

Farey Matrix Product Measure and the Number Theoretic Spin Chain

Arnab Chakraborty

Department of Mathematics and Statistics, McGill University

April 2024

A thesis submitted to McGill University in partial fulfillment of the
requirements of the degree of Masters of Science.

©Arnab Chakraborty

Abstract

This thesis is centered around the number theoretic spin chain, and its connections to the matrix product measures and the Minkowski question mark function. We first introduce shift spaces and measures defined on them, and we proceed to discuss the basic notions of statistical mechanics that will be used later in the thesis. A discussion of various measures associated to Markov processes follows next, and we introduce matrix product measures. It is shown that matrix product measures are equivalent to the Markov measures defined for Markov processes. We then finally discuss the Farey matrix product measure that can be used to describe the number theoretic spin chain. The entropy functionals Z_n are calculated for this measure, and the limit as n tends to infinity is computed explicitly using some results about continued fraction expansions. The function $Z = \lim_{n \rightarrow \infty} Z_n$ is given in terms of the Conway box function, the inverse of the Minkowski question mark function. The explicit description of the number theoretic spin chain in terms of matrix products is then discussed. We conclude the thesis by discussing further developments and possible connections to other topics. This thesis summarizes some of the results in [\[1\]](#) on the number theoretic spin chain.

Résumé

Ce mémoire s'articule autour des chaînes de spin arithmétiques et leurs connexions avec les mesures produit matriciel et la fonction point d'interrogation de Minkowski. On y présente les espaces de décalage et leurs mesures ainsi que les notions de mécanique statistique requises pour ce mémoire. Par la suite, les mesures associées aux processus de Markov et les mesures produit matriciel sont introduites. On y montre que les mesures produit matriciel sont équivalentes aux mesures de Markov définies pour les processus de Markov. Finalement, nous abordons la mesure produit matriciel de Farey qui permet de décrire les chaînes de spin arithmétiques. Les fonctionnelles d'entropie Z_n sont calculées pour cette mesure, et la limite lors que n tend vers l'infini est calculé explicitement en utilisant des résultats sur les développements en fraction continue. La fonction $Z = \lim_{n \rightarrow \infty} Z_n$ est exprimée en termes de la fonction boîte de Conway et de l'inverse de la fonction point d'interrogation de Minkowski. Une description explicite des chaînes de spin arithmétiques est ensuite donnée en termes de produits matriciels. On conclut le mémoire par des pistes d'approfondissement et de potentielles connexions avec d'autres domaines. Cette thèse résume certains des résultats de [1] sur la chaîne de spins théorique des nombres.

Acknowledgments

I would like to extend my heartfelt gratitude to my supervisors Professor Dmitry Jakobson and Professor Vojkan Jakšić, for their generous support and invaluable guidance throughout my Master's degree in mathematics at McGill University. Professor Jakobson and Jakšić have not only been instrumental in my mathematical development, they have also always been extremely kind and understanding. I would like to thank Professors Henri Darmon, Brent Pym, and Joel Kamnitzer for teaching me many beautiful mathematical concepts. A word of thanks also goes to my thesis examiner for carefully reading the thesis and suggesting many corrections. I am extremely fortunate to have had wonderful friends during my years at McGill, and I would like to acknowledge their support and friendship over the years- I will always cherish the time I have spent with them. Last but not least, I would like to thank my parents for supporting me in innumerable ways throughout my life.

Contents

1	Introduction	5
2	Measures on Shift Spaces	7
2.1	Shift Spaces and Shift Map	7
2.2	Entropy	9
2.3	Entropy functionals Z_n	11
2.4	An Overview of Thermodynamic Formalism	15
3	Hidden Markov Measures and the Matrix Product Measure	19
3.1	Markov Chains and Markov Measure	19
3.2	Hidden Markov Measures on Shift Spaces	21
4	Farey Matrix Product Measure	27
4.1	Farey Matrices and Measure	27
4.2	Entropy Functionals Z_n	32
4.3	Continued Fractions	34
4.4	Minkowski Question Mark Function $?(x)$	39
4.5	Stern Diatomic Sequence and Conway Box Function	45
4.6	Relation between Q_n and $?(x)$	51
5	The Number Theoretic Spin Chain via Matrix Products	56
6	Conclusion	61
	Bibliography	66

1. Introduction

Fruitful interactions between mathematics and physics date back at least to the times of Newton in the 16th century, arguably even before that. Nevertheless, number theory, considered one of the purest branches of mathematics, has often had little to say about physics for most part of this rich history of exchange of ideas between mathematics and physics. However, in the last 50 years, we have seen some very beautiful and deep applications in physics of ideas originating in number theory and vice versa. This thesis is primarily a study of some aspects of one such phenomena. In this thesis, we build on lectures given by Vojkan Jaksic at McGill University on Entropic Information Theory [2] centred around the research program on non-equilibrium statistical mechanics developed by Vojkan Jaksic, Claude-Alain Pillet and Armen Shirikyan[3][4][5][6][7][8][9][10], and study some aspects of the number theoretic spin chain - a one-dimensional spin chain in statistical mechanics with particularly interesting number theoretic properties. In particular, we study its relation to matrix product measures and the Minkowski question mark function $?(x)$, and give an equivalent description of this statistical mechanical model in terms of products of Farey matrices. The treatment in this thesis largely follows the exposition in a forthcoming paper [1] on the number theoretic spin chain by T.Benoist, N.Cueno, D.Jakobson, V. Jaksic and C.A. Pillet.

The thesis is organized as follows. In chapter 2, we first define shift spaces, and define measures on them. We briefly discuss basic ideas in thermodynamics like entropy and ergodicity. We then proceed to discuss examples of g-measures, when we also define what are called 'entropy functionals' Z_n for measures following [2]. We conclude the chapter with a brief discussion of some known results in statistical mechanics.

In chapter 3, we discuss Markov processes, and associate a stochastic matrix to it that helps define the Markov, Hidden Markov and multi-step Markov measures. The rest of the chapter is devoted to discussion of the relation such measures with the matrix product measures that we also define in the chapter. We prove in detail the one-to-one correspondence between a Hidden Markov measures and matrix product measure following the exposition in [10], where the authors investigate its

applications in repeated quantum measurements.

Chapter 4 is the heart of the thesis. We define the Farey matrix product measure, a particular example of matrix product measures defined using only two matrices. We then compute the entropy functionals Z_n for the Farey matrix measure, and show that its limit as $n \rightarrow \infty$ can be written in terms of the inverse of the Minkowski question mark function $?(x)$. We also discuss continued fractions, and the Stern diatomic sequence on the way, which helps us write down this relation explicitly. This chapter is essentially a detailed summary of some results in [1].

