathcal{R}_j} \pi_b^j \, \mathbb{P}\big[\vartheta^n(b)_{[1,m]} = v\big]$$

for all $v \in \mathcal{A}^m$, is supported on $\mathcal{L}_m(j)$ and invariant under $\mathcal{M}^{(m)}$. The first claim is obvious. For the second, we first observe that, for $n \geqslant n_m$,

$$\begin{aligned}\sum_{b\in\mathcal{R}_j} \pi_b^j \,\mathbb{P}[\vartheta^{n+1}(b)_{[1,m]} = v] \;&=\; \sum_{b\in\mathcal{R}_j} \pi_b^j \sum_{c\in\mathcal{A}} \mathbb{P}[\vartheta(b)_1 = c]\,\mathbb{P}[\vartheta^n(c)_{[1,m]} = v] \\ &=\; \sum_{c\in\mathcal{R}_j} \pi_c^j \,\mathbb{P}[\vartheta^n(b)_{[1,m]} = v],\end{aligned}$$

where we have used that π^j is invariant under $M^{(1)}$ with entries $M^{(1)}_{b,c} = \mathbb{P}[\vartheta(b)_1 = c]$. Hence, the expression for q is indeed independent of the choice of n. Finally,

$$\begin{aligned}\big(\langle q|M^{(m)}\big)_v \;&=\; \sum_{w\in\mathcal{L}_m(j)} q_w\,\mathbb{P}[\vartheta(w)_{[1,m]} = v] \;=\; \sum_{w\in\mathcal{L}_m(j)} \sum_{b\in\mathcal{R}_j} \pi_b^j\,\mathbb{P}[\vartheta^n(b)_{[1,m]} = w]\,\mathbb{P}[\vartheta(w)_{[1,m]} = v] \\ &=\; \sum_{b\in\mathcal{R}_j} \pi_b^j\,\mathbb{P}[\vartheta^{n+1}(b)_{[1,m]} = v] \;=\; q_v\end{aligned}$$

and the claim follows. □

In the following, we make the additional assumption that ϑ is primitive. Our next aim is to show that, given a ϑ-invariant measure ν on $\mathbb{X}_\vartheta^+$, the word frequencies are well-defined ν-almost surely. For $u \in \mathcal{L}$ and $x \in \mathbb{X}_\vartheta^+$, we define

$$f_u(x) \;=\; \lim_{n\to\infty} \frac{|x_{[1,n]}|_u}{n},$$

provided the limit exists. Similarly, we set $f_u(v) = |v|_u/|v|$ for $v \in \mathcal{A}^+$. It is a classic result that the word frequencies exist almost surely for the stochastic process $(\vartheta^n(a))$. This is a corollary of [Pey80, Thm. 1], compare also [GS20, Prop. 4.3].

Fact 4.1.8. *Let ϑ be a primitive random substitution. Then, for every $u \in \mathcal{L}(\mathbb{X}_\vartheta)$, there exists a word frequency $f_u \in (0,1)$ such that for all $a \in \mathcal{A}$,*

$$\lim_{n\to\infty} f_u(\vartheta^n(a)) \;=\; f_u,$$

holds $\mathbb{P}$-almost surely. ■

Due to the assumed ϑ-invariance of ν, we can decompose sequences x in the support of ν into level-n inflation words for arbitrary levels. Combining this idea with the Borel–Cantelli lemma, we can translate the $\mathbb{P}$-almost sure existence of letter frequencies to a ν-almost sure existence. A first step into this direction is the following [GS20, Lemma 4.2].

Lemma 4.1.9. *Let ϑ be a primitive random substitution, $u \in \mathcal{L}(\mathbb{X}_\vartheta)$ and $\varepsilon > 0$. Then, there exist $k \in \mathbb{N}$, $c > 0$ and $n_0 \in \mathbb{N}$ with the following property. If $v \in \mathcal{A}^n$ for some $n \geqslant n_0$, then*

$$\mathbb{P}\big[\big|f_u\big(\vartheta^k(v)\big) - f_u\big| > \varepsilon\big] \;\leqslant\; \mathrm{e}^{-cn}.$$

Sketch of proof. For the full, somewhat technical proof, we refer to [GS20]. To convey some basic ideas, we sketch a proof for the special case that ϑ is of constant length ℓ. For large enough k, an occurrence of u in $\vartheta^k(v)$ is either contained in $\vartheta^k(v_j)$ for some $1 \leqslant j \leqslant n$ or appears at the boundary of two such inflation words. For the latter, there are at most $(n-1)(|u|-1)$ possibilities. Hence, for $\varepsilon < |f_u(\vartheta^k(v)) - f_u|$, we require

$$\varepsilon n \ell^k \;<\; \bigl||\vartheta^k(v)|_u - n\ell^k f_u\bigr| \;\leqslant\; |u|n + \sum_{j=1}^{n} \bigl||\vartheta^k(v_j)|_u - \ell^k f_u\bigr|. \tag{4.1}$$

For large enough k, $|u|n \leqslant \varepsilon n \ell^k/2$, and (4.1) implies

$$\frac{\varepsilon}{2} n \;\leqslant\; \sum_{j=1}^{n} |f_u(\vartheta^k(v_j)) - f_u|. \tag{4.2}$$

For each $a \in \mathcal{A}$, let $P_a \subset \{1, \dots, n\}$ be the set of positions j with $u_j = a$ and $n_a = \#P_a$. By (4.2), there exists a constant $C > 0$ and a letter $a \in \mathcal{A}$ with $n_a \geqslant Cn$ such that

$$\frac{\varepsilon}{4} \;\leqslant\; \frac{1}{n_a} \sum_{j \in P_a} |f_u(\vartheta^k(v_j)) - f_u|. \tag{4.3}$$

For large enough k, the expected value of $f_u(\vartheta^k(v_j))$ differs from f_u by less than $\varepsilon/8$ due to Fact 4.1.8. Hence, the probability for (4.3) decays exponentially with $n \in \mathbb{N}$ by Cramér's theorem on large deviations [dHo00, Thm. I.4]. As soon as we drop the assumption that ϑ is of constant length, we additionally need some control on the relative sizes of level-k inflation words. Once again, we can use large deviation results to find an exponentially decaying bound for the probability of 'exceptional behaviour'. □

Theorem 4.1.10. *Let ϑ be a primitive random substitution and $\nu \in \mathcal{M}^1(\mathbb{X}_\vartheta^+)$, ϑ^r-invariant for some $r \in \mathbb{N}$. Then, for ν-almost every $x \in \mathbb{X}_\vartheta^+$, we have $f_u(x) = f_u$, for all $u \in \mathcal{L}(\mathbb{X}_\vartheta)$.*

Proof. Without loss of generality, we assume that $r = 1$. By the Borel–Cantelli lemma, it suffices to show that, for every $\varepsilon > 0$ and $u \in \mathcal{A}^+$,

$$\sum_{m \in \mathbb{N}} \nu\bigl(\underbrace{\{x : |f_u(x_{[1,m]}) - f_u| > 2\varepsilon\}}_{=:A_m}\bigr) \;<\; \infty.$$

Let $k, n_0 \in \mathbb{N}$ and $c > 0$ as in Lemma 4.1.9. Since ν is ϑ^k-invariant, we have

$$\nu(A_m) \;=\; \vartheta^k(\nu)(A_m) \;=\; \sum_{v \in \mathcal{A}^m} \nu([v])\, \mathbb{P}\bigl[\bigl|f_u\bigl(\vartheta^k(v)_{[1,m]}\bigr) - f_u\bigr| > 2\varepsilon\bigr].$$

Note that $\vartheta^k(v)_{[1,m]}$ only depends on the smallest prefix v' of v such that $|\vartheta^k(v')| \geqslant m$. The length of this prefix is a random variable. In any case, with $\mu = \max\{|u| : u \in \vartheta(a), a \in \mathcal{A}\}$, we have $|v'| \geqslant m/\mu^k$. By the minimality of v', $m \leqslant |\vartheta^k(v')| < m + \mu^k$. For large enough m, the term μ^k becomes negligible and hence,

$$\bigl|f_u\bigl(\vartheta^k(v)_{[1,m]}\bigr) - f_u\bigr| \;>\; 2\varepsilon \;\implies\; \bigl|f_u\bigl(\vartheta^k(v')\bigr) - f_u\bigr| \;>\; \varepsilon.$$

For $m \geqslant n_0\mu^k$, the probability for this event is bounded by e^{-cm/μ^k}, due to Lemma 4.1.9. Finally, this implies

$$\nu(A_m) \leqslant \sum_{v\in\mathcal{A}^m} \nu([v])\,\mathrm{e}^{-cm/\mu^k} \leqslant \mathrm{e}^{-cm/\mu^k},$$

and the assertion follows. □

Proposition 4.1.11. *Let ϑ be a primitive random substitution, $r \in \mathbb{N}$ and ν a Borel probability measure on $\mathbb{X}_\vartheta$ that restricts to a ϑ^r-invariant measure on $\mathbb{X}_\vartheta^+$. Then, the weak limit*

$$\mu_\vartheta \,=\, \lim_{n\to\infty} \frac{1}{n} \sum_{j=0}^{n-1} \nu \circ S^{-j},$$

exists and defines an S-invariant measure $\mu_\vartheta \in \mathcal{M}^1(\mathbb{X}_\vartheta)$ with $\mu_\vartheta([u]) = f_u$ for all $u \in \mathcal{L}(\mathbb{X}_\vartheta)$. In particular, μ_ϑ has full support on $\mathbb{X}_\vartheta$.

Proof. Since $\mathcal{M}^1(\mathbb{X}_\vartheta)$ is compact in the weak topology, the sequence $(\nu_n)_{n\in\mathbb{N}}$, defined by

$$\nu_n \,=\, \frac{1}{n} \sum_{j=0}^{n-1} \nu \circ S^{-j}$$

has an accumulation point $\mu \in \mathcal{M}^1(\mathbb{X}_\vartheta)$. By construction, this accumulation point is S-invariant. It therefore suffices to show that $\lim_{n\to\infty} \nu_n([u]) = f_u$ for all $u \in \mathcal{L}$. For $u \in \mathcal{L}_m$ and $j \in \mathbb{N}_0$, we have

$$\nu(S^{-j}[u]) \,=\, \nu([x_{[j+1,j+m]} = u]) \,=\, \mathbb{E}_\nu\big[|x_{[j+1,j+m]}|_u\big],$$

for a ν-distributed random variable x on $\mathbb{X}_\vartheta$. Hence,

$$\begin{aligned}\lim_{n\to\infty} \nu_n([u]) \,&=\, \lim_{n\to\infty} \frac{1}{n} \sum_{j=0}^{n-1} \mathbb{E}_\nu\big[|x_{[j+1,j+m]}|_u\big] \,=\, \lim_{n\to\infty} \mathbb{E}_\nu\left[\frac{|x_{[1,n+m-1]}|_u}{n}\right] \\ &=\, \mathbb{E}_\nu\Big[\lim_{n\to\infty} f_u\big(x_{[1,n]}\big)\Big] \,=\, f_u,\end{aligned}$$

where the penultimate step follows by dominated convergence and the last step is due to Theorem 4.1.10. Since $f_u > 0$ for all $u \in \mathcal{L}$ by Fact 4.1.8, it follows that the topological support of μ_ϑ equals $\mathbb{X}_\vartheta$. □

Remark 4.1.12. It should be stressed that the measure μ_ϑ depends neither on the choice of the ϑ^r-invariant measure ν on $\mathbb{X}_\vartheta^+$, nor on its extension to $\mathbb{X}_\vartheta$. Due to the existence of word frequencies, we also obtain μ_ϑ from ν-almost every $x' \in \mathbb{X}_\vartheta^+$ and an arbitrary extension to $x \in \mathbb{X}_\vartheta$, by

$$\mu_\vartheta \,=\, \lim_{n\to\infty} \frac{1}{n} \sum_{j=0}^{n-1} \delta_x \circ S^{-j}.$$

An alternative approach is to *define* $\mu_\vartheta([u]) = f_u$ for all $u \in \mathcal{L}$ and to verify that this extends consistently and uniquely to a shift-invariant probability measure; see [GS20, Mol13] for details. ◊

For a certain family of primitive random substitutions, it was announced in [Mol13] that the corresponding measures μ_ϑ are not only S-invariant but in fact ergodic. However, the original proof contains a small gap. In [GS20], ergodicity of μ_ϑ was shown for all primitive random substitutions. At this point, we get the same result, basically for free.

Theorem 4.1.13. *Let ϑ be a primitive random substitution. Then, the dynamical system $(\mathbb{X}_\vartheta, S, \mu_\vartheta)$ is ergodic.*

Proof. This is a corollary of Proposition 4.1.11 and the proof of Theorem 4.1.10. To establish S-ergodicity of μ_ϑ, it suffices to show that, for μ_ϑ-almost every $x \in \mathbb{X}_\vartheta$, the word frequencies $f_u(x)$ exist and are equal to $\mu_\vartheta[u]$, for all $u \in \mathcal{L}$; compare [Mol13, Prop. 4.21] and [GS20]. Again, this statement on the word frequencies follows by the Borel–Cantelli lemma if, for every $\varepsilon > 0$ and $u \in \mathcal{A}^+$,

$$\sum_{m\in\mathbb{N}} \mu_\vartheta\bigl(\underbrace{\{x : |f_u(x_{[1,m]}) - f_u| > 2\varepsilon\}}_{=:A_m}\bigr) \,<\, \infty.$$

By construction,

$$\mu_\vartheta(A_m) \,=\, \lim_{n\to\infty} \frac{1}{n} \sum_{j=0}^{n-1} \nu(S^{-j}A_m).$$

A slight modification of the proof of Theorem 4.1.10 shows that there are $k, n_0 \in \mathbb{N}$ and $c, \mu > 0$ such that $\nu(S^{-j}A_m) \leqslant \mathrm{e}^{-cm/\mu^k}$, for all $j \in \mathbb{N}_0$ and large enough $m \geqslant n_0\mu^k$. This implies $\mu_\vartheta(A_m) \leqslant \mathrm{e}^{-cm/\mu^k}$ for large enough $m \in \mathbb{N}$, and the claim follows. □

Remark 4.1.14. For a primitive random substitution ϑ and $m \in \mathbb{N}$, we can compute f_u for all $u \in \mathcal{L}_m$, using a finite algorithm. This employs the concept of an *induced* random substitution ϑ_m, defined on $\mathcal{L}_m$ as the corresponding alphabet. There is a primitive substitution matrix M_m associated to ϑ_m with (normalized) PF eigenvector R_m. The values f_u are precisely the entries of R_m. We refer to [GS20, Mol13, Kos12] for details. For deterministic substitutions, this construction is classic [BG13, Que10]. ◊

Ergodicity is in fact the strongest measure-theoretic mixing result that we can expect to hold in such generality for primitive random substitutions. This is because there are primitive random substitutions with a non-trivial pure point part in the diffraction measure, which excludes weak mixing [Mol13, Goh17, BSS18]. We will explore the diffraction measures associated to primitive random substitutions in Section 4.3.

4.2. Entropy

4.2.1. Topological entropy

To obtain general, yet useful, estimates of the entropy associated with a random substitution system requires some control over the random substitution at hand. The task becomes considerably easier if letter frequencies are well-defined and if the powerful machinery of PF theory is available. To avoid being overly repetitive, we fix some notation for the remainder of this section.

Assumption 4.2.1. Let ϑ be a primitive, compatible random substitution. We write M for its substitution matrix and λ for the corresponding PF eigenvalue. The left and right PF eigenvectors of M are denoted by L and R, with $\langle L|R\rangle = ||R||_1 = 1$.

Since topological entropy makes no reference to a measure on $(\mathbb{X}_\vartheta, S)$, we regard ϑ as a set-valued substitution throughout this section. Recall from Section 2.3.2 that the topological entropy of $(\mathbb{X}_\vartheta, S)$ can be calculated by

$$s_\vartheta := h_{\mathrm{top}}(S) = \lim_{n\to\infty} \frac{\log \#\mathcal{L}_n}{n}.$$

That is, we are looking for the exponential growth rate of the set of legal words, as their length increases. On the set of legal words, we unfortunately lack an immediate recursive structure. This is different for the set of *inflation words*. Indeed, we obtain

$$\vartheta^{n+k}(a) = \vartheta^n(\vartheta^k(a)) = \bigcup_{v\in\vartheta^k(a)} \vartheta^n(v), \tag{4.4}$$

for all $k, n \in \mathbb{N}$ and $a \in \mathcal{A}$. For the cardinality of these sets, it is important to assess whether the union in (4.4) is disjoint or contains overlaps. We give a name to the two extremal cases.

Definition 4.2.2. For $k, n \in \mathbb{N}$, we say that ϑ satisfies the *identical set condition of type* (k, n), denoted by $(k$,ISC,$n)$, if

$$u, v \in \vartheta^k(a) \implies \vartheta^n(u) = \vartheta^n(v),$$

for all $a \in \mathcal{A}$. We say that ϑ satisfies the *disjoint set condition of type* (k, n), denoted by $(k$,DSC,$n)$, if

$$u, v \in \vartheta^k(a) \text{ and } u \neq v \implies \vartheta^n(u) \cap \vartheta^n(v) = \varnothing,$$

for all $a \in \mathcal{A}$. Finally, we say that (ISC) [or (DSC)] holds for ϑ, if (1,ISC,n) [or (1,DSC,n)] holds for all $n \in \mathbb{N}$, respectively.

The conditions (ISC) and (DSC) are important because they will enable us to give an explicit formula for the topological entropy s_ϑ.

Example 4.2.3. If ϑ is of constant length, a sufficient condition for (DSC) is $\vartheta(a)\cap\vartheta(b) = \varnothing$, for all $a, b \in \mathcal{A}$ with $a \neq b$ [Goh20, Cor. 20]. This applies in particular to the set-valued period doubling substitution $\vartheta\colon a \mapsto \{ab, ba\}$, $b \mapsto \{aa\}$ from Example 3.2.7. That disjointness of the inflation sets $\vartheta(a)$ and $\vartheta(b)$ is not *necessary* for (DSC) can be seen from the example

$$\vartheta\colon a, b \mapsto \{cac, cca\},\ c \mapsto \{abb\}.$$

Here, $\vartheta^n(a) = \vartheta^n(b)$, but $\vartheta^n(a) \cap \vartheta^n(c) = \emptyset$, for all $n \in \mathbb{N}$. The latter observation also suffices to conclude that $\vartheta^n(cac)$ and $\vartheta^n(cca)$ are disjoint for all $n \in \mathbb{N}$, and hence, (DSC) holds.

Example 4.2.4. The set-valued Fibonacci substitution $\vartheta\colon a \mapsto \{ab, ba\}$, $b \mapsto \{a\}$ from Example 3.2.3 satisfies neither (DSC) nor (ISC). This is because the sets $\vartheta(ab) = \{aba, baa\}$ and

$\vartheta(ba) = \{aab, aba\}$ have a non-trivial intersection. Therefore, both (1,DSC,1) and (1,ISC,1) fail to hold. A slight adaptation, given by

$$\vartheta\colon a, c \mapsto \{ab, ba\},\ b \mapsto \{c\},$$

satisfies (1,DSC,1) but not (1,DSC,2).

Remark 4.2.5. The different types of conditions, listed in Definition 4.2.2, are in fact not independent. For instance, it is a straightforward exercise to show that (1,ISC,1) implies $(k$,ISC,$n)$ for all $k, n \in \mathbb{N}$. Hence, (ISC) is particularly easy to check. Another consequence is that, if ϑ satisfies (ISC), so does ϑ^n for all $n \in \mathbb{N}$.
For (DSC), the situation is in some sense reversed: For all $k, n \in \mathbb{N}$, $(k$,DSC,$n)$ also implies (1,DSC,1). The converse is not true in general, as we can see from Example 4.2.4. With some tedious, yet straightforward work, we can show that (DSC) holds for either all or none of $\{\vartheta^n\}_{n\in\mathbb{N}}$. ◊

Definition 4.2.6. Given $n \in \mathbb{N}$, let θ^n denote the *logarithmic cardinality vector* on $\mathcal{A}$, with entries

$$\theta^n_a \;=\; \log \#\vartheta^n(a),$$

for all $a \in \mathcal{A}$.

The logarithmic cardinality of $\vartheta^n(v)$ for general words $v \in \mathcal{A}^+$ can be obtained from θ^n in a straightforward manner. Indeed, using compatibility, we verify that

$$\log \#\vartheta^n(v) \;=\; \log \prod_{j=1}^{|v|} \#\vartheta^n(v_j) \;=\; \sum_{j=1}^{|v|} \theta^n_{v_j} \;=\; \langle \theta^n | \Phi(v) \rangle, \tag{4.5}$$

for all $n \in \mathbb{N}$ and $v \in \mathcal{A}^+$.

Lemma 4.2.7. *Let $k, n \in \mathbb{N}$. Then,*

$$\langle \theta^n | M^k \;\leqslant\; \langle \theta^{n+k}| \;\leqslant\; \langle \theta^n | M^k + \langle \theta^k|,$$

to be understood elementwise. The lower bound is an equality if and only if ϑ satisfies $(k$,ISC,$n)$. The upper bound is an equality if and only if ϑ satisfies $(k$,DSC,$n)$.

Proof. For $a \in \mathcal{A}$ and $v \in \vartheta^k(a)$, we have by (4.5),

$$\log \#\vartheta^n(v) = \langle \theta^n | \Phi(v) \rangle = \langle \theta^n | M^k | e_a \rangle.$$

In particular, all of the sets $\vartheta^n(v)$ with $v \in \vartheta^k(a)$ have the same cardinality. Hence, taking cardinalities in (4.4), we find

$$\langle \theta^{n+k} | e_a \rangle \;=\; \log \#\vartheta^{n+k}(a) \;\geqslant\; \log \#\vartheta^n(v) \;=\; \langle \theta^n | M^k | e_a \rangle,$$

with equality precisely if all of the sets $\vartheta^n(v)$ with $v \in \vartheta^k(a)$ coincide. Similarly,

$$\langle \theta^{n+k} | e_a \rangle \;=\; \log \#\vartheta^{n+k}(a) \;\leqslant\; \log\big(\#\vartheta^k(a) \#\vartheta^n(v)\big) \;=\; \langle \theta^n | M^k | e_a \rangle + \langle \theta^k | e_a \rangle,$$

with equality precisely if all of the sets $\vartheta^n(v)$ with $v \in \vartheta^k(a)$ are disjoint. □

Proposition 4.2.8. *There is a well-defined limit vector*

$$\theta \;=\; \lim_{k\to\infty} \frac{1}{\lambda^k}\theta^k,$$

which satisfies

$$\frac{\langle\theta^n|R\rangle}{\lambda^n}L \;\leqslant\; \theta \;\leqslant\; \frac{\langle\theta^n|R\rangle}{\lambda^n-1}L, \tag{4.6}$$

for all $n\in\mathbb{N}$. The lower bound is non-decreasing in n.

Proof. We first note that, for all $n\in\mathbb{N}$,

$$\langle\underline{\theta}| \;:=\; \liminf_{k\to\infty}\frac{1}{\lambda^{n+k}}\langle\theta^{n+k}| \;\geqslant\; \liminf_{k\to\infty}\frac{1}{\lambda^n}\langle\theta^n|\frac{1}{\lambda^k}M^k \;=\; \frac{\langle\theta^n|R\rangle}{\lambda^n}\langle L|,$$

due to Lemma 4.2.7. The last expression is non-decreasing in n, since

$$\frac{\langle\theta^{n+1}|R\rangle}{\lambda^{n+1}} \;\geqslant\; \frac{\langle\theta^n|M|R\rangle}{\lambda^{n+1}} \;=\; \frac{\langle\theta^n|R\rangle}{\lambda^n},$$

for all $n\in\mathbb{N}$. By the upper bound in Lemma 4.2.7, we obtain

$$\langle\overline{\theta}| \;:=\; \limsup_{k\to\infty}\frac{\langle\theta^{n+k}|}{\lambda^{n+k}} \;\leqslant\; \limsup_{k\to\infty}\frac{\langle\theta^n|}{\lambda^n}\frac{1}{\lambda^k}M^k + \frac{1}{\lambda^n}\limsup_{k\to\infty}\frac{\langle\theta^k|}{\lambda^k} \;=\; \frac{\langle\theta^n|R\rangle}{\lambda^n}L + \frac{1}{\lambda^n}\langle\overline{\theta}|,$$

and hence,

$$\overline{\theta} \;\leqslant\; \frac{\langle\theta^n|R\rangle}{\lambda^n-1}L.$$

Since the upper and the lower bound in (4.6) converge to the same (positive) vector as $n\to\infty$, we conclude that $\underline{\theta}=\overline{\theta}$, and the limit is well-defined. □

Corollary 4.2.9. *We have $\theta = sL$, with*

$$s \;:=\; \lim_{n\to\infty}\frac{\langle\theta^n|R\rangle}{\lambda^n}.$$

For all $v\in\mathcal{A}^+$, we have

$$s \;=\; \lim_{n\to\infty}\frac{\#\vartheta^n(v)}{|\vartheta^n(v)|}.$$

Proof. The first claim is an immediate consequence of Proposition 4.2.8. For the second claim, let $v\in\mathcal{A}^+$ and recall from (4.5) that $\#\vartheta^n(v)=\langle\theta^n|\Phi(v)\rangle$. By the PF theorem,

$$\lim_{n\to\infty}\frac{|\vartheta^n(v)|}{\lambda^n} \;=\; \sum_{a\in\mathcal{A}} L_a|v|_a \;=\; \langle L|\Phi(v)\rangle.$$

Combined with Proposition 4.2.8, this yields

$$\lim_{n\to\infty}\frac{\#\vartheta^n(v)}{|\vartheta^n(v)|} \;=\; \lim_{n\to\infty}\frac{\lambda^{-n}\langle\theta^n|\Phi(v)\rangle}{\lambda^{-n}|\vartheta^n(v)|} \;=\; \frac{\langle\theta|\Phi(v)\rangle}{\langle L|\Phi(v)\rangle} \;=\; s,$$

and the claim follows. □

We show that this value s is indeed the topological entropy of the dynamical system $(\mathbb{X}_\vartheta, S)$, providing an improvement of the first part of Theorem 4.0.1, stated in the introduction of this chapter.

Theorem 4.2.10. *Let ϑ be a primitive compatible random substitution. The topological entropy of $(\mathbb{X}_\vartheta, S)$ is given by*

$$s_\vartheta \,=\, \lim_{n\to\infty} \frac{\langle \theta^n | R\rangle}{\lambda^n} \,=\, \lim_{n\to\infty} \frac{1}{\lambda^n} \sum_{a\in\mathcal{A}} R_a \, \#\vartheta^n(a).$$

For all $n \in \mathbb{N}$, it satisfies

$$\frac{\langle \theta^n | R\rangle}{\lambda^n} \,\leqslant\, s_\vartheta \,\leqslant\, \frac{\langle \theta^n | R\rangle}{\lambda^n - 1}. \tag{4.7}$$

The lower bound is an equality if and only if ϑ^n satisfies (DSC)*, while the upper bound is an equality if and only if ϑ^n satisfies* (ISC)*.*

Proof. By Corollary 4.2.9, the first claim follows from the second. Since ϑ is primitive, we have $s_\vartheta = s_{\vartheta^n}$, for all $n \in \mathbb{N}$, and hence it suffices to show the second claim for $n = 1$. For an arbitrary $a \in \mathcal{A}$ and $m \in \mathbb{N}$, the legal words of length $|\vartheta^m(a)|$ contain $\vartheta^m(a)$ and hence

$$s_\vartheta \,\geqslant\, \lim_{m\to\infty} \frac{\#\vartheta^m(a)}{|\vartheta^m(a)|} \,\geqslant\, \frac{\langle \theta^1 | R\rangle}{\lambda},$$

again by Corollary 4.2.9 and Proposition 4.2.8. For the upper bound, we use that large words can be decomposed into inflation words. Let $\varepsilon > 0$. In Lemma 3.2.8 we have seen that letter frequencies exist uniformly in $\mathcal{L}$. Hence, there is a number $n_0 \in \mathbb{N}$ such that for all $n \geqslant n_0$ and $u \in \mathcal{L}_n$, we have $|f_a(u) - R_a| < \varepsilon$ for all $a \in \mathcal{A}$. By a similar argument, we can assume that $|\vartheta(u)| \geqslant n(\lambda - \varepsilon)$ for all $u \in \mathcal{L}_n$. Let $m_n = \lfloor n(\lambda - \varepsilon)\rfloor - \ell$, where $\ell = \max_{a\in\mathcal{A}} |\vartheta(a)|$. Then, for every $w \in \mathcal{L}_{m_n}$ there is a $v = v_1 \cdots v_n \in \mathcal{L}_n$ such that w is a subword of some word in $\vartheta(v)$, overlapping the first inflation word $\vartheta(v_1)$. This implies

$$\#\mathcal{L}_{m_n} \,\leqslant\, \ell \#\vartheta(\mathcal{L}_n) \,\leqslant\, \ell \sum_{u\in\mathcal{L}_n} \#\vartheta(u).$$

By (4.5) and the assumption on letter frequencies, we have for $u \in \mathcal{L}_n$ that

$$\log \#\vartheta(u) \,=\, \langle \theta^1 | \Phi(u)\rangle \,\leqslant\, n\langle \theta^1 | R\rangle + n\varepsilon \sum_{a\in\mathcal{A}} \#\vartheta(a) \,=\, n\big(\langle \theta^1 | R\rangle + \varepsilon c\big),$$

for some constant c that is independent of ε and n. We obtain,

$$\log \#\mathcal{L}_{m_n} \,\leqslant\, \log\Big(\ell e^{n(\langle \theta^1 | R\rangle + \varepsilon c)} \#\mathcal{L}_n\Big) \,=\, \log(\ell) + n\big(\langle \theta^1 | R\rangle + \varepsilon c\big) + \log(\#\mathcal{L}_n).$$

Finally, this yields

$$s_\vartheta \,=\, \lim_{n\to\infty} \frac{\log \#\mathcal{L}_{m_n}}{m_n} \,\leqslant\, \frac{\langle \theta^1 | R\rangle + \varepsilon c + s_\vartheta}{\lambda - \varepsilon} \,\xrightarrow{\varepsilon\to 0}\, \frac{\langle \theta^1 | R\rangle}{\lambda} + \frac{s_\vartheta}{\lambda},$$

and hence

$$s_\vartheta \,\leqslant\, \frac{\langle \theta^1 | R\rangle}{\lambda - 1}.$$

If ϑ satisfies (DSC), we have, due to Lemma 4.2.7,

$$\langle \theta^m | R \rangle \;=\; \langle \theta^1 | R \rangle + \lambda \langle \theta^{m-1} | R \rangle = \ldots \;=\; \sum_{k=0}^{m-1} \lambda^k \langle \theta^1 | R \rangle \;=\; \frac{\lambda^m - 1}{\lambda - 1} \langle \theta^1 | R \rangle,$$

for all $m \in \mathbb{N}$, and thereby,

$$s_\vartheta \;=\; \lim_{m \to \infty} \frac{\langle \theta^m | R \rangle}{\lambda^m - 1} \;=\; \frac{\langle \theta^1 | R \rangle}{\lambda - 1}.$$

If ϑ does not satisfy (DSC), there is $m \in \mathbb{N}$ with

$$s_\vartheta \;\leqslant\; \frac{\langle \theta^m | R \rangle}{\lambda^m - 1} \;<\; \frac{\langle \theta^1 | R \rangle}{\lambda - 1}.$$

The claim on the lower bound being an equality precisely if (ISC) holds for ϑ, follows in a similar manner. □

Remark 4.2.11. As we discussed in Remark 4.2.5, (DSC) holds either for all or for none of the powers of ϑ. Hence, either s_ϑ is equal to the upper bound in (4.6) for all $n \in \mathbb{N}$, or it is strictly smaller than the upper bound for all $n \in \mathbb{N}$. For the lower bound, we do not have a similar dichotomy. It is possible to construct an example such that ϑ does not satisfy (ISC), but ϑ^2 does. Comparing with Remark 4.2.5, we observe that (DSC) is in general more difficult to check, but is better behaved under taking powers. ◊

In general, it is not easy to determine whether the subshift $(\mathbb{X}_\vartheta, S)$ contains S-periodic sequences; some criteria were established in [Rus20]. If periodic sequences exist, they are necessarily dense in the subshift [RS18]. In that case, we will show that they can also be used to recover the topological entropy. This is similar to the situation for SFTs [LM95]. In fact, every topologically transitive SFT can be written as a primitive random substitution subshift [GRS19]. Given a length $n \in \mathbb{N}$, we write

$$\mathcal{P}(n) \;=\; \{v \in \mathcal{L}_n \mid v^{\mathbb{Z}} \in \mathbb{X}_\vartheta\},$$

for the set of *periodic seeds* of length n.

Proposition 4.2.12. *Assume that $(\mathbb{X}_\vartheta, S)$ contains an S-periodic sequence. Then,*

$$\limsup_{n \to \infty} \frac{\log \# \mathcal{P}(n)}{n} \;=\; s_\vartheta.$$

Proof. If $\mathbb{X}_\vartheta$ contains an S-periodic sequence, there are $q \in \mathbb{N}$ and $v \in \mathcal{P}(q) \neq \varnothing$. By construction, we have $\vartheta^m(v) \subset \mathcal{P}(|\vartheta^m(v)|)$ for all $m \in \mathbb{N}$. Hence,

$$\limsup_{m \to \infty} \frac{\log \# \mathcal{P}(|\vartheta^m(v)|)}{|\vartheta^m(v)|} \;\geqslant\; \lim_{m \to \infty} \frac{\log \# \vartheta^m(v)}{|\vartheta^m(v)|} \;=\; s \;=\; s_\vartheta,$$

by Corollary 4.2.9 and Theorem 4.2.10. The opposite inequality is immediate since $\mathcal{P}(n) \subset \mathcal{L}_n$ for all $n \in \mathbb{N}$. □

Example 4.2.13. As we have seen in Example 4.2.3, the condition (DSC) is fulfilled by the set-valued period doubling substitution $\vartheta\colon a \mapsto \{ab, ba\}$, $b \mapsto \{aa\}$. The right PF eigenvector is easily determined as $R = (2/3, 1/3)$, and we obtain $\theta^1 = (\log(2), 0)$. Hence, by Theorem 4.2.10, the topological entropy is given by

$$s_\vartheta \,=\, \langle \theta^1 | R \rangle = \frac{2}{3} \log(2).$$

This was already obtained in [BSS18], using different methods. ◊

For more examples and a useful sufficient condition for (DSC) that builds on previous work of Rust [Rus20], we refer to [Goh20].

4.2.2. Measure-theoretic entropy

We still adopt the notation from Assumption 4.2.1. In particular, ϑ denotes a primitive compatible random substitution throughout this section. However, we now regard ϑ as a genuinely random object as introduced in Section 3.2.2 and equip the subshift $(\mathbb{X}_\vartheta, S)$ with the ergodic measure μ_ϑ from Proposition 4.1.11. The measure-theoretic entropy of the dynamical system $(\mathbb{X}_\vartheta, S, \mu_\vartheta)$ is given by

$$h_\vartheta \,:=\, h_{\mu_\vartheta}(S) \,=\, \lim_{n\to\infty} -\frac{1}{n} \sum_{u \in \mathcal{L}_n} \mu_\vartheta([u]) \log \mu_\vartheta([u]) \,=\, \lim_{n\to\infty} \frac{1}{n} \sum_{u \in \mathcal{L}_n} \varphi(\mu_\vartheta([u])),$$

where we have used the notation $\varphi(x) = -x \log x$, which defines a non-negative, concave function on $[0, 1]$. Unfortunately, the exact calculation of $\mu_\vartheta([u]) = f_u$ for long words u is computationally very expensive, even with the algorithm presented in Remark 4.1.14. It is therefore of practical use to find an algorithm for the computation of h_ϑ that works with the Markov process $(\vartheta^n(a))_{n\in\mathbb{N}}$, with $a \in \mathcal{A}$, instead. The construction in this section is inspired by our approach on topological entropy via inflation words, and many of our previous results find their complement in the measurable setting. Instead of taking log-cardinalities of the inflation sets $\vartheta^n(a)$, we consider the entropy of the probability distribution, defined by the random variable $\vartheta^n(a)$. More precisely, we consider

$$H\big(\vartheta^n(a)\big) \,:=\, H_{\mathbb{P}}\big(\vartheta^n(a)\big) \,=\, \sum_{v \in \mathcal{L}} \varphi\big(\mathbb{P}[\vartheta^n(a) = v]\big),$$

which is the entropy of the measurable partition $(\vartheta^n(a))^{-1}\mathcal{A}^+$ of the probability space $(\Omega_a, \mathbb{P}_a)$; compare the discussion in Remark 2.3.3. Note that this is identical to the entropy of the finite probability vector $P_a^n = (P_{a,v}^n)_{v\in\vartheta^n(a)}$. We have $H(\vartheta^n(a)) \leqslant \log \#\vartheta^n(a)$, with equality precisely if P_a^n is a uniform distribution on the set $\vartheta^n(a)$.

Definition 4.2.14. For $n \in \mathbb{N}$, we write H^n for the vector on $\mathcal{A}$, with $H_a^n = H(\vartheta^n(a))$, for all $a \in \mathcal{A}$.

By compatibility, the random variables $\vartheta^n(v_1 \cdots v_{|v|})$ and $(\vartheta^n(v_1), \ldots, \vartheta^n(v_{|v|}))$ induce the same probability vector, up to a renaming of indices. Since neighbouring letters are mapped

independently, we obtain that the entropy functional acts additively, due to Fact 2.3.2, that is,

$$H\big(\vartheta^n(v)\big) \,=\, \sum_{j=1}^{|v|} H\big(\vartheta^n(v_j)\big) \,=\, \langle H^n | \Phi(v)\rangle,$$

for all $n \in \mathbb{N}$ and $v \in \mathcal{A}^+$. Instead of the inflation set recursion (4.4), we get, for all $n, k \in \mathbb{N}$,

$$\begin{aligned} H\big(\vartheta^{n+k}(a)|\vartheta^k(a)\big) \,&=\, \sum_{v \in \vartheta^k(a)} \mathbb{P}\big[\vartheta^k(a) = v\big]\, H\big(\vartheta^n(v)\big) \,=\, \mathbb{E}\big[\big\langle H^n \big| \Phi(\vartheta^k(a))\big\rangle\big] \\ &=\, \langle H^n | M^k | e_a \rangle, \end{aligned} \tag{4.8}$$

where we have used the homogeneity of the Markov process $(\vartheta^n(a))_{n\in\mathbb{N}}$ in the first step.

Lemma 4.2.15. *For all $n, k \in \mathbb{N}$, we have*

$$\langle H^n | M^k \,\leqslant\, \langle H^{n+k}| \,\leqslant\, \langle H^n | M^k + \langle H^k|.$$

The upper bound is an equality if and only if (k,DSC,n) *holds. The lower bound is an equality if and only if* (k,ISC,n) *holds and P_v^n is the same vector for all $v \in \vartheta^k(a)$.*

Proof. By the general properties of (conditional) entropies in Fact 2.3.2,

$$H\big(\vartheta^{n+k}(a)|\vartheta^k(a)\big) \,\leqslant\, H\big(\vartheta^{n+k}(a)\big) \,\leqslant\, H\big(\vartheta^{n+k}(a)|\vartheta^k(a)\big) + H\big(\vartheta^k(a)\big),$$

for all $n, k \in \mathbb{N}$ and $a \in \mathcal{A}$. The lower bound is an equality if and only if $\vartheta^k(a)$ and $\vartheta^{n+k}(a)$ are independent. That is only possible if P_v^n is the same vector for all $v \in \vartheta^k(a)$. The upper bound is an equality if and only if $\vartheta^k(a)$ is completely determined by $\vartheta^{n+k}(a)$. This is a reformulation of (k,DSC,n). The claim now follows from (4.8). □

The result in Lemma 4.2.15 motivates the following modification of the identical set condition.

Definition 4.2.16. We say that ϑ satisfies (ISC) with *identical production probabilities* if for all $a \in \mathcal{A}$ and $n \in \mathbb{N}$,

$$u, v \in \vartheta(a) \;\implies\; P_u^n \,=\, P_v^n.$$

The last expression is equivalent to $\mathbb{P}[\vartheta^n(u) = w] = \mathbb{P}[\vartheta^n(v) = w]$, for all $w \in \mathcal{A}^+$.

The following is a slightly more detailed version of our results in [GMRS21]. The last claim motivates the name *measure-theoretic inflation word entropy* for the quantity h.

Proposition 4.2.17. *There is a well-defined limit vector*

$$H \,=\, \lim_{m\to\infty} \frac{1}{\lambda^m} H^m \,=\, hL,$$

where $h = \lim_{m\to\infty} \lambda^{-m} \langle H^m | R\rangle$ satisfies

$$\frac{\langle H^n | R\rangle}{\lambda^n} \,\leqslant\, h \,\leqslant\, \frac{\langle H^n | R\rangle}{\lambda^n - 1},$$

for all $n \in \mathbb{N}$. The lower bound is an equality if and only if ϑ^n satisfies (ISC) *with identical production probabilities. The upper bound is an equality if and only if ϑ^n satisfies* (DSC)*. Further, for all $v \in \mathcal{A}^+$,*

$$h = \lim_{n\to\infty} \frac{H(\vartheta^n(v))}{|\vartheta^n(v)|}.$$

Proof. Note that the recursion for the sequence $(H^n)_{n\in\mathbb{N}}$ in Lemma 4.2.15 and the recursion for the sequence $(\theta^n)_{n\in\mathbb{N}}$ in Lemma 4.2.7 are precisely the same. The proofs of Proposition 4.2.8 and Corollary 4.2.9 (and the last part of the proof of Theorem 4.2.10) therefore carry over mutatis mutandis. □

Theorem 4.2.18 ([GMRS21])**.** *Let ϑ be a primitive compatible random substitution. The entropy of the measure-theoretic dynamical system $(\mathbb{X}_\vartheta, S, \mu_\vartheta)$ satsifies $h_\vartheta = h$, with h as in Proposition* 4.2.17*.*

Sketch of proof. The full details of the proof can be found in [GMRS21]. We try to convey the basic ideas here. For $n \in \mathbb{N}$, let $\mu^{(n)}$ denote the probability vector on $\mathcal{L}_n$, with $\mu_v^{(n)} = \mu_\vartheta([v])$. It can be shown that $\mu^{(\lfloor \lambda n \rfloor)}$ has a similar distribution to $\vartheta(\mu^{(n)})$, up to modifications that are negligible for the entropy calculation in the limit of large n. Hence,

$$h_\vartheta = \lim_{n\to\infty} \frac{1}{\lambda n} H\big(\mu^{\lfloor \lambda n \rfloor}\big) = \lim_{n\to\infty} \frac{1}{\lambda n} H\big(\vartheta(\mu^{(n)})\big).$$

Also, we can assume that for $u \in \mathcal{L}_n$ and $a \in \mathcal{A}$, we have $|u|_a \approx nR_a$ by the uniform existence of letter frequencies. We obtain, regarding u as a $\mu^{(n)}$-distributed random word,

$$H(\vartheta(\mu^{(n)})|\mu^{(n)}) = \mathbb{E}_{\mu^{(n)}}[H(\vartheta(u))] = \mathbb{E}_{\mu^{(n)}}[\langle H^1|\Phi(u)\rangle] \approx n\langle H^1|R\rangle.$$

For the lower bound, this yields

$$h_\vartheta \geqslant \lim_{n\to\infty} \frac{1}{\lambda n} H\big(\vartheta(\mu^{(n)})|\mu^{(n)}\big) = \frac{1}{\lambda}\langle H^1|R\rangle.$$

For the upper bound, we obtain

$$h_\vartheta \leqslant \lim_{n\to\infty} \frac{1}{\lambda n} H\big(\vartheta(\mu^{(n)})|\mu^{(n)}\big) + \frac{1}{\lambda n} H\big(\mu^{(n)}\big) = \frac{1}{\lambda}\langle H^1|R\rangle + \frac{1}{\lambda} h_\vartheta$$

and it follows that

$$\frac{\langle H^1|R\rangle}{\lambda} \leqslant h_\vartheta \leqslant \frac{\langle H^1|R\rangle}{\lambda - 1}.$$

The corresponding identities for arbitrary $n \in \mathbb{N}$ are found by replacing ϑ with ϑ^n. □

Example 4.2.19. As a random substitution, the set-valued period doubling substitution from Example 3.2.3 takes the form

$$\vartheta\colon a \mapsto \begin{cases} ab & \text{with prob. } q, \\ ba & \text{with prob. } 1-q, \end{cases} \qquad b \mapsto aa,$$

with $q \in (0,1)$. The entropy of the random values $\vartheta(a)$ and $\vartheta(b)$ is given by

$$H(\vartheta(a)) = -q\log q - (1-q)\log(1-q),$$

and $H(\vartheta(b)) = 0$, respectively. As we already discussed in Example 4.2.13, ϑ satisfies (DSC). Hence, we obtain from Theorem 4.2.18 that

$$h_\vartheta = -\frac{2}{3}\big(q\log q + (1-q)\log(1-q)\big),$$

which has a unique maximum for $q = 1/2$, in which case $h_\vartheta = s_\vartheta$. That is, μ_ϑ is a measure of maximal entropy on $(\mathbb{X}_\vartheta, S)$ if and only if $q = 1/2$. ◊

4.3. Diffraction

In this section, we still assume that ϑ is a primitive random substitution, but we slightly weaken the compatibility assumption to that of geometric compatibility; compare Definition 3.2.9. Hence, for every sequence $x \in \mathbb{X}_\vartheta$ and $a \in \mathcal{A}$, we have a corresponding (typed) point set $\Lambda_a(x)$, as defined in Section 3.1.2 and Section 2.4.4.