In Chapter 5, we discuss the number theoretic spin chain, and show that the Farey matrices used in Chapter 4 to define the Farey matrix product measure can also be used to describe the spin chain model.

To conclude, we briefly discuss further work that can be done on the number theoretic spin chain using results discussed in this thesis, and also comment on possible connections to other subfields of mathematics.

2. Measures on Shift Spaces

In this chapter, we introduce shift spaces and discuss measures associated to shift spaces. We introduce notions of entropy and ergodicity, provide examples of g-measures. We then define the important ‘entropy functionals’ Z_n for such measures following [2]. Finally, we survey some standard results in thermodynamic formalism.

2.1 Shift Spaces and Shift Map

Let $\mathcal{A}$ be a finite set, which we call an alphabet. Elements in $\mathcal{A}$ are called letters. The one-sided shift space over $\mathcal{A}$ is the set of all sequences where every term in the sequence is an element of $\mathcal{A}$. i.e.

$$\Omega := \mathcal{A}^{\mathbb{N}} = \{x = (x_k)_{k \in \mathbb{N}} : x_k \in \mathcal{A} \forall k \in \mathbb{N}\}$$

If we denote the set of letters $x_k, x_{k+1}, \dots, x_n$ by x_k^n , we can define the cylinder $[u]$ to be the sequences such that $x_1^n = u$. We now define $\Omega_{fin} = \cup_{n \in \mathbb{N}} \mathcal{A}^n$ to be the set of word of finite length. Finally, we can define a concatenation on words via $ab = a_1 a_2 \dots a_m b_1 b_2 \dots b_n$.

We will denote the concatenation of u, v via uv , and u^n will be used for concatenation of n copies of u . We will use the usual terminology for the length of an alphabet, which is $|u|$.

The discrete topology on $\mathcal{A}$ naturally endows the space of sequences with a product topology. However, it is possible to metrize the product topology on $\mathcal{A}$ via the distance function

$$d(x, y) = |A|^{1 - \inf\{k : x_1^k \neq y_1^k\}}$$

i.e. we look at the smallest k for which the cylinders up to k are different for two letters, and then exponentiate.

We now define the shift map T on Ω , which will be our central object of study.

Definition 2.1. The shift map T is defined via the equation

$$(Tx_n) := x_{n+1}.$$

T translates each sequence to the left by one element and forgets the first element of the sequence. The invariant measures $\mu_{inv}(\Omega)$ are hence defined to be the set of Borel measures on Ω that satisfy $\mu \circ T^{-1} = \mu$. We have the following results.

Lemma 2.1. T is a continuous surjection.

Proof. Let $x \in \Omega$ and a be any letter in $\mathcal{A}$. Then

$$T(a, x_1, x_2, \dots) = (x_1, x_2, \dots) = x$$

and so $T : \Omega \rightarrow \Omega$ is indeed a surjection. For continuity, we observe that if $x, y \in \Omega$ have the same first k - letters, then $d(x, y) \leq |\mathcal{A}|^{-k}$. Hence, $d(Tx, Ty) \leq |\mathcal{A}|^{-k+1}$. $\square$

Lemma 2.2. Functions that depend on finitely many variables are dense in $C(\Omega)$, where $C(\Omega)$ is the Banach space of continuous real-valued functions on Ω .

Proof. Let $f \in C(\Omega)$ and $\epsilon > 0$. As Ω is compact, f is uniformly continuous and so there exists a $\delta > 0$ such that $\forall x, y \in \Omega$, $|f(x) - f(y)| < \epsilon$, whenever $d(x, y) < \delta$. Now, we choose K such that $|\mathcal{A}|^{1-K} < \delta$ and define $g : \Omega \rightarrow \Omega$ via

$$g(x) = f(x_1, x_2, \dots, x_N, a, a, \dots)$$

where $a \in \mathcal{A}$ is chosen arbitrarily. Now, g here depends on only finitely many symbols and we have for all $x \in \Omega$, $|f(x) - g(x)| < \epsilon$ since $d(x, \bar{x}) < |\mathcal{A}|^{1-(N+1)}$ where $\bar{x} = (x_1, x_2, \dots, x_N, a, a, \dots)$. Hence, we finally have $\|f - g\|_\infty \leq \epsilon$, and so functions depending on finitely many symbols are dense in $C(\Omega)$. $\square$

In this thesis, we will denote the marginals of μ on $\mathcal{A}^n$ by μ_n . i.e. $\mu_n(u) = \mu([u])$ for all $u \in \mathcal{A}^n$.

2.2 Entropy

Definition 2.2. The entropy of the measure $\mu = P_1$ on the finite set $\mathcal{A}$ is defined by the non-negative number

$$S(P_1) = - \sum_{u \in \text{supp } P_1} P_1(u) \ln P_1(u)$$

or equivalently

$$S(P_1) = - \sum_{u \in \mathcal{A}} P_1(u) \ln P_1(u)$$

if we use the convention that $0 \ln 0 = 0$.

The entropy function is non-negative as a sum of non-negative terms (since $P_1(u) \leq 1 \forall u \in \mathcal{A}$), and it is also bounded above. We state some of its properties, following [11]

- $0 \leq S(P_1) \leq \ln |\mathcal{A}|$ (equality is achieved in the case of the Dirac measure and the uniform measure)
- the function $P_1 \rightarrow S(P_1)$ is concave. i.e. $S(\lambda P_1 + (1 - \lambda)P_2) \geq \lambda S(P_1) + (1 - \lambda)S(P_2)$ for any two probability measures P_1, P_2 and any real number $\lambda \in [0, 1]$.
- It satisfies a variational principle i.e. for any function $f : A \rightarrow \mathbb{R}$, and a number β , we have

$$\sup_{P_1 \in \mathcal{P}(A)} (S(P_1) - \beta \sum_{a \in A} f(a)P_1(a)) = \ln \sum_{a \in A} e^{-\beta f(a)}$$

The supremum for this function is achieved when $P_1 \propto \sum_{a \in A} e^{-\beta f(a)} \delta_a$.

Similar results for μ_n hold as $\mathcal{A}^n$ can be thought of a finite set of $|\mathcal{A}|^n$ size. we however have a new property called subadditivity,

$$S(\mu_{n+m}) \leq S(\mu_n) + S(\mu_m).$$

We now define a new quantity, called the specific entropy of P as a limit.

Definition 2.3. The specific entropy of μ is defined via the equation

$$h(\mu) := \lim_{n \rightarrow \infty} \frac{1}{n} S(\mu_n)$$

Lemma 2.3. $0 \leq h(\mu) \leq \ln(\mathcal{A})$.