Let $\tau \in \mathbb{C}^{\mathcal{A}}$ denote an arbitrary vector of complex weights. Recall the definition of the weighted Dirac comb $\omega(x) = \sum_{a\in\mathcal{A}} \tau_a \delta_{\Lambda_a(x)}$ from Section 2.4.4. This naturally defines a translation bounded (complex) Radon measure on $\mathbb{R}$. If $x \in \mathcal{A}^{\mathbb{N}}$ is a one-sided sequence, the above quantities are defined analogously and $\omega(x)$ is a Radon measure on $\mathbb{R}^+$. Similarly, for $u \in \mathcal{A}^+$, the left endpoints of the corresponding interval-pattern are given by $\ell_1(u) = 0$ and $\ell_{n+1}(u) = \ell_n(u) + L_{u_n}$, for all $1 \leqslant n \leqslant |u| - 1$, and we define

$$\Lambda_a(u) \,=\, \{\ell_n(u) : u_n = a\},$$

for all $a \in \mathcal{A}$, in obvious analogy to the infinite case. Note that $\omega(u) = \sum_{a\in\mathcal{A}} \tau_a \delta_{\Lambda_a(u)}$ is a finite measure on $\mathbb{R}$ and therefore has a well-defined Fourier transform that is absolutely continuous relative to Lebesgue measure, represented by the density

$$\widehat{\omega(u)}(k) \,=\, \sum_{a\in\mathcal{A}} \tau_a \sum_{r\in\Lambda_a(u)} \mathrm{e}^{-2\pi \mathrm{i} k r}.$$

For notational convenience, we also set $\widehat{\omega}(u) := \widehat{\omega(u)}$ and suppress the explicit dependence on k notationally.

Lemma 4.3.1. *If $(v_n)_{n\in\mathbb{N}}$ is a sequence of words such that $\lim_{n\to\infty} |v_n| \to \infty$, and if, for every $u \in \mathcal{A}^+$, there is a word frequency f_u, such that*

$$\lim_{n\to\infty} \frac{|v_n|_u}{|v_n|} \,=\, f_u,$$

we obtain that the vague limit

$$\gamma \,=\, \lim_{n\to\infty} \gamma(v_n), \quad \textit{with} \quad \gamma(v_n) \,=\, \frac{\omega(v_n) * \widetilde{\omega(v_n)}}{L(v_n)}$$

exists and depends only on $(f_u)_{u\in\mathcal{A}^+}$.

Proof. Let $\Lambda_a^n = \Lambda_a(v_n)$ for all $n \in \mathbb{N}$ and $a \in \mathcal{A}$. Note that the set of obtainable patch sizes $\mathcal{L}_g = \{L(u) : u \in \mathcal{L}\}$ is locally finite and that the distance of any two points in $\Lambda(v_n)$ is in $\mathcal{L}_g$ by construction. The approximation of the autocorrelation takes the form

$$\gamma(v_n) \;=\; \sum_{a,b\in\mathcal{A}} \tau_a \overline{\tau_b} \sum_{z\in\mathcal{L}_g} \eta_{ab}^n(z)\delta_z,$$

where

$$\eta_{ab}^n(z) \;=\; \frac{1}{L(v_n)} \#\big((\Lambda_a^n - z) \cap \Lambda_b^n\big).$$

Note that $r \in (\Lambda_a^n - z) \cap \Lambda_b^n$ precisely if there is a patch in the geometric image of v_n that has an interval of type b positioned at r and an interval of type a positioned at $r+z$. Word frequencies in v_n translate to patch frequencies in its geometric image. Therefore, $\lim_{n\to\infty} \eta_{ab}^n(z)$ exists for every $a, b \in \mathcal{A}$ and $z \in \mathcal{L}_g$ and depends only on the word frequencies. Since $\mathcal{L}_g$ is locally finite, this entails vague convergence of $\gamma(v_n)$. □

In the following, all limits of measures are understood in the vague topology if not explicitly stated otherwise. The observation that the autocorrelation measure depends only on the word frequencies shows that there is a μ_ϑ-almost sure autocorrelation measure associated to $(\mathbb{X}_\vartheta, S)$. It also provides us with several ways to obtain this measure as a vague limit. For the following, recall the notation

$$\gamma_n(x) = \frac{\omega_n(x) * \widetilde{\omega_n(x)}}{|B_n|}$$

for the normalized autocorrelation measure corresponding to $\omega_n(x)$, the restriction of $\omega(x)$ to the ball B_n, compare Definition 2.4.3.

Corollary 4.3.2. *Let ϑ be a primitive, geometrically compatible random substitution and $(\mathbb{X}_\vartheta, S, \mu_\vartheta)$ the associated ergodic dynamical system. Further, assume that ν is a ϑ^p-invariant measure on $\mathbb{X}_\vartheta$. Then, there is a positive definite measure $\gamma_\vartheta = \gamma$, called the* autocorrelation of ϑ*, such that*

$$\gamma \;=\; \lim_{n\to\infty} \gamma_n(x),$$

for μ_ϑ-almost every $x \in \mathbb{X}_\vartheta$ and ν-almost every $x \in \mathbb{X}_\vartheta^+$. It also satisfies

$$\gamma \;=\; \lim_{n\to\infty} \gamma(\vartheta^n(a)),$$

$\mathbb{P}_a$-almost surely, for all $a \in \mathcal{A}$. ■

We call the Fourier transform $\widehat{\gamma}$ of this measure the *diffraction measure of ϑ*. By dominated convergence, we have in particular,

$$\gamma \;=\; \lim_{n\to\infty} \mathbb{E}_\nu[\gamma(x_{[1,n]})],$$

and

$$\gamma \;=\; \lim_{n\to\infty} \mathbb{E}_a\big[\gamma\big(\vartheta^n(a)\big)\big].$$

Note that, although defined on different probability spaces, both relations take a very similar form. The key insight of Godrèche and Luck was that there is an appropriate splitting of the approximants that gives rise to different spectral components of the diffraction measure. This was made rigorous for several examples of primitive, compatible random substitutions in [BSS18, Goh17, Mol14]. The idea of splitting the influence of a randomization on the diffraction measure into its contributions corresponding to first and second moments is not only useful for random substitutions; it has also found applications in the context of point processes and Bernoullization procedures, see for example [BBM10, DV-J05].

Definition 4.3.3. Let v be a random word on some probability space $(\Omega, \mathbb{P})$. We define,

$$\gamma^{\mathrm{ex}}(v) \;=\; \frac{1}{L(v)} \mathbb{E}[\omega(v)] * \widetilde{\mathbb{E}[\omega(v)]},$$

and

$$\gamma^{\mathrm{var}}(v) \;=\; \frac{1}{L(v)} \Big(\mathbb{E}\big[\omega(v) * \widetilde{\omega(v)}\big] - \mathbb{E}[\omega(v)] * \widetilde{\mathbb{E}[\omega(v)]} \Big) ,$$

satisfying $\mathbb{E}[\gamma(v)] = \gamma^{\mathrm{ex}}(v) + \gamma^{\mathrm{var}}(v)$. The Fourier transformed measures are denoted by

$$\widehat{\gamma}^{\mathrm{ex}}(v) \;=\; \frac{1}{L(v)} \big| \mathbb{E}\big[\widehat{\omega}(v)\big] \big|^2 , \quad \widehat{\gamma}^{\mathrm{var}}(v) \;=\; \frac{1}{L(v)} \Big(\mathbb{E}\big[\big|\widehat{\omega}(v)\big|^2 \big] - \big| \mathbb{E}\big[\widehat{\omega}(v)\big] \big|^2 \Big) ,$$

respectively, and satisfy $\mathbb{E}[\widehat{\gamma}(v)] = \widehat{\gamma}^{\mathrm{ex}}(v) + \widehat{\gamma}^{\mathrm{var}}(v)$.

We will use this definition for a sequence of random words $(v_n)_{n\in\mathbb{N}}$, given by either the random inflation words $(\vartheta^n(a))_{n\in\mathbb{N}}$, for some $a \in \mathcal{A}$, or by $(x_{[1,n]})_{n\in\mathbb{N}}$ on $(\mathbb{X}_\vartheta^+, S, \nu)$. Our ultimate goal is to study the diffraction measure $\widehat{\gamma}$ associated with ϑ. Since the Fourier transform is vaguely continuous on positive definite measures [MS17, Lemma 4.11.10], we find that

$$\widehat{\gamma} \;=\; \lim_{n\to\infty} \mathbb{E}[\widehat{\gamma}(v_n)] \;=\; \lim_{n\to\infty} \big(\widehat{\gamma}^{\mathrm{ex}}(v_n) + \widehat{\gamma}^{\mathrm{var}}(v_n) \big), \tag{4.9}$$

in both cases. For $v_n = x_{[1,n]}$, we will write instead $\widehat{\gamma}_\nu^{\mathrm{ex}}(v_n)$ and $\widehat{\gamma}_\nu^{\mathrm{var}}$, when we wish to specify the dependence on the distribution ν.

4.3.1. Absolutely continuous part of the diffraction

Heuristically, $\widehat{\gamma}^{\mathrm{var}}(v_n)$ contains the "fluctuations" introduced by the random substitution, and we therefore expect this part to give an absolutely continuous component in the limit $n \to \infty$. Our next aim is to make this precise for $v_n = \vartheta^n(a)$.

Theorem 4.3.4. *Let ϑ be a primitive, geometrically compatible random substitution. Then, for every $a \in \mathcal{A}$, the vague limit*

$$\widehat{\gamma}^{\mathrm{var}} \;=\; \lim_{n\to\infty} \widehat{\gamma}^{\mathrm{var}}\big(\vartheta^n(a)\big)$$

exists, is independent of $a \in \mathcal{A}$, and is absolutely continuous with respect to Lebesgue measure.

Before we prove this result, we need some preparation. Note that for a random word v, $\widehat{\gamma}^{\mathrm{var}}(v)$ is in fact a rescaled version of the variance

$$\mathbb{V}\big[\widehat{\omega}(v)\big] \,=\, \mathbb{E}\big[\big|\widehat{\omega}(v)\big|^2\big] - \big|\mathbb{E}\big[\widehat{\omega}(v)\big]\big|^2 .$$

Definition 4.3.5. For every $n \in \mathbb{N}_0$, let $V^n = (V_a^n)_{a\in\mathcal{A}}$ denote the vector with

$$V_a^n \,=\, \mathbb{V}\big[\widehat{\omega}(\vartheta^n(a))\big],$$

for all $a \in \mathcal{A}$.

Lemma 4.3.6. *We have $\mathbb{V}[\widehat{\omega}(\vartheta^n(v))] = \langle V^n|\Phi(v)\rangle$, for all $v \in \mathcal{A}^+$ and $n \in \mathbb{N}_0$.*

Proof. For $a, b \in \mathcal{A}$, we have that $\omega(ab) = \omega(a) + \delta_{L_a} * \omega(b)$. Since there is a well-defined inflation factor λ, the action of the random substitution ϑ^n yields

$$\omega(\vartheta^n(ab)) \,=\, \omega(\vartheta^n(a)) + \delta_{\lambda^n L_a} * \omega(\vartheta^n(b)).$$

Note that the convolution with $\delta_{\lambda^n L_a}$ just gives a multiplicative phase under the Fourier transform. Since the summands are *independent* random measures, applying first the Fourier transform and then the variance, we obtain

$$\mathbb{V}\big[\widehat{\omega}\big(\vartheta^n(ab)\big)\big] \,=\, \mathbb{V}\big[\widehat{\omega}\big(\vartheta^n(a)\big)\big] + \mathbb{V}\big[\widehat{\omega}\big(\vartheta^n(b)\big)\big].$$

For general $v \in \mathcal{A}^+$, it follows in a similar manner that

$$\mathbb{V}\big[\widehat{\omega}\big(\vartheta^n(v)\big)\big] \,=\, \sum_{a\in\mathcal{A}} \mathbb{V}\big[\widehat{\omega}\big(\vartheta^n(a)\big)\big]\, |v|_a \,=\, \langle V^n|\Phi(v)\rangle,$$

for arbitrary $n \in \mathbb{N}$, and the claim follows. □

This can be used to derive a recursion for the (function-valued) vectors $(V^n)_{n\in\mathbb{N}}$.

Proposition 4.3.7. *For every $n \in \mathbb{N}_0$, we have $\langle V^{n+1}| = \langle \Delta^{n+1}| + \langle V^n|M$, where*

$$\Delta_a^{n+1} \,=\, \frac{1}{2} \sum_{u,v\in\vartheta(a)} P_{au}P_{av}\, \big|\, \mathbb{E}\big[\widehat{\omega}\big(\vartheta^n(u)\big)\big] - \mathbb{E}\big[\widehat{\omega}\big(\vartheta^n(v)\big)\big]\, \big|^2,$$

for all $a \in \mathcal{A}$. In particular, for all $n \in \mathbb{N}$,

$$\langle V^n| \,=\, \sum_{j=1}^{n} \langle \Delta^j| M^{n-j}. \tag{4.10}$$

Proof. Let $a \in \mathcal{A}$. By the Markov property, we obtain that

$$\mathbb{E}\big[\widehat{\omega}\big(\vartheta^{n+1}(a)\big)\big] \,=\, \sum_{u\in\vartheta(a)} P_{au}\mathbb{E}\big[\widehat{\omega}\big(\vartheta^n(u)\big)\big], \quad \mathbb{E}\big[|\widehat{\omega}\big(\vartheta^{n+1}(a)\big)|^2\big] \,=\, \sum_{u\in\vartheta(a)} P_{au}\mathbb{E}\big[|\widehat{\omega}\big(\vartheta^n(u)\big)|^2\big].$$

From this, a straightforward calculation gives

$$\mathbb{V}[\widehat{\omega}(\vartheta^{n+1}(a))] \,=\, \Delta_a^{n+1} + \sum_{u\in\vartheta(a)} P_{au}\mathbb{V}[\widehat{\omega}(\vartheta^n(u))].$$

Invoking Lemma 4.3.6, we find

$$\sum_{u\in\vartheta(a)} P_{au}\mathbb{V}[\widehat{\omega}(\vartheta^n(u))] \;=\; \sum_{u\in\vartheta(a)} P_{au}\langle V^n|\Phi(u)\rangle \;=\; \langle V^n|M|e_a\rangle,$$

which completes the proof of the first claim. Note that $V_a^0 = 0$, for all $a \in \mathcal{A}$. Iterating the relation $\langle V^n| = \langle\Delta^n| + \langle V^{n-1}|M$ yields the second claim. □

A key observation is that each component Δ^n defines a *non-negative* density function. As an intermediate step, we consider pointwise convergence of the densities defining V^n.

Lemma 4.3.8. *For every $k \in \mathbb{R}$, we have*

$$\lim_{n\to\infty}\frac{1}{\lambda^n}\langle V^n(k)| \;=\; \left(\sum_{m=1}^{\infty}\frac{1}{\lambda^m}\langle\Delta^m(k)|\right)|R\rangle\langle L|.$$

Proof. This follows from (4.10), in conjunction with the PF theorem. If the series on the right hand side converges, it follows by dominated convergence that

$$\lim_{n\to\infty}\frac{1}{\lambda^n}\langle V^n(k)| \;=\; \lim_{n\to\infty}\sum_{m=1}^{n}\frac{\langle\Delta^m(k)|}{\lambda^m}\frac{M^{n-m}}{\lambda^{n-m}} \;=\; \left(\sum_{m=1}^{\infty}\frac{\langle\Delta^m(k)|}{\lambda^m}\right)|R\rangle\langle L|.$$

If $\sum_{m=1}^{n}\lambda^{-m}\langle\Delta_a^m(k)|$ diverges as $n\to\infty$ for some $a\in\mathcal{A}$, then

$$\liminf_{n\to\infty}\sum_{m=1}^{n}\left(\frac{\langle\Delta^m(k)|}{\lambda^m}\frac{M^{n-m}}{\lambda^{n-m}}\right)_b \;\geqslant\; \liminf_{n\to\infty}\left(\sum_{m=1}^{\lfloor n/2\rfloor}\frac{\langle\Delta^m(k)|}{\lambda^m}\right)\frac{|R\rangle L_b}{2} \;=\; \infty,$$

holds for every $b\in\mathcal{A}$. Here, we have used that all entries of $\lambda^{-r}M^r$ are positive for all $r\in\mathbb{N}$. □

Finally, Theorem 4.3.4 is an immediate consequence of the following result, which also provides an expression for the density function corresponding to $\widehat{\gamma}^{\mathrm{var}}$.

Proposition 4.3.9. *The function $g\colon\mathbb{R}\to[0,\infty]$, with*

$$g(k) \;=\; \left(\sum_{m=1}^{\infty}\frac{1}{\lambda^m}\langle\Delta^m(k)|R\rangle\right)$$

is locally integrable with respect to Lebesgue measure, and we have

$$\lim_{n\to\infty}\widehat{\gamma}^{\mathrm{var}}\big(\vartheta^n(a)\big) \;=\; g\lambda_{\mathrm{L}},$$

in the vague topology, for all $a\in\mathcal{A}$.

Proof. Let $a\in\mathcal{A}$ and $g_n\colon\mathbb{R}\to\mathbb{R}$,

$$g_n(k) \;=\; \widehat{\gamma}^{\mathrm{var}}\big(\vartheta^n(a)\big)(k) \;=\; \frac{1}{L_a\lambda^n}V_a^n(k).$$

Similarly, we set $h_n(k) = \widehat{\gamma}^{\mathrm{ex}}(\vartheta^n(a))(k)$. We have established in Lemma 4.3.8 that the pointwise convergence $g_n(k) \to g(k)$ as $n \to \infty$ holds for all $k \in \mathbb{R}$. Let $f \in \mathrm{C}_c(\mathbb{R})$. Recall from (4.9) that

$$\lim_{n\to\infty} \int_{\mathbb{R}} |f(x)|(g_n(k) + h_n(k)) \,\mathrm{d}k \,=\, \widehat{\gamma}(|f|) \,<\, \infty,$$

and that $h_n(k) \geqslant 0$ for all $k \in \mathbb{R}$ by construction. Hence, by Fatou's lemma,

$$\int_{\mathbb{R}} |f(k)|g(k) \,\mathrm{d}k \,=\, \int_{\mathbb{R}} \liminf_{n\to\infty} |f(k)|g_n(k) \,\mathrm{d}k \,\leqslant\, \liminf_{n\to\infty} \int_{\mathbb{R}} |f(k)|g_n(k) \,\mathrm{d}k \,<\, \infty.$$

This implies that g is locally integrable with respect to Lebesgue measure. Additionally, we have for every $n \in \mathbb{N}$ and $k \in \mathbb{R}$ that

$$g_n(k) \,=\, \frac{1}{L_a\lambda^n} \sum_{m=1}^{n} \langle \Delta^m(k)|M^{n-m}|e_a\rangle \,\leqslant\, C\frac{1}{L_a} \sum_{m=1}^{n} \frac{1}{\lambda^m} \langle \Delta^m(k)|R\rangle L_a \,\leqslant\, Cg(k)$$

holds for some $C > 0$ that is independent of n and k. The vague convergence $g_n\lambda_{\mathrm{L}} \to g\lambda_{\mathrm{L}}$ as $n \to \infty$ follows by Lebesgue's dominated convergence theorem. □

4.3.2. Singular part of the diffraction

As a result of (4.9) and Theorem 4.3.4, we immediately obtain the following.

Corollary 4.3.10. *The measure*

$$\widehat{\gamma}^{\mathrm{ex}} \,=\, \lim_{n\to\infty} \widehat{\gamma}^{\mathrm{ex}}(\vartheta^n(a))$$

is well-defined as a vague limit and independent of $a \in \mathcal{A}$. ∎

We hence have found a splitting

$$\widehat{\gamma} \,=\, \widehat{\gamma}^{\mathrm{ex}} + \widehat{\gamma}^{\mathrm{var}},$$

where the last term is an absolutely continuous measure. In this section, we want to investigate the spectral type of the measure $\widehat{\gamma}^{\mathrm{ex}}$. More specifically, we will establish criteria that ensure that $\widehat{\gamma}^{\mathrm{ex}}$ is a pure point measure. This will be done for a more restricted class of random substitutions.

Assumption 4.3.11. We denote by ϑ a primitive random substitution of constant length $\ell \geqslant 2$ throughout this section.

In particular, ϑ is geometrically compatible with $L = (1, \ldots, 1)$ and $\lambda = \ell$. Hence, $\omega(v)$ is supported on $\mathbb{N}$ for all $v \in \mathcal{A}^+$. For each $a \in \mathcal{A}$, the same holds for

$$\begin{aligned} \mathbb{E}[\omega(\vartheta(a))] \,&=\, \sum_{b\in\mathcal{A}} \tau_b \sum_{u\in\vartheta(a)} \mathbb{P}[\vartheta(a) = u]\delta_{\Lambda_b(u)} \,=\, \sum_{n=1}^{\ell} \sum_{b\in\mathcal{A}} \tau_b \,\mathbb{P}[\vartheta(a)_n = b]\,\delta_n, \\ &=:\, \sum_{n=1}^{\ell} \langle \pi_n^a|\tau\rangle \delta_n, \end{aligned} \tag{4.11}$$

which again acquires the form of a weighted Dirac comb, where the weights are determined by the vectors π_n^a, with

$$\langle \pi_n^a | e_b \rangle \;=\; \mathbb{P}[\vartheta(a)_n = b], \tag{4.12}$$

for all $b \in \mathcal{A}$ and $1 \leqslant n \leqslant \ell$. More generally, if v is a random word with well-defined length $|v|$, we define the *positional marginal* of v at n as the probability vector $\pi_n(v)$, with

$$\langle \pi_n(v) | e_a \rangle \;=\; \mathbb{P}[v_n = a], \tag{4.13}$$

for all $a \in \mathcal{A}$ and $1 \leqslant n \leqslant |v|$. With this notation, we easily generalize (4.11) to

$$\mathbb{E}[\omega(v)] \;=\; \sum_{n=1}^{|v|} \langle \pi_n(v) | \tau \rangle \delta_n. \tag{4.14}$$

Before we continue, let us put this observation into a proper framework.

Definition 4.3.12. We denote by $\Delta^{\mathcal{A}}$ the probability simplex indexed by $\mathcal{A}$, that is,

$$\Delta^{\mathcal{A}} \;=\; \Big\{ \pi \in [0,1]^{\mathcal{A}} : \sum_{a \in \mathcal{A}} \langle \pi | e_a \rangle = 1 \Big\}.$$

The weighted Dirac comb corresponding to a word $\pi = \pi_1 \cdots \pi_n \in (\Delta^{\mathcal{A}})^n$ is given by

$$\omega(q) \;=\; \sum_{m=1}^{n} \langle \pi_m | \tau \rangle \delta_m.$$

Definition 4.3.13. Let v be a random variable on some probability space $(\Omega, \mathbb{P})$ such that $|v| = m$ is well defined. The *marginalization* of v maps v to an element $\pi(v) \in (\Delta^{\mathcal{A}})^m$, given by the word

$$\pi(v) \;=\; \pi_1(v) \cdots \pi_m(v),$$

where $\pi_n(v)$ is the positional marginal of v at n, defined in (4.13), for all $1 \leqslant n \leqslant m$.

We will also write $\pi(u)_n$ for $\pi_n(u)$, emphasizing the point of view that $\pi(u)$ should be regarded as a word and $\pi(u)_n$ as one of its letters.

With the new notation, we can rewrite (4.14) in the compact form

$$\mathbb{E}[\omega(v)] \;=\; \omega(\pi(v)).$$

In the next step, we try to understand how the marginalization of a random word v changes if the random substitution ϑ is applied. First, assume $v = a \in \mathcal{A}$ is a (deterministic) letter. Then, $\pi(a) = e_a$ and $\pi(\vartheta(a)) = \pi_1^a \cdots \pi_\ell^a$, with $\pi_n^a = \pi_n(\vartheta(a))$, as defined in (4.12). This motivates us to define a map

$$\overline{\vartheta} \colon e_a \mapsto \pi_1^a \cdots \pi_\ell^a,$$

for all $a \in \mathcal{A}$. Extending this to a linear map in each component leads to the following concept.

Definition 4.3.14. Let ϑ be a primitive random substitution of constant length ℓ. The *average substituton* $\overline{\vartheta}$ is a substitution of constant length ℓ on the compact alphabet $\Delta^{\mathcal{A}}$, which is defined for $\pi \in \Delta^{\mathcal{A}}$ by

$$\overline{\vartheta}(\pi) \,=\, \overline{\vartheta}(\pi)_1 \cdots \overline{\vartheta}(\pi)_\ell,$$

where, for all $1 \leqslant n \leqslant \ell$,

$$\big\langle \overline{\vartheta}(\pi)_n \big| \,=\, \langle \pi | M_n,$$

with

$$M_n \,=\, \sum_{a \in \mathcal{A}} |e_a\rangle \langle \pi_n^a| \,=\, \big(\mathbb{P}[\vartheta(a)_n = b]\big)_{a,b \in \mathcal{A}}.$$

We refer to $\overline{\vartheta}_n \colon \Delta^{\mathcal{A}} \to \Delta^{\mathcal{A}}, \pi \mapsto \overline{\vartheta}(\pi)_n$ as the *n-th column* of the substitution $\overline{\vartheta}$.

Note that M_n is a Markov matrix for all $1 \leqslant n \leqslant \ell$, which ensures that $\overline{\vartheta} \colon \Delta^{\mathcal{A}} \to (\Delta^{\mathcal{A}})^+$ is a well-defined map. It is extended to an endomorphism on $(\Delta^{\mathcal{A}})^+$ by concatenation, as usual. It is straightforward to verify that the consistency equation $\overline{\vartheta}^n = \overline{\vartheta^n}$ holds for all $n \in \mathbb{N}$.

Example 4.3.15. Recall the random period doubling substitution

$$\vartheta \colon a \mapsto \begin{cases} ab & \text{with prob. } p, \\ ba & \text{with prob. } 1-p, \end{cases} \qquad b \mapsto aa,$$

from Example 4.2.19, with

$$M_1 \,=\, \begin{pmatrix} p & 1-p \\ 1 & 0 \end{pmatrix}, \quad M_2 \,=\, \begin{pmatrix} 1-p & p \\ 1 & 0 \end{pmatrix}.$$

The average substitution takes the form $\overline{\vartheta}(\pi) = \big(\langle \pi | M_1\big)\big(\langle \pi | M_2\big)$, for all $\pi \in \Delta^{\mathcal{A}}$. ◊

Lemma 4.3.16. *Let u be a random word on $(\Omega, \mathbb{P})$ such that $|u| \in \mathbb{N}$ is well-defined. Then, the map $\pi \colon u \mapsto \pi(u)$ semi-conjugates the actions of ϑ and $\overline{\vartheta}$, that is,*

$$\pi(\vartheta(u)) = \overline{\vartheta}(\pi(u)).$$

Proof. Let $|u| = 1$. Then, for all $1 \leqslant n \leqslant \ell$ and $a \in \mathcal{A}$, we have

$$\big\langle \pi_n(\vartheta(u)) \big| e_a \big\rangle \,=\, \mathbb{P}\big[\vartheta_n(u) = a\big] \,=\, \sum_{b \in \mathcal{A}} \mathbb{P}[u = b]\, \mathbb{P}\big[\vartheta(b)_n = a\big] \,=\, \langle \pi(u) | M_n | e_a \rangle,$$

implying that $\pi_n(\vartheta(u)) = \overline{\vartheta}_n(\pi(u))$, for all $1 \leqslant n \leqslant \ell$. If $|u| = m \in \mathbb{N}$, we obtain

$$\pi(\vartheta(u)) \,=\, \pi(\vartheta(u_1)) \cdots \pi(\vartheta(u_m)) \,=\, \overline{\vartheta}\big(\pi(u_1)\big) \cdots \overline{\vartheta}\big(\pi(u_n)\big) \,=\, \overline{\vartheta}\big(\pi(u)\big),$$

and the claim follows. □

All of the definitions and arguments presented in this section naturally extend to the infinite setting. In particular, if x is a random sequence, distributed according to some probability measure μ on $\mathcal{A}^{\mathbb{N}}$, we set $\pi(x) = (\pi_n(x))_{n\in\mathbb{N}}$, where $\pi_n(x)$ is the positional marginal satisfying,

$$\langle \pi_n(x) | e_a \rangle = \mu[x_n = a],$$

for all $n \in \mathbb{N}$ and $a \in \mathcal{A}$. For easier reference, we collect in the following result the analogue of the previous observations in the setting of infinite sequences.

Lemma 4.3.17. *Let x be a $\mathcal{A}^{\mathbb{N}}$-valued random sequence, with distribution μ. Then,*

$$\mathbb{E}_\mu[\omega(x)] \,=\, \omega(\pi(x)),$$

and

$$\pi(\vartheta(x)) \,=\, \overline{\vartheta}\big(\pi(x)\big),$$

where $\vartheta(x)$ has distribution $\vartheta(\mu)$. ∎

Corollary 4.3.18. *Let ν be a ϑ-invariant measure on $\mathbb{X}_\vartheta^+$ and let x denote a ν-distributed random sequence. Then, the sequence $\pi(x) \in (\Delta^{\mathcal{A}})^{\mathbb{N}}$ with*

$$\langle \pi_n(x) | e_a \rangle \,=\, \nu[x_n = a],$$

for all $n \in \mathbb{N}$ and $a \in \mathcal{A}$, is invariant under $\overline{\vartheta}$. That is,

$$\langle \pi(x)_{m+s\ell} | \,=\, \big\langle \overline{\vartheta}(\pi(x))_{m+s\ell} \big| \,=\, \langle \pi(x)_{s+1} | M_m,$$

for all $1 \leqslant m \leqslant \ell$, and $s \in \mathbb{N}_0$.

Proof. By Lemma 4.3.17, we have

$$\overline{\vartheta}\big(\pi(x)\big) = \pi(\vartheta(x)).$$

Since ν is ϑ-invariant, the distributions of x and $\vartheta(x)$ coincide, implying $\pi(\vartheta(x)) = \pi(x)$. □

Remark 4.3.19. The sequence $\pi(x)$ is ℓ-automatic as defined in [Que10]. There, it was shown that, if $\overline{\vartheta}$ is primitive, the orbit closure of $\pi(x)$ is strictly ergodic [Que10, Cor. 12.2]. Unfortunately, primitivity of ϑ does *not* necessarily entail that $\overline{\vartheta}$ is also primitive, even if we restrict it to $\mathcal{L}_1(\pi(x)) = \overline{\{\pi(x)_n : n \in \mathbb{N}\}}$. For a counterexample, consider the random substitution

$$\vartheta\colon a \mapsto ab, \quad b \mapsto \begin{cases} bb, & \text{with prob. } 1/2, \\ ba, & \text{with prob. } 1/2, \end{cases}$$

and ν the ϑ-invariant measure with $\nu[x_1 = a] = 1$. Here, M_1 is the identity and the stable distribution of M_2 is given by $\pi = (1/3, 2/3)$. Since $\overline{\vartheta}(\pi) = \pi\pi$ and $(1, 0) \in \mathcal{L}_1$, this substitution is not primitive. Of course, we can also choose a ϑ-invariant measure ν with $(\nu[a], \nu[b]) = \pi$, to obtain $\pi(x) = \pi^{\mathbb{N}}$ and hence a primitive substitution $\overline{\vartheta}$ on the alphabet $\mathcal{L}_1(\pi(x)) = \{\pi\}$. Whether such a construction is possible in general is open at this point and remains to be determined. ◊

Up to replacing ϑ by one of its powers, we can always assume that a ϑ-invariant measure exists on $\mathbb{X}_{\vartheta}^{+}$. This provides us with a way to approach the diffraction of the expected Dirac comb. Since

$$\gamma_{\nu}^{\mathrm{ex}}(x_{[1,n]}) \,=\, \frac{\omega_n * \widetilde{\omega_n}}{n}, \quad \omega_n \,:=\, \mathbb{E}_{\nu}\big[\omega\big(x_{[1,n]}\big)\big],$$

the vague accumulation points of

$$(\widehat{\gamma}_{\nu}^{\mathrm{ex}}(x_{[1,n]}))_{n\in\mathbb{N}}$$

are precisely the diffraction measures of

$$\mathbb{E}_{\nu}(\omega(x)) = \omega(\pi(x)),$$

along $B_n = [0, n]$, as $n \to \infty$. The weighted Dirac comb

$$\omega(\pi(x)) \,=\, \sum_{n=1}^{\infty} \langle \pi(x)_n | \tau \rangle \delta_n = \sum_{n=1}^{\infty} q_n^* \delta_n,$$

is the Dirac comb of the bounded, complex valued sequence $q^* \in \mathbb{C}^{\mathbb{N}}$, where $q_n^* = \langle \pi(x)_n | \tau \rangle$ for all $n \in \mathbb{N}$. As such, it is a translation bounded measure, and pure point diffraction can be characterized in terms of mean-almost periodicity of the sequence q^*.

Definition 4.3.20. Let (A, d) be a metric space and $(n_m)_{m\in\mathbb{N}}$ a non-decreasing sequence of natural numbers. A sequence $q \in A^{\mathbb{N}}$ is called *mean-almost periodic* (m.a.p.) with respect to $(n_m)_{m\in\mathbb{N}}$, if, for every $\varepsilon > 0$, there is a relatively dense set $E_\varepsilon \subset \mathbb{Z}$ such that, for all $k \in E_\varepsilon$,

$$\limsup_{m\to\infty} \frac{1}{n_m} \sum_{n\leqslant n_m} d(q_n, q_{n+k}) \,\leqslant\, \varepsilon,$$

where the assignment of q_{-n} is arbitrary for $n \in \mathbb{N}_0$. A sequence q is m.a.p. if it is m.a.p. with respect to $(n)_{n\in\mathbb{N}}$ and hence with respect to every non-decreasing sequence.

We easily observe that $q^\star$ is m.a.p. for all choices of τ if and only if $\pi(x)$ is m.a.p. in the variation distance d_v, given by

$$d_v(\pi_1, \pi_2) \,=\, \max_{a\in\mathcal{A}} |\langle \pi_1 - \pi_2 | e_a \rangle|,$$

for all $\pi_1, \pi_2 \in \Delta^{\mathcal{A}}$. Due to the constant-length property, a natural candidate for the collection of mean-almost periods is the sequence of sets $(E_n)_{n\in\mathbb{N}}$, with $E_n = \ell^n\mathbb{Z}$. It is a classic result that a (bounded) complex valued sequence q gives rise to a pure point diffraction measure if and only if q is mean-almost periodic. In the uniquely ergodic setting, this can be found in [Que10, Lemma 6.6]. We refer to [LSS20] for a generalization to arbitrary translation bounded measures. The following is an immediate corollary of [LSS20, Thm. 2.13].

Fact 4.3.21. *Let $\omega = \sum_{n=1}^{\infty} q_n \delta_n$, for some bounded $q \in \mathbb{C}^{\mathbb{N}}$. Let $\widehat{\gamma}$ be a diffraction measure of ω with respect to the averaging sequence $([0, n_m])_{m\in\mathbb{N}}$. Then, $\widehat{\gamma}$ is a pure point measure if and only if q is m.a.p. with respect to $(n_m)_{m\in\mathbb{N}}$.* ∎

Our next aim is to give a criterion that ensures that q^* is m.a.p.. For *deterministic* substitutions of constant length, this was solved by Dekking in terms of the *coincidence condition*; we refer to [Dek78] for details. Before we formulate an analogue of this result, we introduce a concept that quantifies how far a Markov matrix is from a projection to its equilibrium state.

Definition 4.3.22. For a Markov matrix M on the state space $\mathcal{A}$, *Dobrushin's coefficient* $\delta(M)$ is defined as

$$\delta(M) \,=\, \frac{1}{2} \max_{a,b} \sum_{c \in \mathcal{A}} |M_{ac} - M_{bc}|.$$

We call a matrix M *scrambling* if $\delta(M) < 1$.

In the following, we assume that every Markov matrix is defined on the state space $\mathcal{A}$. The notion of scrambling Markov matrices was originally introduced by Hajnal [Haj58] for the study of *weak ergodicity* of non-homogeneous Markov chains. There, an equivalent formulation was used, without reference to Dobrushin's coefficient. We state it as a lemma for easier reference and omit the straightforward proof; compare [Sen81, Thm. 2.10].

Lemma 4.3.23. *The Markov matrix M is scrambling if and only if, for all $a, b \in \mathcal{A}$, there exists a letter $c \in \mathcal{A}$ such that $M_{ac} > 0$ and $M_{bc} > 0$.* ∎

Some elementary but useful facts about Dobrushin's coefficient are listed below. They can be found in [Bre20, Ch. 12], to which we also refer for a general introduction to non-homogeneous Markov chains. In fact, restricting the action of a Markov matrix to an appropriate subspace of $\mathbb{R}^{\mathcal{A}}$, δ can be interpreted as an operator norm [NS99], so the following does not come as a surprise.

(1) $0 \leqslant \delta(M) \leqslant 1$ for every Markov matrix M.

(2) $\delta(M) = 0$ precisely if M has identical rows.

(3) $\delta(MN) \leqslant \delta(M)\delta(N)$ for all Markov matrices M, N.

(4) $d_v(\langle\pi_1|M, \langle\pi_2|M) \leqslant d_v(\pi_1, \pi_2)\delta(M)$ for every Markov matrix M and $\pi_1, \pi_2 \in \Delta^{\mathcal{A}}$.

The last property implies that every scrambling matrix acts as a contraction on $(\Delta^{\mathcal{A}}, d_v)$. It should be noted that for a product $N_1 \cdots N_r$ of Markov matrices to be scrambling it is sufficient but *not* necessary, that one of the matrices $N_1, \ldots, N_r$ is scrambling.

Definition 4.3.24. Let ϑ be a primitive random substitution of constant length ℓ, with average substitution $\overline{\vartheta}$. We call ϑ *scrambling* if one of the Markov matrices M_n, defined by $M_n = (\mathbb{P}[\vartheta(a)_n = b])_{a,b \in \mathcal{A}}$ is scrambling.

Proposition 4.3.25. *Let $\pi(x) \in (\Delta^{\mathcal{A}})^{\mathbb{N}}$ be a $\overline{\vartheta}$-invariant sequence. If ϑ is scrambling, the sequence $\pi(x)$ is mean-almost periodic and every diffraction measure of $\omega(\pi(x))$ is pure point.*

Proof. Let $m \in \mathbb{N}$ be a fixed power of the substitution $\overline{\vartheta}$ and let $M_n^{[m]}$ with $1 \leqslant n \leqslant \ell^m$ be the transition matrices, satisfying

$$\langle \overline{\vartheta}^m(\pi)_n| \,=\, \langle \pi | M_n^{[m]},$$

for $\pi \in \Delta^{\mathcal{A}}$. For every such n, there is a multi-index $i = i(n) = (i_1, \dots, i_m)$ in the set $I^m = \{1, \dots, \ell\}^m$ such that

$$M_n^{[m]} \,=\, M(i) \,:=\, M_{i_m} \cdots M_{i_1},$$

and the assignment $n \mapsto i(n)$ is a bijection between $\{1, \dots, \ell^m\}$ and I^m. This construction is classic for automatic sequences; we refer to [Que10, Ch. 12] for details. Since $\pi(x)$ is $\overline{\vartheta}^m$-invariant, we have for all $j \in \{1, \dots, \ell^m\}$ and $k \in \mathbb{N}_0$ that

$$\langle \pi(x)_{j+k\ell^n} | \,=\, \langle \pi(x)_{k+1} | M_j^{[m]},$$

and hence, for all $k_1, k_2 \in \mathbb{N}_0$,

$$d_v(\pi(x)_{j+k_1\ell^m}, \pi(x)_{j+k_2\ell^m}) \,\leqslant\, d_v(\pi(x)_{k_1+1}, \pi(x)_{k_2+1}) \delta(M_j^{[m]}) \,\leqslant\, \delta(M_j^{[m]}).$$

Hence, we need to show that $\delta(M_j^{[m]})$ is small as we average over j. Assume that M_r is scrambling with $\delta = \delta(M_r) < 1$, and let

$$I_k^m \,=\, \{i \in I^m : i_n = r \text{ for precisely } k \text{ numbers } n\}.$$

By the submuliplicativity of Dobrushin's coefficient, we obtain $\delta(M(i)) \leqslant \delta^k$ for every $i \in I_k^m$. This yields

$$\sum_{j=1}^{\ell^m} \delta(M_j^{[m]}) \,=\, \sum_{k=0}^{m} \sum_{i \in I_k^m} \delta(M(i)) \,\leqslant\, \sum_{k=0}^{m} \sum_{i \in I_k^m} \delta^k \,=\, \sum_{k=0}^{m} \binom{m}{k} (\ell - 1)^{m-k} \delta^k \,=\, (\ell - 1 + \delta)^m,$$

and hence, for all $k \in \mathbb{Z}$,

$$\begin{aligned} \limsup_{n\to\infty} \frac{1}{n} \sum_{j=1}^{n} d_v\big(\pi(x)_j, \pi(x)_{j+k\ell^m}\big) \,&\leqslant\, \limsup_{n\to\infty} \frac{1}{n\ell^m} \sum_{s=1}^{n} \sum_{j=1}^{\ell^m} d_v\big(\pi(x)_{j+s\ell^m}, \pi(x)_{j+(s+k)\ell^m}\big) \\ &\leqslant\, \limsup_{n\to\infty} \frac{1}{n\ell^m} \sum_{s=1}^{n} \sum_{j=1}^{\ell^m} \delta(M_j^{[m]}) \,\leqslant\, \frac{(\ell - 1 + \delta)^m}{\ell^m} \,\leqslant\, \varepsilon, \end{aligned}$$

for large enough $m \in \mathbb{N}$. This implies that $\pi(x)$ is mean-almost periodic. The same holds for the sequence q^*, with $q_n^* = \langle \pi(x)_n | \tau \rangle$, for all $n \in \mathbb{N}$. Hence, every diffraction measure of $\omega(\pi(x))$ is a pure point measure by Fact 4.3.21. □

At this point, we have a criterion that ensures that every vague accumulation point of $(\widehat{\gamma}^{\mathrm{ex}}(x_{[1,n]}))_{n\in\mathbb{N}}$ is a pure point measure. It remains to relate this to the spectral type of the measure $\widehat{\gamma}^{\mathrm{ex}} = \lim_{n\to\infty} \widehat{\gamma}^{\mathrm{ex}}(\vartheta^n(a))$, for $a \in \mathcal{A}$. Since the distribution of $\vartheta^n(a)$ converges weakly to a ϑ-invariant measure ν, it seems intuitive that both problems are related. However, weak convergence is not the right notion to ensure convergence of the corresponding diffraction measures.