Proof. Fekete's lemma guarantees existence of $h(\mu_n)$ since the sequence $S(\mu_n)$ is subadditive. We also have that

$$h(\mu) = \lim_{n \rightarrow \infty} \frac{1}{n} S(\mu_n) = \inf_{n \in \mathbb{N}} \frac{S(\mu_n)}{n}$$

Since

$$0 \leq S(\mu_n) \leq \ln |\mathcal{A}^n| = n \ln |\mathcal{A}|,$$

we also get the desired bound. □

We now state an important theorem in the subject, known as *Shannon-McMillan-Breimann* theorem.

Theorem 2.1. If $\mu \in \mu_{inv}(\Omega)$, then the limit

$$h(x) := \lim_{n \rightarrow \infty} \frac{-1}{n} \ln \mu_n(x_1^n)$$

exists for μ -almost all $x \in \Omega$ and satisfies $h(Tx) = h(x)$ for μ -almost all $x \in \Omega$.

The convergence holds in the $L^1(d\mu)$ sense and $\int h d\mu = h(\mu)$.

We do not proof the theorem here, but refer the interested reader to the Masters thesis of Jiacheng Guo, where he gives three different proofs of the result above. We conclude this section by recalling the definition of ergodicity. The reader is referred to [12] for properties of ergodic systems.

Definition 2.4. An element $\mu \in \mu_{inv}(\Omega)$ is ergodic if $\mu(E) \in \{0, 1\}$ whenever $T^{-1}(E) = E$.

2.3 Entropy functionals Z_n

An important quantity to consider when dealing with a probability measure would be the ratio of $\mu_n(x_1^n)$ and $\mu_{n-1}(x_2^n)$. We call them ‘entropy functionals’ following V. Jaksic[2].

Definition 2.5. The entropy functionals Z_n and Y_n associated to the a measure μ are defined via the equations

$$Z_n = \frac{\mu_{n-1}(x_2^n)}{\mu_n(x_1^n)}, \quad Y_n(x) = \frac{\mu_n(x_1^n)}{\mu_{n-1}(x_2^n)}.$$

We have $Z_n = \frac{1}{Y_n}$. The difference between the two is that one is a sub-martingale while the other is a super-martingale. In words, Y_n can be interpreted as the probability of a sequence x starting with x_1 given that he $n - 1$ subsequent letters are $x_2, x_3, \dots, x_n$. Taking $n \rightarrow \infty$, one can interpret $Y(x) = \lim_{n \rightarrow \infty} Y_n(x)$ as the probability of a sequence x starting with x_1 given an infinite tail $x_2, x_3, x_4 \dots$. We compute the functionals Y_n for two simple examples.

- **Bernoulli measures** When $\mu_n = \mu_1^{\otimes n}$ for all $n \in \mathcal{N}$, we have

$$Y_n(x) = \frac{\mu_n(x_1^n)}{\mu_{n-1}(x_2^n)} = \frac{\prod_{k=1}^n \mu_1(x_k)}{\prod_{k=2}^n \mu_1(x_k)} = \mu_1(x_1)$$

We immediately have convergence of Y_n for every $x \in \text{supp } P$ to a function that depends on only one symbol, and we have

$$\mu_n(x_1^n) = e^{\sum_{k=0}^{n-1} \ln Y(T^k x)}$$

- **(Irreducible Markov measures)** A second example are the irreducible Markov measures (to be discussed in greater detail in Section 2). The property of Markov measures we will be using will be that the measures have marginals given by the form

$$\mu_n(x_1^n) = p(x_1) \prod_{k=1}^{n-1} P_{x_k, x_{k+1}}$$

where P is an irreducible stochastic matrix, and p is a left-invariant vector. $pP = p$. We have that

$$Y_n(x) = \frac{\mu_n(x_1^n)}{\mu_{n-1}(x_2^n)} = \frac{p(x_1) \prod_{k=1}^{n-1} P_{x_k, x_{k+1}}}{p(x_1) \prod_{k=2}^{n-1} P_{x_k, x_{k+1}}} = \frac{p(x_1) P(x_1, x_2)}{p(x_2)}$$

for all n . Like in the case for Bernoulli measures we have convergence of Y_n for ever $x \in \text{supp } P$ to a function now depending on only two symbols x_1, x_2 . We have

$$\mu_n(x_1^n) = \mu_1(x_n) e^{\sum_{k=0}^{n-1} \ln Y(T^k x)}.$$

An analysis similar to the Irreducible Markov case holds for the m -step Markov measures where $\mu_n(x_1^n)$ is given by expression of the form

$$p(x_1) P(x_2|x_1) P(x_3|x_1 x_2) \dots P(x_m|x_1 \dots x_{m-1}) P(x_{m+1}|x_1 \dots x_m) \dots P(x_n|x_1 \dots x_n)$$

with $n > m$. In this case also the convergence $Y_n \rightarrow Y$ is very quick, and the limit Y depends on $m + 1$ symbols.

Definition 2.6. A measure μ is a fully supported g-measure if $\text{supp } \mu = \Omega$, and we have the sequences (Y_n) converging to a positive function $Y \in C(\Omega)$.

g-measures are of particular interest because of the following theorem.

Theorem 2.2. If μ is a fully supported g-measure, then $\ln Y_n \rightarrow \ln Y$ and $\ln Z_n \rightarrow \ln Z$ uniformly.

Proof. We assume Y to be everywhere defined as a positive, continuous function. The function $\ln Y$ is then well defined and continuous on Ω . We recall that Y in the

SMB theorem is a priori only defined μ -almost surely. But by assumption, we can take Y to be everywhere defined as a positive, continuous function. The function $\ln Y$ can then well defined and continuous on Ω . Fix $x \in \Omega$ and $n \in N$, otherwise arbitrary. Given $N > n$ and $\epsilon > 0$, define

$$B_{n,\epsilon} := \{u \in \mathcal{A}^N : |\ln Y_{n+N}(x_1^n u) - \ln Y(x)| < \epsilon + \text{var}_n \ln Y\}$$

where var_n is defined via

$$\text{var}_n f := \sup\{|f(x) - f(\eta)| : x, \eta \in \Omega, x_1^n = \eta_1^n\}$$

We have

$$\delta_{n,E} := \sum_{u \notin B_{n,\epsilon}} \mu_{n+N}(x_1^n u)$$

$$\delta_{n,E} \leq \mu\{y \in \Omega : |\ln Y_{n+N}(x_1^n y_1^N) - \ln Y_n(x)| \geq \epsilon + \text{var}_n \ln Y\}$$

We can add and subtract $\ln Y(x_1^n y)$ and use triangle inequality to get

$$\delta_{n,E} \leq \mu\{y \in \Omega : |\ln Y_{n+N}(x_1^n y_1^N) - \ln Y(x_1^n y)| + |\ln Y(x_1^n y) - \ln Y_n(x)| \geq \epsilon + \text{var}_n \ln Y\}$$

we finally have

$$\delta_{n,E} \leq \mu\{y \in \Omega : |\ln Y_{n+N}(x_1^n y_1^N) - \ln Y_n(x_1^n y)| \geq \epsilon\}$$

via the definition of $\text{var}_n \ln Y$. But by the SMB theorem, the last bound converges to 0 as $N \rightarrow \infty$ — almost sure convergence implies convergence in probability. We can say the same for

$$\delta'_{n,E} := \sum_{u \notin B_{n,\epsilon}} \mu_{n+N}(x_2^n u).$$

We have

$$Y_n(x)\mu_{n-1}(x_2^n) = \mu_n(x_1^n) = \sum_{u \in \mathcal{A}^n} \mu_{n+N}(x_1^n u)$$

This can be rewritten as

$$Y_n(x)\mu_{n-1}(x_2^n) = \mu_n(x_1^n) = \sum_{u \in B_{n,\epsilon}} \mu_{n+N}(x_1^n u) + \delta_N$$

and

$$Y_n(x)\mu_{n-1}(x_2^n) = \sum_{u \in B_{n,\epsilon}} Y_{n-1+N}(x_1^n u)\mu_{n-1+N}(x_2^n u) + \delta_N$$

using the expression for $\mu_n(x_1^n)$. Now, we have

$$\frac{Y_n(x)}{Y(x)} - \frac{\delta_N}{Y(x)\mu_{n-1}(x_2^n)} \geq \left(\min_{u \in B_{n,\epsilon}} \frac{Y_{n+N}(x_1^n u)}{Y_n(x)} \right) \frac{\sum_{u \in B_{n,\epsilon}} \mu_{n-1+N}(x_2^n u)}{\mu_{n-1}(x_2^n u)}$$

and

$$\frac{Y_n(x)}{Y(x)} - \frac{\delta_N}{Y(x)\mu_{n-1}(x_2^n)} \leq \left(\max_{u \in B_{n,\epsilon}} \frac{Y_{n+N}(x_1^n u)}{Z_n(x)} \right) \frac{\sum_{u \in B_{n,\epsilon}} \mu_{n-1+N}(x_2^n u)}{\mu_{n-1}(x_2^n u)}$$

Since we have

$$e^{-\epsilon - \text{var}_n \ln Y} \leq \frac{Y_{n+N}(x_1^n u)}{Y(x)} \leq e^{+\epsilon + \text{var}_n \ln Y}$$

for all $u \in B_{n,\epsilon}$. By definition of $B_{n,\epsilon}$, we have that

$$e^{-\epsilon - \text{var}_n \ln Y} \sum_{u \in B_{n,\epsilon}} \frac{\mu_{n+1-N}(x_2^n u)}{\mu_{n-1}(x_2^n)} \leq \frac{Y_n(x)}{Y(x)} - \frac{\delta_N}{Y(x)\mu_{n-1}(x_2^n)} \leq e^{+\epsilon + \text{var}_n \ln Y} \frac{\sum_{u \in B_{n,\epsilon}} \mu_{n+1-N}(x_2^n u)}{\mu_{n-1}(x_2^n)}$$

and therefore also

$$e^{-\epsilon - \text{var}_n \ln Y} \frac{\mu_{n-1}(x_2^n) - \delta'_N}{\mu_{n-1}(x_2^n)} \leq \frac{Y_n}{Y} - \frac{\delta_N}{Y(x)\mu_{n-1}(x_2^n)} \leq e^{+\epsilon + \text{var}_n \ln Y} \frac{\mu_{n-1}(x_2^n) - \delta'_N}{\mu_{n-1}(x_2^n)}$$

Taking $N \rightarrow \infty$, and taking $\epsilon \rightarrow 0$, and taking the logarithm, we get

$$-var_n \ln Y \leq \ln \frac{Y_n(x)}{Y(x)} \leq var_n \ln Y$$

Since x was arbitrary, we have that the convergence is uniform if $\ln Y$ is continuous.

Convergence of $\ln Z_n$ follows immediately. $\square$

Lemma 2.4. If μ is a g -measure, then the function Y satisfies the identity

$$\sum_{a \in \mathcal{A}} Y(ax) = 1$$

for every $x \in \Omega$.

Proof. Let us take any $x \in \Omega$. We have

$$\sum Y_n(ax) = \frac{\sum_a \mu(ax_1^{n-1})}{\mu_{n-1}(x_1^{n-1})} = \frac{\mu_{n-1}(x_1^{n-1})}{\mu_{n-1}(x_1^{n-1})} = 1$$

We can take $n \rightarrow \infty$, and apply pointwise convergence $Y_n(ax) \rightarrow Y(ax)$ for each of the $a \in \mathcal{A}$. Since the alphabet is finite, we have our result directly. $\square$

A positive function is called a *g-function* if it satisfies the identity above.

2.4 An Overview of Thermodynamic Formalism

This chapter is a collection of results in thermodynamic formalism [13] [14], which is a mathematical formalism first established in the works of Gibbs in the 1930s, and then developed by Birkhoff, Sinai, Ruelle and others in the 80s to study equilibrium statistical mechanics. These results are included here just to give the reader an the reader an exposure to some prevalent ideas in the subject, and will not be used later in the thesis.

Theorem 2.3. Given $f \in C(\Omega)$, the limit

$$p_{top}(f) = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{u \in A^n} e^{\sup_{x \in [u]} \sum_{i=0}^{n-1} f(T^i x)}$$

exists and is called the topological pressure of f .