Proposition 4.3.26. *Assume that* $\overline{\vartheta}^n(e_a)$ *converges* in mean *to* $\pi(x)$*, that is,*

$$\lim_{n\to\infty} \frac{1}{\ell^n} \sum_{j=1}^{\ell^n} d_v\big(\overline{\vartheta}^n(e_a)_j, \pi(x)_j\big) \,=\, 0.$$

Then, the diffraction of $\omega(\pi(x))$ *exists, with respect to the averaging sequence* $([0,\ell^n])_{n\in\mathbb{N}}$*, and is given by*

$$\widehat{\gamma}^{\mathrm{ex}} \,=\, \lim_{n\to\infty} \widehat{\gamma}\big(\overline{\vartheta}^n(e_a)\big),$$

where the limit is well-defined due to Corollary 4.3.10. ∎

This is implicitly a condition on the mean convergence of equi-translation bounded measures; compare [BSS18] for background and a more detailed discussion. The proof is immediate from [BSS18, Thm. 8.6] and therefore omitted. Conveniently, the mean convergence of $\overline{\vartheta}^n(e_a)$ also follows if ϑ is scrambling.

Theorem 4.3.27. *Let* ϑ *be a primitive random substitution of constant length* ℓ*, with diffraction measure* $\widehat{\gamma}$*, and assume that some power of* ϑ *is scrambling. Then, the spectral components of* $\widehat{\gamma}$ *are given by the vague limits*

$$\widehat{\gamma}_{\mathrm{pp}} \,=\, \widehat{\gamma}^{\mathrm{ex}} \,=\, \lim_{n\to\infty} \widehat{\gamma}^{\mathrm{ex}}(\vartheta^n(a)) \,=\, \lim_{n\to\infty} \widehat{\gamma}^{\mathrm{ex}}_{\nu}(x_{[1,\ell^n]})$$

and

$$\widehat{\gamma}_{\mathrm{ac}} \,=\, \widehat{\gamma}^{\mathrm{var}} \,=\, \lim_{n\to\infty} \widehat{\gamma}^{\mathrm{var}}(\vartheta^n(a)) \,=\, \lim_{n\to\infty} \widehat{\gamma}^{\mathrm{var}}_{\nu}(x_{[1,\ell^n]}),$$

for all $a \in \mathcal{A}$*, and every measure* ν *on* $\mathbb{X}^+_\vartheta$ *that is invariant under some power of* ϑ*. Additionally,* $\widehat{\gamma}$ *has no singular continuous component.*

Proof. Possibly replacing ϑ by one of its powers, we can assume that ϑ itself is scrambling. Also, without loss of generality assume that ν is ϑ-invariant. By Proposition 4.3.9 and Corollary 4.3.10, the vague limits

$$\widehat{\gamma}^{\mathrm{var}} \,=\, \lim_{n\to\infty} \widehat{\gamma}^{\mathrm{var}}(\vartheta^n(a)), \quad \widehat{\gamma}^{\mathrm{ex}} \,=\, \lim_{n\to\infty} \widehat{\gamma}^{\mathrm{ex}}(\vartheta^n(a)),$$

both exist and are independent of $a \in \mathcal{A}$, where $\widehat{\gamma}^{\mathrm{var}}$ comprises an absolutely continuous measure. Since $\pi(x)$ is invariant under $\overline{\vartheta}$, we have

$$\pi(x)_{[1,\ell^n]} \,=\, \overline{\vartheta}^n(\pi(x)_1),$$

for all $n \in \mathbb{N}$ and hence, for all $1 \leqslant j \leqslant n$,

$$d_v\big(\overline{\vartheta}^n(e_a)_j, \pi(x)_j\big) \,=\, d_v\big(\langle e_a|M_j^{[n]}, \langle \pi(x)_1|M_j^{[n]}\big) \,\leqslant\, \delta\big(M_j^{[n]}\big),$$

with $M_j^{[n]}$ as in the proof of Proposition 4.3.25. Following the same calculations, we find that for some $0 \leqslant \delta < 1$,

$$\limsup_{n\to\infty} \frac{1}{\ell^n} \sum_{j=1}^{\ell^n} d_v\big(\overline{\vartheta}^n(e_a)_j, \pi(x)_j\big) \,\leqslant\, \limsup_{n\to\infty} \frac{1}{\ell^n} \sum_{j=1}^{\ell^n} \delta\big(M_j^{[n]}\big) \,\leqslant\, \lim_{n\to\infty} \frac{(\ell-1+\delta)^n}{\ell^n} \,=\, 0.$$

By Fact 4.3.26, we obtain that

$$\lim_{n\to\infty} \widehat{\gamma}_\nu^{\mathrm{ex}}\bigl(x_{[1,\ell^n]}\bigr) \,=\, \lim_{n\to\infty} \frac{1}{\ell^n}\bigl|\widehat{\omega}\bigl(\pi(x)_{[1,\ell^n]}\bigr)\bigr|^2 \,=\, \widehat{\gamma}^{\mathrm{ex}},$$

which is a pure point measure due to Proposition 4.3.25. It follows that $\widehat{\gamma} = \widehat{\gamma}^{\mathrm{ex}} + \widehat{\gamma}^{\mathrm{var}}$ and

$$\lim_{n\to\infty} \widehat{\gamma}_\nu^{\mathrm{var}}\bigl(x_{[1,\ell^n]}\bigr) \,=\, \widehat{\gamma} - \lim_{n\to\infty} \widehat{\gamma}_\nu^{\mathrm{ex}}\bigl(x_{[1,\ell^n]}\bigr) \,=\, \widehat{\gamma}^{\mathrm{var}},$$

which concludes the proof. □

Remark 4.3.28. In [BSS18], an analogue of Theorem 4.3.27 was shown for the random period doubling substitution, inter alia. There, the proof relied on the observation that the sequence $\pi(x)$ is in fact *strongly* almost periodic. This has the advantage of providing a nice interpretation in terms of a weighted model set and paves the way to find an explicit formula for the diffraction intensities via Fourier–Bohr coefficients. On the other hand, as the name suggests, strong almost periodicity is a much stronger condition than mean almost periodicity. If the almost periods are assumed to be $E_n = \ell^n\mathbb{Z}$, with $n \in \mathbb{N}$, it requires that *every* sequence $(N_n)_{n\in\mathbb{N}}$ with $N_n \in \{M_1, \dots, M_\ell\}$ is *weakly ergodic*, that is, $\delta(N_1 \cdots N_n) \to 0$ as $n \to \infty$. On the other hand, for the proof of Theorem 4.3.27, we only require that the set of matrices $S = \{M_1, \dots, M_\ell\}$ is *mean-weakly ergodic* in the sense that

$$\lim_{n\to\infty} \frac{1}{\ell^n} \sum_{N_1\cdots N_n \in S^n} \delta(N_1 \cdots N_n) \,=\, 0.$$

We also stress that the result in Theorem 4.3.27 is *not* a uniform result for all $x \in \mathbb{X}_\vartheta$ but rather for almost every x, with respect to μ. For example, it applies to the random Thue–Morse substitution $\vartheta\colon a \mapsto \{ab, ba\}, b \mapsto \{ba\}$, while there are certainly points in $\mathbb{X}_\vartheta$ (arising from the deterministic Thue–Morse substitution) with a singular continuous component in the diffraction spectrum. ♢

As the next result shows, the conclusions of Theorem 4.3.27 hold in particular for every primitive, constant-length random substitution ϑ on a binary alphabet $\mathcal{A}$, provided that ϑ is non-deterministic.

Lemma 4.3.29. *Let ϑ be a primitive random substitution of constant length on the alphabet $\mathcal{A} = \{a, b\}$ and assume that* $\max\{\#\vartheta(a), \#\vartheta(b)\} \geqslant 2$. *Then, ϑ is scrambling.*

Proof. Without loss of generality, assume that $\#\vartheta(a) \geqslant 2$ and let $u, v \in \vartheta(a)$ with $u \neq v$. Then, there is a position j such that $u_j \neq v_j$. This implies that $(M_j)_{aa} > 0$ and $(M_j)_{ab} > 0$, that is, M_j has a row of strictly positive entries. It follows that M_j is scrambling. □

In particular, for every primitive random substitution of constant length on a binary alphabet, the almost sure diffraction does not contain a singular continuous component. More generally, if more marginals are added to a primitive, constant-length random substitution, it may gain, but never loose, the scrambling property. Heuristically, it appears that singular continuous diffraction components are not robust under fluctuations.

Remark 4.3.30. If M is a scrambling matrix, it is straightforward to verify that, for some power n, M^n has a strictly positive column. Hence, ϑ has a scrambling power ϑ^m if and only if there is a power ϑ^n, a letter $a \in \mathcal{A}$ and a position j such that $\mathbb{P}[\vartheta^n(b)_j = a] > 0$ for all $b \in \mathcal{A}$. This is a natural generalization of Dekking's coincidence condition for a *deterministic* substitution ϱ, requiring that for some power ϱ^n, $a \in \mathcal{A}$ and position j, we have $\varrho^n(b)_j = a$ for all $b \in \mathcal{A}$ [Dek78]. More precisely, there is a scrambling power of ϑ precisely if there exists a marginal of some power of ϑ that satisfies the coincidence condition. ◊

Example 4.3.31. As soon as we leave the realm of binary alphabets, both scrambling and the conclusion of Theorem 4.3.27 can fail. We illustrate this with an example that is inspired by the classic Thue–Morse substitution. Let ϑ be a random substitution on $\mathcal{A} = \{a, b_1, b_2\}$ with $a \mapsto \{ab_1, ab_2\}$ and $b_i \mapsto \{b_i a\}$ for each $i \in \{1, 2\}$. Choosing a weight vector τ with $\tau_b := \tau_{b_1} = \tau_{b_2}$ makes the distinction between b_1 and b_2 invisible and hence "factors" naturally to the original Thue–Morse sequence, which is known to have a purely singular continuous diffraction measure if $\tau_a = -\tau_b$. This measure will be studied in detail in Chapter 6. ◊

4.3.3. Integer inflation factor

Using mean-almost periodicity to establish that $\widehat{\gamma}^{\mathrm{ex}}$ is a pure point measure is a mechanism that works also outside the constant-length setting. In fact, for deterministic substitutons, there are several algorithms to establish mean-almost periodicity, if the characteristic polynomial of the substitution matrix is irreducible, and if the inflation factor λ is a *Pisot number*, that is, if all the algebraic conjugates of λ are smaller than 1 in modulus. This is related to the famous *Pisot substitution conjecture* and hence there is a vast literature on this and related problems; for an overview we refer to [ABB+15, Sing06] and references therein. There is a natural generalization of the overlap algorithm [SS02, Sol97] to the random setting. However, this requires some additional technical preparation that is outside the scope of this thesis; we refer to the forthcoming work [GMR]. Here, we illustrate how the results on primitive, constant-length random substitutions can be extended beyond the constant-length case if the PF eigenvalue λ is an integer.

Given a geometrically compatible, primitive random substitution ϑ on $\mathcal{A}$, with PF eigenvalue $\lambda \in \mathbb{N}$, we associate a constant-length random substitution ϑ' on a new alphabet $\mathcal{A}'$. We call ϑ' the *pure core* of ϑ, in line with [BBJS12, Sec. 4], where this term was used in the context of deterministic substitution. For this construction, we again rely on the geometric picture; compare Section 3.1.2. First, identify the largest rational number r such that, for every $a \in \mathcal{A}$, there is a natural number $n_a \in \mathbb{N}$ with $L_a = n_a r$. The main idea is to slice the interval I_a of length L_a into n_a smaller intervals of length r each. This replaces each letter $a \in \mathcal{A}$ by a concatenation of letters $a_1 \cdots a_{n_a}$. Formally, we introduce the new alphabet $\mathcal{A}' = \{a_1, \ldots, a_{n_a} : a \in \mathcal{A}\}$ and a morphism $\varphi\colon \mathcal{A}^+ \to (\mathcal{A}')^+$, defined via $\varphi(a) = a_1 \cdots a_{n_a}$ for all $a \in \mathcal{A}$. This construction ensures that the symbolic length of $\varphi(u)$ for some $u \in \mathcal{A}^+$,

$$|\varphi(u)| \,=\, \sum_{n=1}^{|u|} |\varphi(u_n)| \,=\, \frac{1}{r} \sum_{n=1}^{|u|} L_{u_n} \,=\, \frac{1}{r} L(u),$$

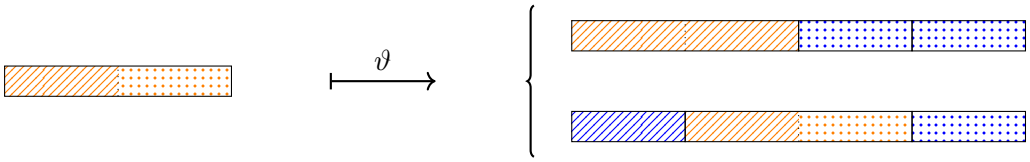

Figure 4.1.: Deriving $\vartheta'(a_1) = \{a_1a_2, b_1a_1\}$ and $\vartheta'(a_2) = \{b_1b_1, a_2b_1\}$ from $\vartheta(a) = \{abb, bab\}$ in Example 4.3.33. Tiles of type a and b are coloured in orange and blue, respectively. Dotted lines indicate the splitting $a \mapsto a_1a_2$. Different types of decorations correspond to a_1 and a_2 on the left hand sides and their respective images on the right hand side.

is the same as the *geometric* length of u, up to a universal factor. By geometric compatibility, each $u \in \vartheta(a)$ satisfies $L(u) = \lambda L_a$ and hence $|\varphi(u)| = \lambda n_a$. We can therefore consistently define a random substitution ϑ' on $\mathcal{A}'$, via

$$\vartheta'(a_m) \,=\, \varphi(\vartheta(a))_{[(m-1)\lambda+1,m\lambda]}, \tag{4.15}$$

for all $a \in \mathcal{A}$ and $1 \leqslant m \leqslant n_a$, to be understood either as an equality of sets or equality of distributions. We emphasize that the set $\varphi(\vartheta(a))$ is contained in $\vartheta'(\varphi(a))$, but it is in general *not* equal. In that sense, the situation is more subtle than for deterministic substitutions.

Remark 4.3.32. Alternatively, we could define (4.15) as an equality of random variables. In that case, ϑ' is modelled on the same probability space as ϑ, but it no longer acts independently on neighbouring letters. Hence, ϑ' could no longer be regarded as a random substitution in the sense defined here, but still fits in the framework provided by Peyriére [Pey80]. This bears some resemblance to the *induced random substitution* [GS20]. For our discussion below, depending only on marginal distributions, both approaches yield the same expressions. ◊

Example 4.3.33. Consider the random substitution

$$\vartheta\colon a \mapsto \begin{cases} abb, & p, \\ bab, & 1-p, \end{cases} \qquad b \mapsto a,$$

which is primitive and geometrically compatible, with $\lambda = 2$ and L a multiple of $(2, 1)$. The diffraction of ϑ was discussed in detail in [Goh17]. Its pure core ϑ' on $\mathcal{A}' = \{a_1, a_2, b_1\}$ is given by

$$\vartheta'\colon a_1 \mapsto \begin{cases} a_1a_2, & p, \\ b_1a_1, & 1-p, \end{cases} \qquad a_2 \mapsto \begin{cases} b_1b_1, & p, \\ a_2b_1, & 1-p, \end{cases} \qquad b_1 \mapsto a_1a_2.$$

For an illustration, compare Figure 4.1. Note that $\vartheta'(\phi(a)) = \vartheta'(a_1a_2)$ contains the words $b_1a_1b_1b_1$ and $a_1a_2a_2b_1$, which do *not* correspond to words in $\vartheta(a)$. We observe that $(\vartheta')^2$ is scrambling. ◊

The pure core ϑ' of ϑ can be used to model the action of ϑ on expected Dirac combs. Since a letter of length 1 in $\mathcal{A}'$ corresponds to a subinterval of length r of some interval I_a, $a \in \mathcal{A}$,

we first need to apply a rescaling. Given a Radon measure ω on $\mathbb{R}$, we denote by $r.\omega$ the Radon measure with

$$(r.\omega)(f) \;=\; \int_{\mathbb{R}} f(rx)\,\mathrm{d}\omega(x),$$

for all $f \in C_c(\mathbb{R})$. With slight abuse of notation, we use the same symbols $\mathbb{P}$ and $\mathbb{E}$ in the context of formally different probability spaces in the following.

Lemma 4.3.34. *Let τ, τ' be weight vectors on $\mathcal{A}, \mathcal{A}'$, respectively, such that $\tau'_{a_m} = \tau_a$ if $m = 1$, and $\tau'_{a_m} = 0$ otherwise, for all $a \in \mathcal{A}$ and $1 \leqslant m \leqslant n_a$. Then, the corresponding weighted Dirac combs ω, ω' satisfy, for every $a \in \mathcal{A}^+$ and $n \in \mathbb{N}$,*

$$\mathbb{E}[\omega(\vartheta^n(a))] \;=\; r.\mathbb{E}[\omega'((\vartheta')^n(\varphi(a)))].$$

Sketch of proof. The pure core ϑ' of ϑ is constructed in such a way that $\vartheta'(a_m)$ corresponds to the restriction of the (typed) point set $\Lambda(\vartheta(a))$ to the interval $(r(m-1)\lambda, rm\lambda]$. At the same time, it determines the restriction of $\Lambda(\vartheta'(\phi(a)))$ to $((m-1)\lambda, m\lambda]$ by the constant-length property. The same holds for higher powers of the substitution ϑ. Although $(\vartheta^n)'$ and $(\vartheta')^n$ do not coincide as random substitutions, they have the same marginal distributions, that is,

$$\mathbb{P}[(\vartheta')^n(a)_j = b] \;=\; \mathbb{P}[(\vartheta^n)'(a)_j = b],$$

for all $a, b \in \mathcal{A}'$, $n \in \mathbb{N}$, and $1 \leqslant j \leqslant \lambda^n n_a$. Note that the choice of τ' erases all control points that do not correspond to the starting point of an interval I_a, with $a \in \mathcal{A}$. From these observations, the relation follows in a straightforward manner. The full details of the proof will appear in [GMR]. □

Corollary 4.3.35. *Let ϑ be a geometrically compatible primitive random substitution with integer PF eigenvalue λ. If some power of the pure core ϑ' is scrambling, then $\widehat{\gamma} = \widehat{\gamma}_{\mathrm{pp}} + \widehat{\gamma}_{\mathrm{ac}}$ and*

$$\widehat{\gamma}_{\mathrm{pp}} \;=\; \lim_{n\to\infty} \widehat{\gamma}^{\mathrm{ex}}(\vartheta^n(a)), \quad \widehat{\gamma}_{\mathrm{ac}} \;=\; \lim_{n\to\infty} \widehat{\gamma}^{\mathrm{var}}(\vartheta^n(a)),$$

for all $a \in \mathcal{A}$.

Proof. This follows in principle from Lemma 4.3.34 and Theorem 4.3.27. We want to apply the reasoning of Theorem 4.3.27 to the constant-length random substitution ϑ' to obtain that the expected Dirac combs of large inflation words give rise to a pure point diffraction measure. The only subtlety is that, in Lemma 4.3.34, we start iterating from a word $\varphi(a) = a_1 \cdots a_{n_a}$ instead of a single letter. However, we verify that the proof of Theorem 4.3.27 readily generalizes to that case. □

5. Schrödinger operators

5.1. Schrödinger operators and symbolic dynamics

In this chapter, we consider discrete Schrödinger operators $H_V \colon \ell^2(\mathbb{Z}) \to \ell^2(\mathbb{Z})$, defined by

$$(H_V\psi)_n \;=\; \psi_{n+1} + \psi_{n-1} + V_n\psi_n,$$

for some bounded potential $V = (V_n)_{n\in\mathbb{Z}} \in \mathbb{R}^{\mathbb{Z}}$. The operator H_V is self-adjoint and bounded by the assumption that V is bounded. Its spectrum $\sigma(H_V)$ is hence a compact subset of $\mathbb{R}$. The potential V models the influence of a surrounding material. We will mostly be concerned with the case that this material exhibits some kind of aperiodic order. More precisely, we start with a symbolic subshift $(\mathbb{X}, S)$, where each sequence $x \in \mathbb{X}$ serves as a choice for some configuration. To obtain the corresponding potential $V = V(x)$, we choose a bounded continuous function $f \colon \mathbb{X} \to \mathbb{R}$ and set $V_n(x) = f(S^n(x))$, for all $n \in \mathbb{Z}$. It is natural to assume that $V_0(x) = f(x)$ is most sensitive to changes in x that happen near the origin. Often, it is even assumed that $f(x) = g(x_0)$, for some $g \colon \mathcal{A} \to \mathbb{R}$, depends only on one coordinate. In general, we make the somewhat weaker assumption that f depends only on a finite window around the origin, that is, it is locally constant.
To summarize, we are concerned with a family of Schrödinger operators $\{H_{V(x)}\}_{x\in\mathbb{X}}$ associated to a symbolic subshift $(\mathbb{X}, S)$, satisfying

$$(H_{V(x)}\psi)_n \;=\; \psi_{n+1} + \psi_{n-1} + V_n(x)\psi_n,$$

for all $x \in \mathbb{X}$ and $\psi \in \ell^2(\mathbb{Z})$, where $V_n(x) = f(S^n x)$ for some locally constant function $f \colon \mathbb{X} \to \mathbb{R}$. More generally, the subshift $(\mathbb{X}, S)$ can be replaced by any (topological) dynamical system. This puts us in the framework of Schrödinger operators with dynamically defined potentials; compare [Dam17] for a review on this topic. With the exception of Section 5.3.1, we will stick to the case that $(\mathbb{X}, S)$ is a subshift. For a comprehensive introduction to the spectral theory of discrete Schrödinger operators, we refer the interested reader to the upcoming monographs [DFa, DFb].

The pivotal role of Schrödinger operators in quantum mechanics comes from the fact that they drive the equation of motion for quantum states. In our current formulation, the sequence $\psi \in \ell^2(\mathbb{Z})$ models the *wave function* of a valence electron at a given point in time ($t = 0$) if the normalization condition $||\psi||_2 = 1$ is satisfied. This normalization reflects that, for $n \in \mathbb{N}$, the value ψ_n is a *probability amplitude*, that is, $|\psi_n|^2$ is the probability of finding the particle at position n. We are interested in the *time evolution* $\{\psi(t)\}_{t\in\mathbb{R}}$, where $t \in \mathbb{R}$ is a time parameter, $\psi(0) = \psi$ and $\psi(t) \in \ell^2(\mathbb{Z})$ for all $t \in \mathbb{R}$ satisfies the time-dependent Schrödinger equation

$$\mathrm{i}\partial_t\psi(t) \;=\; H_V\psi(t),$$

which implicitly depends on the potential V. The long-term behaviour of $\psi(t)$ can be characterized in terms of spectral properties of H_V. This is summarized in the celebrated *RAGE* theorem, termed this way by Barry Simon due to the important contributions of Ruelle [Rue69a], Amrein–Georgescu [AG73] and Enss [Ens78]. In the following formulation, it is taken from [Dam17, Thm. 2.1].

Theorem 5.1.1. *Let $\psi(0) = \psi \in \ell^2(\mathbb{Z})$ and $\psi(t)$ a solution to the Schrödinger equation $\mathrm{i}\partial_t\psi(t) = H_V\psi(t)$. If μ_ψ is the spectral measure corresponding to the pair (H_V, ψ), we have*

(1) *$\mu_\psi = \mu_{\psi,\mathrm{pp}}$ iff, for all $\varepsilon > 0$, there is $N \in \mathbb{N}_0$ such that, for all $t \in \mathbb{R}$,*

$$\sum_{|n|\geqslant N} |\psi_n(t)|^2 \, < \, \varepsilon.$$

(2) *$\mu_\psi = \mu_{\psi,c}$ iff, for all $N \in \mathbb{N}_0$,*

$$\lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} \sum_{|n|\leqslant N} |\psi_n(t)|^2 \, \mathrm{d}t \, = \, 0.$$

(3) *$\mu_\psi = \mu_{\psi,\mathrm{ac}}$ implies that*

$$\lim_{|t|\to\infty} \sum_{|n|\leqslant N} |\psi_n(t)|^2 \, = \, 0,$$

for all $N \in \mathbb{N}_0$. ∎

In words, eigenvalues lead to localized solutions, continuous spectral measures correspond to solutions that flow out of every given box on time average, and if the spectral measure is absolutely continuous, the corresponding solution eventually escapes to infinity.

This explains why we are not only interested in the spectrum as a set but also want to understand which components are present in the Lebesgue decomposition of the measure of maximal spectral type ν. Arguably, the best quantitative understanding of the spectral data has been achieved in the case that V is a periodic sequence. In this case, $\nu = \nu_{\mathrm{ac}}$ and $\sigma(H_V)$ is a finite union of closed intervals that can be calculated explicitly. On the other hand, if $(V_n)_{n\in\mathbb{N}}$ are iid random variables, drawn from a finite set, $\sigma(H_V)$ still has finitely many gaps but $\nu = \nu_{\mathrm{pp}}$ for almost-every choice of the potential. This nourishes the heuristic that the spectral type becomes more singular the more disorder is introduced to the potential (note that this is in stark contrast to diffraction). There is a long standing quest to make this intuition more precise, relating structural properties of V to spectral properties of the associated Schrödinger operator H_V.

Coming back to the family $\{H_{V(x)}\}_{x\in\mathbb{X}}$, we can either start from structural information on individual points x or from properties that pertain to the whole subshift $(\mathbb{X}, S)$. In the latter case, we would like to establish uniform spectral results on $H_{V(x)}$, either for all $x \in \mathbb{X}$ or at least for elements $x \in \mathbb{X}$ that are in some sense *typical* for the subshift. Standard strong-approximation arguments [ReS80, Thm. VIII.24] readily provide the following, well-known tool for the comparison of the sets $\{\sigma(H_{V(x)})\}_{x\in\mathbb{X}}$.

Lemma 5.1.2. *Assume that $(\mathbb{X}, S)$ is a symbolic subshift and that $x, y \in \mathbb{X}$ are such that x is contained in the orbit closure of y under the shift map. Then, $\sigma(H_{V(x)}) \subset \sigma(H_{V(y)})$.* ∎

In particular, we have $\sigma(H_{V(x)}) \subset \sigma(H_{V(y)})$ for all $x \in \mathbb{X}$ if the orbit of y is dense in $\mathbb{X}$ and $\sigma(H_{V(x)})$ is uniformly constant if $(\mathbb{X}, S)$ is minimal. A similar relation exists for the ac-part, with the inclusion in the opposite direction [DFa, Ch. 2]. However, none of these relation extends to the spectral components $\sigma_{\mathrm{pp}}(H_{V(x)})$ and $\sigma_{\mathrm{sc}}(H_{V(x)})$. Still, given an ergodic probability measure μ on $(\mathbb{X}, S)$, it is a classic result that each of the spectral components is constant for μ-almost every $x \in \mathbb{X}$. This was generalized recently to *infinite* ergodic measures, which will become important in the next section. The following is a corollary of a more general theorem [BDFL21], adapted to our setting.

Theorem 5.1.3. *Let μ be a σ-finite, non-atomic ergodic measure on a subshift $(\mathbb{X}, S)$. Then, there are compact sets $\Sigma, \Sigma_{\mathrm{pp}}, \Sigma_{\mathrm{sc}}, \Sigma_{\mathrm{ac}} \subset \mathbb{R}$ such that for μ-almost every $x \in \mathbb{X}$, we have $\sigma_\bullet(H_{V(x)}) = \Sigma_\bullet$ for each $\bullet \in \{\varnothing, \mathrm{pp}, \mathrm{sc}, \mathrm{ac}\}$.* ∎

We call $E \in \mathbb{R}$ a *generalized eigenvalue* of $H_{V(x)}$ if there exists a polynomially bounded solution $\phi \in \mathbb{C}^{\mathbb{Z}}$ to the difference equation

$$\phi_{n-1} + \phi_{n+1} + f(S^n x)\phi_n \;=\; E\phi_n, \tag{5.1}$$

for all $n \in \mathbb{Z}$. We call such a sequence ϕ a *generalized eigensolution*. The spectrum $\sigma(H_{V(x)})$ coincides with the closure of the set of generalized eigenvalues of $H_{V(x)}$. The sets $\sigma_{\mathrm{pp}}(H_{V(x)}), \sigma_{\mathrm{sc}}(H_{V(x)}), \sigma_{\mathrm{ac}}(H_{V(x)})$ can be characterized in a similar manner, using more refined control on the growth behaviour of the generalized eigensolutions [DFa]. With a slight shift of perspective, we can rewrite (5.1) as a matrix equation

$$\begin{pmatrix} \phi_{n+1} \\ \phi_n \end{pmatrix} \;=\; A_E(S^n x) \begin{pmatrix} \phi_n \\ \phi_{n-1} \end{pmatrix}, \quad A_E(S^n x) \;=\; \begin{pmatrix} E - f(S^n x) & -1 \\ 1 & 0 \end{pmatrix},$$

for all $n \in \mathbb{Z}$. Hence, in order to determine the growth behaviour of ϕ, we can alternatively study the product of matrices $A_E^n(x)$, where

$$A_E^n(x) \;=\; \begin{cases} A_E(S^{n-1}x) \cdots A_E(x), & \text{for } n > 0, \\ \mathrm{Id}, & \text{for } n = 0, \\ A_E(S^{-n}x)^{-1} \cdots A_E(S^{-1}x)^{-1}, & \text{for } n < 0. \end{cases}$$

This observation is at the heart of many results that relate the structure of x to spectral properties of $H_{V(x)}$. The matrix product $A_E^n(x)$ can be obtained by sampling the single $\mathrm{SL}(2, \mathbb{R})$-valued matrix function $A_E \colon x \mapsto A_E(x)$ along the orbit of x. This leads to the following concept.

Definition 5.1.4. Let $E \in \mathbb{R}$. The corresponding (Schrödinger) *cocycle* is given by the skew product

$$(S, A_E) \colon \mathbb{X} \times \mathbb{R}^2 \to \mathbb{X} \times \mathbb{R}^2, \quad (x, v) \mapsto (Sx, A_E(x)v).$$

We call (S, A_E) *uniformly hyperbolic* if there are $c > 0$ and $r > 1$ such that $||A_E^n(x)|| \geqslant cr^{|n|}$, for all $n \in \mathbb{Z}$ and $x \in \mathbb{X}$. Further, we set

$$\mathcal{UH} \;=\; \{E \in \mathbb{R} : (S, A_E) \text{ is uniformly hyperbolic}\}.$$

Cocycles are a classic tool in the spectral theory of Schrödinger operators [DFa]. They have also been shown to be useful in the context of diffraction theory recently [Mañ19].

It is easily verified that $(S, A_E)^n = (S^n, A_E^n)$ holds for all $n \in \mathbb{Z}$ and hence that uniform hyperbolicity is genuinely a property of the cocycle (S, A_E). There are several useful equivalent conditions for (S, A_E) to be uniformly hyperbolic; we refer to [DFa] for an overview. A hyperbolic cocycle (S, A_E) corresponds to solutions to the difference equation (5.1) that are exponentially increasing in at least one direction, ruling out E as a Schrödinger eigenvalue. In fact, we have the following useful characterization, going back to Johnson [Joh86]; compare also [DEG15, Thm. 4.6].

Theorem 5.1.5. *For every $x \in \mathbb{X}$ with a dense orbit in $\mathbb{X}$, we have $\sigma(H_{V(x)}) = \mathbb{R} \setminus \mathcal{UH}$.* ∎

In general, a particularly simple description is available for the absolutely continuous component $\sigma_{\mathrm{ac}}(H_{V(x)})$ of the spectrum. First, let us assume that x, and hence $V(x)$, is a periodic sequence. If p denotes the period of $V(x)$, we have by standard arguments [DFa, DFb] that

$$\sigma(H_{V(x)}) \,=\, \sigma_{\mathrm{ac}}(H_{V(x)}) \,=\, \bigl\{E \in \mathbb{R} : \mathrm{tr}\bigl(A_E^p(x)\bigr) \in [-2,2]\bigr\}.$$

Due to a celebrated result by Remling [Rem11], $\sigma_{\mathrm{ac}}(H_{V(x)})$ is non-empty precisely if $V(x)$ is eventually periodic. If $V(x)$ is eventually periodic to the right, there is an eventual period $p \in \mathbb{N}$, such that $V^+ = \lim_{n\to\infty} S^{np}V$ is periodic. Similarly, $V^- = \lim_{n\to\infty} S^{-nq}V$ exists and is periodic for a unique $q \in \mathbb{N}$, if $V(x)$ is eventually periodic to the left. We have

- $\sigma_{\mathrm{ac}}(H_{V(x)}) = \sigma(H_{V^+})$, if $V(x)$ is only eventually periodic to the right.
- $\sigma_{\mathrm{ac}}(H_{V(x)}) = \sigma(H_{V^-})$, if $V(x)$ is only eventually periodic to the left.
- $\sigma_{\mathrm{ac}}(H_{V(x)}) = \sigma(H_{V^+}) \cup \sigma(H_{V^-})$, if $V(x)$ is eventually periodic to both sides.

Often, we consider $V(x)$ that are not eventually periodic and hence $\sigma_{\mathrm{ac}}(H_{V(x)}) = \varnothing$ will be the typical situation. This holds uniformly if $(\mathbb{X}, S)$ is a non-periodic subshift that arises from a primitive substitution ϱ. Since $(\mathbb{X}, S)$ is minimal, we have $\Sigma = \sigma(H_{V(x)})$ for all $x \in \mathbb{X}$ in this situation. A remarkable feature of this model for aperiodic order is that Σ is always a Cantor set of Lebesgue measure 0. This was originally proved using *trace maps* [BG94, LTWW02]. Later, it was shown by Damanik and Lenz [DL06a, DL06c] that this phenomenon holds for the more general class of Boshernitzan subshifts. Recall from Definition 2.2.9 that $(\mathbb{X}, S)$ is said to satisfy *Boshernitzan's condition* if there exists an invariant measure μ on $(\mathbb{X}, S)$ such that

$$\limsup_{n\to\infty} n\underline{\mu}(n) \,>\, 0, \qquad \text{with} \qquad \underline{\mu}(n) \,=\, \min\{\mu([v]) : v \in \mathcal{L}_n(\mathbb{X})\}.$$

and that $(\mathbb{X}, S)$ is called a *Boshernitzan subshift* if, in addition, it is also minimal. Boshernitzan's condition has strong implications on the growth behaviour of $A_E^n(x)$ for every $x \in \mathbb{X}$ and $E \in \sigma(H_{V(x)})$. This can be used to conclude the following.

Theorem 5.1.6 ([DL06a])**.** *Let $(\mathbb{X}, S)$ be a Boshernitzan subshift. Then, for every $x \in \mathbb{X}$, the spectrum $\Sigma = \sigma(H_{V(x)})$ is a Cantor set of Lebesgue measure 0 for all $x \in \mathbb{X}$, unless $V(x)$ is periodic.* ∎

Remark 5.1.7. Primitive subshifts satisfy the even stronger assumption $\lim_{n\to\infty} n\underline{\mu}(n) > 0$. This is consistent with the observation that the complexity function corresponding to such a subshift grows at most linearly. In general, this need not hold for a Boshernitzan subshift, since $n_k\underline{\mu}(n_k)$ needs to be bounded away from 0 only on a *subsequence* $(n_k)_{k\in\mathbb{N}}$ of natural numbers. In fact, it was shown recently that there are Boshernitzan subshifts such that the corresponding complexity function grows faster than any polynomial on an appropriate subsequence [CK21]. Boshernitzan's criterion covers the vast majority of subshifts for which Cantor spectrum of Lebesgue measure 0 is known. Notable exceptions occur in the context of *Toeplitz sequences* [LQ11, LQ12] and for the recently introduced class of *leading sequences* [GLNS19]. ◊

For the Fibonacci substitution subshift, the trace map formalism [AP86, KKT83], together with tools from hyperbolic dynamics, offers a route to a more refined analysis of the spectrum. This includes estimates for the Hausdorff dimension of the spectrum Σ [DG11], as well as establishing the uniform absence of Schrödinger eigenvalues [DL99]. To the best of the author's knowledge, there are no known examples of sequences x from a primitive substitution subshift $(\mathbb{X}, S)$ such that $H_{V(x)}$ admits an eigenvalue. Still, proving absence of eigenvalues for general primitive substitution subshifts turns out to be notoriously difficult. It is customary to resort to structural criteria that guarantee the absence of Schrödinger eigenvalues at least for a subset of $\mathbb{X}$. We present two particularly prominent conditions in the following. These are most naturally formulated for the potential sequence $V = V(x)$ instead of x. Since f is assumed to be locally constant, the set $R = f(\mathbb{X})$ is finite and hence $V(x) \in R^{\mathbb{Z}}$ is a sequence over the finite alphabet $R \subset \mathbb{R}$, for each $x \in \mathbb{X}$. In fact, the function $V\colon x \mapsto V(x)$ is a sliding block code; compare Definition 2.1.7.

Given a finite alphabet $\mathcal{A}$, and a word $u = u_1 \cdots u_n \in \mathcal{A}^+$, we define the reflected word $\widetilde{u} = u_n \cdots u_1 \in \mathcal{A}^+$ by reversing the order of the letters in u. We call u a *palindrome* if $u = \widetilde{u}$. If $x_{[\ell,k]} = u$ for some sequence x and $\ell, k \in \mathbb{Z}$, we say that u is *centered* at $(\ell + k)/2$ in x.

Definition 5.1.8. Let $B > 1$. A sequence $y \in \mathcal{A}^{\mathbb{Z}}$ is called *B-strongly palindromic* if there exists a sequence $(P_n)_{n\in\mathbb{N}}$ of palindromes, centered at $c_n > 0$ in y, satisfying $c_n \to \infty$ as $n \to \infty$, and

$$\lim_{n\to\infty} \frac{B^{c_n}}{|P_n|} = 0.$$

We call y *strongly palindromic* if it is B-strongly palindromic for all $B > 1$.

Theorem 5.1.9 ([HKS95])**.** *If $V \in \mathcal{R}^{\mathbb{Z}}$ is a strongly palindromic sequence, H_V has no eigenvalues.* ■

If V is B-strongly palindromic for some, but not all $B > 1$, it is sometimes still possible to exclude eigenvalues on a subset of $\sigma(H_V)$; compare our discussion in [EG21] for details. Another useful symmetry that helps to exclude eigenvalues is a local threefold repetition property on arbitrary scales.

Definition 5.1.10. We call $V \in R^{\mathbb{Z}}$ a *Gordon potential*, if there exists a sequence of words $(v_n)_{n\in\mathbb{N}} \in R^+$ such that $\lim_{n\to\infty} |v_n| = \infty$ and for all $n \in \mathbb{N}$,

$$V \in [v_n.v_nv_n].$$

The following statement, known as the *Gordon lemma*, goes back to a result by Gordon [Gor76], and is textbook material [DFa] by now.

Lemma 5.1.11. *For every Gordon potential V, the corresponding Schrödinger operator H_V admits no eigenvalues.* ∎

Remark 5.1.12. Often, if $H_{V(x)}$ has no eigenvalues for a *single* $x \in \mathbb{X}$, this property can be extended to a relevant subset of $\mathbb{X}$, especially if $(\mathbb{X}, S)$ is strictly ergodic. To explain this, we first note that the set of eigenvalues of H_V is invariant under a shift of the potential. Hence, the set

$$\mathcal{E} \;=\; \{x \in \mathbb{X}_\varrho : H_{V(x)} \text{ has no eigenvalues}\}$$

is shift-invariant and therefore is either null or conull with respect to every ergodic measure μ on $(\mathbb{X}, S)$. That is, μ-almost everywhere absence of eigenvalues is equivalent to $\mu(\mathcal{E}) > 0$, which is sometimes easier to establish. On a different note, $\mathcal{E}$ is a G_δ set due to a classic result by Simon [Sim95]. Hence, in order to show that $\mathcal{E}$ is *generic* (a dense G_δ set), it suffices to show that $H_{V(x)}$ has no eigenvalues for some $x \in \mathbb{X}$ with a dense S-orbit. ◊

5.1.1. Main results of this chapter

Let us come back to the heuristic that eigenvalues of $H_{V(x)}$ are indicative of structural disorder in x. The apparent absence of eigenvalues for primitive substitution subshifts $(\mathbb{X}, S)$ is then maybe not too surprising, given that $(\mathbb{X}, S)$ is of low complexity in more than one direction: in a topological sense (minimal), a measure-theoretic sense (uniquely ergodic), as well as in a combinatorial sense (at most linear growth of the complexity function). Almost minimal substitutions provide a class of subshifts that are slightly more complex in all of these senses. They are therefore natural candidates to probe the details of the aforementioned heuristic. In fact, somewhat surprisingly, it is possible to produce Schrödinger eigenvalues in this framework.

Theorem 5.1.13. *There is an almost minimal substitution ϱ, with an associated ergodic subshift $(\mathbb{X}_\varrho, S, \mu)$ and a sliding block potential function V on $\mathbb{X}_\varrho$, with the following properties.*

- ***(1)*** *The almost sure spectral sets satisfy $\Sigma = \Sigma_{\mathrm{sc}}$ and $\Sigma_{\mathrm{pp}} = \Sigma_{\mathrm{ac}} = \varnothing$.*
- ***(2)*** *There exists a point $x \in \mathbb{X}_\varrho$ such that $\sigma(H_{V(x)}) = \Sigma$, and $H_{V(x)}$ admits an eigenvalue that is an accumulation point of the spectrum.*

In Section 5.3, we investigate the influence of a periodic background, added to the potential arising from a subshift $(\mathbb{X}, S)$. This is modelled via a dynamical system of the form $(\mathbb{X}\times\mathbb{Z}_p, T)$, where $\mathbb{Z}_p = \mathbb{Z}/p\mathbb{Z}$ is the cyclic group, equipped with addition modulo p, and T is defined as $T(x, m) = (Sx, m + 1)$, for all $(x, m) \in \mathbb{X} \times \mathbb{Z}_p$. At first, we show that if $(\mathbb{X}, S)$ is a Boshernitzan subshift, the system $(\mathbb{X}\times\mathbb{Z}_p, T)$ inherits some of its essential properties, leading to the following result.

Theorem 5.1.14. *Let $(\mathbb{X}, S)$ be a Boshernitzan subshift and $f\colon \mathbb{X}\times\mathbb{Z}_p \to \mathbb{R}$ a locally constant function, with corresponding potential $V(y) = (f(T^n y))_{n\in\mathbb{Z}}$, for all $y \in \mathbb{X} \times \mathbb{Z}_p$. Then, for*

every $y \in \mathbb{X} \times Z_p$, the spectrum $\sigma(H_{V(y)})$ is a Cantor set of Lebesgue measure 0, unless $V(y)$ is a periodic sequence. Further,

$$\#\{\sigma(H_{V(y)}) : y \in \mathbb{X} \times \mathbb{Z}_p\} \leqslant s(p),$$

where $s\colon \mathbb{N} \to \mathbb{N}$ denotes the D-function of the minimal system $(\mathbb{X}, S)$.

In Lemma 5.3.6, we obtain a modification of Gordon's criterion that covers the product space $(\mathbb{X} \times \mathbb{Z}_p, T)$. This provides some examples, where eigenvalues can be excluded almost surely.

If $\mathbb{X} = \mathcal{A}^{\mathbb{Z}}$ is the full shift, we expect the Schrödinger operators on $(\mathcal{A}^{\mathbb{Z}} \times \mathbb{Z}_p, T)$ to give rise to spectra with only a finite number of gaps, almost surely. In the case $p = 2$, we show the stronger result that the almost sure spectrum can be obtained from finitely many periodic points.

Theorem 5.1.15. *Let $\mathcal{A}$ be a finite alphabet and $(\mathcal{A}^{\mathbb{Z}}, S)$ the full shift on $\mathcal{A}$. Assume that $f\colon \mathcal{A}^{\mathbb{Z}} \times \mathbb{Z}_p \to \mathbb{R}$, satisfies $f(x, m) = f(x', m)$ for all $x, x' \in \mathcal{A}^{\mathbb{Z}}$ with $x_0 = x'_0$ and $m \in \mathbb{Z}_p$. Let $V(y) = (f(T^n y))_{n\in\mathbb{Z}}$ for all $y \in \mathcal{A}^{\mathbb{Z}} \times \mathbb{Z}_p$. Then, for every point y with a dense T-orbit in $\mathbb{X} \times \mathbb{Z}_p$,*

$$\sigma(H_{V(y)}) \;=\; \bigcup_{a,b\in\mathcal{A}} \sigma(H_{V((ab)^{\mathbb{Z}},0)}).$$

This uses a characterization of uniformly hyperbolic cocycles that was presented in [ABY10].

5.2. Almost minimal substitutions

In this section, we are interested in Schrödinger operators associated with almost minimal substitutions. More precisely, we consider the collection Sams of simple almost minimal substitutions, as defined in Definition 3.1.16. For the following, let $p, r \geqslant 2$ and $1 \leqslant j \leqslant r-1$ be arbitrary but fixed and let $\varrho = \varrho_{p,r,j}$, with

$$\varrho_{p,r,j}\colon a \mapsto a^p, b \mapsto b^j a b^{r-j}$$

be the corresponding substitution in Sams. Iterating ϱ on a yields $\varrho^n(a) = a^{p^n}$ for all $n \in \mathbb{N}_0$. Starting from b, the substitution rule satisfies the recursion

$$\varrho^{n+1}(b) \;=\; (\varrho^n(b))^j a^{p^n} (\varrho^n(b))^{r-j},$$

for all $n \in \mathbb{N}_0$. Let μ be the unique non-atomic ergodic measure on $(\mathbb{X}_\varrho, S)$. Given $x \in \mathbb{X}_\varrho$, recall that $V(x) \in \mathbb{R}^{\mathbb{Z}}$ satisfies $V_n(x) = f(S^n x)$, for all $n \in \mathbb{N}$ and some locally constant $f\colon \mathbb{X} \to \mathbb{R}$. To avoid trivialities, we assume the following throughout.

Assumption 5.2.1. The function $f\colon \mathbb{X} \to \mathbb{R}$ is non-constant, that is, $R = f(\mathbb{X})$ consists of more than one point.

Lemma 5.2.2. *The subshift $(V(\mathbb{X}_\varrho), S)$ is almost minimal and infinite. Further, the eventually periodic sequences in $V(\mathbb{X}_\varrho)$ are in fact eventually constant and precisely the images of the eventually constant sequences in $\mathbb{X}_\varrho$.*

Proof. Since V is a topological factor map from $(\mathbb{X}, S)$ to $(V(\mathbb{X}), S)$, almost minimality is clearly preserved. As f is non-constant, we have $\#V(\mathbb{X}_\varrho) > 1$. For a moment, suppose that $\#V(\mathbb{X}_\varrho) = n > 1$ is finite. Then, almost minimality implies that $V(\mathbb{X}_\varrho)$ consists of a single n-periodic orbit under the shift map, contradicting the fact that $V(a^{\mathbb{Z}})$ is a fixed point under S. By the same argument, all eventually periodic sequences in $V(\mathbb{X}_\varrho)$ are in fact eventually constant. Finally, if $x \in \mathbb{X}_\varrho$ is *not* eventually constant, it contains every allowed word infinitely often to both sides of the origin. This forbids $V(x)$ to be eventually constant. □

Corollary 5.2.3. *There are compact sets $\Sigma, \Sigma_{\mathrm{pp}}, \Sigma_{\mathrm{sc}}, \Sigma_{\mathrm{ac}} \subset \mathbb{R}$ such that, for μ-almost every $x \in \mathbb{X}$, we have $\sigma_\bullet(H_{V(x)}) = \Sigma_\bullet$ for each $\bullet \in \{\varnothing, \mathrm{pp}, \mathrm{sc}, \mathrm{ac}\}$. Further, we have the following.*

- ***(1)*** *The almost sure spectrum is in fact almost uniform, in the sense that $\Sigma = \sigma(H_{V(x)})$ for all $x \in \mathbb{X}_\varrho \setminus \{a^{\mathbb{Z}}\}$.*
- ***(2)*** *We have $\sigma_{\mathrm{ac}}(H_{V(x)}) = \varnothing$, unless $x \in \mathrm{orb}(x^*_{ab}) \cup \mathrm{orb}(x^*_{ba}) \cup \{a^{\mathbb{Z}}\}$. In particular, the almost sure* ac*-part of the spectrum Σ_{ac} is empty.*
- ***(3)*** *The spectrum of the point $a^{\mathbb{Z}}$ is given by $\sigma(H_{V(a^{\mathbb{Z}})}) = \sigma_{\mathrm{ac}}(H_{V(a^{\mathbb{Z}})}) = f(a^{\mathbb{Z}}) + [-2, 2]$. This set is* strictly *contained in Σ.*
- ***(4)*** *If $x \in \mathrm{orb}(x^*_{ab}) \cup \mathrm{orb}(x^*_{ba})$ is an eventually constant point, the* ac*-part of its spectrum is given by $\sigma_{\mathrm{ac}}(H_{V(x)}) = f(a^{\mathbb{Z}}) + [-2, 2]$.*

Proof. The first claim is immediate from Theorem 5.1.3. By almost minimality, every point $x \in \mathbb{X}_\varrho \setminus \{a^{\mathbb{Z}}\}$ has a dense orbit in $\mathbb{X}_\varrho$ and since $V \colon \mathbb{X} \to \mathbb{R}^{\mathbb{Z}}$ is continuous, the same holds for $V(x)$ in $V(\mathbb{X}_\varrho)$. Together with Lemma 5.1.2, this implies the first item. From Lemma 3.1.17, we know that the eventually periodic points in $\mathbb{X}_\varrho$ are precisely the orbits of x^*_{ab}, x^*_{ba} and $a^{\mathbb{Z}}$. Again, this is preserved as we apply V, due to Lemma 5.2.2. The discussion in Section 5.1 therefore yields the remaining three items, except for the *strict* inclusion $f(a^{\mathbb{Z}}) + [-2, 2] \subsetneq \Sigma$. The inequality holds because $\sigma(H_V) = r + [-2, 2]$ is only possible for the constant potential V with $V_n = r$ for all $n \in \mathbb{Z}$ [KP03, Thm. 8]. □

5.2.1. Generic absence of eigenvalues

The aim of this section is to establish criteria for the generic absence of eigenvalues via strongly palindromic potentials. Although the main conclusion of this section will be superseded by a stronger result later on, we believe that this approach is still instructive because it extends to more general almost minimal substitutions. First, we need a criterion that ensures that $V(x)$ is strongly palindromic as soon as x is strongly palindromic. Without loss of generality, assume that $f(x) = g(x_{[0,d]})$, for some $g \colon \mathcal{L}_{d+1} \to \mathbb{R}$, depends only on the first $d+1$ coordinates for some $d \in \mathbb{N}_0$.

Definition 5.2.4. We say that $g \colon \mathcal{L}_{d+1} \to \mathbb{R}$ is *reflection invariant* if $g(u) = g(\widetilde{u})$ for all $u \in \mathcal{L}_{d+1}$.

If $d = 0$, the function g is automatically reflection invariant. If g is reflection invariant, it is straightforward to verify that the corresponding sliding block code maps palindromic words

of length $m + d$ to palindromic words of length m for all $m \in \mathbb{N}_0$. In particular, the image $V(x)$ of a strongly palindromic sequence x is again strongly palindromic in this case.

Clearly, $a^{\mathbb{Z}} \in \mathbb{X}_\varrho$ is strongly palindromic. However, to obtain generic absence of Schrödinger eigenvalues we need a strongly palindromic sequence that has a dense orbit in $\mathbb{X}_\varrho$. Thus, we are led to ask about strongly palindromic sequences in $\mathbb{X}_\varrho \setminus \{a^{\mathbb{Z}}\}$.

Lemma 5.2.5. *If there exists a strongly palindromic sequence $x \in \mathbb{X}_\varrho \setminus \{a^{\mathbb{Z}}\}$, it follows that $\varrho(b)$ is a palindrome.*

Proof. Let $x \in \mathbb{X}_\varrho \setminus \{a^{\mathbb{Z}}\}$ be strongly palindromic and let u be a finite subword of x. Strong palindromicity implies that every part of x is eventually covered by some palindrome. Hence, there exists a palindrome $P \triangleleft x$ such that $u \triangleleft P$ and we obtain $\widetilde{u} \triangleleft \widetilde{P} = P \triangleleft x$. Decomposing x into inflation words, we can write it as a bi-infinite concatenation of words of the form $\varrho(a) = a^p$ and $\varrho(b) = b^j a b^{r-j}$. From this structure it follows that every occurrence of the word ba^pb in x can be extended to the word $ab^{r-j}a^pb^ja$. Since x has a dense orbit, it indeed contains $u = ab^{r-j}a^pb^ja$. We have seen above that this implies that $ab^ja^pb^{r-j}a = \widetilde{u} \triangleleft x$. This is only possible if $j = r - j$ and hence $\varrho(b) = b^j a b^j$ is a palindrome. □

Conversely, if $\varrho(b)$ is a palindrome, it follows by induction that $\varrho^n(b)$ is a palindrome for all $n \in \mathbb{N}$. Hence, we obtain a sequence of palindromes that eventually contains every legal word and we might be inclined to think that strongly palindromic sequences exist in $\mathbb{X}_\varrho \setminus \{a^{\mathbb{Z}}\}$. However, the situation is a bit more subtle.

Lemma 5.2.6. *If $\varrho = \varrho_{p,r,j}$ with $r = 2$ and $j = 1$, there are no B-strongly palindromic sequences in $\mathbb{X}_\varrho \setminus \{a^{\mathbb{Z}}\}$ for all $B > 1$.*

Proof. In this situation, we have $\varrho\colon a \mapsto a^p, b \mapsto bab$. Let $B > 1$ and for the sake of establishing a contradiction assume that $x \in \mathbb{X}_\varrho \setminus \{a^{\mathbb{Z}}\}$ is B-strongly palindromic. Since B-strong palindromicity is preserved under the shift map, we can assume that $x_0 = b$ without loss of generality. Let $P \triangleleft x$ be a palindrome centered at $c > 0$ in x and let $n \in \mathbb{N}$ be maximal with the property that $u = ba^{p^n}b \triangleleft P$. Note that $\varrho^{n+1}(b)$ contains precisely one occurrence of u, positioned at its center. Decomposing x into words of the form $\varrho^{n+1}(b)$ and $\varrho^{n+1}(a) = a^{p^{n+1}}$ we conclude that P is contained in $a^{p^k}\varrho^{n+1}(b)a^{p^k}$, for some $k \geqslant n + 1$, with u at its center. Since $\varrho^{n+1}(b)$ can be extended to the word $\varrho^{n+2}(b) = \varrho^{n+1}(b)a^{p^{n+1}}\varrho^{n+1}(b)$ either to its right or its left, we have in fact $k = n + 1$. We get an upper bound for $|P|$ by

$$|P| \leqslant |\varrho^{n+1}(b)| + 2p^{n+1} \leqslant CM^n$$

for some $M, C > 0$. On the other hand, since $x_0 = b$ and P has the word $ba^{p^n}b$ at its center, it follows that

$$c \geqslant \frac{1}{2}p^n.$$

In particular, there is a *monomial function* g such that $|P| \leqslant g(c)$. This contradicts the assumption that there exists a sequence of palindromes P_n, centered at c_n, such that $|P_n|$ grows *exponentially* faster than c_n. □

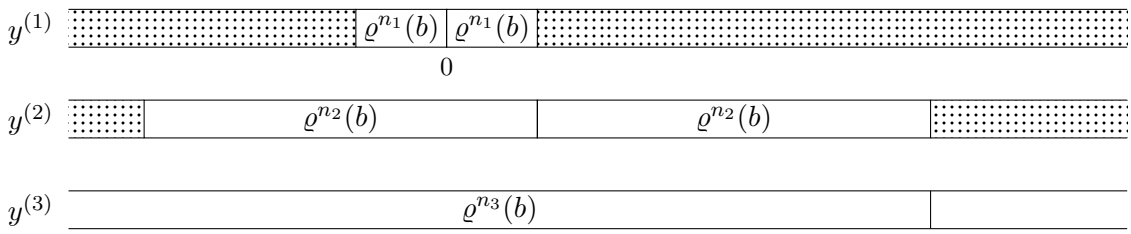

Figure 5.1.: Illustration of the nesting property of the predetermined parts of the sequences $y^{(1)}$, $y^{(2)}$ and $y^{(3)}$. The undetermined parts of the sequences are displayed as dotted regions.

Within the class of simple almost minimal substitutions, the substitution in Lemma 5.2.6 is in fact the only such counterexample.

Proposition 5.2.7. *Let $\varrho = \varrho_{p,r,j} \in \mathsf{Sams}$, and assume that $\varrho(b)$ is a palindrome and that $j \geqslant 2$. Then, there is a strongly palindromic sequence in $\mathbb{X}_\varrho \setminus \{a^{\mathbb{Z}}\}$.*

Proof. The decisive difference to the example in Lemma 5.2.6 is that we are now in a situation where the word bb is legal because $bb \triangleleft \varrho(b) = b^j a b^j$ if $j \geqslant 2$. This enables us to find a nested inclusion of palindromes such that the length of the palindromes grows arbitrarily faster than the shift in their center. A strongly palindromic sequence $y \in \mathbb{X}_\varrho$ emerges as an appropriate limit of this construction. The details follow.
First recall that $\varrho^n(b)$ is a palindrome for all $n \in \mathbb{N}$ and note that for $n_1, n_2 \in \mathbb{N}$ with $n_1 < n_2$, we have that $\varrho^{n_1}(b)\varrho^{n_1}(b)$ is a suffix of $\varrho^{n_2}(b)$. Let $(n_m)_{m\in\mathbb{N}_0}$ be a strictly increasing sequence of natural numbers, to be specified later. For each $m \in \mathbb{N}$, let $x^{(m)} \in \mathbb{X}_\varrho$ be a sequence that is of the form

$$x^{(m)} \;=\; \cdots \varrho^{n_m}(b).\varrho^{n_m}(b) \cdots$$

around the origin. Since $\varrho^{n_m}(b)\varrho^{n_m}(b)$ is a suffix of $\varrho^{n_{m+1}}(b)$, we obtain that $S^{-|\varrho^{n_m}(b)|}x^{(m+1)}$ takes the same form as $x^{(m)}$ around the origin but specifies a larger patch. Inductively, we define $\ell_1 = 0$ and $\ell_{m+1} = \ell_m + |\varrho^{n_m}(b)|$, for all $m \in \mathbb{N}$. By construction, the sequence, defined by $y^{(m)} = S^{-\ell_m}x^{(m)}$, for all $m \in \mathbb{N}$, converges to some $y \in \mathbb{X}_\varrho \subset \{a^{\mathbb{Z}}\}$ as $m \to \infty$; compare Figure 5.1. For each $m \in \mathbb{N}$ the sequence y contains the palindrome $P_{m+1} = \varrho^{n_{m+1}}(b)\varrho^{n_{m+1}}(b)$, with its center shifted away from the origin by $\sum_{j=1}^{m} |\varrho^{n_j}(b)|$. Choosing the sequence $(n_m)_{m\in\mathbb{N}}$ appropriately, we can ensure that

$$\lim_{m\to\infty} \frac{B^{\sum_{j=1}^{m} |\varrho^{n_j}(b)|}}{2|\varrho^{n_{m+1}}(b)|} \;=\; 0,$$

for all $B > 1$, and hence that y is strongly palindromic. □

Remark 5.2.8. The ideas presented in the proofs of Lemma 5.2.6 and Proposition 5.2.7 can be generalized to obtain a characterization of strong palindromicity for *all* almost minimal substitutions on $\mathcal{A} = \{a, b\}$ [EG21, Thm. 4.36]. There, an additional phenomenon can occur that is not covered by the results in this section. More specifically, recall the substitution

$\varrho\colon a \mapsto a, b \mapsto bba$ from Section 3.1.3. There are uncountably many B-strongly palindromic sequences in $\mathbb{X}_\varrho \setminus \{a^{\mathbb{Z}}\}$ precisely if $B < 4$ and no B-strongly palindromic sequences in $\mathbb{X}_\varrho \setminus \{a^{\mathbb{Z}}\}$ if $B \geqslant 4$. The existence of such a critical value for the parameter B seems to be a novel phenomenon. In the cases where strongly palindromic sequences exist, a little bit of extra effort shows that they form an uncountable set of measure 0 with respect to the non-atomic S-ergodic measure μ on $\mathbb{X}_\varrho$ [EG21]. ◊

Corollary 5.2.9. *Assume that g is reflection invariant, $\varrho = \varrho_{p,r,j} \in \mathsf{Sams}$, with $j \geqslant 2$ and that $\varrho(b)$ is a palindrome. Then, for all x in a generic subset of $\mathbb{X}$, $H_{V(x)}$ has no eigenvalues*
.

Proof. This follows by combining Proposition 5.2.7 with Theorem 5.1.9 and Remark 5.1.12. □

5.2.2. Almost sure absence of eigenvalues

In order to show μ-almost everywhere absence of Schrödinger eigenvalues, it suffices that $\mu(\mathcal{G}) > 0$ holds for

$$\mathcal{G} \;=\; \{x \in \mathbb{X}_\varrho : V(x) \text{ is a Gordon potential}\};$$

compare Remark 5.1.12. Again, there is no loss of generality in assuming that $f(x) = g(x_{[0,d]})$ depends only on the first $d+1$ coordinates. Given $n \geqslant d$, the sliding block structure implies that $V(x) \in [w.ww]$ for some $w \in R^n$, provided that $x \in [v.vvv_{[1,d]}]$ for some $v \in \mathcal{L}_n$ (with the convention that $v_{[1,0]} = \epsilon$ is the empty word). The classic approach to show that $\mu(\mathcal{G}) > 0$ goes as follows; compare [DL06b, Thm. 3] for the special case $d = 0$. First,

$$\mathcal{G} \;\supset\; \bigcap_{k \geqslant d} \bigcup_{n \geqslant k} \mathcal{G}_n,$$

where

$$\mathcal{G}_n \;=\; \bigcup_{v \in \mathcal{L}_n} [v.vvv_{[1,d]}],$$

for all $n \in \mathbb{N}$ with $n \geqslant d$. This is just a reformulation of the observation that $V(x)$ is a Gordon potential if $x \in \mathcal{G}_n$ for infinitely many values of $n \in \mathbb{N}$. If μ is a *finite* measure, we have

$$\mu(\mathcal{G}) \;\geqslant\; \mu(\limsup_{n\to\infty} \mathcal{G}_n) \;\geqslant\; \limsup_{n\to\infty} \mu(\mathcal{G}_n),$$

and it suffices to show that $\mu(\mathcal{G}_n)$ remains bounded away from 0 on a subsequence. If μ is infinite, the relation $\mu(\limsup_{n\to\infty} A_n) \geqslant \limsup_{n\to\infty} \mu(A_n)$ no longer holds for arbitrary sequences of measurable sets $(A_n)_{n\in\mathbb{N}}$. Indeed, for $A_n = [a^n.a^{2n+d}]$, we obtain that $\limsup_{n\to\infty} A_n = \{a^{\mathbb{Z}}\}$, and hence

$$0 \;=\; \mu(\{a^{\mathbb{Z}}\}) \;=\; \mu(\limsup_{n\to\infty} A_n) \;<\; \limsup_{n\to\infty} \mu(A_n) \;=\; \infty.$$

This also shows that the observation $A_n \subset \mathcal{G}_n$ is not particularly useful for our purposes. Since $\mu([b]) = 1$, a natural workaround is to restrict everything to the set $[b]$. We obtain

$$\mu(\mathcal{G} \cap [b]) \;=\; \mu(\limsup(G_n \cap [b])) \;\geqslant\; \limsup_{n\to\infty} \mu(\mathcal{G}_n \cap [b]),$$

which also holds if μ is an infinite measure. This can be used to show the following adaptation of [DL06b, Thm. 3] to the almost minimal setting.

Lemma 5.2.10. *Assume that there is a word $u \in \mathcal{L}^b := \{v \in \mathcal{L} : v_1 = b\}$ such that $uuub \in \mathcal{L}$. Then, $H_{V(x)}$ has no eigenvalues for μ-almost every $x \in \mathbb{X}_\varrho$.*

Proof. By the discussion above, we have

$$\mu(\mathcal{E}) \geqslant \limsup_{n\to\infty} \mu(\mathcal{G}_n \cap [b]),$$

and it thus suffices to show that the μ-measure of the sets

$$\mathcal{G}_n \cap [b] = \bigcup_{v\in\mathcal{L}_n^b} [v.vvv_{[1,d]}]$$

is bounded away from 0 on a subsequence. Recall the definition of μ on cylinder sets from Proposition 3.1.14. Setting $n_k = |\varrho^k(u)|$, we obtain

$$\mu(\mathcal{G}_{n_k} \cap [b]) = \sum_{v\in\mathcal{L}_{n_k}^b} \mu[v^3 v_{[1,d]}] = \lim_{m\to\infty} \sum_{v\in\mathcal{L}_{n_k}^b} \frac{|\varrho^m(b)|_{v^3 v_{[1,d]}}}{r^m}, \tag{5.2}$$

for all $k \in \mathbb{N}$. Let $j \in \mathbb{N}$ be such that $uuub \triangleleft \varrho^j(b)$. Then, for every $k \in \mathbb{N}$, we have $w_k := \varrho^k(uuub) \triangleleft \varrho^{j+k}(b)$. Note that w_k starts and ends with the word $\varrho^k(b)$. It therefore contains at least $|\varrho^k(b)|_b - d = r^k - d$ words $v^3 v_{[1,d]} \in \mathcal{L}^b$ with $|v| = n_k$ (counted with multiplicities). That is,

$$\sum_{v\in\mathcal{L}_{n_k}^b} |\varrho^{j+k}(b)|_{v^3 v_{[1,d]}} \geqslant r^k - d,$$

for all $k \in \mathbb{N}$. On the other hand, we have, provided $m > j + k$,

$$|\varrho^m(b)|_w \geqslant |\varrho^{m-j-k}(b)|_b \, |\varrho^{j+k}(b)|_w = r^{m-j-k} |\varrho^{j+k}(b)|_w,$$

for all $w \in \mathcal{L}$. Together with (5.2), this yields

$$\mu(\mathcal{G}_{n_k} \cap [b]) \geqslant \lim_{m\to\infty} \frac{1}{r^m} \sum_{v\in\mathcal{L}_{n_k}^b} r^{m-j-k} |\varrho^{j+k}(b)|_{v^3 v_{[1,d]}} \geqslant \frac{1}{r^j} - \frac{d}{r^{j+k}},$$

for all $k \in \mathbb{N}$, and the claim follows. □

The *index* of a subshift is the largest (not necessarily integer) power of a word that can appear in its language. In Lemma 5.2.10, it was further assumed that u starts with the letter b. This motivates us to define the index for the subset $\mathcal{L}^b$ as follows.

Definition 5.2.11. The index of $\mathcal{L}^b = \{v \in \mathcal{L} : v_1 = b\}$ is given by

$$\mathrm{Ind}(\mathcal{L}^b) = \sup\{s \in \mathbb{Q} : v^s \in \mathcal{L}^b \text{ for some } v \in \mathcal{L}^b\}.$$

With this notation, the requirement in Lemma 5.2.10 can be stated as $\mathrm{Ind}(\mathcal{L}^b) > 3$. In [EG21, Prop. 4.46], we give an algorithm to determine $\mathrm{Ind}(\mathcal{L}^b)$ for arbitrary almost minimal substitutions on $\mathcal{A} = \{a, b\}$. For the class Sams, this takes a particularly simple form.

Proposition 5.2.12. *If $\varrho = \varrho_{p,r,j} \in \mathsf{Sams}$, we have $\mathrm{Ind}(\mathcal{L}^b) = r$.*

Proof. We first show that $\mathrm{Ind}(\mathcal{L}^b) \geqslant r$. For $r = 2$, this follows from $\varrho\colon b \mapsto bab \mapsto babaabab$ and hence $baba \in \mathcal{L}^b$. If $r \geqslant 3$, we have that bb is legal and so is $\varrho(bb) = b^j a b^r a b^{r-j}$, implying that $b^r \in \mathcal{L}^b$.
Now, assume conversely that $u^s \in \mathcal{L}^b$ for some $s \geqslant r$. Let $k \in \mathbb{N}$ be maximal with the property that $a^k \triangleleft u^s$. Since u starts with the letter b, this implies that $ba^k \triangleleft u$ and $ba^kb \in \mathcal{L}$. Hence, $k = p^n$ for some $n \in \mathbb{N}$. We can find u^s as a subword of x, for all $x \in \mathbb{X}_\varrho \setminus \{a^{\mathbb{Z}}\}$. We write x as a concatenation of the words $\varrho^{n+1}(b)$ and $\varrho^{n+1}(a)$, with at most r copies of $\varrho^{n+1}(b)$ next to each other. Note that $\varrho^{n+1}(a) = a^{p^{n+1}}$ and $\varrho^{n+1}(b)$ contains a single occurrence of a^{p^n}. Hence, each occurrence of u overlaps a word $\varrho^{n+1}(b)$ in this decomposition and u^s cannot contain $\varrho^{n+1}(a)$ by the maximality of $k = p^n$. We distinguish two cases. If $r = 2$, $\varrho^{n+1}(b) = va^{p^n}v$, for some $v \in \mathcal{L}^b$ and it follows that $s = 2$ and $u = va^{p^n}$. If $r \geqslant 3$, having $r-1$ occurrences of $ba^{p^n}b$ in u^r enforces that u is a cyclic permutation of the word $\varrho^{n+1}(b)$. In fact, it follows that $u = \varrho^{n+1}(b)$ and $r = s$. □

Remark 5.2.13. The proof of Proposition 5.2.12 also allows us to determine the words u in $\mathcal{L}^b$ such that u^r is legal. For $r = 2$, these are precisely the words of the form $u = \varrho^n(b)a^{p^n}$, with $n \in \mathbb{N}_0$. For $r \geqslant 3$, these are the words of the form $u = \varrho^n(b)$, with $n \in \mathbb{N}_0$. ◊

Corollary 5.2.14. *Let $\varrho = \varrho_{p,r,j} \in \mathsf{Sams}$ with $r \geqslant 4$. Then, for μ-almost every $x \in \mathbb{X}_\varrho$, $H_{V(x)}$ admits no eigenvalues.*

Proof. This follows by combining Proposition 5.2.12 with Lemma 5.2.10. □

5.2.3. Eigenvalues for eventually constant sequences

Consider the substitution $\varrho = \varrho_{p,5,1}$, with $p > 5$, that is,

$$\varrho\colon a \mapsto a^p, \quad b \mapsto bab^4.$$

Since $p > r$, the ergodic measure μ is infinite by Proposition 3.1.14. Also, $r = 5$ implies that the spectrum of $H_{V(x)}$ is singular continuous for μ-almost every $x \in \mathbb{X}_\varrho$, by Corollary 5.2.14. In this section, we show that this result is sharp in the sense that there exists a sequence $x^* \in \mathbb{X}_\varrho$ and a choice of the potential function f such that $H_{V(x^*)}$ admits an eigenvalue. More precisely, recalling the notation from Lemma 3.1.17, we set

$$x^* = x^*_{ab} = a^\infty.\varrho^\infty(b).$$

We proved the following in [EG21, Prop. 5.13].

Proposition 5.2.15. *Let $\varrho = \varrho_{p,5,1}$, with $p > 5$ and $x^* = x^*_{ab}$. There exists a choice of the potential function f and $E \in \mathbb{R}$ such that E is an eigenvalue of the operator $H_{V(x^*)}$. The corresponding eigenstate ψ is exponentially decaying to both sides.*

Sketch of proof. We provide here an outline of the essential arguments. The full details can be found in [EG21]. We can decompose $\varrho^\infty(b)$ as

$$\varrho^\infty(b) \;=\; \varrho(b)a^{m_1}\varrho(b)a^{m_2}\varrho(b)a^{m_3}\cdots\,,$$

for some sequence $(m_n)_{n\in\mathbb{N}}$. For notational convenience, we write A^n for $A_E^n(x^*)$ in the following. We will construct an exponentially decaying solution $\psi \in \ell^2(\mathbb{Z})$ of the equation $H_{V(x^*)}\psi = E\psi$ explicitly. Setting $v = (\psi_0, \psi_{-1})$, this requires that both $A^{-n}|v\rangle$ and $A^n|v\rangle$ decay exponentially as $n \to \infty$. Assuming that the function $f(x) = g(x_0)$ depends only on the first coordinate, we obtain $A^{-n} = T_E(a)^{-n}$ and $A^{|\varrho(b)|} = T_E(b)^4 T_E(a) T_E(b)$, where

$$T_E(c) \;=\; \begin{pmatrix} E-g(c) & -1 \\ 1 & 0 \end{pmatrix},$$

for $c \in \{a, b\}$. A straightforward calculation shows that there are parameters $g(a), g(b), E \in \mathbb{R}$ with the property that $T_E(a)$ is hyperbolic and

$$T_E(a)^m\, A^{|\varrho(b)|} \;=\; A^{|\varrho(b)|}\, T_E(a)^{-m} \tag{5.3}$$

for all $m \in \mathbb{N}$ [EG21, Lemma 5.15]. Intuitively, (5.3) means that $A^{|\varrho(b)|}$ acts as a 'switch' between the stable and unstable eigenspace of $T_E(a)$. Choosing v in the unstable eigendirection of $T_E(a)$, we immediately obtain that $A^{-n}|v\rangle$ decays exponentially. It remains to show that $A^n|v\rangle$ also decays exponentially as $n \to \infty$. As a first step, we consider the subsequence $n_k \in \mathbb{N}$, defined by

$$x^*_{[0,n_k-1]} \;=\; \varrho(b)a^{m_1}\cdots\varrho(b)a^{m_k},$$

for all $k \in \mathbb{N}$. By (5.3),

$$A^{n_k} \;=\; T_E(a)^{m_k} A^{|\varrho(b)|} \cdots T_E(a)^{m_1} A^{|\varrho(b)|} \;=\; (A^{|\varrho(b)|})^k\, T_E(a)^{-h(k)}, \tag{5.4}$$

where $h(k) = \sum_{j=1}^k (-1)^{j+1} m_j$. Hence, we need to show that $h(k) \to \infty$ as $k \to \infty$, sufficiently fast. By induction, it follows that x^* starts with the word

$$\varrho^{n+1}(b) \;=\; \varrho^n(b)a^{m_{r^{n-1}}}(\varrho^n(b))^4,$$

for all $n \in \mathbb{N}$, where $m_{r^{n-1}} = p^n$ denotes the largest power of a in $\varrho^{n+1}(b)$. Since r^{n-1} is odd, this shows that the first occurrence of a new largest number in the sequence $(m_n)_{n\in\mathbb{N}}$ always happens at an odd position. A more careful analysis yields that $h(k) \geqslant p^n$, whenever $r^n \leqslant k < r^{n+1}$ [EG21, Lemma 5.16]. This bound on k also implies that $x^*_{[0,n_k-1]}$ is a prefix of $\varrho^{n+1}(b)$ and hence that $n_k \leqslant Cp^n$, for some constant $C > 0$. That is, $h(k) \sim n_k$, and since $p > r$, the number k becomes negligible in comparison. It follows that $A^{n_k}v$ decays exponentially fast in n_k, due to (5.4).

Finally, assume that $n \in \mathbb{N}$ with $n_k < n \leqslant n_{k+1}$. That is $x^*_{[0,n-1]} = x^*_{[0,n_k-1]}u$, for some prefix u of $\varrho(b)a^{n_{k+1}}$. First, if u is a prefix of $\varrho(b)$, we have $A^n = M_u A^{n_k}$, for some matrix M_u whose norm is uniformly bounded and the asymptotic behaviour is unaltered. Otherwise, $u = \varrho(b)a^s$ for some $0 \leqslant s \leqslant n_{k+1}$ and

$$A^n \;=\; (A^{|\varrho(b)|})^{k+1}\, T_E(a)^{-h_n}$$

for some h_n between $h(k)$ and $h(k+1)$. We can therefore bound $A^n|v\rangle$ by interpolating between $A^{n_k}|v\rangle$ and $A^{n_{k+1}}|v\rangle$. It follows that $A^n|v\rangle$ decays exponentially in n. □

Remark 5.2.16. The point $x^* = a^\infty.\varrho^\infty(b)$ in Proposition 5.2.15 represents a fusion of a non-trivial structure to the right with a constant potential to the left. The eigenfunction ψ that was constructed in the proof of Proposition 5.2.15 is localized around the origin, which constitutes the boundary point of $\varrho^\infty(b)$. It is therefore reminiscent of a *surface state*, a common phenomenon that appears at the boundary of a solid; compare [KP97] for background and the history of this concept.

Remark 5.2.17. The results in Section 5.2.1 and Section 5.2.2 are persistent in the sense that they hold irrespective of a linear scaling of the function f. This is not the case for Proposition 5.2.15, as the method of proof singles out distinct values of the *coupling constant* $g(b)-g(a)$. It remains an open question whether a Schrödinger eigenvalue occurs for arbitrary (or at least an infinite set of) coupling constants. The solutions of the equations of the form

$$T_E(a)A^{|\varrho^n(b)|} \;=\; A^{|\varrho^n(b)|}T_E(a)^{-1} \tag{5.5}$$

are nested, in the sense that all parameters $(g(b)-g(a), E)$ solving (5.5) for a fixed $n = n_0$, also form a solution for all $n \geqslant n_0$. A natural next step would be to investigate if the number of solutions diverges as $n \to \infty$. The proof of Proposition 5.2.15 can be adapted to show that for each such solution, there is an exponentially decaying eigensolution ψ to the equation $H_{V(x^*)}\psi = E\psi$. ◊

Corollary 5.2.18. *Let ϱ be an almost minimal substitution as defined in Proposition* 5.2.15. *There exists a potential function f such that the operator $H_{V(x)}$*

- ***(1)*** *has purely singular continuous spectrum, for μ-almost every $x \in \mathbb{X}_\varrho$,*
- ***(2)*** *admits an eigenvalue that is not an isolated point of the spectrum, for a dense set of points $x \in \mathbb{X}_\varrho$,*
- ***(3)*** *has purely absolutely continuous spectrum if $x = a^{\mathbb{Z}} \in \mathbb{X}_\varrho$.*

Proof. The first item follows from Corollary 5.2.14. By Proposition 5.2.15, $H_{V(x^*)}$ admits an eigenvalue E for an appropriate choice of f. The same holds for all x in the shift-orbit of x^*, which lies dense in $\mathbb{X}_\varrho$. This also implies that $E \in \sigma(H_{V(x^*)}) = \Sigma = \Sigma_{\mathrm{sc}}$ and hence E cannot be isolated because a singular continuous measure supports no isolated points. □

This, together with the general observations in Corollary 5.2.3, proves Theorem 5.1.13.

Remark 5.2.19. The precise spectral nature of the operator $H_{V(x^*)}$ from Proposition 5.2.15 remains open. Since x^* is eventually constant, we know that $\sigma_{\mathrm{ac}}(H_{V(x^*)}) = g(a) + [-2, 2]$. There are two possibilities. Either $\Sigma \setminus \sigma_{\mathrm{ac}}(H_{V(x^*)})$ consists entirely of (infinitely many) eigenvalues, or $\sigma_{\mathrm{sc}}(H_{V(x^*)}) \neq \varnothing$, in which case all spectral components appear for $H_{V(x^*)}$. Both cases would be interesting and somewhat surprising. ◊

An analogue of Proposition 5.2.15 holds for some, but not all, almost minimal substitutions on $\mathcal{A} = \{a, b\}$. For example, a similar construction works for $\varrho\colon a \mapsto a^p, b \mapsto bab^2$, with $p > 3$.

In contrast, for $\varrho\colon a \mapsto a^p, b \mapsto bba$, with $p \geqslant 2$, one can show explicitly that there are no eigenvalues for $H_{V(x^*)}$, with $x^* = a^\infty.\varrho^\infty(b)$, at least as long as $f(x) = g(x_0)$ depends only on the first coordinate [EG21, Prop. 5.18]. Hence, the fact that x^* is eventually constant to the left certainly does not suffice alone to guarantee the occurrence of Schrödinger eigenvalues.

5.3. Mixed potentials

So far, we were concerned with Schrödinger operators with potentials that reflect the structural properties of a single sequence. As a next step, we wish to investigate the case that there are two sources for the potential with different structural properties. This models situations such as an aperiodically ordered material (mathematical quasicrystal) that is combined with a periodic background potential. Regarding the spectrum of the corresponding Schrödinger operator, we thus create a test for the stability of certain properties (like Cantor spectra of Lebesgue measure 0 or absence of eigenvalues) under periodic perturbations. This will be our main focus in Section 5.3.1. In fact, rather than testing directly the stability of spectral properties, we instead analyze the stability of certain structural properties that *ensure* the spectral property in question. Heuristically, our finding is that periodic backgrounds do not seem to alter much of the qualitative behaviour of the spectrum for many of the well-studied models of aperiodic order. In Section 5.3.2, we turn to the combination of a random and a periodic potential. If the periodic potential has period 2, we show how to calculate the corresponding spectrum explicitly.

5.3.1. Minimal subshifts and periodic potentials

Assume that x is an element of a minimal subshift $(\mathbb{X}, S)$ over the alphabet $\mathcal{A}$ and that x' is an element of a periodic subshift $(\mathbb{X}', S)$ with period p. We model the combined influence of x and x' via a potential $V(x, x')$ that acts as a sliding block code on both x and x'. Since $(\mathbb{X}', S)$ is p-periodic, it is in fact topologically conjugate to the cyclic rotation $(\mathbb{Z}_p, +1)$, where $\mathbb{Z}_p = \mathbb{Z}/p\mathbb{Z}$ and addition is understood modulo p. As V depends only on the dynamical properties of x', there is no loss of generality in assuming that V is a function on $(\mathbb{X} \times \mathbb{Z}_p, T)$, where $T\colon (x, m) \mapsto (Sx, m + 1)$, for all $(x, m) \in \mathbb{X} \times \mathbb{Z}_p$. More precisely, $V(x, m) = (V_n(x, m))_{n\in\mathbb{Z}}$, where

$$V_n(x, m) \;=\; f(T^n(x, m)) \;=\; f(S^n x, m + n),$$

for all $n \in \mathbb{Z}$ and some locally constant function f. Recall that, due to the minimality of $(\mathbb{X}, S)$, the spectrum $\sigma(H_{V(x)})$ does not depend on $x \in \mathbb{X}$. Can we expect the same behaviour for $(\mathbb{X} \times \mathbb{Z}_p, T)$? In general, the answer is no, but we can give a bound on the number of different spectral sets that can occur.

The space $\mathbb{X} \times \mathbb{Z}_p$ consists of p disjoint copies of $\mathbb{X}$ that are visited in a cyclical order under the action of T. Let us fix one of these copies, say $\mathbb{X} \times \{0\}$. The first return of a point $(x, 0) \in \mathbb{X}\times\{0\}$ in the corresponding T-orbit to the set $\mathbb{X}\times\{0\}$ is given by $T^p(x, 0) = (S^p x, 0)$. Similar reasoning applies to $\mathbb{X} \times \{m\}$ for all $m \in \mathbb{Z}_p$. Hence, the first return map induced by T on $\mathbb{X} \times \{m\}$ is naturally identified with S^p on $\mathbb{X}$ for all $m \in \mathbb{Z}_p$. This allows us to

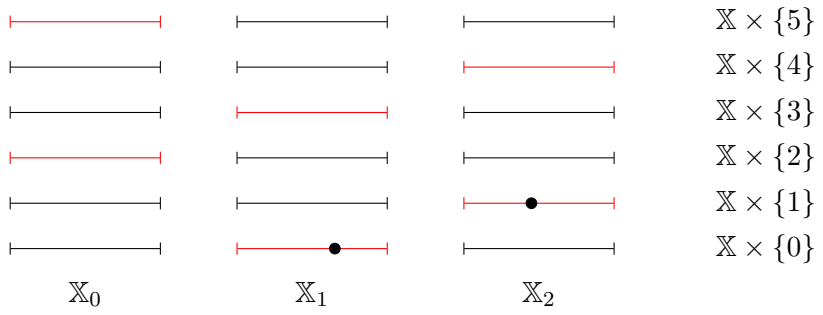

Figure 5.2.: Schematic for the decomposition of $\mathbb{X} \times \mathbb{Z}_p$, for $p = 6$ and $s(p) = 3$. The set $\mathbb{X}_1^T$ is highlighted in red. The dots indicate the action of T on some element in $\mathbb{X}_1 \times \{0\}$.

infer the orbit structure of $(\mathbb{X} \times \mathbb{Z}_p, T)$ from corresponding results on $(\mathbb{X}, S^p)$. Recall that $(\mathbb{X}, S^p)$ decays into a disjoint cycle of S^p-minimal sets $\mathbb{X}_0, \cdots, \mathbb{X}_{s(p)-1}$, where $s(p)$ denotes the D-function corresponding to the minimal subshift $(\mathbb{X}, S)$.

Lemma 5.3.1. *$(\mathbb{X} \times \mathbb{Z}_p)$ can be written as a union of $s(p)$ disjoint, T-minimal subsets $\mathbb{X}_0^T, \cdots, \mathbb{X}_{s(p)-1}^T$, where, for all $0 \leqslant j \leqslant s(p) - 1$,*

$$\mathbb{X}_j^T = \bigcup_{k=0}^{p-1} T^k(\mathbb{X}_j \times \{0\}).$$

Proof. First, the sets $\mathbb{X}_j^T$ indeed cover the whole space since

$$\bigcup_{j=0}^{s(p)-1} \mathbb{X}_j^T = \bigcup_{k=0}^{p-1} T^k(\mathbb{X} \times \{0\}) = \mathbb{X} \times \mathbb{Z}_p.$$

Note that $T^k(\mathbb{X}_j \times \{0\}) = (\mathbb{X}_{j+k} \times \{k\})$. From this, disjointness is straightforward to verify. We refer to Figure 5.2 for an illustration. It remains to show that $\mathbb{X}_j^T$ is minimal for all $0 \leqslant j \leqslant s(p) - 1$. The elements in $\mathbb{X}_j^T$ are precisely those of the form $(S^k x, k)$ for some $x \in \mathbb{X}_j$ and $k \in \mathbb{Z}_p$. Let $(S^k x, k)$ and $(S^m y, m)$ be two such elements. Since x and y belong to the same S^p-minimal component of $\mathbb{X}$, there is a sequence $(n_m)_{m \in \mathbb{N}}$ such that $\lim_{m \to \infty} S^{n_m p} x = y$. Consequently,

$$\lim_{m \to \infty} T^{n_m p + m - k}(S^k x, k) = \lim_{m \to \infty} (S^m S^{n_m p} x, m) = (S^m y, m).$$

This shows that every point in $\mathbb{X}_j^T$ has a dense orbit in $\mathbb{X}_j^T$. □

Corollary 5.3.2. *For all $p \in \mathbb{N}$, we have*

$$\#\{\sigma(H_{V(x,m)}) : (x, m) \in \mathbb{X} \times \mathbb{Z}_p)\} \leqslant s(p),$$

where $s \colon \mathbb{N} \to \mathbb{N}$ is the D-function on $(\mathbb{X}, S)$. ■

In general, this only gives an *upper* bound for the number of different spectra $\sigma(H_{V(x,m)})$, with $(x, m) \in \mathbb{X} \times \mathbb{Z}_p$. Showing that the spectra arising from different T-minimal components are indeed different is a difficult problem, in general. This has been achieved for a specific example, with $\mathbb{X}$ the Thue–Morse substitution subshift and $p = 4$ [CDFG21], thanks to a construction by Fillman.