Proof. Let us fix $f \in C(\Omega)$, and let n, m arbitrary. If $u \in A^n$ and $v \in A^m$, we have $uv \in A^{n+m}$, and

$$\begin{aligned} \sup_{x \in [uv]} \sum_{i=0}^{n+m-1} f(T^i x) &\leq \sup_{x \in [uv]} \sum_{i=0}^{n-1} f(T^i x) + \sup_{x \in [uv]} \sum_{i=n}^{n+m-1} f(T^i x) \\ &\leq \sup_{x \in [u]} \sum_{i=0}^{n-1} f(T^i x) + \sup_{x \in T^{-n}[v]} \sum_{i=n}^{n+m-1} f(T^i x) \\ &\leq \sup_{x \in [u]} \sum_{i=0}^{n-1} f(T^i x) + \sup_{y \in [v]} \sum_{j=0}^{m-1} f(T^j y) \end{aligned}$$

If we exponentiate, we get the exponentials as a product on the right hand side.

$$\sum_{w \in A^{n+m}} e^{\sup_{x \in [w]} \sum_{i=0}^{n+m-1} f(T^i x)} \leq \sum_{u \in A^n} e^{\sup_{x \in [u]} \sum_{i=0}^{n-1} f(T^i x)} \sum_{v \in A^m} e^{\sup_{y \in [v]} \sum_{j=0}^{m-1} f(T^j y)}$$

Taking logarithms simply gives

$$\ln \sum_{w \in A^{n+m}} e^{\sup_{x \in [w]} \sum_{i=0}^{n+m-1} f(T^i x)} \leq \ln \sum_{u \in A^n} e^{\sup_{x \in [u]} \sum_{i=0}^{n-1} f(T^i x)} + \ln \sum_{v \in A^m} e^{\sup_{y \in [v]} \sum_{j=0}^{m-1} f(T^j y)}$$

By Fekete's lemma, the limit must exist and hence the topological pressure is well-defined. $\square$

Theorem 2.4. If $f \in C(\Omega)$, then

$$p_{top}(f) \geq \int f d\mu + h(\mu)$$

for every $\mu \in \mu_{inv}$.

Proof. We are only required to prove the result for f depending on finitely many symbols r . We have

$$\sum_{u \in \mathcal{A}^n} e^{\sup_{x \in [u]} \sum_{i=0}^n f(T^i x)} \geq \sum_{u \in \mathcal{A}^n} e^{\sum_{i=0}^{n-r-1} f(T^i x) - r|f|}$$

for $n > r$. We have

$$\frac{1}{n} \sum_{u \in \mathcal{A}^n} e^{\sup_{x \in [u]} \sum_{i=0}^n f(T^i x)} \geq -\frac{r|f|}{n} + \frac{1}{n} \ln \sum_{u \in \mathcal{A}^n} e^{\sum_{i=0}^{n-r-1} f(T^i x)}$$

We can use the variational principle on finite sets to get

$$\begin{aligned} \ln \sum_{u \in \mathcal{A}^n} e^{\sum_{i=0}^{n-r-1} f(T^i x)} &\geq S(\mu_n) + \sum_{u \in \mathcal{A}^n} \sum_{i=0}^{n-r-1} f(T^i u) \mu_n(u) \\ \ln \sum_{u \in \mathcal{A}^n} e^{\sum_{i=0}^{n-r-1} f(T^i x)} &\geq S(\mu_n) + (n-r) \sum_{u \in \mathcal{A}^n} f(u) \mu_r(u) \end{aligned}$$

where we have used μ -invariance in the second line. Finally we can divide throughout by n to get

$$\frac{1}{n} \ln \sum_{u \in \mathcal{A}^n} e^{\sum_{i=0}^n f(T^i x)} \geq -\frac{r|f|}{n} + \frac{1}{n} S(\mu_n) + \frac{n-r}{n} \sum_{u \in \mathcal{A}^n} f(u) \mu_r(u)$$

We can take $n \rightarrow \infty$ to get the desired bound. $\square$

Now we state the Variational principle, which gives us a definition of the topological pressure in terms of invariant measures.

Theorem 2.5. One has the following expression for the topological pressure due to Walters [15].

$$p_{top}(f) = \sup \left\{ \int f d\mu + h(\mu) : \mu \in \mu_{inv}(\Omega) \right\}$$

We now state a theorem following Palmers-Parry-Walters [16] to conclude the chapter.

Theorem 2.6. Suppose that $g \in C(\Omega)$ is a g-function. Let $\mu \in \mu_{inv}(\Omega)$ and Y_n be as above for this measure μ . Then, the following are equivalent:

- $Y_n \rightarrow g$ in the μ -almost sure sense. (i.e. μ is a fully supported g-measure.)
- $Y_n \rightarrow g$ uniformly on Ω .

3. Hidden Markov Measures and the Matrix Product Measure

In this chapter, we discuss Markov processes, and define Markov, Hidden Markov and multi-step Markov measures. The main result in this chapter is the equivalence of Hidden Markov and Matrix Product measures, which follows the exposition in [\[10\]](#).

3.1 Markov Chains and Markov Measure

In mathematics literature, a Markov chain is defined as a stationary process that only depends on the previous state of the system and nothing else. Or, in terms of conditional probability,

$$P[X_{n+1} = a_{n+1} | x_1 = a_1, x_2 = a_2, \dots, x_n = a_n] = P[X_{n+1} = a_{n+1} | x_n = a_n]$$

We first define a stochastic matrix associated to a Markov process, and its corresponding probability vector.

Definition 3.1. A stochastic matrix P is a matrix with entries P_{ij} such that $P_{ij} \geq 0$ for all i, j and that

$$\sum_j P_{ij} = 1.$$

The reason behind this definition is that the entries of the stochastic matrix are used to represent transition probabilities from one state to another in the Markov process. This imposes the condition that the values must all be between 0 and 1 since it is a probability, and that the sum of transition probabilities from one state to all the possible states should add to 1. We now define probability vector p for a stochastic matrix P .

Definition 3.2. A probability vector for a stochastic matrix is a row vector $(p_1, \dots, p_l)$ on R^l such that all its entries are positive, and the entries sum to 1. Given a stochastic matrix, we say a probability vector is an invariant vector of the matrix if

$$pP = p.$$

We show that every stochastic matrix has at least one invariant vector. We take the vector $p = (1/n \ 1/n \ 1/n \ \dots \ 1/n)$ and consider the transpose of the stochastic matrix P . We have

$$\begin{aligned} \left(1/n \ 1/n \ \dots \ 1/n\right) P^T &= \left(1/n \ 1/n \ \dots \ 1/n\right) \begin{pmatrix} P_{11} & P_{21} & \dots & P_{n1} \\ P_{12} & P_{22} & \dots & P_{n2} \\ \dots & \dots & \dots & \dots \\ P_{1n} & P_{2n} & \dots & P_{nn} \end{pmatrix} \\ &= \left(1/n \sum_j P_{1j} \ 1/n \sum_j P_{2j} \ \dots \ 1/n \sum_j P_{nj}\right) = \left(1/n \ 1/n \ \dots \ 1/n\right) \end{aligned}$$

where the last equality holds because P is a stochastic matrix and hence $\sum_j P_{ij} = 1$ for all i . So we have found an eigenvector for P^T with eigenvalue 1. However, P and P^T have the same eigenvalues, so P should also have an eigenvector with eigenvalue 1.