Remark 5.3.3. If $(\mathbb{X}, S)$ admits no root of unity as a topological eigenvalue, we obtain that $s(p) = 1$ for all $p \in \mathbb{N}$. In this case, all elements in $\mathbb{X} \times \mathbb{Z}_P$ give rise to the same Schrödinger spectrum. This applies for instance to the subshift $(\mathbb{X}_F, S)$ arising from the Fibonacci substitution ϱ_F; compare Example 3.1.5 and [BG13]. We discuss the D-function for some more examples in [CDFG21]. ◊

In the following, we specialize to the case that $(\mathbb{X}, S)$ is a Boshernitzan subshift. Since Boshernitzan's condition is formulated for symbolic subshifts, it will be convenient to regard $(\mathbb{X} \times \mathbb{Z}_p)$ as a symbolic subshift over the alphabet $\mathcal{A}' = \mathcal{A} \times \mathbb{Z}_p$. More precisely, we define a function $h\colon \mathbb{X} \times \mathbb{Z}_p \to (\mathcal{A}')^{\mathbb{Z}}$ with $h(x, m)_n = (x_n, m + n) \in \mathcal{A}'$ for all $n \in \mathbb{Z}$ and we set $\mathbb{Y} = h(\mathbb{X} \times \mathbb{Z}_p)$. This corresponds to "merging" x coordinatewise with the periodic sequence in $\mathbb{Z}_p$ that describes the orbit of m. Then, h defines a topological conjugation between $(\mathbb{X} \times \mathbb{Z}_p, T)$ and $(\mathbb{Y}, S)$. In particular, $(\mathbb{Y}, S)$ also decays into a disjoint union of minimal subsets $\mathbb{Y}_0, \cdots \mathbb{Y}_{s(p)-1}$, where $\mathbb{Y}_j = h(\mathbb{X}_j^T)$ for all $0 \leqslant j \leqslant s(p) - 1$.

Proposition 5.3.4. *The subshift $(\mathbb{Y}_j, S)$ is a Boshernitzan subshift for each $0 \leqslant j \leqslant s(p)-1$.*

Proof. Let μ be the unique ergodic probability measure on $(\mathbb{X}, S)$ and μ' the uniform probability measure on $\mathbb{Z}_p$. Then, $\mu \times \mu'$ is a T-invariant probability measure on $\mathbb{X} \times \mathbb{Z}_p$, and its pushforward $\nu = (\mu \times \mu') \circ h^{-1}$ is an S-invariant measure on $(\mathbb{Y}, S)$. This measure assigns the same value to each of the S-minimal subsets. Indeed,

$$\nu(\mathbb{Y}_j) \;=\; (\mu \times \mu')(\mathbb{X}_j^T) \;=\; \sum_{k=0}^{p-1} \mu(\mathbb{X}_j)\mu'(\{0\}) \;=\; \frac{1}{s(p)},$$

where we have used that $\mu(\mathbb{X}_j) = \mu(\mathbb{X}_0) = s(p)^{-1}$, for all $0 \leqslant j \leqslant s(p) - 1$. We claim that, for each j, the conditional probability measure

$$\nu_j \;=\; s(p)\nu|_{\mathbb{Y}_j}$$

satisfies Boshernitzan's condition on $(\mathbb{Y}_j, S)$. Since the sets $\mathbb{Y}_0, \cdots, \mathbb{Y}_{s(p)-1}$ are disjoint and compact, there is a positive distance between pairwise different sets $\mathbb{Y}_j$ and $\mathbb{Y}_k$. Hence, there is a number n_0 such that for each $n \geqslant n_0$ and $u \in \mathcal{L}_n(\mathbb{Y})$, the set $[u]$ is completely contained in precisely one of the sets $\mathbb{Y}_0, \ldots, \mathbb{Y}_{s(p)-1}$. In other words, $\mathcal{L}_n(\mathbb{Y})$ decays into a disjoint union

$$\mathcal{L}_n(\mathbb{Y}) \;=\; \bigsqcup_{j=0}^{s(p)-1} \mathcal{L}_n(\mathbb{Y}_j),$$

for all $n \geqslant n_0$. The set $\mathcal{L}_n(\mathbb{Y}_j)$ consists precisely of the words $u = (a_1, m) \cdots (a_n, m+n)$ with $a_1 \cdots a_n \in \mathcal{L}_n(\mathbb{X}_j)$ and $m \in \mathbb{Z}_p$. For such a word u, we have

$$\nu([u]) \;=\; (\mu \times \mu')\big(h^{-1}([u])\big) \;=\; \mu([a_1 \cdots a_n])\mu'(\{m\}) \;=\; \frac{1}{p}\mu([a_1 \cdots a_n]).$$

From $n \geqslant n_0$, it follows that $[u] \subset \mathbb{Y}_j$. This yields $\nu_j([u]) = s(p)\nu([u])$ and hence,

$$\underline{\nu_j}(n) \,=\, \inf\bigl\{\nu_j([u]) : u \in \mathcal{L}_n(\mathbb{Y}_j)\bigr\} \,=\, \frac{s(p)}{p} \inf\bigl\{\mu([v]) : v \in \mathcal{L}_n(\mathbb{X}_j)\bigr\} \,\geqslant\, \frac{s(p)}{p}\underline{\mu}(n).$$

Since $(\mathbb{X}, S)$ is a Boshernitzan subshift with unique S-invariant measure μ, this yields that $\limsup_{n\to\infty} n\,\underline{\nu_j}(n) > 0$ and the claim follows. □

From this, the main result of this Section follows as a Corollary.

Theorem 5.3.5. *Let $(\mathbb{X}, S)$ be a Boshernitzan subshift and $p \in \mathbb{N}$. Then, for every point $(x, m) \in \mathbb{X} \times \mathbb{Z}_p$, the spectrum $\sigma(H_{V(x,m)})$ is a Cantor set of Lebesgue measure 0, unless $V(x, m)$ is a periodic sequence.*

Proof. The idea of proof is to use that $(\mathbb{X} \times \mathbb{Z}_p, T)$ is topologically conjugate to the subshift $(\mathbb{Y}, S)$, which decays into a disjoint union of Boshernitzan subshifts due to Proposition 5.3.4. More specifically, the function

$$\widetilde{f} \,=\, f \circ h^{-1} \colon \mathbb{Y} \to \mathbb{R},$$

inherits the property of being locally constant from the function f. The corresponding sliding block code (potential)

$$\widetilde{V} \colon y \mapsto \bigl(\widetilde{f}(S^n y)\bigr)_{n\in\mathbb{Z}}$$

satisfies $\widetilde{V}(h(x, m)) = V(x, m)$, for all $(x, m) \in \mathbb{X} \times \mathbb{Z}_p$. By Poposition 5.3.4, $h(x, m)$ pertains to a Boshernitzan subshift and the claim follows from the general criterion, given in Theorem 5.1.6. □

Together with Corollary 5.3.2, this proves Theorem 5.1.14. Next, we turn to the question of whether the absence of Schrödinger eigenvalues carries over from $(\mathbb{X}, S)$ to the system $(\mathbb{X} \times \mathbb{Z}_p, T)$. Often, a slight modification of the conditions or methods that are used to rule out Schrödinger eigenvalues for $(\mathbb{X}, S)$ can be adapted to cover $(\mathbb{X} \times \mathbb{Z}_p, T)$ as well. We illustrate this at the example of Gordon's criterion; compare the discussion in Section 5.2.2.

Lemma 5.3.6. *Assume that $f(x, m) = g(x_{[0,d]}, m)$ and let $\mathcal{G}_n = \bigcup_{v\in\mathcal{L}_n} [v.vvv_{[1,d]}] \subset \mathbb{X}$, for all $n \in \mathbb{N}$. If $\limsup_{n\to\infty} \mu(\mathcal{G}_{np}) > 0$, then*

$$\mathcal{E} \,=\, \{(x, m) \in \mathbb{X} \times \mathbb{Z}_p : H_{V(x,m)} \text{ has no eigenvalues}\}$$

is a set of full measure for every T-invariant probability measure on $\mathbb{X} \times \mathbb{Z}_p$.

Proof. First, we characterize the T-ergodic measures on $\mathbb{X} \times \mathbb{Z}_p$. By Proposition 5.3.4, each of the subshifts $(\mathbb{Y}_j, S)$ is strictly ergodic and hence, the same holds for the conjugate subshifts $(\mathbb{X}_j^T, T)$, for each $0 \leqslant j \leqslant s(p) - 1$. It follows that the ergodic measures on $(\mathbb{X} \times \mathbb{Z}_p, T)$ are precisely those of the form

$$\mu_j \,=\, s(p)(\mu \times \mu')|_{\mathbb{X}_j^T},$$

with $0 \leqslant j \leqslant s(p) - 1$. Hence, it suffices to show that $\mu_j(\mathcal{E}) = 1$ in order to obtain the same conclusion for each T-invariant probability measure on $\mathbb{X} \times \mathbb{Z}_p$. In fact, due to the ergodicity

of μ_j, we only need to show $\mu_j(\mathcal{E}) > 0$. Let j be fixed in the following. The additional factor p in the sequence $(\mathcal{G}_{np})_{n\in\mathbb{N}}$ harmonizes the 3-block structure of elements in $\mathcal{G}_{np} \subset \mathbb{X}$ with the p-periodicity of the factor $(\mathbb{Z}_p, +1)$. Indeed, the assumption on f ensures that, for $(x, m) \in \mathcal{G}_{np} \times \mathbb{Z}_p$, we have $V(x, m) \in [u.uu]$, for some $u \in R^{np}$. Hence, by Gordon's lemma 5.1.11, $\mathcal{E}$ contains the set

$$\mathcal{G} \;=\; \limsup_{n\to\infty} (\mathcal{G}_{np} \times \mathbb{Z}_p).$$

The intersection with $\mathbb{X}_j^T$ yields

$$(\mathcal{G}_{np} \times \mathbb{Z}_p) \cap \mathbb{X}_j^T \;=\; \bigcup_{k=0}^{p-1} (\mathbb{X}_{j+k} \cap \mathcal{G}_{np}) \times \{k\}$$

and hence

$$\mu_j(\mathcal{G}_{np} \times \mathbb{Z}_p) \;=\; \frac{s(p)}{p} \sum_{k=0}^{p-1} \mu(\mathbb{X}_{j+k} \cap \mathcal{G}_{np}) = \mu(\mathcal{G}_{np}),$$

where the last step follows from the fact that the sets $\mathbb{X}_0, \ldots, \mathbb{X}_{s(p)-1}$ form a partition of $\mathbb{X}$. We get the desired relation via $\mu_j(\mathcal{E}) \geqslant \mu_j(\mathcal{G}) \geqslant \limsup_{n\to\infty} \mu(\mathcal{G}_{np})$ which is larger than 0 by assumption. □

If $(\mathbb{X}, S)$ is a substitution subshift, the condition $\limsup_{n\to\infty} \mu(\mathcal{G}_{np}) > 0$ can be established using the self-similar structure. The following result is similar to Lemma 5.2.10 in spirit.

Proposition 5.3.7. *Let ϱ be a primitive substitution with corresponding subshift $(\mathbb{X}, S)$ and $p \in \mathbb{N}$. Assume that there exists a word $u \in \mathcal{L}(\mathbb{X})$ with $uuuu_1 \in \mathcal{L}(\mathbb{X})$ and the property that $|\varrho^n(u)|$ is divisible by p for infinitely many $n \in \mathbb{N}$. Then, given any T-invariant probability measure ν on $\mathbb{X} \times \mathbb{Z}_p$, for ν-almost every $(x, m) \in \mathbb{X} \times \mathbb{Z}_p$, the operator $H_{V(x,m)}$ has no eigenvalues.*

Sketch of proof. Without loss of generality, $f(x, m) = g(x_{[0,d]}, m)$ for some $d \in \mathbb{N}_0$. By Lemma 5.3.6, it suffices to show that $\limsup_{n\to\infty} \mu(\mathcal{G}_{np}) > 0$. The proof for this statement follows similar lines as for Lemma 5.2.10 and is essentially provided in [DL06b]. We also refer to [CDFG21] for details. □

Example 5.3.8. Proposition 5.3.7 applies in particular if ϱ is a primitive substitution of constant length ℓ, $p = \ell^k$ for some $k \in \mathbb{N}_0$, and $\mathrm{Ind}(\mathbb{X}) > 3$. This holds for example for the period doubling substitution $\varrho\colon a \mapsto ab, b \mapsto aa$ if $p = 2^k$, with $k \in \mathbb{N}_0$. In fact, if $f(x, m) = g(x_0, m)$ depends only on one coordinate, it is possible, for this particular example, to show that $H_{V(x,m)}$ has no eigenvalues for *any* $(x, m) \in \mathbb{X} \times \mathbb{Z}_p$ [CDFG21]. The proof of this result generalizes the methods in [Bel90, BBG91, Dam01], by modifying the trace map formalism to incorporate periodic decorations. ♢

Example 5.3.9. Consider the Fibonacci substitution $\varrho\colon a \mapsto ab, b \mapsto a$ from Example 3.1.5, with subshift $(\mathbb{X}_F, S)$, and let $p \in \mathbb{N}$ be arbitrary. It is classic that $\varrho^n(a)\varrho^n(a)\varrho^n(a)a \in \mathcal{L}(\mathbb{X}_F)$ for all large enough $n \in \mathbb{N}$ [Ber99, Prop. 4]. The substitution matrix M has determinant -1 and is hence invertible over the integers. By the pigeon hole principle, there are $j, k \in \mathbb{N}$

with $j \neq k$ such that $M^j|\Phi(a)\rangle \equiv M^k|\Phi(a)\rangle \mod p$. Since equality modulo p is preserved under applications of M and M^{-1} (both integer matrices), this shows that $\big(M^n|\Phi(a)\rangle\big)_{n\in\mathbb{Z}}$ is periodic modulo p. This periodic cycle of vectors includes $M^{-2}|\Phi(a)\rangle = (1,-1)$. Hence, there is a lattice of integers n with

$$|\vartheta^n(a)| \,=\, \langle(1,1)|M^n|\Phi(a)\rangle \,\equiv\, \langle(1,1)|(1,-1)\rangle \,=\, 0 \mod p.$$

We can thus apply Proposition 5.3.7 and obtain almost sure absence of Schrödinger eigenvalues with respect to every T-invariant measure on $\mathbb{X}_F \times \mathbb{Z}_p$. ◊

Remark 5.3.10. The Fibonacci subshift $(\mathbb{X}, S)$ is an instance of a *Sturmian subshift*. These are precisely the minimal subshifts on a binary alphabet that have minimal word complexity ($p(n) = n+1$ for all $n \in \mathbb{N}$). They also arise from codings of the translation action by an irrational number $\alpha \in (0,1)$ on the one-dimensional torus. We were able to show in [CDFG21] that Proposition 5.3.7 applies to Sturmian subshifts for Lebesgue-almost every $\alpha \in (0,1)$, uniformly for all $p \in \mathbb{N}$. More precisely, the result holds for all $\alpha \in (0,1)$ with an unbounded continued fraction expansion. This is only a sufficient condition and does *not* cover the Fibonacci subshift, for which different arguments are available; compare our discussion in Example 5.3.9. ◊

5.3.2. Random and periodic potentials

Here, we take $\mathcal{A}$ to be a finite alphabet of arbitrary cardinality and $(\mathbb{X}, S)$, with $\mathbb{X} = \mathcal{A}^{\mathbb{Z}}$ the full shift. A natural choice for an ergodic probability measure on $(\mathbb{X}, S)$ is the Bernoulli measure $\mu = \varrho^{\mathbb{Z}}$, where ϱ denotes the uniform distribution on $\mathcal{A}$. If $f(x) = g(x_0)$ depends only on the first coordinate, this constitutes an instance of the famous Anderson model. In this case, the spectral type is pure point with exponentially decaying eigensolutions for μ-almost every $x \in \mathbb{X}$; compare [BDF+19] and references therein. The spectrum is easy to determine.

Fact 5.3.11. *For the Bernoulli shift $(\mathcal{A}^{\mathbb{Z}}, S, \mu)$, with $f(x) = g(x_0)$, the almost sure spectrum is given by*

$$\Sigma \,=\, g(\mathcal{A}) + [-2,2] \,=\, \bigcup_{a\in\mathcal{A}} \sigma(H_{V(a^{\mathbb{Z}})}).$$

This result is well-known and there are several ways to prove it. We present here a line of thought that will also be fruitful for the analysis of mixed potentials later on. First, $\Sigma = \sigma(H_{V(x)})$ for every $x \in \mathcal{A}^{\mathbb{Z}}$ with a dense orbit and hence, $\sigma(H_{V(a^{\mathbb{Z}})}) \subset \Sigma$ follows from Lemma 5.1.2, for all $a \in \mathcal{A}$. On the other hand, invoking Theorem 5.1.5, we know that $\mathbb{R} \setminus \Sigma = \mathcal{UH}$ is the set of energies $E \in \mathbb{R}$ such that the associated cocycle (S, A_E) is uniformly hyperbolic. It then remains to show that (S, A_E) is uniformly hyperbolic for E in the complement of $g(\mathcal{A}) + [-2,2]$.

In the context of $\mathrm{SL}(2,\mathbb{R})$-valued cocycles, it is often convenient to view a matrix A as a map on the real projective space $\mathbb{RP}_1$. That is, identifying a point $v \in \mathbb{R}^2$ with its equivalence class in $\mathbb{RP}_1$, we view A as a function

$$A\colon \mathbb{RP}_1 \to \mathbb{RP}_1, \quad v \mapsto Av,$$

and, by linearity of A, this map is well-defined. The advantage of this point of view is that the application of a *hyperbolic* $A \in \mathrm{SL}(2,\mathbb{R})$ contracts the complement of the contracting eigendirection towards the expanding eigendirection in $\mathbb{RP}_1$. More precisely, for every connected open set $I \subset \mathbb{RP}_1$ that contains the expanding eigendirection and does not contain the contracting eigendirection, we have $AI \Subset I$, meaning that the closure of AI is contained in the interior of I. From now on, with a slight change of perspective, we also regard (S, A_E) as a map

$$(S, A_E)\colon \mathbb{X} \times \mathbb{RP}_1 \to \mathbb{X} \times \mathbb{RP}_1.$$

For the full shift, there is a useful characterization of uniformly hyperbolic cocycles; in the following form, it is taken from [ABY10, Thm. 2.2].

Theorem 5.3.12. *Let (S, A) be an $\mathrm{SL}(2,\mathbb{R})$-valued cocycle over the full shift $\mathcal{A}^{\mathbb{Z}}$, where $A(x) = T(x_0)$ for all $x \in \mathcal{A}^{\mathbb{Z}}$. Then, (S, A) is uniformly hyperbolic if and only if there is a non-empty open subset $M \subset \mathbb{RP}_1$ with $\overline{M} \neq \mathbb{RP}_1$ such that $T(a)M \Subset M$ holds for all $a \in \mathcal{A}$. It is possible to take M with finitely many connected components, and these components with disjoint closures.* ∎

Proof of Fact 5.3.11. We apply Theorem 5.3.12 to (S, A_E), with $E \notin g(\mathcal{A}) + [-2, 2]$ and $A_E(x) = T_E(x_0)$. In this case, $E \notin \sigma(H_{V(a^{\mathbb{Z}})})$, which is equivalent to $T_E(a)$ being hyperbolic with eigenvalue λ_a, with $|\lambda_a| > 1$, for all $a \in \mathcal{A}$. The corresponding expanding and contracting eigendirections are given by $(\lambda_a, 1)$ and $(\lambda_a^{-1}, 1)$, respectively. In particular, all contracting eigendirections are of the form $(x, 1)$, with $|x| < 1$, and all expanding eigendirections are of the form $(x, 1)$, with $|x| > 1$. In other words, the connected open set

$$M \;=\; \{(x, 1) : |x| > 1\} \subset \mathbb{RP}_1$$

contains all of the expanding and none of the contracting eigendirections. Thus, $T(a)M \Subset M$ for all $a \in \mathcal{A}$, and we conclude with the help of Theorem 5.3.12 that (S, A_E) is uniformly hyperbolic. □

Again, we modify this model via a p-periodic background for some $p \in \mathbb{N}$. Hence, we consider $(\mathbb{X} \times \mathbb{Z}_p, T)$ with $T(x, m) = (Sx, m + 1)$, where addition is taken modulo p, and denote by μ' the Haar measure on $\mathbb{Z}_p$. It is a straightforward exercise to verify that $\mu \times \mu'$ is ergodic on $(\mathbb{X} \times \mathbb{Z}_p, T)$. We maintain that f depends only on the first coordinate of x, that is, $f(x, m) = g(x_0, m)$, for some $g\colon \mathcal{A} \times \mathbb{Z}_p \to \mathbb{R}$. The corresponding sliding block code gives the potential

$$V(x, m) \;=\; (g(x_n, m + n))_{n \in \mathbb{Z}},$$

for all $(x, m) \in \mathbb{X} \times \mathbb{Z}_p$.

Remark 5.3.13. The space $(\mathbb{X} \times \mathbb{Z}_p, T)$ is topologically conjugate to an SFT. In fact, if g is injective on $\mathcal{A} \times \mathbb{Z}_p$, the space $(V(\mathbb{X} \times \mathbb{Z}_p), S)$ is itself an SFT. It is conceivable that this model still exhibits Anderson localization (pure point spectrum with exponentially decaying eigensolutions). Substantial steps into this direction have been made recently [ADZ20]. ◊

In what follows, we show how to calculate the $\mu \times \mu'$-almost sure spectrum Σ explicitly, if $p = 2$. Here, we will build on our discussion of the corresponding result for the Anderson model, given in Fact 5.3.11. As a first step, we relate (T, A_E) to a cocycle on a full shift.

Lemma 5.3.14. *Let $E \in \mathbb{R}$. The cocycle (T, A_E) is uniformly hyperbolic precisely if the cocycle $(S^2, \widetilde{A}_E)$ is uniformly hyperbolic on the full shift $(\mathbb{X}, S^2)$ over the alphabet $\mathcal{A}^2$, where $\widetilde{A}_E(x) = T_E(x_0, x_1)$, with*

$$T_E(x_0, x_1) \;=\; \begin{pmatrix} (E - g(x_0, 0))(E - g(x_1, 1)) - 1 & g(x_1, 1) - E \\ E - g(x_0, 0) & -1 \end{pmatrix}, \tag{5.6}$$

defines a locally constant $\mathrm{SL}(2, \mathbb{R})$-valued map on $\mathbb{X}$.

Proof. Since the norm of A_E (and its inverse) is bounded on $\mathbb{X}$, it follows that (T, A_E) is uniformly hyperbolic, precisely if (T^2, A_E^2) is uniformly hyperbolic. Under the action of T^2, $\mathbb{X} \times \mathbb{Z}_2$ decays into the invariant components $\mathbb{X} \times \{m\}$, with $m \in \mathbb{Z}_2$. Since $T(\mathbb{X} \times \{0\}) = \mathbb{X} \times \{1\}$, it suffices to consider the cocycle restricted to $(\mathbb{X} \times \{0\}, T^2) \cong (\mathbb{X}, S^2)$, where the map $(x, 0) \mapsto x$ conjugates T^2 and S^2. On this subspace,

$$A_E^2((x, 0)) \;=\; \begin{pmatrix} E - g(x_1, 1) & -1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} E - g(x_0, 0) & -1 \\ 1 & 0 \end{pmatrix} \;=\; T_E(x_0, x_1),$$

with $T_E(x_0, x_1)$ as defined in (5.6). □

We obtain the following reformulation of Theorem 5.1.15.

Proposition 5.3.15. *The $\mu \times \mu'$-almost sure Schrödinger spectrum on $(\mathbb{X} \times \mathbb{Z}_2, T)$ is*

$$\Sigma \;=\; \bigcup_{a,b \in \mathcal{A}} \sigma(H_{V((ab)^{\mathbb{Z}}, 0)}).$$

Proof. Again, $\Sigma = \mathbb{R} \setminus \mathcal{UH} = \sigma(H_{V(x,m)})$ for those points (x, m) that have a dense T-orbit in $\mathbb{X} \times \mathbb{Z}_2$, and the inclusion $\sigma(H_{V((ab)^{\mathbb{Z}}, 0)}) \subset \Sigma$ follows by strong approximation, for all $a, b \in \mathcal{A}$. Conversely, assume that $E \notin \sigma(H_{V((ab)^{\mathbb{Z}}, 0)})$, for all $a, b \in \mathcal{A}$. In this case, $T_E(a, b)$ as defined in (5.6), is hyperbolic for all $a, b \in \mathcal{A}$, and we employ similar methods as in the proof of Fact 5.3.11 to obtain that $(S^2, \widetilde{A}_E)$ is uniformly hyperbolic. More precisely, we show that there is an open connected subset $M \subset \mathbb{RP}_1$ which contains all of the expanding eigendirections and none of the contracting eigendirections, and conclude using Theorem 5.3.12.

Let $x_a = E - g(a, 0)$ and $y_b = E - g(b, 1)$. The expanding and contracting eigendirections of $T_E(a, b)$ are given by $(v^+, 1)$ and $(v^-, 1)$, respectively, where

$$v^{\pm} \;=\; v^{\pm}(x_a, y_b) \;=\; \frac{x_a}{2} \left(1 \pm \sqrt{\frac{x_a y_b - 4}{x_a y_b}} \right),$$

and $T_E(a, b)$ being hyperbolic is equivalent to $x_a y_b \notin [0, 4]$. For a moment, let us fix $y_b \neq 0$. A direct calculation yields that v^- is monotonically decreasing in x_a and that

$$\lim_{x_a \to \pm\infty} v^-(x_a, y_b) \;=\; 1/y_b.$$

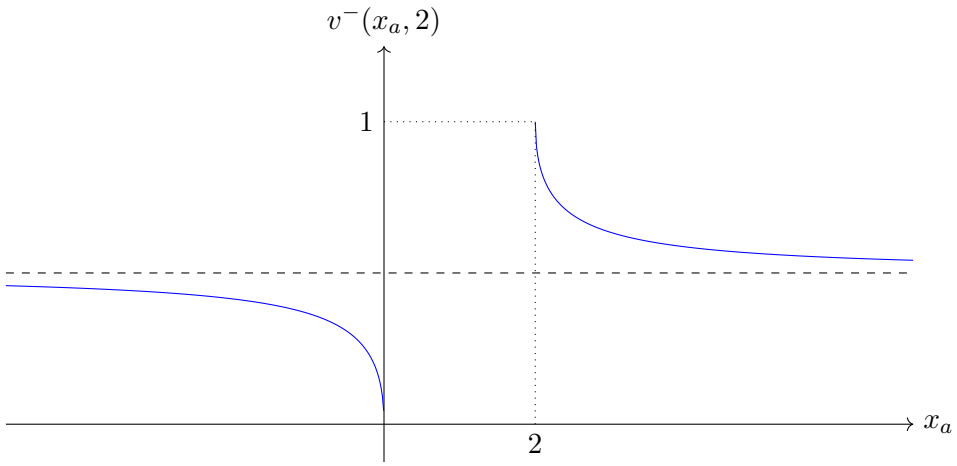

Figure 5.3.: Plot of v^- as a function of x_a, for fixed $y_b = 2$, in the region where $x_a y_b \notin [0, 4]$. The dashed line indicates the asymptotic value as $|x_a| \to \infty$.

From the boundary cases $v^-(x_a, y_b) = x_a/2 = 2/y_b$ for $x_a y_b = 4$ and $v^-(x_a, y_b) = 0$ for $x_a = 0$, we infer that $v^-(x_a, y_b)$ lies strictly between 0 and $2/y_b$ for all x_a with $x_a y_b \notin [0, 4]$. This is illustrated in Figure 5.3. Let

$$y_+ \;=\; \inf\{y_b \mid b \in \mathcal{A}, y_b > 0\} \quad \text{and} \quad y_- \;=\; \sup\{y_b \mid b \in \mathcal{A}, y_b < 0\}$$

be the smallest positive and the largest negative value of y_b, respectively. If $y_b = y_+$, then $T_E(a, b)$ being hyperbolic implies that $x_a < 0$ or $x_a > 4/y_+$ for all $a \in \mathcal{A}$. Similarly, we obtain that $x_a > 0$ or $x_a < 4/y_-$. This is still valid for the degenerate cases $y_+ = \infty$ and $y_- = -\infty$. In summary, we have

$$x_a \;<\; \frac{4}{y_-} \quad \text{or} \quad x_a \;>\; \frac{4}{y_+},$$

for all $a \in \mathcal{A}$. For a moment, assume that $x_a, y_b > 0$. Then,

$$v^+(x_a, y_b) \;>\; \frac{x_a}{2} \;>\; \frac{2}{y_+} \quad \text{and} \quad 0 \;<\; v^-(x_a, y_b) \;<\; \frac{2}{y_b} \;\leqslant\; \frac{2}{y_+}.$$

Exhausting all possible cases in a similar manner, we obtain that

$$v^-(x_a, y_b) \in \left(\frac{2}{y_-}, \frac{2}{y_+}\right) \quad \text{and} \quad v^+(x_a, y_b) \in \left[\frac{2}{y_-}, \frac{2}{y_+}\right]^C \tag{5.7}$$

for all $a, b \in \mathcal{A}$. It follows that there is an open connected subset $M \subset \mathbb{RP}_1$ such that $T_E(a, b)M \Subset M$, for all $a, b \in \mathcal{A}$. Theorem 5.3.12 in conjunction with Lemma 5.3.14 implies that $E \in \mathcal{UH} = \mathbb{R} \setminus \Sigma$. □

The result in Proposition 5.3.15 allows us to explicitly compute the $\mu \times \mu'$-almost sure spectrum Σ. This is because $\sigma(H_{V((ab)^{\mathbb{Z}},0)})$ coincides with the set of those $E \in \mathbb{R}$ such that

$$\operatorname{tr}\big(T_E(a, b)\big) \;=\; \big(E - g(a, 0)\big)\big(E - g(b, 1)\big) - 2 \;\in\; [-2, 2],$$

which is the union of two closed intervals. The intervals for different choices of $ab \in \mathcal{A}^2$ may or may not overlap, depending on the choice of the function g.

Remark 5.3.16. Comparing the result in Proposition 5.3.15 with Fact 5.3.11, we observe that, in both cases, the almost sure spectrum can be built from those sequences which are constant under an associated full shift. In the situation of Proposition 5.3.15, this full shift is given by S^2, the square of the original shift action. This is the reason for the more complicated form of the corresponding cocycle, as detailed in (5.6), and thereby for the additional steps required in the proof of Proposition 5.3.15. More generally, the space $(\mathbb{X} \times \mathbb{Z}_p, T)$ may be regarded as a discrete suspension of the full shift $(\mathbb{X}, S^p)$ over the alphabet $\mathcal{A}^p$, for all $p \in \mathbb{N}$, and it seems natural to ask about a generalization of Proposition 5.3.15. However, already for $p = 3$, the inclusion $\Sigma \subset \cup_{a,b,c \in \mathcal{A}} \sigma(H_{V((abc)^{\mathbb{Z}},0)})$ does *not* hold in general, as can be shown by numerical calculations. Hence, hyperbolicity of the matrices $T_E(a,b,c) = A_E^3((abc)^{\mathbb{Z}}, 0)$ no longer suffices to guarantee uniform hyperbolicity of the associated cocycle on the full shift. It would be desirable to understand—at least on a heuristic level—why the mechanism for showing uniform hyperbolicity breaks down for $p = 3$. However, the dependence of the eigendirections of $T_E(a,b,c)$ on the spectral parameter E is already quite involved, such that an analysis similar to that in the proof of Proposition 5.3.15 would become tedious. It therefore seems that some additional insight is required in order to proceed. $\Diamond$

6. The TM measure as a g-measure

The guiding theme of this chapter is an in-depth study of the classic *Thue–Morse measure*,

$$\mu_{\mathrm{TM}} \,=\, \prod_{k=0}^{\infty} (1-\cos(2\pi 2^k x)), \tag{6.1}$$

to be understood as a limit of absolutely continuous measures on $\mathbb{R}$ in the vague topology. The Thue–Morse (TM) measure is a paradigmatic example of a singular continuous *Riesz product*, for which the infinite product notation employed in (6.1) is the usual convention [Zyg02, Ch. V.7]. From the point of view of aperiodic order, the TM measure is important because it arises as a diffraction measure from the extensively studied TM substitution $\varrho\colon a \mapsto ab, b \mapsto ba$ [BG13, Ch. 10]. It is in fact a measure of maximal spectral type for the associated shift-dynamical system in the orthocomplement of the pure point component [Que10]. It is possibly the easiest (and certainly the most famous) example of a singular continuous measure arising from a substitution dynamical system; compare [Mah27] for the historic origins, and [Que18] for a recent review.

While the TM measure is certainly the main interest and the guiding example throughout this chapter, we sometimes take a detour into more general (or related) classes of measures. The reason for this are twofold. On the one hand, this offers new perspectives on the TM measure and makes the link to other mathematical fields such as stochastic processes and the thermodynamic formalism more transparent. On the other hand, it allows us to formulate several results in proper generality and paves the way for an analogous treatment of a larger class of singular continuous (diffraction) measures.

We start by collecting some of the characteristic features of the TM measure. As the diffraction measure of a weighted Dirac comb on the lattice, it is 1-periodic and can hence be decomposed as $\mu_{\mathrm{TM}} = \delta_{\mathbb{Z}} * \mu_{\mathrm{TM}}|_{[0,1)}$. In the following we therefore regard μ_{TM} as a probability measure on the torus $\mathbb{T}$, represented by the interval $[0,1)$. This is the standard approach for the diffraction of constant length substitutions; compare [Que10]. The inflation factor $\lambda = 2$ of the TM substitution also leaves an imprint on its diffraction measure μ_{TM}. More precisely, the TM measure is strongly mixing for the doubling map

$$T_2\colon x \mapsto 2x \mod 1,$$

on the torus $\mathbb{T}$ [Kea72]. The self-similar structure of μ_{TM} with respect to the doubling map also manifests itself in the infinite product representation (6.1). Indeed, denoting by λ_{L} the Lebesgue measure on $\mathbb{T}$, the TM measure is the weak limit of the probability measures $2^n g_n \lambda_{\mathrm{L}}$, where

$$g_n(x) \,=\, \prod_{k=0}^{n-1} g(T_2^k x), \quad g(x) \,=\, \frac{1}{2}(1-\cos(2\pi x)). \tag{6.2}$$

for all $n \in \mathbb{N}$. It is an easy but crucial observation that

$$\sum_{y \in T_2^{-1}(x)} g(y) \;=\; g\big(\tfrac{x}{2}\big) + g\big(\tfrac{x+1}{2}\big) = 1, \tag{6.3}$$

for all $x \in \mathbb{T}$. This allows us to interpret g as the transition probabilities of a stationary random walk on $\mathbb{T}$, where x is mapped to $x/2$ with probability $g(x/2)$ and to $(x+1)/2$ with probability $g((x+1)/2)$ for all $x \in \mathbb{T}$. The measure μ_{TM} turns out to be the unique invariant (Borel) probability measure under this process. More generally, given a measurable function g that satisfies (6.3), the invariant probability measures under the corresponding random walk became known as *g-measures* [Kea72]. In Section 6.1, we will give a more detailed introduction to g-measures and discuss under which condition they are unique, singular continuous and can be represented as an infinite product.

Since singular continuous measures assign full measure to certain sets of Lebesgue measure 0, but no measure to individual points, it is often difficult to assess or illustrate their mass distribution. We approach this problem by analyzing the local scaling behaviour, quantifying how much mass a measure μ accumulates around a given point x. This is done by determining how fast the measure decays on shrinking neighbourhoods $B_r(x)$ as $r \to 0$. More precisely, we are interested in the *upper and lower local dimension* of a measure μ at x, given by

$$\underline{d}_\mu(x) \;=\; \liminf_{r\to 0} \frac{\log(\mu(B_r(x)))}{\log(r)}, \quad \overline{d}_\mu(x) \;=\; \limsup_{r\to 0} \frac{\log(\mu(B_r(x)))}{\log(r)},$$

respectively, and denote by $d_\mu(x)$ their common value if the limit exists. In this case, $d_\mu(x)$ is called the *local dimension* of μ at x, and $\mu(B_r(x))$ scales roughly like $r^{d_\mu(x)}$ as $r \to 0$. Obviously, $d_\mu(x) = \infty$ if $B_r(x)$ is eventually empty, that is, if x is not in the support of μ. However, even if $x \in \operatorname{supp}(\mu)$, it is possible that $d_\mu(x) = \infty$ due to a decay of $\mu(B_r(x))$ that is faster than any polynomial function in r. As we shall see in Section 6.2, this is indeed the case for μ_{TM} and every dyadic point $x \in \mathbb{T}$. In fact, we show that this behaviour is universal for g-measures, if the corresponding function g has a unique zero at $x = 0$, and is polynomially bounded in a neighbourhood of 0. In these cases, we provide a more refined analysis of the scaling behaviour at the dyadic points in Theorem 6.2.2.

Although $d_\mu(x)$ reveals no useful information about $\mu = \mu_{\mathrm{TM}}$ if x is a dyadic point, there are still many points $x \in \mathbb{T}$ such that $d_\mu(x)$ is a well-defined real number. In fact, we will see that for all α larger than some minimal scaling exponent, the level sets

$$\mathcal{F}(\alpha) \;=\; \{x \in \mathbb{T} : d_\mu(x) = \alpha\}$$

have positive Hausdorff dimension. Investigating the corresponding dimension

$$f(\alpha) \;=\; \dim_{\mathrm{H}} \mathcal{F}(\alpha),$$

as a function of α is a typical exercise in multifractal analysis, which we pursue in Section 6.3. The corresponding theory is well-developed if μ is a g-measure corresponding to a *strictly* positive function g. However, many of the classical methods break down if g is allowed to have a 0, as is the case for the TM measure. We circumvent this problem by exhausting $\mathbb{T}$ with T_2-invariant subsets of points that avoid a small neighbourhood of the critical point 0. The chapter culminates with the following result; compare [BGKS19, Thm. 1].

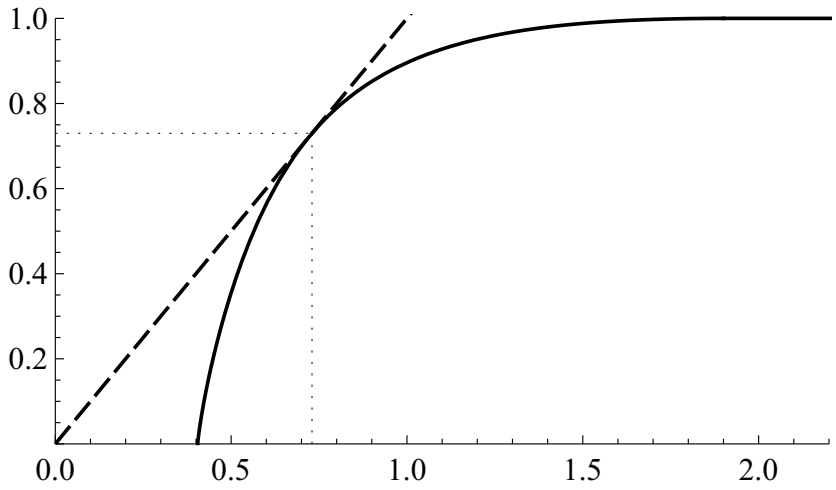

Figure 6.1.: The graph of the dimension spectrum $\alpha \mapsto f(\alpha)$ (solid line). The dotted lines indicate the position of the fixed point $\alpha^* = f(\alpha^*)$.

Theorem 6.0.1. *Let* $\alpha_0 = 2 - \log(3)/\log(2)$ *and* $\alpha_1 = 2$. *The level sets* $\mathcal{F}(\alpha)$ *are empty for* $\alpha < \alpha_0$ *and they are dense for* $\alpha \geqslant \alpha_0$. *The exceptional set*

$$\widehat{\mathcal{F}} = \{x \in \mathbb{T} : d_\mu(x) \text{ is not defined}\}$$

has full Hausdorff dimension $\dim_{\mathrm{H}} \widehat{\mathcal{F}} = 1$. *The function* f *is equal to* 0 *on* $(-\infty, \alpha_0]$, *it is concave on* $[\alpha_0, \infty)$ *and it is equal to* 1 *on* $[\alpha_1, \infty)$. *On the interval* (α_0, α_1), *there is a fixed point* $\alpha^* = f(\alpha^*)$, *satisfying*

$$\alpha^* = \frac{1}{\log(2)} h_\mu,$$

where $h_\mu = h_\mu(T_2)$ *denotes the entropy of* μ *on* $(\mathbb{T}, T_2)$.

Remark 6.0.2. The multifractal analysis of the TM measure was listed as an open problem in Queffelec's recent review [Que18], despite its early appearance in the physics literature [GL90]. The entropy h_μ can be calculated numerically with very high precision by interpreting μ as an *equilibrium measure* (compare Section 6.3.1) and using ideas from [ZPK99]; see [BGKS19, Rem. 9.4]. This answers affirmatively a conjecture stated by Queffelec in her review paper [Que18, Sec. 4.4.1]. ◊

6.1. Some background on g-measures

There are several ways to introduce the notion of a g-measure. Doing so via an associated Markov chain is a matter of taste and can certainly be avoided—compare the classic expositions in [Kea72] and [Led74]. However, this perspective offers an additional interpretation of g-measures as invariant distributions, and makes powerful tools from martingale theory available. These have been used to great avail by Conze and Raugi in [CR90], which is an

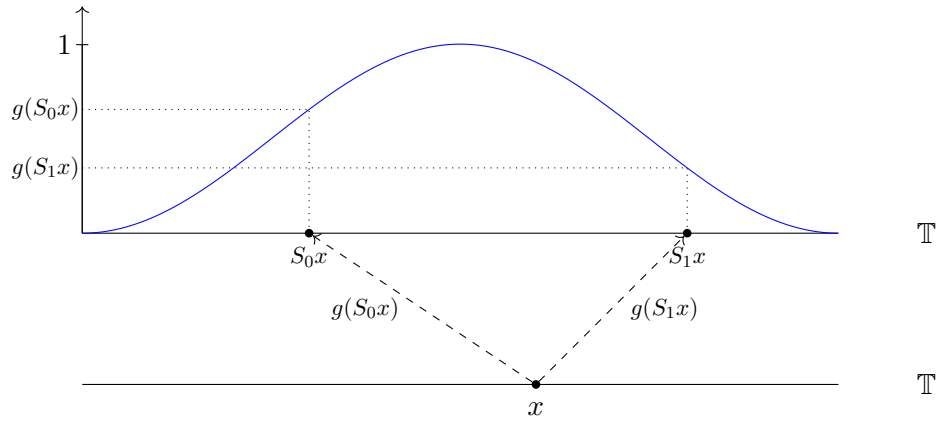

Figure 6.2.: Illustration of the random walk induced by the g-function on $(\mathbb{T}, T_2)$ that was given in (6.2).

invaluable resource for the classification of g-measures in terms of the zeros of the corresponding g-function; we present some of their main results in Section 6.1.2. We only mention in passing that—after the original work by Keane, motivated by the analysis of diffraction measures [Kea72]—the study of g-measures has found applications in several fields, such as multifractal analysis [BGKS19, Fan97, Oli99], wavelet theory [CDF92, CR90, FL98] and learning models [BK93].

Although we are ultimately interested in invariant measures under the doubling map $(\mathbb{T}, T_2)$, some useful methods implicitly require to consider powers of the map T_2, given by $T_2^k = T_{2^k}$. We therefore take a step back and consider a more general expansive transformation on the torus $\mathbb{T}$, given by

$$T_p \colon x \mapsto px \mod 1,$$

for some $p \in \mathbb{N} \setminus \{1\}$.

The map T_p is p-to-1 and hence certainly not invertible. In fact, there are p different (right) inverse branches, given by

$$S_j \colon x \mapsto \frac{x+j}{p},$$

for $0 \leqslant j \leqslant p-1$, each satisfying $T \circ S_j(x) = x$ on $\mathbb{T}$. We construct a "stochastic inverse" of T_p as follows. For each $x \in \mathbb{T}$, map x to a point in $T_p^{-1}(x)$ according to some prescribed probability vector on

$$T_P^{-1}(x) \;=\; \{S_0x, S_1x, \cdots, S_{p-1}x\}.$$

That is, x maps to $S_jx \in T_p^{-1}(x)$ with probability $g(S_jx)$ for some $g \colon \mathbb{T} \to [0,1]$, satisfying

$$\sum_{j=0}^{p-1} g(S_jx) \;=\; 1, \tag{6.4}$$

for all $x \in \mathbb{T}$; compare Figure 6.2.