One can use the Perron-Frobenius theorem [17] to show that the invariant vector is unique when the stochastic matrix is irreducible and has all entries that are strictly positive.

Theorem 3.1. (Perron-Frobenius) Let P be an irreducible $n \times n$ matrix with positive entries. Then P has a dominant eigenvalue α of multiplicity one such that $\alpha > 0$ and its associated eigenvector has all positive entries. Furthermore, any other eigenvalue k of P satisfies $|k| \leq \alpha$.

When working with shift spaces, we can define a measure μ on the shift space

A^N to be Markov if the marginals are given by the form

$$\mu_n(a_1, a_2, \dots, a_n) = p_{a_1} P_{a_1, a_2} \dots P_{a_{n-1}, a_n}$$

where we have a probability vector p that is invariant for the stochastic matrix P i.e. $pP = p$. One can think of the condition of marginals as giving the probability of starting at a particular state a_1 , and then transitioning to a_2 via P_{a_1, a_2} and to a_3 via P_{a_2, a_3} , all the way to a_n .

The Markov measure defined in this manner satisfies *Kolmogorov Consistency Theorem*, and the probability measure is shift invariant if and only if p is the invariant vector.

3.2 Hidden Markov Measures on Shift Spaces

We now define Hidden Markov measures on a shift space A^N .

Definition 3.3. We say a measure μ is Hidden Markov if there is an underlying pair (s, S) defining a Markov measure on some other shift space C^N with another stochastic matrix R such that

$$\mu_n(a_1, a_2, \dots, a_n) = \sum_{c_1, c_2, \dots, c_n \in C} s_{c_1} R_{c_1, a_1} S_{c_1, c_2} R_{c_2, a_2} S_{c_2, c_3} \dots S_{c_{n-1}, c_n} R_{c_n, a_n}$$

The matrix R is referred to as the exit matrix, as it quantifies the extent to which the shift space on C^N affects the state on A^N . We now do a first characterization of Hidden Markov measures.

Theorem 3.2. We say a measure μ on A^N is hidden markov iff there exists another Markov measure ν on B^N and a function $f : B \rightarrow A$ such that the marginals are related via

$$\mu_n = \nu_n \circ f_n^{-1} \quad \forall n,$$

where we define $f_n : B^n \rightarrow A^n$ via

$$f_n(b_1, \dots, b_n) = (f(b_1), f(b_2), \dots, f(b_n)).$$

Proof. $\Leftarrow$ Let (q, Q) be the probability vector and stochastic matrix defining ν .

We have that

$$(\nu_n \circ f_n^{-1})(a_1, a_2, \dots, a_n) = \sum_{b_1 \in f^{-1}(a_1), b_2 \in f^{-1}(a_2), \dots, b_n \in f^{-1}(a_n)} \nu_n(b_1, \dots, b_n)$$

Hence

$$(\nu_n \circ f_n^{-1})(a_1, a_2, \dots, a_n) = \sum_{b_1 \in f^{-1}(a_1), b_2 \in f^{-1}(a_2), \dots, b_n \in f^{-1}(a_n)} q_{b_1} Q_{b_1, b_2} \dots Q_{b_{n-1}, b_n}$$

We now define our exit matrix R via

$$R_{b,a} := \begin{cases} 1 & f(b) = a \\ 0 & \text{otherwise} \end{cases}$$

Now we have

$$\nu_n \circ f_n^{-1}(a_1, a_2, \dots, a_n) = \sum_{b_1, b_2, \dots, b_n \in B} q_{b_1} R_{b_1, a_1} Q_{b_1, b_2} R_{b_2, a_2} \dots Q_{b_{n-1}, b_n} R_{b_n, a_n}$$

Hence, we have μ as a Hidden Markov measure with the same underlying Markov chain as ν . $\Rightarrow$ For the other direction, we start with a Hidden Markov measure μ with an underlying chain (s, S) on C with exit matrix R . We have

$$\mu_n = \sum_{c_1, c_2, \dots, b_n \in C} s_{c_1} R_{c_1, a_1} S_{c_1, c_2} R_{c_2, a_2} \dots S_{c_{n-1}, c_n} R_{c_n, a_n}$$

The tricky part is to define the shift space B on which the process is Markov. We set $B = C \times A$, and let

$$q_{c,a} := s_c R_{c,a}$$

and

$$Q_{(c,a),(c',a')} = S_{c,c'} R_{c',a'}$$

It is easy to see we now have a stochastic matrix Q with invariant vector q . if we let

$$f : B \rightarrow A | (c, a) \rightarrow a$$

We have,

$$\begin{aligned} \sum_{b_1 \in f^{-1}(a_1), b_2 \in f^{-1}(a_2), \dots, b_n \in f^{-1}(a_n)} q_{b_1} Q_{b_1, b_2} \dots Q_{b_{n-1}, b_n} = \\ \sum_{c_1, c_2, \dots, c_n \in C} q_{c_1, a_1} Q_{(c_1, a_1), (c_2, a_2)} \dots Q_{(c_{n-1}, a_{n-1}), (c_n, a_n)} = \mu_n(a_1, a_2, \dots, a_n) \end{aligned}$$

□

We now define a new class of measures on shift spaces called Matrix product measures, which are more number-theoretic in flavour.

Definition 3.4. We define a measure to be a Matrix product measure if there is a family of matrices $M_a, a \in A$, a probability row vector p such that

$$\mu_n(a_1, a_2, \dots, a_n) = p M_{a_1} \dots M_{a_n} \mathbf{1}$$

for all n such that $\sum M_a$ is a stochastic matrix, and p is its invariant vector. i.e.

$$p \sum_a M_a = p.$$

It is possible to form a one-to-one correspondence between Matrix product measures and Hidden Markov measures. What this does for us is it gives us another way of

looking at Hidden Markov processes, oftentimes simplifying some of the computations that show up. We provide the correspondence in the proof of the statements below, following [10].

Theorem 3.3. Every Hidden Markov measure μ with underlying chain (s, S) and exit matrix R is equivalent to a Matrix product measure.