Definition 6.1.1. A (Borel-)measurable function $g\colon \mathbb{T} \to [0,1]$ is called a *g-function* on $(\mathbb{T}, T_p)$ if it satisfies (6.4) for all $x \in \mathbb{T}$.

Every g-function $(\mathbb{T}, T_p)$ induces a random walk on $\mathbb{T}$ as described above. These processes have been studied as *chains with infinite connections* [DoF37] or *chains of infinite order* [Har55] in the early literature. The random walk can be described via a Markov kernel on $\mathbb{T}$, defined by

$$\kappa(x, \cdot) \,=\, \sum_{j=0}^{p-1} g(jx)\delta_{jx}(\cdot),$$

for all $x \in \mathbb{T}$. Given a probability measure μ on $\mathbb{T}$ (initial distribution) and a function $f \in C(\mathbb{T})$ (observable), the expected value of f is described by $\mu(f)$. After one iteration of the Markov process, the expected value of f changes to

$$\int_{\mathbb{T}} \int_{\mathbb{T}} f(y)\, \kappa(x, \mathrm{d}y)\, \mathrm{d}\mu(x).$$

It is a matter of perspective whether we think of the time evolution as acting on the observable f or on the initial distribution μ. The former point of view is reflected by the transfer (or time evolution) operator φ_g on $C(\mathbb{T})$, defined by

$$\varphi_g f\colon x \mapsto \int_{\mathbb{T}} f(y)\, \kappa(x, \mathrm{d}y) \,=\, \sum_{j=0}^{p-1} g(jx) f(jx),$$

for all $f \in C(\mathbb{T})$. Alternatively, the time evolution of μ is induced by the dual of φ_g (which is also the dual operator of the transition kernel κ), acting on μ via

$$(\varphi_g)_*\mu\colon A \mapsto \int_{\mathbb{T}} \kappa(x, A)\, \mathrm{d}\mu(x),$$

which satisfies $((\varphi_g)_*\mu)(f) = \mu(\varphi_g f)$ for all $f \in C(\mathbb{T})$ by construction. Functions and probability measures that are invariant under the random walk are of particular interest.

Definition 6.1.2. Let g be a g-function on $(\mathbb{T}, T_p)$. A function $f \in C(\mathbb{T})$ is called *g-harmonic* if it satisfies $\varphi_g f = f$. We call $\mu \in \mathcal{M}^1(\mathbb{T})$ a *g-measure* if $(\varphi_g)_*\mu = \mu$.

By the defining relation of g-functions (6.4), every constant function f is g-harmonic. This also implies that $(\varphi_g)_*$ maps probability measures to probability measures (as it should), due to $((\varphi_g)_*\mu)(1) = \mu(\varphi_g 1) = \mu(1)$.

The fact that the Markov process is a one-sided inverse of T_p finds its complement in operator-theoretic terms. Indeed, the Koopman operator $U_{T_p}\colon f \mapsto f \circ T_p$, is a right inverse of φ_g, that is,

$$(\varphi_g \circ U_{T_p})(f) \,=\, \varphi_g(f \circ T_p) = f,$$

for all $f \in C(\mathbb{T})$. Similarly, $((\varphi_g)_*\mu) \circ T_p^{-1} = \mu$ for all $\mu \in \mathcal{M}^1(\mathbb{T})$. In particular, every g-measure on $(\mathbb{T}, T_p)$ is automatically T_p-invariant.

Remark 6.1.3. Conversely, every T_p-invariant measure μ is a g-measure for some appropriate g-function, that is, T_p-invariant measures and g-measures are in 1-to-1 correspondence. The corresponding g-function is the Radon–Nikodym derivative of μ with respect to the *local* lift $\mu \circ T_p$ [Kea72]. This is consistent with the idea that g describes how much of the mass at $T_p x$ is transported to x—instead of any of the other points in $T_p^{-1}(T_p(x))$—under the Markov chain. ◊

For later applications, it will be convenient to have a closed form expression for the powers of the operator φ_g.

Lemma 6.1.4. *Let g be a g-function on $(\mathbb{T}, T_p)$. For every $n \in \mathbb{N}$, we have $\varphi_g^n = \varphi_{g_n}$, where g_n is a g-function on $(\mathbb{T}, T_{p^n})$, given by*

$$g_n \colon x \mapsto \prod_{k=0}^{n-1} g(T_p^k x). \tag{6.5}$$

Sketch of proof. First, note that $T_{p^n} = T_p^n$. Iterating the random walk n times, the point $T_p^n x$ follows the trajectory $T_p^n x \mapsto T_p^{n-1} x \mapsto \ldots \mapsto x$ with probability

$$g_n(x) \;=\; g(T_p^{n-1}x) g(T_p^{n-2}x) \cdots g(x).$$

Hence, g_n is the g-function corresponding to the n-th power of the original Markov chain. □

6.1.1. Lebesgue decomposition of g-measures

Given $\mu \in \mathcal{M}^1(\mathbb{T})$, we denote by $\mu = \mu_{\mathrm{pp}} + \mu_{\mathrm{sc}} + \mu_{\mathrm{ac}}$ the Lebesgue decomposition into the spectral components of μ. Conveniently, φ_g acts separately on each of the spectral components.

Lemma 6.1.5. *We have $(\varphi_g)_*(\mu_\bullet) = ((\varphi_g)_*\mu)_\bullet$ for $\bullet \in \{\mathrm{pp}, \mathrm{sc}, \mathrm{ac}\}$.*

Proof. By the uniqueness of the Lebesgue decomposition, it suffices to show that φ_g leaves the classes of pure point/continuous/absolutely continuous and singular measures invariant. This follows readily from the identity $((\varphi_g)_*\mu) \circ T_p^{-1} = \mu$. First, if μ is supported on a countable set A, then $(\varphi_g)_*\mu$ is supported on the countable set $T_p^{-1}(A)$. Conversely, if μ is continuous, so is $(\varphi_g)_*\mu$ since,

$$((\varphi_g)_*\mu)(\{x\}) \;\leqslant\; ((\varphi_g)_*\mu) \circ T_p^{-1}(\{T_p(x)\}) \;=\; \mu(\{T_p(x)\}) = 0,$$

for all $x \in \mathbb{T}$. Assume that $\mu \ll \lambda_{\mathrm{L}}$ and let $\lambda_{\mathrm{L}}(A) = 0$. Then, $\lambda_{\mathrm{L}}(T_p A) \leqslant p\lambda_{\mathrm{L}}(A) = 0$ and hence,

$$((\varphi_g)_*\mu)(A) \leqslant \mu(T_p A) \;=\; 0,$$

implying that $(\varphi_g)_*\mu \ll \lambda_{\mathrm{L}}$. On the other hand, if $\mu \perp \lambda_{\mathrm{L}}$, there is a Borel-measurable set B with $\mu(B) = 1$ and $\lambda_{\mathrm{L}}(B) = 0$. It follows that $((\varphi_g)_*\mu)(T_p^{-1}B) = \mu(B) = 1$ and $\lambda_{\mathrm{L}}(T_p^{-1}(B)) = \lambda_{\mathrm{L}}(B) = 0$, since λ_{L} is T_p-invariant. Hence, $(\varphi_g)_*\mu \perp \lambda_{\mathrm{L}}$, and all cases are complete. □

Lemma 6.1.5 immediately implies the following classic result on g-measures; compare [BDEG88, Thm. 1.2] and [DuF66, Lemma 2.2].

Corollary 6.1.6. *For every g-measure μ, we have $(\varphi_g)_*(\mu_\bullet) = \mu_\bullet$, for each $\bullet \in \{\mathrm{pp}, \mathrm{sc}, \mathrm{ac}\}$. In particular, if the g-measure μ is unique, it is of pure type.* ∎

Given a g-function, both the absolutely continuous g-measures and the pure point g-measures are easy to determine. We start by discussing the pure point case. Strong restrictions already follow from the requirement that every g-measure is also T_p-invariant. Let us denote by $\mathbb{T}(\mathrm{per})$ the set of T_p-periodic points in $\mathbb{T}$ and by $\mathrm{orb}(x) = \{T_p^n(x) : n \in \mathbb{N}_0\}$ the T_p-orbit of $x \in \mathbb{T}$. The following result is probably standard, we give an elementary proof for the reader's convenience; compare also [BCEG21, Prop. 3.2] for the special case $p = 2$.

Lemma 6.1.7. *Let μ be a T_p-invariant, finite pure point measure on $\mathbb{T}$. Then, μ is a linear combination of measures in $\{\mu_x : x \in \mathbb{T}(\mathrm{per})\}$, where μ_x is the uniform distribution on the orbit of x, given by*

$$\mu_x \,=\, \frac{1}{\#\,\mathrm{orb}(x)} \sum_{y \in \mathrm{orb}(x)} \delta_y .$$

Proof. Since $x \in T_p^{-1}(T_p x)$, we have

$$\mu(\{T_p x\}) \,=\, \mu \circ T_p^{-1}(\{T_p x\}) \,\geqslant\, \mu(\{x\}),$$

for all $x \in \mathbb{T}$. Hence, if $\mu(\{x\}) > 0$, then $\mu(\{y\}) \geqslant \mu(\{x\})$ for all $y \in \mathrm{orb}(x)$. Because μ has finite mass, this requires that $\mathrm{orb}(x)$ is a finite set and that μ is eventually constant on $\mathrm{orb}(x)$. Hence, there are $n, m \in \mathbb{N}$ with $n > m$ such that $T_p^m x = T_p^n x$. If $m > 0$, $T_p^{m-1} x$ and $T_p^{n-1} x$ are both contained in $T_p^{-1}(T_p^n x)$. Assuming $T_p^{m-1} x \neq T_p^{n-1} x$, would imply

$$\mu(\{T_p^{n-1} x\}) + \mu(\{T_p^{m-1} x\}) \,\leqslant\, \mu(\{T_p^{-1}(T_p^n x)\}) \,=\, \mu(\{T_p^n x\}) \,=\, \mu(\{T_p^{n-1} x\}),$$

and hence $\mu(\{T_p^{m-1} x\}) = 0$, contradicting $\mu(\{T_p^{m-1} x\}) \geqslant \mu(\{x\}) > 0$. It follows that $\mathrm{orb}(x)$ is in fact periodic and that the restriction of μ to this orbit is a multiple of μ_x. □

Proposition 6.1.8. *Given a g-function g, the pure point g-measures are precisely the convex combinations of measures in*

$$\big\{\mu_x : x \in \mathbb{T}(\mathrm{per}) \text{ and } \mathrm{orb}(x) \subset g^{-1}(1)\big\}.$$

Proof. By Lemma 6.1.7, every g-measure μ is a linear combination of measures of the form μ_x with $x \in \mathbb{T}(\mathrm{per})$, and μ being a probability measure requires that this combination is *convex*. Since the Markov chain maps $T_p y$ with probability $g(y)$ to y, it is straightforward to verify that

$$((\varphi_g)_* \mu)(\{y\}) \,=\, g(y) \mu(\{T_p y\}),$$

for all $y \in \mathbb{T}$. Given $x \in \mathbb{T}(\mathrm{per})$, this implies

$$((\varphi_g)_* \mu_x)(\{y\}) \,=\, g(y) \mu_x(\{T_p y\}) \,=\, g(y) \mu_x(\{y\}),$$

and hence μ_x is a g-measure precisely if $g(y) = 1$ for all $y \in \mathrm{orb}(x)$. □

This settles the classification of pure point g-measures. Within the class of absolutely continuous probability measures on $\mathbb{T}$, only the Lebesgue measure is T_p-invariant.

Lemma 6.1.9. *Let $\mu \in \mathcal{M}^1(\mathbb{T})$ be a T_p-invariant measure with $\mu \ll \lambda_{\mathrm{L}}$. Then, $\mu = \lambda_{\mathrm{L}}$.*

Proof. For $n \in \mathbb{Z}$, let $f_n \in C(\mathbb{T})$ be the function $f_n \colon x \mapsto \mathrm{e}^{2\pi \mathrm{i} n x}$. By the T_p-invariance of μ, we obtain

$$\widehat{\mu}(n) \,=\, \mu(f_n) \,=\, \mu(f_n \circ T_p) \,=\, \mu(f_{pn}) \,=\, \widehat{\mu}(pn),$$

for all $n \in \mathbb{Z}$. If $n \neq 0$, this yields

$$\widehat{\mu}(n) \,=\, \lim_{k \to \infty} \widehat{\mu}(p^k n) = 0,$$

due to the Riemann–Lebesgue lemma. Further, $\widehat{\mu}(0) = \mu(1) = 1$, since μ is assumed to be a probability measure. By the uniqueness of the Fourier–Stieltjes coefficients $(\widehat{\mu}(n))_{n \in \mathbb{Z}}$, see [Rud62, Thm. 1.3.6], this implies that $\mu = \lambda_{\mathrm{L}}$. □

Proposition 6.1.10. *Given a g-function g, there exists a unique absolutely continuous g-measure, given by λ_{L}, precisely if $g(x) = \frac{1}{p}$ for λ_{L}-almost every $x \in \mathbb{T}$. Otherwise, there is no absolutely continuous g-measure.*

Proof. By Lemma 6.1.9, λ_{L} is the only candidate for an absolutely continuous g-measure. For every $f \in C(\mathbb{T})$, we obtain

$$((\varphi_g)_* \lambda_{\mathrm{L}})(f) \,=\, \lambda_{\mathrm{L}}(\varphi_g f) \,=\, \sum_{j=0}^{p-1} \int_{\mathbb{T}} g(S_j x) f(S_j x) \, \mathrm{d}x \,=\, \sum_{j=0}^{p-1} \int_{S_j \mathbb{T}} p \, g(y) f(y) \, \mathrm{d}y \,=\, (pg\lambda_{\mathrm{L}})(f),$$

and hence $(\varphi_g)_* \lambda_{\mathrm{L}} = \lambda_{\mathrm{L}}$ precisely if $pg(x) = 1$ for λ_{L}-almost every $x \in \mathbb{T}$. □

Example 6.1.11. Let $p = 2$. Consider the g-function given by $g(0) = 1$, $g(\frac{1}{2}) = 0$ and $g(x) = \frac{1}{2}$, otherwise. Then, δ_0 and λ_{L} are both g-measures, according to Propisition 6.1.8 and Proposition 6.1.10. In particular, the g-measure is not unique and there are in fact g-measures of different spectral types. ♢

Remark 6.1.12. The idea behind the construction in Example 6.1.11 can easily be generalized. Recall from Proposition 6.1.8 that the uniform distribution μ_x on a periodic orbit $\mathrm{orb}(x)$ is a g-measure precisely if $g(y) = 1$ for all $y \in \mathrm{orb}(x)$. This is a condition on only finitely many points. Hence, given any g-function g and a periodic point x, there is a g-function g' that differs from g only on finitely many points such that μ_x is a g-measure for g'. On the other hand, if μ is a *continuous* g-measure for g, then it is also a g-measure for g', since individual points carry no weight. Hence, continuous g-measures can be combined with arbitrary pure point g-measures for the same g-function, unless we restrict the class of g-functions by further regularity assumptions. ♢

The singular continuous g-measures are much harder to assess. In general, the Lebesgue measure of their topological support can be arbitrarily small and even 0.

Example 6.1.13. Consider the g-function on $(\mathbb{T}, T_3)$, given by $g(S_0 x) = g(S_2 x) = \frac{1}{2}$ and $g(S_1 x) = 0$, for all $x \in \mathbb{T}$. It is straightforward to verify that the uniform distribution on the classical middle-third Cantor set is a g-measure for g. ◊

Conversely, it is a classic result that every g-measure has full support if g is strictly positive [Wal75, Lemma 2.1]. If the g-measure is assumed to be continuous, this conclusion also holds under weaker assumptions.

Proposition 6.1.14. *Let g be a g-function with at most countably many zeros and let μ be a continuous g-measure for g. Then, the topological support of μ equals $\mathbb{T}$.*

Proof. For every $n \in \mathbb{N}$, the sets $S_{j_n} \cdots S_{j_1} \mathbb{T}$, with $0 \leqslant j_i \leqslant p-1$, are half-open intervals of length p^{-n} that form a disjoint partition of $\mathbb{T}$. Due to their arbitrarily small diameter, it suffices to show that μ assigns positive measure to each such interval. For $J = S_{j_n} \cdots S_{j_1} \mathbb{T}$, we obtain by Lemma 6.1.4 (and extending the action of φ_g to step functions),

$$(\varphi_g^n \mathbb{1}_J)(x) \,=\, \sum_{y \in T_p^{-n} x} g_n(y) \mathbb{1}_J(y) \,=\, g_n(S_{j_n} \cdots S_{j_1} x) \,=:\, f_J(x),$$

for all $x \in \mathbb{T}$. Since μ is assumed to be invariant under $(\varphi_g)_*$, we obtain

$$\mu(J) \,=\, \mu(\varphi_g^n \mathbb{1}_J) \,=\, \mu(f_J).$$

Since g has at most countably many zeros, the same holds for g_n and hence for f_J. For every $k \in \mathbb{N}$, let $A_k = \{x \in \mathbb{T} : f_J(x) \geqslant 1/k\}$ and $B = f_J^{-1}(0)$. Then, $\bigcup_{k \in \mathbb{N}} A_k \cup B = \mathbb{T}$ and since μ is continuous, it assigns no mass to B, yielding $1 = \mu(\mathbb{T}) \leqslant \sum_{k \in \mathbb{N}} \mu(A_k)$. That is, there is some k with $\mu(A_k) > 0$. Since f_J is non-negative, we obtain

$$\mu(f_J) \,\geqslant\, \frac{1}{k} \mu(A_k) \,>\, 0,$$

and the claim follows. □

6.1.2. Existence, uniqueness and product representation

The aim for this section is to determine conditions on a g-function such that there is a unique corresponding g-measure that can be written as an infinite product, as is the case for the Thue–Morse measure. Indeed, a product form naturally arises by iterating $(\varphi_g)_*$ on the Lebesgue measure; compare [FL98, Prop. 1].

Lemma 6.1.15. *For every $n \in \mathbb{N}$, we have $(\varphi_g^n)_* \lambda_{\mathrm{L}} = p^n g_n \lambda_{\mathrm{L}}$, with g_n as defined in* (6.5).

Proof. By Lemma 6.1.4, we have $\varphi_g^n = \varphi_{g_n}$. The identity $(\varphi_{g_n})_* \lambda_{\mathrm{L}} = p^n g_n \lambda_{\mathrm{L}}$ follows just like in the proof of Proposition 6.1.10. □

That is, $(\varphi_g^n)_* \lambda_{\mathrm{L}}$ is an absolutely continuous measure with probability density given by $\prod_{k=0}^{n-1} p g(T_p^k x)$. If the weak limit $\mu = \lim_{n \to \infty} (\varphi_g^n)_* \lambda_{\mathrm{L}}$ exists, we denote this measure by

$$\prod_{k=0}^{\infty} p\, g(T_p^k x),$$

with slight abuse of notation. This is what we call the *product representation* of μ. Such a measure may or may not be a g-measure. It certainly *is* if we assume that g is continuous and hence that $(\varphi_g)_*$ is a continuous operator on $\mathcal{M}^1(\mathbb{T})$, endowed with the weak topology.

In fact, there are more reasons to restrict our attention to *continuous* g-functions. As we discussed in Remark 6.1.12, we cannot hope to obtain a unique g-measure for general g-functions. In fact, there are even g-functions such that no associated g-measure exists [Kea72]. Conversely, the existence of g-measures is automatic if g is continuous.

Fact 6.1.16 ([Kea72])**.** *If g is a continuous g-function, there is at least one associated g-measure.* ∎

However, even in the much studied case that g is strictly positive, continuity of g does not suffice to obtain *uniqueness* of the associated g-measure [Kea72]. It is therefore common to demand some control over the modulus of continuity of g.

Definition 6.1.17. For $f \in C(\mathbb{T})$ and $\delta > 0$, let

$$f[\delta] = \sup_{|x-y|\leqslant\delta} |f(x) - f(y)|.$$

We say that f is of *bounded variation* over $(\mathbb{T}, T_p)$ if $\sum_{j=0}^{\infty} f[p^{-j}\delta] < \infty$ for some (equivalently all) $\delta > 0$.

It is easy to verify that every Hölder continuous function is of bounded variation. If $g > 0$, bounded variation of g suffices to conclude that there is a unique g-measure [Wal75]. Finding weaker regularity assumptions on *strictly* positive g-functions that ensure uniqueness of the corresponding g-measure is an active area of research; compare for example [FGP20, JOP12] and references therein. If g is allowed to have zeros, bounded variation does not suffice to ensure uniqueness, as is obvious from Proposition 6.1.8. However, even in this case, bounded variation is still a useful concept; compare our discussion in [BCEG21].

Remark 6.1.18. To appreciate the difference between the cases $g > 0$ and $g \geqslant 0$, we invite the reader to recall the situation for *finite* state Markov chains. If the Markov matrix is strictly positive, there is a unique, attractive invariant distribution. Conversely, for arbitrary Markov matrices, every set of states that is invariant under the process supports an invariant distribution. Even if the matrix is irreducible, it is possible to approach a limit-cycle of distributions that are invariant under some power of the Markov matrix; compare for example [Bre20] for background. Similar phenomena occur in the context of g-functions. ◊

Following the notation in [CR90], we call a closed subset of $\mathbb{T}$ invariant, if it is invariant under the Markov chain induced by g. In more formal terms, this takes the following form.

Definition 6.1.19. A closed subset $A \subset \mathbb{T}$ is called (g, T_p)-*invariant*, if, for all $x \in A$ and $y \in T_p^{-1}(x)$ with $g(y) > 0$, it follows that $y \in A$. A closed (g, T_p)-invariant set is said to be (g, T_p)-*minimal* if it contains no non-trivial closed (g, T_p)-invariant subsets. We say that g is *proximal* on $(\mathbb{T}, T_p)$ if there are no two disjoint and closed, (g, T_p)-invariant subsets of $\mathbb{T}$.

We just speak of invariant and minimal subsets if the context is clear. Since invariant sets can be nested, there can be non-trivial minimal subsets of $\mathbb{T}$, even if g is proximal. However, since two minimal subsets are either identical or disjoint, g being proximal is equivalent to having precisely one minimal subset. Just like for finite state Markov chains, the *extremal* invariant distributions are supported by disjoint invariant (even minimal) subsets of the state space—under some regularity assumptions on g. The following is a slight adaptation of [CR90, Thm. 5.4].

Theorem 6.1.20. *Let g be a continuous g-function of bounded variation on $(\mathbb{T}, T_p)$. Then, every (g, T_p)-minimal subset supports precisely one g-measure and these are exactly the extremal g-measures corresponding to g. In particular, there is a* unique *g-measure if and only if g is proximal on $(\mathbb{T}, T_p)$.*

Sketch of proof. This is essentially [CR90, Thm. 5.4], with a few modifications. In [CR90], the authors consider the doubling map T_2 instead of T_p. As they remark in the introduction, this is only for the sake of an easier exposition. Also, they consider $[0, 1]$ instead of $\mathbb{T}$ as the state space, but an identification of the endpoints does not alter their arguments. Finally, the authors do not use the term *minimal* but formulate the theorem for a maximal set of closed invariant subsets. However, it becomes clear from their proof that the extremal measures are indeed supported on minimal subsets. □

Remark 6.1.21. The difference between taking $([0, 1], T_p)$ or $(\mathbb{T}, T_p)$ as the base space merits a closer inspection. From the defining relation of g-functions, it is apparent that $g(0) = g(1)$, even if 0 and 1 are considered to be separate points. Hence, we consider the same function g on both spaces. The g-measures on $([0, 1], T_p)$ that give no mass to $\{0, 1\}$ are in 1-to-1 correspondence to g-measures on $(\mathbb{T}, T_p)$ without a point mass in 0. The same holds for minimal subsets that do not contain the boundary points. Conversely, the following are all equivalent to $g(0) = 1$,

- $\{0\}$ is minimal on $(\mathbb{T}, T_p)$,
- δ_0 is a g-measure on $(\mathbb{T}, T_p)$,
- both δ_0 and δ_1 are g-measures on $([0, 1], T_p)$,
- $\{0\}$ and $\{1\}$ are both minimal on $([0, 1], T_p)$.

This provides an alternative way to verify that Theorem 6.1.20 is valid for $(\mathbb{T}, T_p)$ precisely if it holds for $([0, 1], T_p)$. ◊

Conveniently, there can be only finitely many minimal subsets of $\mathbb{T}$ if the g-function is continuous [CR90, Prop. 3.1]. Even if there is a unique g-measure, this measure need not be attractive. In general, it is possible that we approach a limit cycle of measures that are φ_g^d-invariant for some $d \in \mathbb{N}$ [CR90, Thm. 6.2]. If there is only one φ_g^d-invariant measure for all $d \in \mathbb{N}$ (coinciding with the unique g-measure), this cycle is necessarily contracted to a single measure and we get convergence. More precisely, we get the following corollary of [CR90, Thm. 6.2].

Theorem 6.1.22. *Assume that g is a continuous g-function of bounded variation. Then, the following statements are equivalent.*

(1) *The function g_n is proximal on $(\mathbb{T}, T_p^n)$ for all $n \in \mathbb{N}$.*

(2) *The sequence $(\varphi_g^n f)_{n\in\mathbb{N}}$ converges uniformly to a constant function for all $f \in C(\mathbb{T})$.*

(3) *The sequence $\big((\varphi_g^n)_*\nu\big)_{n\in\mathbb{N}}$ converges weakly to the unique g-measure for all $\nu \in \mathcal{M}^1(\mathbb{T})$.*

In this case, the unique g-measure is strongly mixing on $(\mathbb{T}, T_p)$. In particular, if g_n is proximal for all $n \in \mathbb{N}$, the unique g-measure has a product representation, given by

$$\mu \;=\; \prod_{k=0}^{\infty} p\, g(T_p^k x).$$

Sketch of proof. By [CR90, Thm. 6.2], the first point is equivalent to the weak convergence of $(\varphi_g^n)_*\delta_x$ to the unique g-measure μ, for all $x \in \mathbb{T}$. Hence, (3) $\Longrightarrow$ (1).

Using that $(\varphi_g^n f)(x) = ((\varphi_g^n)_*\delta_x)(f)$, we get that (1) implies the pointwise convergence of $(\varphi_g^n f)$ to $\mu(f)$. From the fact that g is of bounded variation, it follows that the sequence $(\varphi_g^n f)_{n\in\mathbb{N}}$ is uniformly equicontinuous; compare our discussion in [BCEG21]. Uniform convergence then follows by the Arzela–Ascoli theorem, and we obtain (1) $\Longrightarrow$ (2).

Under this condition, for all $\nu \in \mathcal{M}^1(\mathbb{T})$ and $f \in C(\mathbb{T})$,

$$\lim_{n\to\infty} ((\varphi_g^n)_*\nu)(f) \;=\; \lim_{n\to\infty} \nu(\varphi_g^n f) \;=\; \nu(c_f) \;=\; c_f,$$

for some constant c_f that does not depend on ν. It is straightforward to verify that the functional $\mu\colon f \mapsto c_f$ is indeed a measure and that this measure is invariant under $(\varphi_g)_*$. That is, (2) $\Longrightarrow$ (3), completing the proof of the equivalent conditions.

The strong mixing property goes back to a simple, but elegant argument provided by Keane [Kea72]. The final statement follows from Lemma 6.1.15 and the discussion thereafter. □

To summarize, we obtain a unique g-measure if and only if g is proximal, and this measure is (globally) attractive if and only if the g-function remains proximal for all accelerations of the underlying Markov chain. This is an analogue of aperiodicity in the context of irreducible, finite state Markov chains.

We get one step closer to the TM measure by requiring that g has only finitely many zeros. In this case, Conze and Raugi provide the following useful criterion [CR90, Cor. 3.4].

Lemma 6.1.23. *Let g be a continuous g-function with finitely many zeros. The function g is proximal precisely if $g^{-1}(1)$ contains at most one finite T_p-orbit. If $g^{-1}(1)$ contains more than one finite T_p-orbit, these orbits are precisely the minimal subsets of $\mathbb{T}$.* ■

This criterion ensures uniqueness of the associated g-measure due to Theorem 6.1.20. We provide some sufficient conditions for g_n being proximal for all $n \in \mathbb{N}$. For the sake of definiteness, we restrict our attention to the doubling map $(\mathbb{T}, T_2)$. The following is a slight improvement of a result by Keane [Kea72] that we announced in [BCEG21].

Proposition 6.1.24. *Let g be a continuous g-function on $(\mathbb{T}, T_2)$ of bounded variation and assume that one of the following conditions holds.*

(1) *$g^{-1}(1)$ is finite and does not contain a complete T_2-orbit.*

(2) *All zeros of g are contained in $(\frac{1}{6}, \frac{5}{6}] \cup \{0\}$ or in $[\frac{1}{6}, \frac{5}{6}) \cup \{0\}$.*

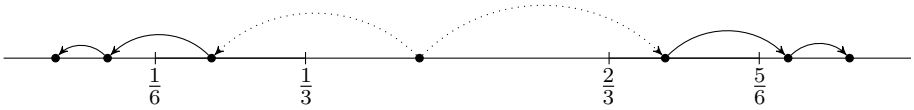

Figure 6.3.: Possible trajectories of the random walk, provided that condition (2) in Proposition 6.1.24 holds. At least one of the dotted lines is viable.

Then, there is a unique, globally attractive g-measure corresponding to g, given by the infinite product

$$\mu \,=\, \prod_{k=0}^{\infty} 2\, g(T_2^k x).$$

The measure μ is strongly mixing on $(\mathbb{T}, T_2)$. It is singular continuous precisely if $g(0) \neq 1$ and $g \not\equiv \frac{1}{2}$.

Proof. By Theorem 6.1.22, all conclusions, apart from the spectral type of μ, follow if g_n is proximal on $(\mathbb{T}, T_2^n)$ for all $n \in \mathbb{N}$.

First, assume that $g^{-1}(1)$ is finite and does not contain a complete T_2-orbit and let $n \in \mathbb{N}$. Note that $g_n(x) = 1$ requires that $g(T_2^k x) = 1$ for all $0 \leqslant k \leqslant n-1$. It follows that $g_n^{-1}(1)$ is also finite and cannot contain a complete T_2^n-orbit. Hence, g_n is proximal by Lemma 6.1.23.

Second, assume that all zeros of g are contained in $(\frac{1}{6}, \frac{5}{6}] \cup \{0\}$. Given $n \in \mathbb{N}$, we show that every closed (g_n, T_2^n)-invariant subset A contains the point 0. Let $x \in A \setminus \{0\}$. If $x \leqslant \frac{1}{3}$, it follows that $S_0^k x \leqslant \frac{1}{6}$ and hence $g(S_0^k x) > 0$ for all $k \in \mathbb{N}$. Thereby, $g_n(S_0^{kn} x) > 0$ for all $k \in \mathbb{N}$. By the invariance of A, this shows inductively that $S_0^{kn} x \in A$ for all $k \in \mathbb{N}$ and so is the limit point 0, since A is assumed to be closed. If $x > \frac{2}{3}$, we observe that $S_1^k x > \frac{5}{6}$ for all $k \in \mathbb{N}$. Since 1 and 0 are identified on the torus, the sequence $(S^{kn} x)_{k \in \mathbb{N}}$ converges to 0 and a similar argument as above yields $0 \in A$. Finally, if $\frac{1}{3} < x \leqslant \frac{2}{3}$, we have $S_0 x \leqslant \frac{1}{3}$ and $S_1 x > \frac{2}{3}$. Since $g(S_0 x) > 0$ or $g(S_1 x) > 0$, we can repeat the argument above. Hence, 0 is contained in every closed (g_n, T_2^n)-invariant subset, implying that g_n is proximal; compare Figure 6.3 for an illustration. The claim for the case that all zeros are in $[\frac{1}{6}, \frac{5}{6}) \cup \{0\}$ follows in the same manner, with obvious modifications.

Finally, we show that the unique g-measure μ is singular continuous. Due to Proposition 6.1.10 and the assumption that g is continuous with $g \not\equiv \frac{1}{2}$, it follows that μ is a singular measure. On the other hand Proposition 6.1.8 shows that μ is continuous, given that $g^{-1}(1)$ does not support a finite T_2-orbit. Indeed, condition (2) implies that $g^{-1}(1)$ is contained in $(\frac{1}{3}, \frac{2}{3}]^C$ or in $[\frac{1}{3}, \frac{2}{3})^C$. If $x \neq 0$, its T_2-orbit contains at least one point in $(\frac{1}{3}, \frac{2}{3}]$ and at least one point in $[\frac{1}{3}, \frac{2}{3})$. That is, $g^{-1}(1)$ cannot support a T_2-orbit and the claim follows. □

Remark 6.1.25. The second condition in Proposition 6.1.24 is sharp in the following sense. If we only assume that all zeros of g are contained in the compact interval $[\frac{1}{6}, \frac{5}{6}]$, it is possible to choose $g(\frac{1}{6}) = g(\frac{5}{6}) = 0$ and hence $g(\frac{1}{3}) = g(\frac{2}{3}) = 1$. In this case, the conclusions of Proposition 6.1.24 fail. Indeed, for such a choice of g,

$$\mu \,=\, \frac{1}{2}\big(\delta_{\frac{1}{3}} + \delta_{\frac{2}{3}}\big),$$

is a g-measure that is neither singular continuous nor globally attractive. In fact, we easily verify that the sequence $(\varphi_g^n)_*\delta_{1/3}$ oscillates between the measures $\delta_{1/3}$ and $\delta_{2/3}$. ◊

Example 6.1.26. The continuous g-function $g = \frac{1}{2}(1-\cos(2\pi x))$ is clearly Hölder continuous and thereby of bounded variation. It satisfies $g^{-1}(1) = \frac{1}{2}$ and hence the assumptions of Proposition 6.1.24. It follows that the infinite product

$$\mu \;=\; \prod_{k=0}^{\infty}(1-\cos(2\pi 2^k x))$$

is singular continuous, strongly mixing on $(\mathbb{T}, T_2)$, and that μ is the unique, attractive g-measure corresponding to g. This is precisely the TM measure, introduced in (6.1). ◊

Example 6.1.27. Fan, Schmeling and Shen investigate in [FSS21] the scaling behaviour of the L^∞-norm of

$$p_n(x) \;=\; \prod_{k=0}^{n-1}(1+\cos(2\pi(2^k x + c))),$$

for arbitrary $c \in \mathbb{T}$ as $n \to \infty$. These trigonometric polynomials arise from a family of generalized TM sequences. Indeed, it is straightforward to verify that

$$\mu \;=\; \prod_{k=0}^{\infty}(1+\cos(2\pi(2^k x + c)))$$

is the unique diffraction measure associated to the substitution

$$\varrho\colon z \mapsto z\,(\mathrm{e}^{2\pi \mathrm{i} c} z),$$

on the compact alphabet $\{z \in \mathbb{C} : |z| = 1\}$. By Proposition 6.1.24, μ is the unique, strongly mixing and attractive g-measure associated to $g(x) = 1 + \cos(2\pi(x+c))$. It is singular continuous unless $c = 0$, and coincides with the TM measure if $c = \frac{1}{2}$. ◊

Example 6.1.28. Consider the g-function $g(x) = \frac{1}{2}(1+\cos(6\pi x))$ on $(\mathbb{T}, T_2)$, which is again of bounded variation. In this case, $g^{-1}(1) = \{0, \frac{1}{3}, \frac{2}{3}\}$ contains two finite T_2-orbits and the extremal g-measures are precisely δ_0 and $\frac{1}{2}(\delta_{1/3} + \delta_{2/3})$, due to Lemma 6.1.23 and Theorem 6.1.20. Although we cannot apply Proposition 6.1.24, we still get that $\lim_{n\to\infty} 2^n g_n \lambda_{\mathrm{L}}$ exists and is given by

$$\prod_{k=0}^{\infty}(1+\cos(6\pi 2^k x)) \;=\; \frac{1}{3}\Big(\delta_0 + \delta_{\frac{1}{3}} + \delta_{\frac{2}{3}}\Big). \tag{6.6}$$

This is because each of the measures $2^n g_n \lambda_{\mathrm{L}}$ is invariant under the rotation $x \mapsto x + \frac{1}{3}$, a symmetry that carries over to each of the weak accumulation points. In fact, the Riesz product (6.6) describes the Fourier transform of the Dirac comb $\delta_{3\mathbb{Z}}$, restricted to the 1-torus. We therefore arrive at the same conclusion by Poisson's summation formula; compare [BG13, Ch. 9]. ◊

6.2. Super-polynomial scaling

As was discussed in the introduction of this chapter, the TM measure is particularly "thin" in the neighbourhood of dyadic point. This was already noted by Godrèche and Luck in [GL90]. Here, we discuss this phenomenon for a class of g-measures that includes the TM measure, and give quantitative upper bounds. Super-polynomial scaling behaviour of diffraction measures at the *origin* has received some attention recently [BCEG21, BG19], and is intimately connected to strong assumptions on the balancedness of pattern distributions that were discussed under the term *hyperuniformity* [OSST17, OSST19, TS03].

For convenience of notation, we restrict our attention to g-functions over the doubling map $(\mathbb{T}, T_2)$, but emphasize that the results in this section easily generalize to the covering map $(\mathbb{T}, T_p)$, with $p \in \mathbb{N}$. As a first step, we bound the mass that a g-measure assigns to small intervals by the product of rescaled copies of the g-function.

Lemma 6.2.1. *Let μ be a g-measure, associated to some g-function on $(\mathbb{T}, T_2)$, $n \in \mathbb{N}$ and $j_1, \dots, j_n \in \{0, 1\}$. Then,*

$$\mu(S_{j_1} \cdots S_{j_n} \mathbb{T}) \;=\; \int_{\mathbb{T}} g_n(S_{j_1} \cdots S_{j_n} x) \, \mathrm{d}\mu(x) \;\leqslant\; \max_{x \in \mathbb{T}} g_n(S_{j_1} \cdots S_{j_n} x).$$

Sketch of proof. The first step is contained in the proof of Proposition 6.1.14 and the inequality follows because μ is a probability measure. □

If $y \in \mathbb{T}$ is a dyadic point, it has exactly two binary representations. More precisely, there are two sequences $(y_n)_{n \in \mathbb{N}}$ and $(\overline{y}_n)_{n \in \mathbb{N}}$ in $\{0, 1\}^{\mathbb{N}}$, as well as $n_0 \in \mathbb{N}$ with the property that $y_n = 0$ and $\overline{y}_n = 1$ for all $n > n_0$, and

$$y \;=\; \sum_{n \in \mathbb{N}} y_n 2^{-n} \;=\; \sum_{n \in \mathbb{N}} \overline{y}_n 2^{-n}.$$

For each $n \in \mathbb{N}$, y is the left endpoint of the interval $S_{y_1} \cdots S_{y_n} \mathbb{T}$ and the right endpoint of the interval $S_{\overline{y}_1} \cdots S_{\overline{y}_n} \mathbb{T}$. The union of these two intervals coincides with the closed ball $B_{2^{-n}}(y)$. For a moment, assume that $y = 0$. Then, the product

$$g_n(S_{y_1} \cdots S_{y_n} x) \;=\; \prod_{k=1}^{n} g(S_0^k x) \tag{6.7}$$

consists of many factors that are close to 0 if g is continuous at 0 and satisfies $g(0) = 0$. If y is dyadic, we have $T_2^m y = 0$ for some $m \in \mathbb{N}$ and a similar observation holds. In order to get a quantitative bound on products of the form (6.7), we make additional assumptions on the modulus of continuity of g at 0. In the following, we denote by $d(x, y)$ the natural distance of points $x, y \in \mathbb{T}$.

Theorem 6.2.2. *Let μ be a g-measure associated to some g-function on $(\mathbb{T}, T_2)$, satisfying $g(0) = 0$. Assume that there are $C, \theta > 0$ such that*

$$g(x) \;\leqslant\; C d(x, 0)^{\theta},$$

for all x in some neighbourhood of 0. Then, for every dyadic point $y \in \mathbb{T}$,

$$\log_2(\mu(B_r(y))) \leqslant -\frac{\theta}{2}\log_2(r)^2 + O(\log_2(r)),$$

as $r \to 0^+$.

Proof. Let $y \in \mathbb{T}$ be a dyadic point and let $m \in \mathbb{N}$ such that $T^m y = 0$. It suffices to prove the claim for $r = 2^{-n}$ and $n \in \mathbb{N}$, the more general statement follows by straightforward interpolation arguments. In this case, $\log_2(r) = -n$ and

$$B_{2^{-n}}(y) = S_{y_1} \cdots S_{y_n}\mathbb{T} \cup S_{\overline{y}_1} \cdots S_{\overline{y}_n}\mathbb{T}.$$

Recall from Lemma 6.2.1 that

$$\mu(S_{y_1} \cdots S_{y_n}\mathbb{T}) \leqslant \max_{x\in\mathbb{T}} g_n(S_{y_1} \cdots S_{y_n}x) \leqslant \prod_{k=1}^{n} \max_{x\in\mathbb{T}} g(S_{y_k} \cdots S_{y_n}x). \tag{6.8}$$

For $k \geqslant m+1$, we have $y_k = 0$, and hence $S_{y_k} \cdots S_{y_n}x = S_0^{n-k+1}x$. By assumption, there is a number n_0 such that for all $j \geqslant n_0$, we have

$$g(S_0^j x) \leqslant C d(S_0^j x, 0)^\theta \leqslant C\, 2^{-j\theta}.$$

This gives a bound for the factors in (6.8), as long as $k \geqslant m+1$ and $n-k+1 \geqslant n_0$. Using the uniform bound 1 for the remaining factors, we obtain

$$\mu(S_{y_1} \cdots S_{y_n}\mathbb{T}) \leqslant \prod_{j=n_0}^{n-m} \max_{x\in\mathbb{T}} g(S_0^j x) \leqslant \prod_{j=n_0}^{n-m} C 2^{-j\theta}.$$

By similar reasoning, we get the same bound for $\mu(S_{\overline{y}_1} \cdots S_{\overline{y}_n}\mathbb{T})$. Taking logarithms, we thus obtain

$$\log_2(\mu(B_{2^{-n}}(y))) \leqslant -\theta \sum_{j=n_0}^{n-m} j + O(n) = -\frac{\theta}{2}n^2 + O(n),$$

and the claim follows. □

Remark 6.2.3. The proof of Theorem 6.2.2 can be refined to obtain some quantitative control on the constants that are hidden in the error term $O(\log_2(r))$. These depend both on y and on the size of the neighbourhood of 0 where we have control over the modulus of continuity of g. We leave the details to the interested reader. ◊

Remark 6.2.4. Theorem 6.2.2 should be compared to a similar result that we provided in [BCEG21, Thm. 1.4]. However, there are some differences. In [BCEG21, Thm. 1.4], we only consider a (one-sided) neighbourhood of $y = 0$ instead of general dyadic points and obtain both upper and lower bound for the associated measure. This comes at the cost of additional assumptions on g. We emphasize that a non-trivial lower bound is harder to obtain and requires some *global* control over g. Indeed, consider a g-function with $g(0) = 0$ and $g(\frac{1}{3}) = g(\frac{2}{3}) = 1$. Then, $\mu = (\delta_{1/3} + \delta_{2/3})/2$ is a g-measure that has a trivial scaling around 0. This shows that controlling g only in a neighbourhood of 0 is insufficient to obtain non-trivial lower bounds. For the special case of the TM measure, a more refined analysis of the scaling at 0 is possible; compare [BG19]. ◊

6.3. Multifractal analysis

The general philosophy of a multifractal analysis can be summarized as follows; compare [Pes97, Ch. 6]. We start from a a set X and some characteristic $h\colon X' \to \mathbb{R}$ that is well-defined on some $X' \subset X$. The *multifractal decomposition* of X,

$$X \,=\, X \setminus X' \cup \bigcup_{\alpha\in\mathbb{R}} X(\alpha), \quad X(\alpha) \,=\, \bigl\{x \in X' : h(x) = \alpha\bigr\},$$

splits X' into the level sets with respect to h. Although the sets $X(\alpha)$ might have a very intricate structure, the "size" of the sets $X(\alpha)$ can often be made more explicit and accessible to numerical approximations. More precisely, we analyze the *spectrum* $f(\alpha) = \dim X(\alpha)$ for some dimensional characteristic on X, often taken to be the Hausdorff dimension.

In this section, we establish such a "multifractal miracle" for the *dimension spectrum*

$$f(\alpha) \,=\, \dim_{\mathrm{H}} \mathcal{F}(\alpha), \quad \mathcal{F}(\alpha) \,=\, \{x \in \mathbb{T} : \dim_{\mu}(x) = \alpha\}, \tag{6.9}$$

corresponding to the TM measure $\mu = \mu_{\mathrm{TM}}$. This is done in two steps. First, we relate the dimension of the measure at x to the scaling behaviour of $g_n(x)$, with the help of Lemma 6.2.1. Then, the multifractal analysis of the scaling exponents is related to a *globally* defined quantity—the topological pressure function, which is one of the central quantities in the *thermodynamic formalism*. This route is classic for g-measures corresponding to *strictly* positive g-functions [PW01]. We therefore proceed cautiously and first restrict our attention to T_2-invariant subspaces of $\mathbb{T}$ that avoid some forbidden neighbourhood of 0. As we send the size of this neighbourhood to 0, we recover the full dimension spectrum.

Such an approach is more easily formulated in a symbolic setup. Indeed, the system $(\mathbb{T}, T_2)$ is intimately related to the one-sided full shift $(\mathbb{X}, S)$, with $\mathbb{X} = \{0,1\}^{\mathbb{N}}$. More precisely, there is a natural projection

$$\pi\colon (x_n)_{n\in\mathbb{N}} \mapsto \sum_{k=1}^{\infty} x_n 2^{-n},$$

that induces a topological factor map from $(\mathbb{X}, S)$ to $(\mathbb{T}, T_2)$. This map is invertible on the complement of dyadic points. Since the countable set of dyadic sets has vanishing Hausdorff dimension, we identify $(\mathbb{X}, S)$ and $(\mathbb{T}, T_2)$ in the following.

Remark 6.3.1. The identification of $(\mathbb{T}, T_2)$ and $(\mathbb{X}, S)$ also entails that we regard functions and measures on $\mathbb{T}$ as being defined on $\mathbb{X}$ in the following. More specifically, given $g\colon \mathbb{T} \to \mathbb{R}$, we consider $\widetilde{g} = g \circ \pi$ on $\mathbb{X}$. The property of being a g-function is preserved, that is, for all $x \in \mathbb{X}$,

$$\sum_{y\in S^{-1}x} \widetilde{g}(y) \,=\, 1,$$

whenever g satisfies (6.3). Conversely, whenever the set of dyadic points forms a null set for $\mu \in \mathcal{M}(\mathbb{T})$, there is a unique lift $\widetilde{\mu} \in \mathcal{M}(\mathbb{X})$ with $\mu = \widetilde{\mu} \circ \pi^{-1}$. With some abuse of notation, we will drop the distinction between g and $\widetilde{g}$, as well as the distinction between μ and $\widetilde{\mu}$. $\Diamond$

Remark 6.3.2. There is still a subtlety that complicates the identification of $(\mathbb{X}, S)$ with $(\mathbb{T}, T_2)$. That is because the spaces carry different metric structures, which are relevant for both the definition of the pointwise dimension $d_\mu(x)$ and the Hausdorff dimension. Heuristically, the relationship between these metrics becomes more distorted the closer we get to dyadic points. However, for the SFTs avoiding the point 0 that we consider in the following, both metrics are bi-Lipschitz equivalent. Hence, for these subspaces, both the pointwise dimension $d_\mu(x)$ and the Hausdorff dimension of subsets are invariant under the identification due to a classic result [Pes97, Thm. 6.3]. For a more careful notational distinction of the metric structures, we refer the reader to [BGKS19]. ♢

Given $m \in \mathbb{N} \setminus \{1\}$, we define $(\mathbb{X}_m, S)$ as the subshift of $(\mathbb{X}, S)$ that avoids both of the cylinders $[0^{m+1}]$ and $[1^{m+1}]$, that is,

$$\mathbb{X}_m \;=\; \big\{x \in \mathbb{X} : x_{[j,j+m]} \notin \{0^{m+1}, 1^{m+1}\} \text{ for all } j \in \mathbb{N}\big\}. \tag{6.10}$$

This is a SFT, with $\mathcal{F} = \{0^{m+1}, 1^{m+1}\}$ as the set of forbidden words; compare Example 2.1.5. For $m = 1$, the definition in (6.10) would yield a finite orbit, consisting of two points. The restriction $m \geqslant 2$ is hence a non-triviality assumption, guaranteeing that $\mathbb{X}_m$ is an uncountable set. Conveniently, both the thermodynamic formalism and multifractal analysis are well-developed for SFTs as the underlying dynamical system.

6.3.1. Thermodynamic formalism

Variational principles are ubiquitous in physics. Both the modern formulation of classical mechanics and the description of elementary particles via quantum field theory are based on variational principles. The basic idea underlying both theories is that the equations of motion can be found by minimizing some *action functional*. A similar approach is taken in thermodynamics to characterize those distributions in the space of configurations that are stable and hence physically observable. Here, depending on the physical setup, the role of the action is taken by some macroscopic quantity that depends on the microscopic distribution, and a measure minimizing (or maximising) this macroscopic quantity is called an *equilibrium measure*. More concretely, let us consider a thermodynamic system that is in exchange with a heat bath regulating its temperature, and is subject to some potential (energy) function ψ on the space of configurations X. In this so-called *canonical ensemble*, the role of the action is taken by the *free energy*. If the configuration space is finite, the weight of the equilibrium state at $x \in X$ is known to be proportional to $\exp(-\beta\psi(x))$, where β depends on the temperature. These measures (and their generalizations to infinite systems) became known as *Gibbs measures*, thanks to the pivotal contribution of Gibbs to this area of thermodynamics [Gib02].

In the mathematics literature, the contributions of Ruelle [Rue68, Rue69b] led to a revival and further development of these ideas; we also refer to the monograph by Keller [Kel98] for an insightful and readable introduction. Compared to the original context in physics, some changes in notation and conventions took place—some quantities changed their sign, and the free energy essentially changed its name to *pressure*—but the basic formalism remained relatively unchanged.

Definition 6.3.3. Let (X, T) be a compact dynamical system and $\mathcal{M}_T^1(X)$ the set of T-invariant probability measures on X. Further, let $\psi\colon X \to [-\infty, \infty)$ be an upper semi-continuous function, called the (thermodynamic) *potential*. The (variational) *pressure* of ψ is given by

$$\mathcal{P}(\psi) \;=\; \sup_{\nu \in M_T^1(X)} \Bigl(h_\nu(T) + \int_X \psi(x)\, \mathrm{d}\nu(x) \Bigr). \tag{6.11}$$

A measure $\mu \in \mathcal{M}_T^1(X)$ that attains the supremum in (6.11) is called an *equilibrium measure* for the potential ψ.

For ease of notation, we write $\nu(\psi)$ for the integral of ψ with respect to ν, even if it evaluates to $-\infty$. In the following, we specialize to the case that (X, T) is in fact a subshift, denoted by $(\mathbb{X}, S)$. Conveniently, it is possible to characterize g-measures on the full shift $(\mathbb{X}, S)$ (over a finite alphabet $\mathcal{A}$) as equilibrium measures due to a result by Ledrappier [Led74, Thm. 1]. We restrict to the case that g is continuous although Ledrappier's original result is formulated in greater generality.

Theorem 6.3.4 ([Led74])**.** *Let g be a continuous g-function on the full shift $(\mathbb{X}, S)$. Then, $\mathcal{P}(\log(g)) = 0$ and $\mu \in \mathcal{M}_S^1(\mathbb{X})$ is a g-measure if and only if it is an equilibrium measure for the potential $\psi = \log(g)$.* ∎

Since g-measures need not be unique, we immediately see that equilibrium measures need not be unique either. Since an equilibrium measure μ maximises the sum of the entropy and the potential energy, it is reasonable to assume that it gives larger weight to regions, where ψ takes large values. Since μ is also assumed to be invariant, it should in fact favour those points, where the corresponding *orbit* remain in regions of high energy. Sampling ψ along finite parts of the orbit, leads to the quantity

$$\psi_n(x) \;=\; \sum_{k=0}^{n-1} \psi(S^k x),$$

for all $n \in \mathbb{N}$ and $x \in \mathbb{X}$.

Definition 6.3.5. A measure $\mu \in \mathcal{M}_S^1(\mathbb{X})$ is called a *Gibbs measure* with respect to some potential ψ on $\mathbb{X}$ if there are constants $C_1, C_2 > 0$ and $P \in \mathbb{R}$ such that

$$C_1 \;\leqslant\; \frac{\mu([x_1 \cdots x_n])}{\exp(\psi_n(x) - nP)} \;\leqslant\; C_2, \tag{6.12}$$

for all $x \in \mathbb{X}$ and $n \in \mathbb{N}$.

In fact, we have already encountered Gibbs measures in the last section, although the defining relation (6.12) was slightly hidden.

Lemma 6.3.6. *Let $(\mathbb{X}, S)$ be the full shift and g a strictly positive and continuous g-function on $(\mathbb{X}, S)$. Further, assume that g is of bounded variation. Then, its unique g-measure is a Gibbs measure with respect to the potential $\psi = \log(g)$.*

Proof. Let us fix $n \in \mathbb{N}$ and $x_1 \cdots x_n \in \{0,1\}^n$. Translating Lemma 6.2.1 to the symbolic setting yields

$$\mu([x_1 \cdots x_n]) \;=\; \int_{\mathbb{X}} g_n(x_1 \cdots x_n y)\, \mathrm{d}\mu(y). \tag{6.13}$$

For $\psi = \log(g)$, we easily verify $\exp(\psi_n(x)) = \prod_{k=0}^{n-1} g(S^k x) = g_n(x)$, for all $x \in \mathbb{X}$. Hence, we obtain from (6.13),

$$\inf_{y \in [x_1 \cdots x_n]} \exp(\psi_n(y)) \;\leqslant\; \mu([x_1 \cdots x_n]) \;\leqslant\; \sup_{y \in [x_1 \cdots x_n]} \exp(\psi_n(y)).$$

It remains to verify that the variation of $\exp(\psi_n(y))$ on $[x_1 \cdots x_n]$ is bounded by some uniform multiplicative constant. Since the values of g are contained in a compact interval that is strictly bounded away from 0, it follows that $\psi = \log(g)$ is also of bounded variation. For $y, z \in [x_1 \cdots x_n]$, we observe that $|S^k y - S^k z| \leqslant 2^{-n+k}$, and hence

$$|\psi_n(y) - \psi_n(z)| \;\leqslant\; \sum_{k=0}^{n-1} |\psi(S^k y) - \psi(S^k z)| \;\leqslant\; \sum_{j=1}^{n} \psi[2^{-j}] \;\leqslant\; \sum_{j=0}^{\infty} \psi[2^{-j}] \;=:\; c \;<\; \infty.$$

Hence, the maximal and the minimal value of $\exp(\psi_n(x))$ on $[x_1 \cdots x_n]$ differ by at most a factor e^c, yielding

$$\mathrm{e}^{-c} \;\leqslant\; \frac{\mu([x_1 \cdots x_n])}{\exp(\psi_n(x))} \;\leqslant\; \mathrm{e}^{c},$$

irrespective of the choice of $x \in [x_1 \cdots x_n]$, and the claim follows with $P = 0$. □

Hence, we have seen that, for a strictly positive, continuous g-function of bounded variation, the g-measure is both an equilibrium measure and a Gibbs measure to the same potential $\psi = \log(g)$. It is noteworthy that the parameter P, entering the defining relation of a Gibbs measure, is given by $P = 0 = \mathcal{P}(\psi)$. This is not a coincidence. Indeed, Bowen established quite general results in this direction in his classic monograph [Bow75]. Combining [Bow75, Thm. 1.4] and [Bow75, Thm. 1.22] yields the following result.

Theorem 6.3.7 ([Bow75])**.** *Let $(\mathbb{X}, S)$ be a topologically mixing SFT and ψ a Hölder continuous function on $\mathbb{X}$. Then, there exists a unique Gibbs measure $\mu_\psi \in \mathcal{M}^1_S(\mathbb{X})$ to the potential ψ, with parameter $P = \mathcal{P}(\psi)$. This measure is also the unique equilibrium measure for the potential ψ. Further, $\mathcal{P}(\psi)$ is equal to the* topological pressure*, defined as*

$$\mathcal{P}_{\mathrm{top}}(\psi) \;=\; \lim_{n\to\infty} \frac{1}{n} \log \sum_{w \in \{0,1\}^n} \sup_{x \in [w] \cap \mathbb{X}} \exp(\psi_n(x)), \tag{6.14}$$

and the limit is well-defined. ■

Remark 6.3.8. Just like g-measures are fixed points under the operator φ_g, the Gibbs measure μ_ψ in Theorem 6.3.7 is an eigenmeasure of the dual of the *Ruelle–Perron–Frobenius operator* $\mathcal{L}_\psi$, defined via

$$(\mathcal{L}_\psi f)(x) \;=\; \sum_{y \in S^{-1}x} \mathrm{e}^{\psi(y)} f(y),$$

for all $f \in C(\mathbb{X})$. In fact, there is a powerful convergence result under the assumptions in Theorem 6.3.7, parallel to our discussion in Section 6.1.2. More precisely, there are $\lambda > 0$ and $h \in C(\mathbb{X})$, such that $(\mathcal{L}_\psi)_* \mu_\psi = \lambda \mu_\psi$, $\mathcal{L}_\psi h = \lambda h$ and

$$\lim_{n\to\infty} \frac{1}{\lambda^n} \mathcal{L}_\psi^n \;=\; |h\rangle\langle \mu_\psi|,$$

using the intuitive bra-ket notation for elements in $C(\mathbb{X})$ and its dual space. Convergence is understood in the strong operator topology. Due to its similarity to the classic PF theorem, this result became known as the *Ruelle–Perron–Frobenius theorem*; for details compare [Bow75, Rue68]. The eigenvalue λ is related to the pressure via $\mathcal{P}(\psi) = \log(\lambda)$ under the assumptions in Theorem 6.3.7 [Bow75]. ◊

6.3.2. Birkhoff spectrum and dimension spectrum

Regular potentials

If μ is a Gibbs measure for the potential ψ, the defining relation (6.12) yields

$$\overline{\psi}(x) \;:=\; \lim_{n\to\infty} \frac{\psi_n(x)}{n} \;=\; P - \log(2)\, d_\mu(x), \tag{6.15}$$

provided that any (equivalently both) of the limits exist. That is, the multifractal analysis of the local dimensions of μ is exchangeable with the multifractal analysis of the Birkhoff averages $\overline{\psi}$. We set

$$b(\beta) \;=\; \dim_{\mathrm{H}} \mathcal{B}(\beta), \quad \mathcal{B}(\beta) \;=\; \{x \in \mathbb{X} : \overline{\psi}(x) = \beta\},$$

and call $b(\cdot)$ the *Birkhoff spectrum* of ψ. Due to (6.15), this is related to the dimension spectrum in (6.9) via $\mathcal{F}(\alpha) = \mathcal{B}(P - \log(2)\alpha)$ and

$$f(\alpha) \;=\; b(P - \log(2)\alpha), \tag{6.16}$$

for all $\alpha \in \mathbb{R}$. It is a folklore in the area of multifractal analysis that both the Birkhoff spectrum and the dimension spectrum can be related to the (numerically better accessible) pressure function

$$p\colon \mathbb{R} \to \mathbb{R}, \quad p(t) \;=\; \mathcal{P}_{\mathrm{top}}(t\psi).$$

We denote the *Legendre(–Fenchel) transform* of p by

$$p^*(q) \;=\; \sup_{t\in\mathbb{R}} (qt - p(t)).$$

The following result is classic. Most of it is stated in this form in [PW01, Thm. 1], the explicit formula for b in terms of p can be extracted from the proof; compare also [JK11, Schm99].

Fact 6.3.9. *Let $(\mathbb{X}, S)$ be a topologically mixing, one-sided SFT, ψ a Hölder continuous potential on $\mathbb{X}$ and μ the corresponding equilibrium measure. Assume further that μ is not the measure of maximal entropy on $(\mathbb{X}, S)$. Then, there is a non-empty interval $(\beta^-, \beta^+) \subset \mathbb{R}$ such that,*

(1) *b is a strictly concave, real analytic function on (β^-, β^+), given by*

$$b(\beta) \,=\, -\frac{p^*(\beta)}{\log(2)},$$

where $p\colon t \to \mathcal{P}_{\mathrm{top}}(t\psi)$ is a strictly convex, real analytic function on $\mathbb{R}$.

(2) *$[\beta^-, \beta^+] = \mathrm{dom}(p^*) := \{\beta \in \mathbb{R} : p^*(\beta) < \infty\}$ is the domain of p^*.*

(3) *The level sets $\mathcal{B}(\beta)$ are dense in $\mathbb{X}$ if $\beta \in [\beta^-, \beta^+]$ and $\mathcal{B}(\beta) = \varnothing$ otherwise.*

(4) *The exceptional set $\widehat{\mathcal{B}} = \{x \in \mathbb{X} : \overline{\psi}(x)$ does not exist$\}$ has full Hausdorff dimension, that is, $\dim_{\mathrm{H}} \widehat{\mathcal{B}} = \dim_{\mathrm{H}} \mathbb{X}$.* ■

Remark 6.3.10. In the situation of Fact 6.3.9, it is clear that $b(\beta) = 0$ as soon as β lies outside of $[\beta_1, \beta_2]$. A priori, it is possible that $b(\beta_1)$ and $b(\beta_2)$ can be strictly positive, although $b(\beta_1) = b(\beta_2) = 0$ is the *typical* situation (for a residual set of potentials) [Schm99]. ◊

The corresponding relations for $\mathcal{F}(\alpha)$ and the dimension spectrum $f(\alpha)$ are easily obtained from (6.15) and (6.16). Our main goal for the next section is to find an analogue of Fact 6.3.9, as well as a relation between the Birkhoff and dimension spectrum for the TM measure.

The TM potential

For the remainder of this chapter, we specialise to the TM measure, that is, we set

$$\mu \,=\, \mu_{\mathrm{TM}} \,=\, \prod_{k=0}^{\infty} 2g(S^k x), \quad g \,=\, \frac{1}{2}(1 - \cos(2\pi x)),$$

and $\psi = \log(g)$ for the *TM potential*. We emphasize that both Theorem 6.3.7 and Fact 6.3.9 do *not* cover the TM measure because g has a zero and, consequently, the potential ψ has a singularity at $x = 0$. In fact, although μ_{TM} is an equilibrium measure, due to Theorem 6.3.4, it cannot be a Gibbs measure. This is because the function $\exp(\psi_n(x) - nP)$ is 0 at one of the boundary points of $[x_1 \cdots x_n]$, irrespective of the choice of P. Nevertheless, we encounter a Gibbs-like property if we follow the orbit of points x in one of the SFTs $\mathbb{X}_m$, defined in (6.10).

Lemma 6.3.11. *For every $m \in \mathbb{N}$, there exists a constant $K = K(m)$ such that*

$$K^{-1} \,\leqslant\, \frac{\mu([x_1 \cdots x_n])}{\exp(\psi_n(x))} \,\leqslant K,$$

for all $x \in \mathbb{X}_m$ and $n \in \mathbb{N}$.

Proof. The basic ideas for this proof are similar to those employed in the proof of Lemma 6.3.6, but some additional care is needed to deal with the singularities in ψ. We start by proving the upper bound. Due to Lemma 6.2.1,

$$\mu([x_1 \cdots x_n]) \,\leqslant \sup_{y \in [x_1 \cdots x_n]} \exp(\psi_n(y)), \tag{6.17}$$

for all $x \in \mathbb{X}$ and $n \in \mathbb{N}$. Given $m \in \mathbb{N}$, denote by $\psi^{[m]}$ the restriction of ψ to the clopen subset $\mathbb{X}_{[m]} = \mathbb{X} \setminus \{[0^{m+1}], [1^{m+1}]\}$. This function is Hölder continuous, and we denote by

$$C = C(m) = \sum_{k=0}^{\infty} \psi^{[m]}[2^{-k}] < \infty$$

its bounded variation constant. Since $\mathbb{X}_m \subset \mathbb{X}_{[m]}$, we have $\psi(y) = \psi^{[m]}(y)$ for $y \in \mathbb{X}_m$. Due to the fact that ψ is increasing on $[0]$ and decreasing on $[1]$, finding the supremum of ψ on some cylinder set $[w]$ that overlaps $\mathbb{X}_m$ amounts to finding the supremum on $[w] \cap \mathbb{X}_{[m]}$. Hence, for all $k \in \mathbb{N}$, $w \in \{0,1\}^k$ and $z \in [w] \cap \mathbb{X}_m \neq \varnothing$, we obtain

$$\sup_{y \in [w]} \psi(y) - \psi(z) \leqslant \sup_{y \in [w] \cap \mathbb{X}_{[m]}} \psi^{[m]}(y) - \psi^{[m]}(z) \leqslant \psi^{[m]}[2^{-k}].$$

Given $x \in \mathbb{X}_m$ and $n \in \mathbb{N}$, this yields

$$\sup_{y \in [x_1 \cdots x_n]} \psi_n(y) - \psi_n(x) \leqslant \sum_{k=1}^{n} \Big(\sup_{y \in [x_k \cdots x_n]} \psi(y) - \psi(S^{k-1}x) \Big) \leqslant C,$$

and hence the upper bound follows, with $K = \exp(C)$.

For the lower bound, we need to be slightly more careful because the infimum of $\exp(\psi_n(y))$ on $[x_1 \cdots x_n]$ yields only the trivial bound 0. However, due to the symmetry of g, we have $\mu([0]) = \mu([1]) = 0.5$, which yields, for $j \in \{0,1\}$,

$$\mu([x_1 \cdots x_n]) \geqslant \int_{[j]} g_n(x_1 \cdots x_n y) \, \mathrm{d}\mu(y) \geqslant \frac{1}{2} \inf_{y \in [x_1 \cdots x_n j]} g_n(y).$$

Setting $j = x_n + 1 \mod 2$, and using $x \in \mathbb{X}_m$, we observe that $[x_k \cdots x_n j] \subset \mathbb{X}_{[m]}$, for all $1 \leqslant k \leqslant n$. Thus, by similar reasoning as above,

$$\psi_n(x) - \inf_{y \in [x_1 \cdots x_n j]} \psi_n(y) \leqslant \sum_{k=1}^{n} \Big(\psi(S^{k-1}x) - \inf_{y \in [x_k \cdots x_n j]} \psi(y) \Big) \leqslant C,$$

and the lower bound follows, with the modified constant $K = \exp(2C)$. □

Corollary 6.3.12. *Given $m \in \mathbb{N}$, we have*

$$\overline{\psi}(x) := \lim_{n \to \infty} \frac{\psi_n(x)}{n} = -\log(2) \dim_\mu(x),$$

for all $x \in \mathbb{X}_m$, provided the limit exists. ■

Another reason why it is useful to restrict our attention to the subshifts $(\mathbb{X}_m, S)$ is that it allows us to refer to the powerful machinery presented in the last section. Indeed, Theorem 6.3.7 and Fact 6.3.9 *do* apply to the subshifts of finite type $(\mathbb{X}_m, S)$, together with the restricted potential

$$\psi^{(m)} := \psi|_{\mathbb{X}_m}.$$

Indeed, it is a straightforward exercise to verify that $(\mathbb{X}_m, S)$ is topologically mixing, $\psi^{(m)}$ is Hölder continuous and the corresponding equilibrium measure μ_m is not equal to the measure

of maximal entropy (the Parry measure) on $(\mathbb{X}_m, S)$. It should be noted that $\exp(\psi^{(m)}(\cdot))$ does not define a g-function on $(\mathbb{X}_m, S)$ because the sum $\sum_{y\in S^{-1}x} \exp(\psi^{(m)}(y))$ can be strictly smaller than 1. Hence, μ_m is not a g-measure, but it is a Gibbs measure by Theorem 6.3.7. We denote the associated pressure function by

$$p_m \colon \mathbb{R} \to \mathbb{R}, \quad p_m(t) \,=\, \mathcal{P}_{\mathrm{top}}\big(t\psi^{(m)}\big),$$

with $\mathcal{P}_{\mathrm{top}}$ as in (6.14), being implicitly defined over $(\mathbb{X}_m, S)$. By Fact 6.3.9, it follows that

$$b_m(\beta) \,:=\, \dim_{\mathrm{H}} \mathcal{B}_m(\beta), \quad \mathcal{B}_m(\beta) \,:=\, \mathcal{B}(\beta) \cap \mathbb{X}_m$$

satisfies

$$b_m(\beta) \,=\, -\frac{p_m^*(\beta)}{\log(2)}, \tag{6.18}$$

for all $\beta \in (\beta_m^-, \beta_m^+)$, where $[\beta_m^-, \beta_m^+] = \mathrm{dom}(p_m^*)$. Further, due to Corollary 6.3.12, it immediately follows that

$$\mathcal{F}_m(\alpha) \,:=\, \mathcal{F}(\alpha) \cap \mathbb{X}_m \,=\, \mathcal{B}_m(-\log(2)\alpha) \tag{6.19}$$

and

$$f_m(\alpha) \,:=\, \dim_{\mathrm{H}} \mathcal{F}_m(\alpha) \,=\, b_m(-\log(2)\alpha) \tag{6.20}$$

hold for all $m \geqslant 2$.

Proposition 6.3.13. *We have $\beta_m^+ = \beta^+ = \log(3/4)$ for all $m \geqslant 2$, and $(\beta_m^-)_{m\geqslant 2}$ is a non-increasing sequence, with $\lim_{m\to\infty} \beta_m^- = -\infty$.*

Proof. Due to a classic result by Gelfond [Gel68], the maximal Birkhoff average for ψ is given by

$$\sup_{x\in\mathbb{X}} \overline{\psi}(x) \,=\, \psi\big((01)^{\mathbb{N}}\big) \,=\, \log\!\left(\frac{3}{4}\right),$$

compare also [BGKS19, Prop. 3.1] for an alternative proof of this relation. Hence, $\mathcal{B}_m(\beta) = \varnothing$ for all $\beta > \log(3/4)$. On the other hand, since the alternating sequence $(01)^{\mathbb{N}}$ is contained in $\mathbb{X}_m$ for all $m \geqslant 2$, it follows that $\mathcal{B}_m(\beta) \neq \varnothing$ for $\beta = \log(3/4)$. That is, $\beta_m^+ = \log(3/4)$ is the critical value for the onset of empty level sets $\mathcal{B}_m(\beta)$.

On the other hand, since $(\mathcal{B}_m(\beta))_{m\geqslant 2}$ is an increasing sequence of sets for all $\beta \in \mathbb{R}$, so is the sequence of intervals $([\beta_m^-, \beta_m^+])_{m\geqslant 2}$ that encode the obtainable levels. Hence, $(\beta_m^-)_{m\geqslant 2}$ is non-increasing. Given $m \geqslant 2$, we consider the periodic sequence $y_m = \overline{0^m 1} \in \mathbb{X}_m$. For large $m \in \mathbb{N}$, the S-orbit of y_m is close to the singularity of ψ for a positive fraction of the orbit. This suffices to conclude that

$$\lim_{m\to\infty} \overline{\psi}(y_m) \,=\, -\infty.$$

By definition, $y_m \in \mathcal{B}_m\big(\overline{\psi}(y_m)\big) \neq \varnothing$ and hence $\beta_m^- \leqslant \overline{\psi}(y_m)$, and the last claim follows. □

Corollary 6.3.14. *The level set $\mathcal{B}(\beta)$ is dense in $\mathbb{X}$ for each $\beta \leqslant \log(3/4)$, and $\mathcal{B}(\beta) = \varnothing$ if $\beta > \log(3/4)$. Also, the exceptional set*

$$\widehat{\mathcal{B}} \,=\, \{x \in \mathbb{X} : \overline{\psi}(x) \text{ does not exist}\},$$

has full Hausdorff dimension $\dim_{\mathrm{H}} \widehat{\mathcal{B}} = 1$.

Proof. The fact that $\mathcal{B}(\beta) = \varnothing$ for $\beta > \beta^+ = \log(3/4)$ follows again by Gelfond's result, compare the proof of Proposition 6.3.13. On the other hand, for $\beta \leqslant \beta^+$, there exists a $m_0 \in \mathbb{N}$ such that $\beta \in [\beta_m^-, \beta_m^+]$ for all $m \geqslant m_0$, due to Proposition 6.3.13. Hence $\mathcal{B}_m(\beta)$ is dense in $\mathbb{X}_m$ by Fact 6.3.9. Since the union of all $\mathbb{X}_m$ with $m \geqslant m_0$ is dense in $\mathbb{X}$, so is the set

$$\mathcal{B}(\beta) \supset \bigcup_{m \geqslant m_0} \mathcal{B}_m(\beta).$$

Finally, again by Fact 6.3.9, it follows that

$$\dim_{\mathrm{H}}(\widehat{\mathcal{B}} \cap \mathbb{X}_m) = \dim_{\mathrm{H}} \mathbb{X}_m,$$

which yields

$$\dim_{\mathrm{H}} \widehat{\mathcal{B}} \geqslant \dim_{\mathrm{H}} \bigcup_{m \geqslant 2} (\widehat{\mathcal{B}} \cap \mathbb{X}_m) = \sup_{m \geqslant 2} \dim_{\mathrm{H}} \mathbb{X}_m,$$

by standard properties of the Hausdorff dimension [Pes97, Thm. 6.1]. Hence, it suffices to show that $\dim_{\mathrm{H}} \mathbb{X}_m$ converges to 1 as $m \to \infty$. There are certainly several ways to do so. We sketch one version in the following. The SFT is topologically conjugate to a Markov subshift with a primitive transition matrix A; compare Section 2.1.3 and [LM95]. By [Pes97, Thm. A2.9], we have

$$\dim_{\mathrm{H}} \mathbb{X}_m = \frac{\log(\rho_m)}{\log(2)},$$

where ρ_m is the PF eigenvalue of the transition matrix A. Using the methods provided in [HA21], we obtain that ρ_m is the largest root of the polynomial

$$q_m(z) = z^{m+1} - 2z^m + 1.$$

Note that $q_m(2) = 1$ and that $q_m'(2) \to \infty$ as $m \to \infty$. From this, it is straightforward to show that $\rho_m \to 2$ as $m \to \infty$. We obtain $\dim_{\mathrm{H}} \widehat{\mathcal{B}} \geqslant \lim_{m\to\infty} \dim_{\mathrm{H}} \mathbb{X}_m = 1$, and the proof is complete. □

The proof of Corollary 6.3.14 makes use of the idea that, in order to obtain certain relations on $(\mathbb{X}, S)$, it is possible to first analyze the corresponding relations on $(\mathbb{X}_m, S)$, with $m \geqslant 2$, and then to "exhaust" $(\mathbb{X}, S)$ by the spaces $(\mathbb{X}_m, S)$ in an appropriate sense. We build on this idea in order to push the relation between b_m, f_m and p_m in (6.18) and (6.20) towards a relation between b, f and the pressure function

$$p\colon \mathbb{R} \to (-\infty, \infty], \quad p(t) = \mathcal{P}_{\mathrm{top}}(t\psi).$$

By definition, it is straightforward to see that each of the sequences of functions $(b_m)_m$, $(f_m)_m$ and $(p_m)_m$ are non-decrasing and bounded by b, f and p, respectively. In fact, we will see that the sequences converge pointwise to b, f and p, respectively. This requires some additional control, ensuring that the level sets of $\overline{\psi}(x)$ and $\dim_\mu(x)$ in $\mathbb{X}$ take up a similar "proportion" (in the sense of Hausdorff dimension) when restricted to $\mathbb{X}_m$, for large enough m. The main technical tool in this direction was provided by Schindler in [BGKS19, Lemma 5.2] and [BGKS19, Lemma 5.3], relating the behaviour of ψ_n on cylinders in $\mathbb{X}$ to the behaviour of ψ_n on appropriately chosen cylinders in $\mathbb{X}_m$.

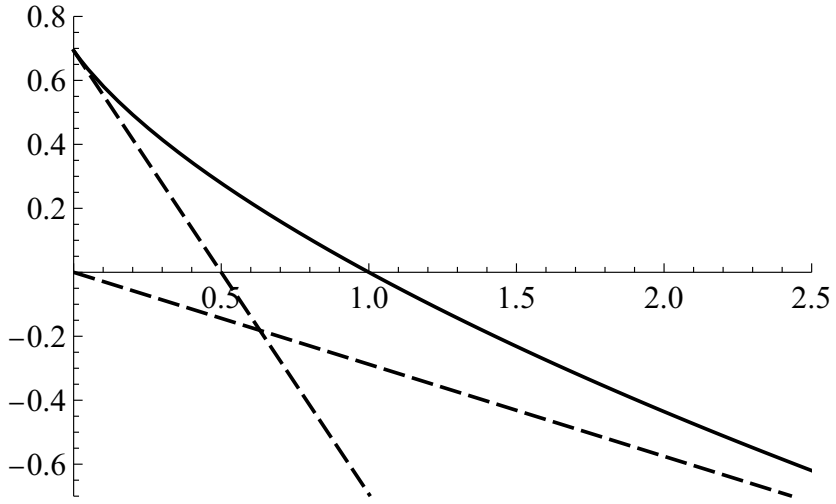

Figure 6.4.: Illustration of the pressure function $t \mapsto p(t)$ (solid line) with the asymptotes $t \mapsto (1-2t)\log(2)$ and $t \mapsto \log(3/4)t$ (dashed lines).

Lemma 6.3.15 ([BGKS19])**.** *Let $m \geqslant 2$ and $\mathcal{A} = \{0,1\}$. Then, there exists a function $h_m \colon \mathcal{A}^+ \to \mathcal{A}^+$ with the following properties.*

(1) *Given $n \in \mathbb{N}$, we have $h_m(\mathcal{A}^n) \subset \mathcal{A}^n$ and $\# h_m^{-1}(v) \leqslant 2^{\lfloor n/m \rfloor}$ for all $v \in \mathcal{A}^n$.*

(2) *For every $w \in \mathcal{A}^n$, the word $v = h_m(w)$ satisfies $[v] \cap \mathbb{X}_m \neq \varnothing$, and*

$$\sup_{y \in [w]} \psi_n(y) \leqslant \sup_{x \in [v] \cap \mathbb{X}_m} \psi_n(x) + 4\lfloor n/m \rfloor + 2^{m+2},$$

for all $n \in \mathbb{N}$. ■

Remark 6.3.16. In fact, we considered a slightly different potential function in [BGKS19], given by $\widetilde{\psi}(x) = \psi(x) + \log(2)$, for all $x \in \mathbb{X}$. Since μ is the unique equilibrium measure to both ψ and $\widetilde{\psi}$, this choice is somewhat arbitrary. It is straightforward to verify that the corresponding pressure function is related to p via

$$\widetilde{p}(t) := \mathcal{P}_{\mathrm{top}}\big(t\widetilde{\psi}\big) = p(t) + t\log(2),$$

for all $t \in \mathbb{R}$, and that the corresponding Legendre transform is obtained from p^* by a shift in the argument of magnitude $\log(2)$. ◊

Proposition 6.3.17. *For every $t \in \mathbb{R}$, we have $\lim_{m\to\infty} p_m(t) = p(t)$. The convergence is uniform on every compact subset of $\mathbb{R}_{\geqslant 0}$. In particular, p is convex and continuous on its domain $\mathrm{dom}(p) = \mathbb{R}_{\geqslant 0}$.*

Proof. We start by showing pointwise convergence. Since $(p_m(t))_{m\geqslant 2}$ is non-decrasing and bounded by $p(t)$ for all $t \in \mathbb{R}$, it suffices to show that $p(t) \leqslant \lim_{m\to\infty} p_m(t)$. Recall that

$$p_m(t) = \mathcal{P}_{\mathrm{top}}(t\psi^{(m)}) = \lim_{n\to\infty} \frac{1}{n} \log \sum_{w \in \{0,1\}^n} \sup_{x \in [w] \cap \mathbb{X}_m} \exp(t\psi_n(x)).$$

For $t < 0$, set $y_m = \overline{0^m 1} \in \mathbb{X}_m$ as in the proof of Proposition 6.3.13, and recall that $\lim_{m\to\infty} \overline{\psi}(y_m) = -\infty$. Hence,

$$p_m(t) \geqslant \lim_{n\to\infty} \frac{1}{n} \log \exp(t\psi_n(y_m)) = t\overline{\psi}(y_m) \xrightarrow{m\to\infty} \infty.$$

This also shows that $p(t) = \infty$ for all $t < 0$. For $t \geqslant 0$, we group the words in $\{0,1\}^n$ according to their images under h_m and obtain, using Lemma 6.3.15,

$$p(t) = \lim_{n\to\infty} \frac{1}{n} \log \sum_{v\in h_m(\mathcal{A}^n)} \sum_{w\in h_m^{-1}(v)} \sup_{x\in[w]} \exp(t\psi_n(x)) \leqslant \frac{4t + \log(2)}{m} + p_m(t),$$

for all $m \geqslant 2$. Performing the limit $m \to \infty$ yields the desired convergence. Clearly, the convergence is uniform on every compact subset of $\mathbb{R}_{\geqslant 0}$. Hence, the property of being convex and continuous on $\mathbb{R}_{\geqslant 0}$ is inherited from the functions p_m. □

Corollary 6.3.18. *The sequence of functions $(p_m^*)_{m\geqslant 2}$ converges pointwise to p^* on $\mathrm{dom}(p^*)$.*

Proof. This follows by standard arguments in convex analysis. Since p is convex and continuous with a closed domain, it is *closed* in the sense that $p^{-1}(-\infty, y]$ is a closed subset of $\mathbb{R}$, for all $y \in \mathbb{R}$. The same holds for each of the functions p_m, with $m \geqslant 2$. This is enough to obtain $p_m^* \to p^*$ from the pointwise convergence $p_m \to p$ [JK11, Prop. 4.1]; compare also [Roc70, SW77] for background on convex analysis and the convergence of epigraphs. □

Combining Corollary 6.3.18 with (6.18), we observe that

$$-\frac{p^*(\beta)}{\log 2} = \lim_{m\to\infty} b_m(\beta) \leqslant b(\beta), \tag{6.21}$$

for all $\beta \in \mathrm{dom}(p^*)$. Hence, in order to get equality, it suffices to bound $b(\beta)$ from above by $-p^*(\beta)/\log(2)$. Similar reasoning applies to the sequence $(f_m(\alpha))_{m\geqslant 2}$. The proof of the remaining step is due to a construction by Kesseböhmer in [BGKS19, Sec. 6], which we reproduce in the following, with minor changes in the presentation.

Proposition 6.3.19. *We have $\lim_{m\to\infty} f_m(\alpha) = f(\alpha)$ and $\lim_{m\to\infty} b_m(\beta) = b(\beta)$ for all $\alpha, \beta \in \mathbb{R}$. Further, for all $\beta \leqslant \log(3/4)$,*

$$b(\beta) = -\frac{p^*(\beta)}{\log(2)},$$

and $b(\beta) = 0$, otherwise. The function f satisfies $f(\alpha) = b(-\log(2)\alpha)$ for all $\alpha \in \mathbb{R}$.

Proof. We focus on the relation for b, the corresponding result for f is obtained by similar arguments. By Corollary 6.3.14, $b(\beta) = 0$ for $\beta > \log(3/4)$ is immediate. Hence, we assume $\beta \leqslant \log(3/4)$ in the following. Our aim is to show that, given $\varepsilon > 0$ and $s > -p^*(\beta-\varepsilon)/\log(2)$, the s-dimensional Hausdorff measure $m_{\mathrm{H}}^s(\mathcal{B}(\beta))$ vanishes. We then obtain

$$b(\beta) = \dim_{\mathrm{H}} \mathcal{B}(\beta) \leqslant \limsup_{\varepsilon\to 0} -\frac{p^*(\beta-\varepsilon)}{\log(2)} \leqslant -\frac{p^*(\beta)}{\log(2)},$$

where the last step is due to the fact that p^* is automatically lower semi-continuous, by standard properties of the Legendre transform. In particular, $(-\infty, \log(3/4)] \subset \mathrm{dom}(p^*)$. In conjunction with (6.21), this yields the desired relations.

It remains to show that $m_H^s(\mathcal{B}(\beta)) = 0$ in the situation above. As a first step, note that $\overline{\psi}(x) = \beta$ requires that $\psi_n(x) > n(\beta - \varepsilon)$ eventually, and hence,

$$\mathcal{B}(\beta) \,\subset\, \limsup_{n\to\infty} \mathcal{G}(n, \alpha - \varepsilon), \quad \mathcal{G}(n, \alpha - \varepsilon) \,=\, \{x \in \mathbb{X} : \psi_n(x) > n(\beta - \varepsilon)\},$$

implying that

$$m_{\mathrm{H}}^s(\mathcal{B}(\beta)) \,\leqslant\, \lim_{k\to\infty} \sum_{n\geqslant k} m_{\mathrm{H}}^s\big(\mathcal{G}(n, \beta - \varepsilon)\big). \tag{6.22}$$

In the following, we cover $\mathcal{G}(n, \beta - \varepsilon)$ by cylinders of length n and obtain an upper bound for the number of such cylinders. This will yield an upper bound for the s-dimensional Hausdorff measure that decays exponentially with n. Let $\Gamma_n = \{w \in \{0,1\}^n : [w] \cap \mathcal{G}(n, \beta - \varepsilon) \neq \varnothing\}$. Since $\beta \leqslant \log(3/4)$, the set Γ_n is non-empty. We bound $\#\Gamma_n$ in terms of s. First note that by the assumption on s, there exists a $t \geqslant 0$ such that

$$s \log(2) \,>\, p(t) - t(\beta - \varepsilon).$$

Given $\delta > 0$, we can find n_0 such that for all $n \geqslant n_0$,

$$p(t) + \delta \,\geqslant\, \frac{1}{n} \log \sum_{w\in\Gamma_n} \exp(tn(\beta - \varepsilon)) \,=\, \frac{\log(\#\Gamma_n)}{n} + t(\beta - \varepsilon).$$

Choosing $\delta > 0$ small enough, we obtain

$$s \log(2) > \frac{\log(\#\Gamma_n)}{n}, \tag{6.23}$$

for all $n \geqslant n_0$. Using $\mathcal{G}(n, \alpha - \varepsilon) \subset \bigcup_{w\in\Gamma_n} [w]$, we get for $n \geqslant n_0$,

$$m_{\mathrm{H}}^s(\mathcal{G}(n, \alpha - \varepsilon)) \,\leqslant\, \sum_{w\in\Gamma_n} m_{\mathrm{H}}^s([w]) \,=\, \#\Gamma_n \mathrm{e}^{-ns\log(2)} \,=\, \mathrm{e}^{-n(s\log(2) - \log(\#\Gamma_n)/n)}.$$

Because of (6.23), there exists a number $r > 0$ such that $s\log(2) - \log(\#\Gamma_n)/n \geqslant r$, for all $n \geqslant n_0$. Finally, due to (6.22),

$$m_{\mathrm{H}}^s(\mathcal{B}(\beta)) \,\leqslant\, \lim_{k\to\infty} \sum_{n\geqslant k} \mathrm{e}^{-nr} \,=\, 0,$$

and the claim follows. □

In the proof of Proposition 6.3.19, we have already seen that $(-\infty, \log(3/4)] \subset \mathrm{dom}(p^*)$. In fact, with a bit more work, it is possible to show that $(-\infty, \log(3/4)] = \mathrm{dom}(p^*)$ [BGKS19]. Hence, we can express b on $\mathbb{R}$ as

$$b(\beta) \,=\, \max\left\{-\frac{p^*(\beta)}{\log(2)}, 0\right\}. \tag{6.24}$$

Analyzing the properties of p, it is possible to be more explicit about the form of the Birkhoff spectrum b.