Proof. We set

$$[M_a]_{i,j} := R_{i,a} S_{i,j}$$

with $p := s$. To check the equivalence we simply compute the Matrix product measure ν given by this family of matrices, and show it is equal to μ . We have

$$\nu_n(a_1, a_2, \dots, a_n) = p M_{a_1} \dots M_{a_n} \mathbf{1} = \sum_{i_0, i_1, \dots, i_n} p_{i_0} [M_{a_1}]_{i_0, i_1} \dots [M_{a_n}]_{i_{n-1}, i_n}$$

We have

$$\nu_n(a_1, a_2, \dots, a_n) = \sum_{c_1, c_2, \dots, c_n \in C} s_{c_1} R_{c_1, a_1} S_{c_1, c_2} \dots S_{c_{n-1}, c_n} R_{c_n, a_n}$$

Hence, we have

$$\nu_n(a_1, a_2, \dots, a_n) = \mu_n(a_1, a_2, \dots, a_n)$$

and we are good. A subtlety that needs to be checked is that the family of matrices sum to a stochastic matrix. Since R and S are stochastic, the entries of $\sum M$ is nonnegative. We also have

$$\sum_{j=1} \sum_a [M_a]_{i,j} = \sum_j S_{ij} \sum_a R_{i,a} = 1$$

as both sums are equal to 1 since R and S are themselves stochastic matrices. $\square$

To establish the one-to-one correspondence, we need to prove the other direction. The primary difficulty is finding the correct underlying shift space. Setting

$C := \{1, \dots, I\} \times A$ solves the problem for us.

Theorem 3.4. Every matrix product measure ν with matrices M_a and probability vector p is a Hidden Markov measure.

Proof. On $C = \{1, \dots, I\} \times A$, we define the Markov chain (s, S) via

$$S_{(i,a),(i',a')} := [M_{a'}]_{i,i'}$$

Our probability vector p however does not work as the underlying shift space C is much bigger. To solve this, we set

$$s_{(i,a)} = \sum_{j=1}^I p_j [M_a]_{j,i}$$

Our exit matrix R is then just

$$R_{(i,a),a'} = \delta_{a,a'}$$

We now have

$$\begin{aligned} \mu_n &= \sum_{c_1, c_2, \dots, c_n \in C} s_{c_1} R_{c_1, a_1} S_{c_1, c_2} R_{c_2, a_2} \dots S_{c_{n-1}, c_n} R_{c_n, a_n} \\ &\quad \sum_{i_0, i_1, \dots, i_n=1}^I s_{(i_1, a_1)} [M_{a_2}]_{i_1, i_2} \dots [M_{a_n}]_{i_{n-1}, i_n} \\ &= \sum_{i_0, i_1, \dots, i_n=1}^I p_{i_0} [M_{a_2}]_{i_0, i_1} [M_{a_2}]_{i_1, i_2} \dots [M_{a_n}]_{i_{n-1}, i_n} = \nu_n(a_1, a_2, \dots, a_n). \end{aligned}$$

□

We now define a generalization of the Markov process above, called Multistep Markov process. The conditional probability for a Multistep Markov process is given

by

$$\begin{aligned} & P[X_{n+1} = a_{n+1} | x_1 = a_1, x_2 = a_2, \dots, x_n = a_n] \\ &= P[X_{n+1} = a_{n+1} | x_n = a_n, x_{n-1} = a_{n-1}, \dots, x_{n-k} = a_{n-k}] \end{aligned}$$

i.e. the process is dependent on the last k steps only. In terms of marginals, we have a rather complicated formula.

$$\mu_n(a_1, a_2, \dots, a_n) = p_{a_1, a_2, \dots, a_k} \prod_{j=0}^{n-k-1} p_{a_{k+j+1} | a_{1+j}, a_{2+j}, \dots, a_{k+j}}$$

where the numbers are all nonnegative and satisfy

$$\sum_{a_1, a_2, \dots, a_k \in A} p_{a_1, a_2, \dots, a_k} = 1$$

$$\sum_{a_{k+1} \in A} p_{a_{k+1} | a_1, a_2, \dots, a_k} = 1$$

and finally

$$\sum_{a_1 \in A} p_{a_1, a_2, \dots, a_k} p_{a_{k+1} | a_1, a_2, \dots, a_k} = p_{a_2, a_2, \dots, a_{k+1}}$$

The definition of the marginals has a similar interpretation as that in the Markov case, where we start with the $a_1, a_2, \dots, a_k$ and then the transition probabilities are given by $p_{a_{1+j}, a_{2+j}, \dots, a_{k+j} | a_{k+j+1}}$. It can be shown that this multistep measure is also a matrix product measure by performing computations in the basis formed by taking the tensor product $\{e_{a_1, a_2, \dots, a_k} := e_{a_1} \otimes e_{a_2} \otimes \dots \otimes e_{a_k}\}_{a_1, a_2, \dots, a_k \in A}$.

4. Farey Matrix Product Measure

Chapter 4 contains several important results in this thesis related to the Farey Matrix Product measure. The entropy functionals Z_n of this measure are computed following [2]. We then proceed to show that the limit $Z = \lim Z_n$ can be expressed in terms of the Conway Box function. The main result is Theorem 4.8, which helps express the function Q_n (which appears in Section 4.2) in terms of Minkowski question mark function $?(x)$. In our exposition, we follow [2] and [1].

4.1 Farey Matrices and Measure

We now study a particular example of a matrix product measure, defined via Farey matrices, following [2]. It will be later seen that this matrix product measure can also be used to describe a one-dimensional spin chain [1], which we discuss further in the later chapter.

We choose our alphabet to be the simplest alphabet consisting of two letters 0, 1, i.e. $A = \{0, 1\}$, and define the Farey matrices F_0, F_1 via

$$F_0 := \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix} \quad F_1 := \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$$

If we let $M_0 = \frac{1}{3}F_0$ and $M_1 = \frac{1}{3}F_1$, it is easy to see that

$$M_0 + M_1 = \begin{pmatrix} \frac{2}{3} & \frac{1}{3} \\ \frac{1}{3} & \frac{2}{3} \end{pmatrix}$$

defines a stochastic matrix. The invariant probability vector is easily determined to be $p = [\frac{1}{2}, \frac{1}{2}]$. i.e. $pP = p(M_0 + M_1) = p$. Given our stochastic matrix P and a probability vector, we can now define the marginals μ_n of this matrix product measure. We have

$$\mu_n(x_1^n) := \mu_n(x_1, x_2, \dots, x_n) = p M_{x_1} M_{x_2} \dots M_{x_n} \begin{pmatrix} 1 \\ 1 \end{pmatrix}.$$

We can rewrite our M_{x_i} in terms of the corresponding Farey matrices, to get

$$\mu_n(x_1^n) = \frac{1}{2} \frac{1}{3^n} \begin{pmatrix} 1 & 1 \end{pmatrix} F_{x_1} F_{x_2} \dots F_{x_n} \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

We have that for any invariant measure $\mu \in P_1(\Omega)$,

$$\mu([x_1^n]) = \mu_n(x_1, x_2, \dots, x_n) \geq 0$$

and hence μ is mixing and ergodic. To analyze our simple example of the matrix product measure, we now denote

$$F_0 F_{x_1} F_{x_2} \dots F_{x_n} = \begin{pmatrix} r_n(x_1^n) & p_n(x_1^n) \\ s_n(x_1^n) & q_n(x_1^n) \end{pmatrix},$$

and we define a map, which we will call the Q -map associated with each such matrix similar to the Mobius transformation in complex analysis.