Proposition 6.3.20. *The function b is concave on $(-\infty, \log(3/4)]$, identical to 1 on the interval $(-\infty, \log(1/4)]$, and strictly smaller than 1 otherwise. Further, $b(\beta) = 0$ for all $\beta \geqslant \log(3/4)$.*

Sketch of proof. Using (6.24), all of the announced properties of b follow from properties of the pressure function p, that were established by Kesseböhmer and Schindler in [BGKS19, Prop. 7.1]; compare also [BGKS19, Sec. 8]. Let us sketch an alternative proof for the fact that $b(\beta) = 1$ for all $\beta \leqslant \log(1/4)$ that sheds a different light on this probably most uncommon feature of the Birkhoff spectrum. By an application of Birkhoff's ergodic theorem, it follows that for Lebesgue almost every $x \in \mathbb{T}$,

$$\overline{\psi}(x) \,=\, \int_{\mathbb{T}} \psi(x)\, \mathrm{d}\lambda_{\mathrm{L}}(x) \,=\, \log(1/4),$$

by direct calculation, compare [BGN14]. This implies that the points $x \in \mathbb{X}$ with $\overline{\psi}(x) = \log(1/4)$ have full Hausdorff dimension, that is,

$$b(\log(1/4)) \,=\, \dim_{\mathrm{H}} \mathcal{B}(\log(1/4)) \,=\, 1.$$

On the other hand, $b_m(\log(1/4)) \to 1$ as $m \to \infty$ due to Proposition 6.3.19, and $b_m(\beta_m^-) \geqslant 0$, with $\beta_m^- \to -\infty$, as we have seen in Proposition 6.3.13. Since b_m is concave, the graph of b_m lies above the line segment connecting the points $(\beta_m^-, b_m(\beta_m^-))$ and $(\log(1/4), b_m(\log(1/4)))$. For all $\beta < \log(1/4)$, this yields

$$1 \,\geqslant\, b(\beta) \,=\, \lim_{m\to\infty} b_m(\beta) \,=\, 1,$$

and the claim follows. □

Collecting the results and procedures above, we are now in a position to prove Theorem 6.0.1. Again, we gloss over the subtle differences between $(\mathbb{T}, T_2)$ and $(\mathbb{X}, S)$; compare Remark 6.3.2 and [BGKS19] for a more careful account.

Proof of Theorem 6.0.1. The statements on the form of the dimension spectrum follow from Proposition 6.3.20, together with the relation $f(\alpha) = b(-\log(2)\alpha)$, which was stated in Proposition 6.3.19. Using the relation between restricted level sets in (6.19), the proof of Corollary 6.3.14 is easily adapted to yield the corresponding results for the level sets $\mathcal{F}(\alpha)$ and the exceptional set $\widehat{\mathcal{F}}$. Here, in order to obtain that $\mathcal{F}(\alpha) = \varnothing$ for $\alpha < \alpha_0$, we use Gelfond's result in conjunction with the upper bound of the measure on cylinder sets given in (6.17).

Finally, by the Shannon–MacMillan–Breiman theorem, stated in Theorem 2.3.7, the local dimension $d_\mu(x)$ is μ-almost surely given by

$$d_\mu \,=\, \frac{h_\mu}{\log(2)},$$

which is often called the *information dimension* [GL90]. This means just $\mu(\mathcal{F}(d_\mu)) = 1$. By standard arguments in dimension theory, this implies that $d_\mu = f(d_\mu)$ [Pes97, Ch. 5]. □

Remark 6.3.21. The fact that the dimension spectrum $f(\alpha)$ is constant to 1 for all $\alpha \geqslant \alpha_1$ can be traced back to the singularity of $\psi(x)$ at $x = 0$; compare also the discussion in [GL90]. In fact, as we already discussed in Section 6.2, $\mu(B_r(x))$ decays faster than any polynomial at dyadic points. Due to approximation ideas, it is therefore not surprising that arbitrarily fast polynomial scaling is typical in the sense of full Hausdorff dimension. In a sense, the pointwise dimension of μ does not offer enough "resolution" to get a meaningful characterization of the thin part of μ from the corresponding spectrum. The fastest scaling of $\mu(B_r(x))$ is to be exprected at $x = 0$, where it satisfies

$$\log \mu(B_r(x)) \sim -\theta \log^{\gamma}(r), \tag{6.25}$$

with $\theta = 1/\log(2)$ and $\gamma = 2$, see [GL90, BCEG21, BG19]. It might well be that we obtain a meaningful spectrum if we regard the level sets of points x such that (6.25) holds for $\gamma = 2$ and some $\theta \in [0, 1/\log(2)]$ (alternatively for $\gamma \in [1, 2]$ and arbitrary θ). This could lead to a multifractal analysis that captures the behaviour of μ at points where the scaling is super-polynomial and seems to be a promising direction for future research. However, the subshifts $\mathbb{X}_m$ are clearly blind to the super-polynomial scaling, calling for new methods. ◊

Remark 6.3.22. It is a special feature of the potential ψ that the singularity is placed at a fixed point under the doubling map. It seems natural to inquire what happens if the singularity if shifted to a different point. A multifractal analysis for the more general family of potentials

$$\psi_b(x) = \psi(x + b),$$

with $b \in \mathbb{T}$, is performed in upcoming work of Fan, Schmeling and Shen [FSS]. It is also worth noting that, for each $b \in \mathbb{T}$, the equilibrium measure that corresponds to ψ_b is the diffraction measure of a generalized TM substitution on a compact alphabet, as defined in Example 6.1.27. ◊

Outlook

One of the striking features of random substitutions is the coexistence of positive entropy and non-trivial Bragg peaks in the diffraction image. The latter also signify the existence of non-trivial dynamical eigenvalues. In [BSS18], it was shown for several examples that the corresponding eigenfunctions are discontinuous, but that the points of discontinuity form a null set for the ergodic measure. In this situation, the maximal equicontinuous factor of the dynamical system is trivial and the pure point part of the dynamical spectrum is captured by its measure-theoretic analogue, the *Kronecker factor*. In fact, the Kronecker factor is the same for all ergodic measures of interest in this case, and it coincides with the *maximal equicontinuous generic factor*, which is a purely topological object, defined on the subset of points with a dense orbit; compare [HY12, Kel20] for background. It would be interesting to see if these observations hold in general for compatible random substitutions (possibly under mild extra assumptions), just like some of the other characteristics discussed in Chapter 4.

A natural analogue of the classic Pisot substitution conjecture in the random setting is to assume absence of a singular continuous component in the diffraction spectrum of (irreducible) compatible Pisot random substitutions. In full generality, this is probably difficult to establish, but it should be possible to construct effective algorithms that check this property for specific examples. This is already work in progress [GMR].

Concerning the algorithms for computing the topological and measure-theoretic entropy of random substitutions, a generalization to higher dimensions seems to be well within reach, at least for particularly simple examples like random stone inflations (compare also the randomization of the Penrose tiling, discussed in [GL89]). This would already suffice to cover the models of random percolation investigated by Dekking and Meester in [DM90]. There, it was shown that varying the underlying probability parameter can lead to several phase transitions. Maximising the entropy with respect to this parameter might be a way to single out a physically reasonable phase.

A general motive, underlying the treatment of random substitutions in this thesis, is that characteristics that are defined for the limiting objects (bi-infinite sequences) can be most efficiently calculated or determined by following iterations of the random substitution on finite words. A similar idea, in the form of so called trace maps, has proved to be fruitful for the discussion of Schrödinger operators that are associated to *deterministic* substitutions [BG94]. Although a generalization to random substitutions is certainly desirable, it seems to come with serious technical obstacles. This is partly due to the fact that spectral properties of Schrödinger operators are much less robust to small changes in the base dynamics than characteristics like entropy or diffraction.

The existence of periodic points in a subshift directly yields that the Schrödinger operator associated to a point with dense orbit exhibits intervals in its spectrum. It is an open question

whether the same property is implied by positive entropy of the underlying subshift. Many classic examples in the positive entropy regime like Bernoulli subshifts or subshifts of finite type also contain periodic points and hence do not help to settle this question. In this regard, the random Fibonacci subshift, which combines positive entropy with the absence of periodic points, provides a useful example. The same holds for (aperiodic) Boshernitzan subshifts with a random background. Unfortunately, for both cases, the study of the Schrödinger spectrum has remained inconclusive so far.

Regarding the multifractal analysis of the Thue–Morse measure, provided in Chapter 6, several further research directions suggest themselves. First, the diffraction of more general substitutions of constant length naturally leads us to consider Matrix-valued Riesz products [Que10]. In this context, we are dealing with non-commuting objects in general, and hence the scaling analysis can be expected to become more involved. A non-additive version of the thermodynamic formalism might be of help; compare [Pes97, Ch. 4] and [Bar96]. On a more conceptual level, the study of potentials with singularities in the framework of thermodynamic formalism certainly deserve further, and more systematic, attention. A first step is provided in [FSS]. Finally, revisiting the idea presented in Remark 6.3.21, it would be desirable to find a multifractal analysis of the TM measure that accounts for those points of the unit interval where the measure decays faster than any polynomial.

A. Conditional probabilities

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and assume that f is a random variable in $L^1(\Omega, \mathcal{F}, \mathbb{P})$. We fix some σ-algebra $\mathcal{F}_0 \subset \mathcal{F}$.

Definition A.0.1. A *conditional expectation* of f with respect to $\mathcal{F}_0$ is an $\mathcal{F}_0$-measurable random variable f_0 such that

$$\int_B f \, \mathrm{d}\mathbb{P} \;=\; \int_B f_0 \, \mathrm{d}\mathbb{P},$$

for all $B \in \mathcal{F}_0$.

A conditional expectation of f with respect to $\mathcal{F}_0$ always exists and it is unique up to $\mathbb{P}$-null sets. We denote by $\mathbb{E}[f|\mathcal{F}_0]$ the corresponding equivalence class of functions.

Definition A.0.2. Let $A \in \mathcal{F}$. The *conditional probability* of A with respect to $\mathcal{F}_0$ is given by

$$\mathbb{P}[A|\mathcal{F}_0] \;=\; \mathbb{E}[\mathbb{1}_A|\mathcal{F}_0].$$

We emphasize that a conditional probability is an equivalence class of functions that is well-defined almost surely. To see the connection with the more elementary notion of conditional probability, assume for a moment that $\mathcal{F}_0$ is the σ-algebra generated by some finite partition $\{B_1, \ldots, B_n\}$ of Ω, composed of sets with positive $\mathbb{P}$-measure. Then, we can choose a version of the conditional probability by

$$\mathbb{P}[A|\mathcal{F}_0](\omega) \;=\; \mathbb{P}[A|B_j],$$

whenever $\omega \in B_j$ and $1 \leqslant j \leqslant n$. Similarly, if Z is a random variable on $(\Omega, \mathcal{F})$ that takes only finitely many values with positive probability, we get, for ω satisfying $Z(\omega) = z$,

$$\mathbb{P}[A|\sigma(Z)](\omega) \;=\; \mathbb{P}[A|Z = z].$$

Even for more general random variables Z, $\mathbb{P}[A|\sigma(Z)]$ can be written as a function that depends only on the values of Z. This is due to the Doob–Dynkin lemma [Bob05, Ch. 2.1].

The same observation holds for conditional expectations. If Y, Z are $\mathcal{F}$-measurable random variables, it is customary to employ the notation

$$\mathbb{E}[Y|Z] \;:=\; \mathbb{E}[Y|\sigma(Z)],$$

which is implicitly a function on Z.

Bibliography

[Aar97] J. Aaronson, *An Introduction to Infinite Ergodic Theory*, AMS, Providence, RI (1997).

[AKM65] R.L. Adler, A.G. Konheim and M.H. McAndrew, Topological entropy, *Trans. Amer. Math. Soc.* **114** (1965) 309–319.

[AC12] E. Akin, J. Carlson, Conceptions of topological transitivity, *Topology Appl.* **159** (2012) 2815–2830.

[ABB+15] S. Akiyama, M. Barge, V. Berthé, J.-Y. Lee and A. Siegel, On the Pisot substitution conjecture, in *Mathematics of Aperiodic Order*, eds. J. Kellendonk, D. Lenz and J. Savinien, Birkhäuser, Basel (2015) pp. 33–72.

[AP86] J.-P. Allouche and J. Peyrière, Sur une formule de récurrence sur les traces de produits de matrices associés à certaines substitutions, *C. R. Acad. Sci. Paris* **302** (1986) 1135–1136.

[AG73] W.O. Amrein and V. Georgescu, On the characterization of bound states and scattering states in quantum mechanics, *Helv. Phys. Acta* **46** (1973/74) 635–658.

[ABM19] N. Aubrun, S. Barbieri, and E. Moutot, The domino problem is undecidable on surface groups, *44th International Symposium on Mathematical Foundations of Computer Science*, eds. P. Rossmanith, P. Heggernes and J.-P. Katoen, Art. No. 46 (2019); `arXiv:1811.08420`.

[ABY10] A. Avila, J. Bochi and J.-C. Yoccoz, Uniformly hyperbolic finite-valued $\mathrm{SL}(2;\mathbb{R})$-cocycles, *Comment. Math. Helv.* **85** (2010) 813–884; `arXiv:0808.0133`.

[ADZ20] A. Avila, D. Damanik and Z. Zhang, Schrödinger operators with potentials generated by hyperbolic transformations: I. Positivity of the Lyapunov exponent, *preprint* (2020); `arXiv:2011.10146`.

[BBM10] M. Baake, M. Birkner, and R.V. Moody, Diffraction of stochastic point sets: explicitly computable examples, *Commun. Math. Phys.***293** (2010) 611–660; `arXiv:0803.1266`.

[BCEG21] M. Baake, M. Coons, J. Evans and P. Gohlke, On a family of singular continuous measures related to the doubling map, *Indag. Math.* **32** (2021) 847–860; `arXiv:2006.09755`.

[BGM19] M. Baake, F. Gähler and N. Mañibo, Renormalisation of pair correlation measures for primitive inflation rules and absence of absolutely continuous diffraction, *Commun. Math. Phys.* **370** (2019) 591–635; `arXiv:1805.09650`.

[BGKS19] M. Baake, P. Gohlke, M. Kesseböhmer and T. Schindler, Scaling properties of the Thue–Morse measure, *Discr. Cont. Dynam. Syst. A* **39** (2019) 4157–4185. `arXiv:1810.06949`.

[BG13] M. Baake and U. Grimm, *Aperiodic Order. Vol. 1: A Mathematical Invitation*, Cambridge University Press, Cambridge (2013).

[BG19] M. Baake and U. Grimm, Scaling of diffraction intensities near the origin: some rigorous results, *J. Stat. Mech.: Theory Exp.* **5** (2019) 054003:1–25; `arXiv:1905.04177`.

[BGN14] M. Baake, U. Grimm and J. Nilsson, Scaling of the Thue–Morse diffraction measure, *Acta Phys. Pol. A*, **126** (2014) 431–434; `arXiv:1311.4371`.

[BL04] M. Baake and D. Lenz, Dynamical systems of translation bounded measures: Pure point dynamical and diffraction spectra, *Ergodic Theory Dynam. Systems* **24** (2004) 1867–1893; `arXiv:math.DS/0302061`.

[BLvE15] M. Baake, D. Lenz and A.C.D. van Enter, Dynamical versus diffraction spectrum for structures with finite local complexity, *Ergodic Theory Dynam. Systems* **35** (2015) 2017–2043; `arXiv:1307.7518`.

[BSS18] M. Baake, T. Spindeler and N. Strungaru, Diffraction of compatible random substitutions in one dimension, *Indag. Math.* **29** (2018) 1031–1071; `arXiv:1712.00323v2`.

[BBJS12] M. Barge, H. Bruin, L. Jones and L. Sadun, Homological Pisot substitutions and exact regularity, *Israel J. Math.* **188** (2012) 281–300; `arXiv:1001.2027`.

[BDEG88] M.F. Barnsley, S.G. Demko, J.H. Elton and J.S. Geronimo, Invariant measures for Markov processes arising from iterated function systems with place-dependent probabilities, *Ann. Inst. H. Poincarè Probab. Statist.* **24** (1988) 367–394.

[Bar96] L.M. Barreira, A non-additive thermodynamic formalism and applications to dimension theory of hyperbolic dynamical systems, *Ergodic Theory Dynam. Systems* **16** (1996) 871–927.

[Bel90] J. Bellissard, Spectral properties of Schrödinger's operator with a Thue–Morse potential, in *Number Theory and Physics*, eds. J.-M. Luck, P. Moussa and M. Waldschmidt, Springer, Berlin (1990), pp. 140–150.

[BBG91] J. Bellissard, A. Bovier and J.-M. Ghez, Spectral properties of a tight binding Hamiltonian with period doubling potential, *Commun. Math. Phys.* **135** (1991) 379–399.

[Ber99] J. Berstel, On the index of Sturmian words, in *Jewels are Forever*, eds. J. Karhumäki, H. Maurer, G. Păun and G. Rozenberg, Springer, Berlin (1999), pp. 287–294.

[BSTY19] V. Berthé, W. Steiner, J.M. Thuswaldner and R. Yassawi, Recognizability for sequences of morphisms, *Ergodic Theory Dynam. Systems* **39** (2019) 2896–2931; `arXiv:1705.00167`.

[Bil99] P. Billingsley, *Convergence of Probability Measures*, 2nd ed., Wiley, New York (1999).

[Bil95] P. Billingsley, *Probability and Measure*, 3rd ed., Wiley, New York (1995).

[Bir31] G. Birkhoff, Proof of the ergodic theorem, *Proc. Nat. Acad. Sci. USA* **17** (1931) 656–660.

[Bob05] A. Bobrowski, *Functional Analysis for Probability and Stochastic Processes. An Introduction*, Cambridge University Press, Cambridge (2005).

[Bos84] M. Boshernitzan, A condition for unique ergodicity of minimal symbolic flows with linear block growth, *J. Analyse Math.* **44** (1984/85) 77–96.

[Bos92] M. Boshernitzan, A condition for unique ergodicity of minimal symbolic flows, *Ergodic Theory Dynam. Systems* **12** (1992) 425–428.

[BDFL21] M. Boshernitzan, D. Damanik, J. Fillman and M. Lukić, Ergodic Schrödinger operators in the infinite measure setting, *J. Spectr. Theory* **11** (2021) 873–902; arXiv:1907.12471.

[BG94] A. Bovier and J.-M. Ghez, Spectral properties of one-dimensional Schrödinger operators with potentials generated by substitutions, *Commun. Math. Phys.* **158** (1993) 45–66; Erratum: *Commun. Math. Phys.* **166** (1994) 431–432.

[Bow73] R. Bowen, Entropy for group endomorphisms and homogeneous spaces, *Trans. Amer. Math. Soc.* **153** (1971) 401–414; Erratum: *Trans. Amer. Math. Soc.* **181** (1973) 509–510.

[Bow75] R. Bowen, *Equilibrium States and the Ergodic Theory of Anosov Diffeomorphisms*, LNM 470, Springer, Berlin (1975).

[Bre20] P. Brémaud, *Markov Chains. Gibbs fields, Monte Carlo Simulation and Queues*, 2nd ed., Springer, Cham (2020).

[BDF+19] V. Bucaj, D. Damanik, J. Fillman, V. Gerbuz, T. VandenBoom, F. Wang and Z. Zhang, Localization for the one-dimensional Anderson model via positivity and large deviationsfor the Lyapunov exponent, *Trans. Amer. Math. Soc.* **372** (2019) 3619–3667; `arXiv:1706.06135`.

[BK93] R.M. Burton and G. Keller, Stationary measures for randomly chosen maps, *J. Theoret. Probab.* **6** (1993) 1–16.

[Cas97] J. Cassaigne, Complexité et facteur spéciaux, *Bull. Belg. Math. Soc.* **4** (1997) 67–88.

[Cec37] E. Čech, On bicompact spaces, *Ann. of Math.* **38** (1937) 823–844.

[CDFG] J. Chaika, D. Damanik, J. Fillman and P. Gohlke, Schrödinger operators with mixed potentials, *in preparation.*

[CDFG21] J. Chaika, D. Damanik, J. Fillman and P. Gohlke, Zero measure spectrum for multi-frequency Schrödinger operators, *J. Spectr. Theory*, to appear; `arXiv:2009.11946`.

[CDF92] A. Cohen, I. Daubechies and J.-C. Feauveau, Biorthogonal bases of compactly supported wavelets, *Commun. Pure Appl. Math.* **45** (1992) 485–560.

[Con90] J.B. Conway, *A Course in Functional Analysis*, 2nd ed., GTM 96, Springer, New York (1990).

[CR90] J.-P. Conze and A. Raugi, Fonctions harmoniques pour un opérateur de transition et applications, *Bull. Soc. Math. France* **118** (1990) 273–310.

[CFS82] I.P. Cornfeld, S.V. Fomin and Y.G. Sinai, *Ergodic Theory*, Springer, New York (1982).

[CK21] V. Cyr, B. Kra, Boshernitzan's condition, factor complexity, and an application, *Proc. Amer. Math. Soc.*, to appear; `arXiv:2006.01931`.

[DV-J05] D.D. Daley and D. Vere-Jones, *An Introduction to the Theory of Point Processes I: Elementary Theory and Methods*, 2nd ed., Springer, New York (2005).

[Dam17] D. Damanik, Schrödinger operators with dynamically defined potentials, *Ergodic Theory Dynam. Systems* **37** (2017) 1681–1764; `arXiv:1410.2445`.

[Dam01] D. Damanik, Uniform singular continuous spectrum for the period doubling Hamiltonian, *Ann. Henri Poincaré* **2** (2001), 101–118.

[DEG15] D. Damanik, M. Embree and A. Gorodetski, Spectral properties of Schrödinger operators arising in the study of quasicrystals, in *Mathematics of Aperiodic Order*, eds. J. Kellendonk, D. Lenz and J. Savinien, Birkhäuser (2015) pp. 307–370; `arXiv:1210.5753`.

[DFa] D. Damanik and J. Fillman, *One-Dimensional Ergodic Schrödinger Operators, I. General Theory*, Monograph in preparation.

[DFb] D. Damanik and J. Fillman, *One-Dimensional Ergodic Schrödinger Operators, II. Special Classes*, Monograph in preparation.

[DG11] D. Damanik and A. Gorodetski, Spectral and quantum dynamical properties of the weakly coupled Fibonacci Hamiltonian, *Commun. Math. Phys.* **305** (2011) 221–277; `arXiv:1001.2552`.

[DL06a] D. Damanik and D. Lenz, A condition of Boshernitzan and uniform convergence in the Multiplicative Ergodic Theorem, *Duke Math. J.* **133** (2006) 95–123; `arXiv:math/0403190`.

[DL06b] D. Damanik and D. Lenz, Substitution dynmical systems: Characterization of linear repetitivity and applications, *J. Math. Anal. Appl.* **321** (2006) 766–780; `arXiv:math/0302231`.

[DL99] D. Damanik and D. Lenz, Uniform spectral properties of one-dimensional quasicrystals, I. Absence of eigenvalues, *Commun. Math. Phys.* **207** (1999) 687–696; `arXiv:math-ph/9903011`.

[DL06c] D. Damanik and D. Lenz, Zero-measure Cantor spectrum for Schrödinger operators with low-complexity potentials, *J. Math. Pures Appl.* **85** (2006) 671–686.

[Dek78] F.M. Dekking, The spectrum of dynamical systems arising from substitutions of constant length, *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete* **41** (1978) 221–239.

[DM90] F.M. Dekking and R.W.J. Meester, On the structure of Mandelbrot's percolation process and other random Cantor sets, *J. Stat. Phys.* **58** (1990) 1109–1126.

[DvdW01] F.M. Dekking and P. vd Wal, Fractal percolation and branching cellular automata, *Probab. Theory Related Fields* **120** (2001) 277–308.

[dOL02] C.R. de Oliveira and M.V. Lima, Singular continuous spectrum for a class of nonprimitive substitution Schrödinger operators, *Proc. Amer. Math. Soc.* **130** (2002) 145–156.

[dHo00] F. den Hollander, *Large Deviations*, AMS, Providence, RI (2000).

[Din70] E.I. Dinaburg, A correlation between topological entropy and metric entropy, (Russian) *Dokl. Akad. Nauk SSSR* **190** (1970) 19–22; Translation: *Soviet Math. Dokl.* **11** (1969) 13–16.

[DoF37] W. Doeblin and R. Fortet, Sur des chaînes à liaisons complètes, *Bull. Soc. Math. France* **65** (1937) 132–148.

[DuF66] L.E. Dubins and D.A. Freedman, Invariant probabilities for certain Markov processes, *Ann. Math. Stat.* **37** (1966) 837–848.

[Dur98] F. Durand, A characterization of substitutive sequences using return words, *Discr. Math.* **179** (1998) 89–101; `arXiv:0807.3322`.

[DLR13] F. Durand, J. Leroy and G. Richomme, Do the properties of an S-adic representation determine factor complexity?, *J. Integer Seq.* **16** (2013) Article 13.2.6; `arXiv:1208.6376`.

[DOP18] F. Durand, N. Ormes and S. Petite, Self-induced systems, *J. Anal. Math.* **135** (2018) 725–756; `arXiv:1511.01320`.

[EG21] B. Eichinger and P. Gohlke, Spectral properties of Schrödinger operators associated with almost minimal substitution systems, *Ann. Henri Poincaré* **22** (2021) 1377–1427; `arXiv:2002.04552`.

[EFHN15] T. Eisner, B. Farkas, M. Haase and R. Nagel, *Operator Theoretic Aspects of Ergodic Theory*, GTM 272, Springer, Cham (2015).

[Ens78] V. Enss, Asymptotic completeness for quantum mechanical potential scattering. I. Short range potentials, *Commun. Math. Phys.* **61** (1978) 285–291.

[EMM21] G.B. Escolano, N. Mañibo and E.D. Miro, Mixing properties and entropy bounds of a family of Pisot random substitutions, *preprint* (2021); `arXiv:2103.04866`.

[Fan97] A.H. Fan, Multifractal analysis of infinite products, *J. Stat. Phys.* 86 (1997) 1313–1336.

[FL98] A.H. Fan and K.-S. Lau, Asymptotic behavior of multiperiodic functions $G(x) = \prod_{n=1}^{\infty} g(x/2n)$, *J. Fourier Anal. Appl.* **4** (1998) 129–150.

[FSS] A.H. Fan, J. Schmeling and W. Shen, Generalized Thue–Morse trigonometric polynomials II. Multifractal analysis, *in preparation.*

[FSS21] A.H. Fan, J. Schmeling and W. Shen, L^∞-estimation of generalized Thue–Morse trigonometric polynomials and ergodic maximization, *preprint*; `arXiv:1903.09425`.

[Fek23] M. Fekete, Über die Verteilung der Wurzeln bei gewissen algebraischen Gleichungen mit ganzzahligen Koeffizienten, *Math. Z.* **17** (1923) 228–249.

[FGP20] R.F. Ferreira, S. Gallo and F. Paccaut, Non-regular g-measures and variable length memory chains, *Nonlinearity* **33** (2020) 6026–6052. `arXiv:1907.02442`.

[Gel68] A.O. Gelfond, Sur les nombres qui ont des propriétés additives et multiplicatives données, *Acta Arithm.* **13** (1968) 259–265.

[Gib02] J.W. Gibbs, *Elementary Principles in Statistical Mechanics, developed with especial reference to the rational foundation of thermodynamics*, Charles Scribner's Sons, New York (1902).

[GL90] C. Godrèche and J.M. Luck, Multifractal analysis in reciprocal space and the nature of the Fourier transform of self-similar structures, *J. Phys. A: Math. Gen.* **23** (1990) 3769–3797.

[GL89] C. Godrèche and J.M. Luck, Quasiperiodicity and randomness in tilings of the plane, *J. Stat. Phys.* **55** (1989) 1–28.

[Goh20] P. Gohlke, Inflation word entropy for semi-compatible random substitutions, *Monatsh. Math.* **192** (2020) 93–110; `arXiv:1901.07044`.

[Goh17] P. Gohlke, *On a Family of Semi-Compatible Random Substitutions*, Masters Thesis, Bielefeld University (2017).

[GMR] P. Gohlke, N. Mañibo and D. Rust, Diffraction of primitive random substitutions, *in preparation.*

[GMRS21] P. Gohlke, M. Mitchell, D. Rust and T. Samuel, Measure theoretic entropy of random substitution subshifts, *preprint* (2021); `arXiv:2105.11224`.

[GRS19] P. Gohlke, D. Rust and T. Spindeler, Shifts of finite type and random substitutions, *Discr. Cont. Dyn. Syst.* **9** (2019), 5085–5103; `arXiv:1712.05340`.

[GS20] P. Gohlke and T. Spindeler, Ergodic frequency measures for random substitutions, *Studia Math.* **255** (2020) 265–301; `arXiv:1810.11122`.

[Goo71] T.N.T. Goodman, Relating topological entropy and measure entropy, *Bull. London Math. Soc.* **3** (1971) 176–180.

[Goo72] L.W. Goodwyn, Comparing topological entropy with measure-theoretic entropy, *Amer. J. Math.* **94** (1972) 366–388.

[Gor76] A.Y. Gordon, The point spectrum of the one-dimensional Schrödinger operator, (Russian) *Uspehi Math. Nauk.* **31** (1976) 257–258.

[GH55] W.H. Gottschalk and G.A. Hedlund, *Topological Dynamics*, AMS, Providence, RI (1955).

[GLNS19] R. Grigorchuk, D. Lenz, T. Nagnibeda and D. Sell, Subshifts with leading sequences, uniformity of cocycles and spectra of Schreier graphs, *preprint* (2019); `arXiv:1906.01898`.

[Haj58] J. Hajnal, Weak ergodicity in non-homogeneous Markov chains, *Proc. Cambridge Philos. Soc.* **54** (1958) 233–246.

[Hal74] P. Halmos, *Measure Theory*, GTM 18, Springer, New York (1974).

[HY11] M. Hama and H. Yuasa, Invariant measures for subshifts arising from substitutions of some primitive components, *Hokkaido Math. J.* **40** (2011) 279–312; `arXiv:1003.3364`.

[HA21] C. Haritha and N. Agarwal, On the Perron root and eigenvectors associated with a subshift of finite type, *preprint*; `arXiv:2005.03282`.

[Har55] T.E. Harris, On chains of infinite order, *Pacific J. Math.* **5** (1955) 707–724.

[HKS95] A. Hof, O. Knill and B. Simon, Singular continuous spectrum for palindromic Schrödinger operators, *Commun. Math. Phys.* **174** (1995) 149–159.

[HY12] W. Huang and X. Ye, Generic eigenvalues, generic factors and weak disjointness, in *Dynamical Systems and Group Actions*, eds. L. Bowen, R. Grigorchuk and Y. Vorobets, AMS, Providence, RI (2012) pp. 119–142.

[JK11] J. Jaerisch and M. Kesseböhmer, Regularity of multifractal spectra of conformal iterated function systems, *Trans. Amer. Math. Soc.* **363** (2011) 313–330; `arXiv:0902.2473`.

[JOP12] A. Johansson, A. Öberg and M. Pollicott, Unique Bernoulli g-measures, *J. Eur. Math. Soc.* **14** (2012) 1599–1615. `arXiv:1004.0650`.

[Joh86] R. Johnson, Exponential dichotomy, rotation number, and linear differential operators with bounded coefficients, *J. Differential Equations* **61** (1986) 54–78.

[Kam72] T. Kamae, A topological invariant of substitution minimal sets, *J. Math. Soc. Japan* **24** (1972) 285–306.

[Kea72] M. Keane, Strongly mixing g-measures, *Inv. Math.* **16** (1972) 309–324.

[Kel98] G. Keller, *Equilibrium States in Ergodic Theory*, Cambridge University Press, Cambridge (1998).

[Kel20] G. Keller, Maximal equicontinuous generic factors and weak model sets, *Discr. Contin. Dyn. Syst.* **40** (2020) 6855–6875; `arXiv:1610.03998`.

[KP97] B.A. Khoruzenko and L. Pastur, The localization of surface states: An exactly solvable model, *Phys. Rep.* **288** (1997) 109–126.

[KP03] R. Killip and B. Simon, Sum rules for Jacobi matrices and their applications to spectral theory, *Ann. of Math.* **158** (2003) 253–321; `arXiv:math-ph/0112008`

[Kle14] A. Klenke, *Probability Theory. A Comprehensive Course*, 2nd ed., Springer, London (2014).

[KKT83] M. Kohmoto, L.P. Kadanoff and C. Tang, Localization problem in one dimension: mapping and escape. *Phys. Rev. Lett.* **50** (1983) 1870–1872.

[Kol59] A.N. Kolmogorov, Entropy per unit time as a metric invariant of automorphisms, (Russian) *Dokl. Akad. Nauk SSSR* **124** (1959) 754–755.

[Koo31] B.O. Koopman, Hamiltonian systems and transformations in Hilbert space, *Proc. Natl. Acad. Sci. USA* **17** (1931) 315–318.

[Kos12] D. Koslicki, *Substitution Markov Chains with Applications to Molecular Evolution*, PhD thesis, The Pennsylvania State University (2012).

[KD16] D. Koslicki and M. Denker, Substitution Markov chains and Martin boundaries, *Rocky Mountain J. Math.* **46** (2016) 1963–1985; `arXiv:1705.10830`.

[Led74] F. Ledrappier, Principe variationnel et systèmes dynamiques symboliques, *Z. Wahrscheinlichkeitsth. Verw. Gebiete* **30** (1974) 185–202.

[LSS20] D. Lenz, T. Spindeler and N. Strungaru, Pure point diffraction and mean, Besicovitch and Weyl almost periodicity, *preprint* (2020); `arXiv:2006.10821`.

[Li89] W. Li, Spatial l/f Spectra in open dynamical systems, *Europhys. Lett.* **10** (1989) 395–400.

[LQ11] Q.-H. Liu and Y.-H. Qu, Uniform convergence of Schrödinger cocycles over simple Toeplitz subshift, *Ann. Henri Poincaré* **12** (2011) 153–172.

[LQ12] Q.-H. Liu and Y.-H. Qu, Uniform convergence of Schrödinger cocycles over bounded Toeplitz subshift, *Ann. Henri Poincaré* **13** (2012) 1483–1500.

[LTWW02] Q.-H. Liu, B. Tan Z.-X. Wen and J. Wu, Measure zero spectrum of a class of Schrödinger operators, *J. Stat. Phys.* **106** (2002) 681–691.

[LM95] D. Lind and B. Marcus, *An Introduction to Symbolic Dynamics and Coding*, Cambridge University Press, Cambridge (1995).

[Mah27] K. Mahler, The spectrum of an array and its application to the study of the translation properties of a simple class of arithmetical functions. Part II: On the translation properties of a simple class of arithmetical functions, *J. Math. Massachusetts* **6** (1927) 158–163.

[MT-JU18] C. Maldonado, L. Trejo-Valencia and E. Ugalde, Constant-length random substitutions and Gibbs measures, *J. Stat. Phys.* **171** (2018) 269–287; `arXiv:1712.00586`.

[MR18] G.R. Maloney and D. Rust, Beyond primitivity for one-dimensional substitution subshifts and tiling spaces, *Ergodic Theory Dynam. Systems* **38** (2018) 1086–1117; `arXiv:1604.01246`.

[Mañ87] R. Mañé, *Ergodic Theory and Differentiable Dynamics*, Springer, Berlin (1987).

[Mañ19] N. Mañibo, *Lyapunov Exponents in the Spectral Theory of Primitive Inflation Systems*, PhD thesis, Univ. Bielefeld (2019); available at `urn:nbn:de:0070-pub-29359727`.

[MMRT20] N. Mañibo, E. Miro, D. Rust and G.S. Tadeo, Zeckendorf representations and mixing properties of sequences, *Tsukuba J. Math.* **44** (2020) 251–269; `arXiv:1912.01573`.

[MRW21] N. Mañibo, D. Rust and J.J. Walton, Spectral properties of substitutions on compact alphabets, *preprint* (2021); `arXiv:2108.01762`.

[MRW] N. Mañibo, D. Rust and J.J. Walton, Substitutions on compact alphabets, *in preparation*.

[MRST21] E. Miro, D. Rust, L. Sadun and G.S. Tadeo, Topological Mixing of Random Substitutions, *preprint* (2021); `arXiv:2103.02361`.

[Mol14] M. Moll, Diffraction of random noble means words, *J. Stat. Phys.* **156** (2014) 1221–1236; `arXiv:1404.7411`.

[Mol13] M. Moll, *On a Family of Random Noble Means Substitutions*, PhD thesis, Univ. Bielefeld (2013); available at `urn:nbn:de:hbz:361-26378078`.

[MS17] R.V. Moody, N. Strungaru, Almost periodic measures and their Fourier transforms, in *Aperiodic Order. Vol. 2: Crystallography and Almost Periodicity*, eds. M. Baake and U. Grimm, Cambridge University Press, Cambridge (2017) pp.173–270.

[MH38] M. Morse and G.A. Hedlund, Symbolic dynamics, *Amer. J. Math.* **60** (1938) 815–866.

[Mos92] B. Mossé, Puissances de mots et reconnaissabilité des points fixes d'une substitution, *Theor. Comput. Sci.* **99** (1992) 327–334.

[NS99] M. Neumann and H. Schneider, The convergence of general products of matrices and the weak ergodicity of Markov chains, *Linear Algebra Appl.* **287** (1999) 307–314.

[Nil12] J. Nilsson, On the entropy of a family of random substitutions, *Monatsh. Math.* **168.3** (2012) 563–577; `arXiv:1103.4777`.

[Nil13] J. Nilsson, On the entropy of a two step random Fibonacci substitution, *Entropy* **15** (2013) 3312–3324; `arXiv:1303.2526`.

[OSST17] E.C. Oğuz, J.E.S. Socolar, P.J. Steinhardt and S. Torquato, Hyperuniformity of quasicrystals, *Phys. Rev. B* **95** (2017) 054119:1–10. `arXiv:1612.01975`.

[OSST19] E.C. Oğuz, J.E.S. Socolar, P.J. Steinhardt and S. Torquato, Hyperuniformity and anti-hyperuniformity in one-dimensional substitution tilings, *Acta Cryst. A* **75** (2019) 3–13. `arXiv:1806.10641`.

[Oli99] E. Olivier, Multifractal analysis in symbolic dynamics and distribution of pointwise dimension for g-measures, *Nonlinearity* **12** (1999) 1571–1585.

[Pan84] J.-J. Pansiot, Complexité des facteurs des mots infinis engendrés par morphismes itérés, in *Automata, Languages and Programming*, ed. J. Paredaens, Springer, Berlin (1984) pp. 380–389.

[Par64] W. Parry, Intrinsic Markov chains, *Trans. Amer. Math. Soc.* **112** (1964) 55–66.

[Pes97] Y. Pesin, *Dimension Theory in Dynamical Systems: Contemporary Views and Applications*, Chicago Lectures in Mathematics, University of Chicago Press, Chicago (1997).

[PW01] Y. Pesin and H. Weiss, The multifractal analysis of Birkhoff averages and large deviations, in *Global Analysis of Dynamical Systems*, eds. H.W. Broer, B. Krauskopf and G. Vegter, IoP Publishing, Bristol and Philadelphia (2001) pp. 419–431.

[Pet83] K. Petersen, *Ergodic Theory*, Cambridge University Press, Cambridge (1983).

[Pey80] J. Peyrière, Substitutions aléatoires itérées, *Sémin. Théor. Nombres Bordeaux* **17** (1980–1981), 1–9.

[PF02] N. Pytheas Fogg, *Substitutions in Dynamics, Arithmetics and Combinatorics*, LNM 1794, Springer, Berlin (2002).

[Que10] M. Queffélec, *Substitution Dynamical Systems—Spectral Analysis*, 2nd ed., LNM 1294, Springer, Berlin (2010).

[Que18] M. Queffélec, Questions around the Thue–Morse sequence, *Unif. Distrib. Th.* **13** (2018) 1–25.

[ReS80] M. Reed and B. Simon, *Methods of Modern Mathematical Physics. I. Functional Analysis*, 2nd ed., Academic Press, New York (1980).

[Rem11] C. Remling, The absolutely continuous spectrum of Jacobi matrices, *Ann. of Math.* **174** (2011) 125–171; `arXiv:0706.1101`.

[Roc70] R.T. Rockafellar, *Convex Analysis*, Princeton University Press, Princeton (1970).

[RoS76] G. Rozenberg and A. Salomaa, *The Mathematical Theory of L Systems*, Plenum Press, New York (1976).

[Rud62] W. Rudin, *Fourier Analysis on Groups*, Wiley, New York (1962).

[Rud87] W. Rudin, *Real and Complex Analysis*, 3rd ed., McGraw-Hill Book Co., New York (1987).

[Rue69a] D. Ruelle, A remark on bound states in potential-scattering theory, *Nuovo Cimento A* **61** (1969) 655–662.

[Rue68] D. Ruelle, Statistical mechanics of a one-dimensional lattice gas, *Commun. Math. Phys.* **9** (1968) 267–278.

[Rue69b] D. Ruelle, *Statistical Mechanics: Rigorous Results*, Benjamin, New York (1969).

[Rus20] D. Rust, Periodic points in random substitution subshifts, *Monatsh. Math.* **193** (2020) 683–704; `arXiv:1808.05934`.

[RS18] D. Rust and T. Spindeler, Dynamical systems arising from random substitutions, *Indag. Math.* **29** (2018) 1131–1155; `arXiv:1707.09836`.

[SW77] G. Salinetti and R.J.-B. Wets, On the relations between two types of convergence for convex functions, *J. Math. Anal. Appl.* **60** (1977) 211–226.

[Scha99] H.H. Schaefer, *Banach Lattices and Positive Operators*, Springer, Berlin (1974).

[Schm99] J. Schmeling, On the completeness of multifractal spectra, *Ergodic Theory Dynam. Systems* **19** (1999) 1595–1616.

[Sen81] E. Seneta, *Non-negative Matrices and Markov Chains*, 2nd ed., Springer, New-York (1981).

[SBGC84] D. Shechtman, I. Blech, D. Gratias and J.W. Cahn, Metallic phase with long-range orientational order and no translational symmetry, *Phys. Rev. Lett.* **53** (1984) 1951–1953.

[Sil92] S. Silverman, On maps with dense orbits and the definition of chaos, *Rocky Mountain J. Math.* **22** (1992) 353–375.

[Sim95] B. Simon, Operators with singular continuous spectrum: I. General operators, *Ann. of Math.* **141** (1995) 131–145.

[Sina59] Y.G. Sinai, Flows with finite entropy, (Russian) *Dokl. Akad. Nauk SSSR* **125** (1959) 1200–1202.

[Sing06] B. Sing, *Pisot Substitutions and Beyond*, PhD thesis, Univ. Bielefeld (2006); available at `urn:nbn:de:hbz:361-11555`.

[SS02] V.F. Sirvent and B. Solomyak, Pure discrete spectrum for one-dimensional substitution systems of Pisot type, *Canad. Math. Bull.* **45** (2002) 697–710.

[Sol97] B. Solomyak, Dynamics of self-similar tilings, *Ergodic Theory Dynam. Systems* **17** (1997) 695–738; Erratum: *Ergodic Theory Dynam. Systems* **19** (1999) 1685.

[TS03] S. Torquato and F. Stillinger, Local density fluctuations, hyperuniformity, and order metrics, *Phys. Rev. E* **68** (2003) 041113:1–25 and 069901 (Erratum). `arXiv:cond-mat/0311532`.

[Tyc30] A. Tychonoff, Über die topologische Erweiterung von Räumen, *Math. Ann.* **102** (1930) 544–561.

[Wal82] P. Walters, *An Introduction to Ergodic Theory*, Springer, New York (1982).

[Wal75] P. Walters, Ruelle's operator theorem and g-measures, *Trans. Amer. Math. Soc.* **214** (1975) 375–387.

[Yua07] H. Yuasa, Invariant measures for the subshifts arising from non-primitive substitutions, *J. Anal. Math.* **102** (2007) 143–180.

[Ye92] X. Ye, D-function of a minimal set and an extension of Sharkovskii's theorem to minimal sets, *Ergodic Theory Dynam. Systems* **12** (1992) 365–376.

[ZPK99] M.A. Zaks, A.S. Pikovsky and J. Kurths, On the generalized dimensions for the Fourier spectrum of the Thue–Morse sequence, *J. Phys. A: Math. Gen.* **32** (1999) 1523–1530.

[Zyg02] A. Zygmund, *Trigonometric Series*, combined 3rd ed., Cambridge University Press, Cambridge (2002).