Definition 4.1. The Q_n -map associates to every sequence $(x_1, x_2, \dots, x_n) \in \{0, 1\}^n$ and every vector $(a, b) \in \mathbb{R}^2$ the ratio

$$Q_n(x_1^n : a, b) := \begin{pmatrix} r_n(x_1^n) & p_n(x_1^n) \\ s_n(x_1^n) & q_n(x_1^n) \end{pmatrix} \cdot \begin{pmatrix} a \\ b \end{pmatrix} = \frac{ar_n + bp_n}{as_n + bq_n}.$$

We sometimes omit the cylinder (x_1^n) and write $Q_n(a, b)$ instead of $Q_n(x_1^n : a, b)$ for clarity. When $a = b = 1$, we also sometimes write Q_n instead of $Q_n(1, 1)$ if there is no possibility of confusion as to which vector Q_n acts on. The Q_n -function uses the matrix product $F_0 F_{x_1} F_{x_2} \dots F_{x_n}$ associated to each sequence to define the ratio. We now prove some simple results involving the Q_n function.

Lemma 4.1.

$$|Q_n(x_1^n : a, b) - Q_n(x_1^n : 1, 1)| \leq \frac{|a - b|}{\min(a, b)} \frac{1}{n^2}$$

Proof. To prove the result, we simply compute. We have

$$Q_n(a, b) - Q_n(1, 1) = \frac{ar_n + bp_n}{as_n + bq_n} - \frac{r_n + p_n}{s_n + q_n}$$

$$Q_n(a, b) - Q_n(1, 1) = \frac{(r_n q_n - p_n s_n)(a - b)}{(as_n + bq_n)(s_n + q_n)}$$

Now, since our Farey matrices all have determinant 1, we have

$$\det(F_{x_1} F_{x_2} \dots F_{x_n}) = \det F_{x_1} \det F_{x_2} \dots \det F_{x_n} = 1,$$

and hence

$$\det(F_{x_1} F_{x_2} \dots F_{x_n}) = \det \begin{pmatrix} r_n(x_1^n) & p_n(x_1^n) \\ s_n(x_1^n) & q_n(x_1^n) \end{pmatrix} = 1 \implies r_n q_n - s_n p_n = 1.$$

Also, we note that all the entries in our Farey matrices are non-negative, so we have $s_n + q_n \geq n$ via a simple induction. We get

$$|Q_n(a, b) - Q_n(1, 1)| \leq \frac{|a - b|}{\min(a, b)} \frac{|r_n q_n - p_n s_n|}{(s_n + q_n)^2} \leq \frac{|a - b|}{\min(a, b)} \frac{1}{n^2}$$

□

Lemma 4.2. We have that

$$|Q_{n+1} - Q_n| \leq \frac{1}{n^2}$$

and in particular

$$\lim_{n \rightarrow \infty} Q_n(x_1^n) = \lim_{n \rightarrow \infty} Q_n(x).$$

exists in the Banach space of continuous functions $C(\Omega)$. We call the limit $Q(x)$.

Proof. We first note that

$$Q_{n+1}(x_1^{n+1}) = \begin{cases} Q_n(x_1^n : 1, 2) & x_{n+1} = 0 \\ Q_n(x_1^n : 2, 1) & x_{n+1} = 1 \end{cases}$$

We only do the case $x_{n+1} = 0$, the other case is done similarly. We have

$$Q_n(x_1^{n+1} : 1, 1) = \begin{pmatrix} r_n & p_n \\ s_n & q_n \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix} \cdot \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} r_n + p_n & p_n \\ s_n + q_n & q_n \end{pmatrix} \cdot \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \frac{r_n + 2p_n}{s_n + 2q_n},$$

where the last equality follows from computing the action on the vector $(1, 1)$. However, this is equal to $Q_n(x_1^n : 1, 2)$ since

$$\begin{pmatrix} r_n & p_n \\ s_n & q_n \end{pmatrix} \cdot \begin{pmatrix} 1 \\ 2 \end{pmatrix} = \frac{r_n + 2p_n}{s_n + 2q_n}.$$

Since $|a - b|^2 / \min(a, b) = 1/1 = 1$ in both cases, the desired inequality follows by applying Lemma 4.1.

To prove existence of $Q(x)$ we simply show Cauchy convergence of the Q_n s. We have for $n, m \geq N, n \geq m$,

$$\begin{aligned} |Q_n - Q_m| &= |Q_n - Q_{n-1} + Q_{n-1} - Q_{n-2} + \dots + Q_{m+1} - Q_m| \\ &\leq |Q_n - Q_{n-1}| + |Q_{n-1} - Q_{n-2}| + \dots + |Q_{m+1} - Q_m| = \sum_{j=m}^{n-1} \frac{1}{j^2} \end{aligned}$$

using the Lemma above. By the integral test, we have

$$\sum_{j=m}^{n-1} \frac{1}{j^2} \leq \int_{x=m}^{\infty} \frac{1}{x^2} dx = \frac{1}{m} \leq \frac{1}{N}.$$

Hence $\{Q_n\}$ is a Cauchy sequence, and therefore converges. $\square$

Lemma 4.3. Finally,

$$\sup_{x \in \Omega} |Q(x) - Q_n(x)| \leq \frac{1}{n}.$$

Proof. We use the inequality from Lemma 4.2

$$|Q_n - Q_m| \leq \frac{1}{m}$$

Taking $n \rightarrow \infty$ gives us the bound for $|Q - Q_m|$, and hence for its supremum. $\square$

We can use the ratios r_n/s_n and p_n/q_n to give approximations to Q_n and Q_n . In particular, we have the following inequalities.

Lemma 4.4. The ratios r_n/s_n and p_n/q_n satisfy

$$\left| \frac{r_n}{s_n} - Q_n \right| \leq \frac{1}{n}, \quad \left| \frac{r_n}{s_n} - Q \right| \leq \frac{2}{n} \quad s_n \neq 0.$$

$$\left| \frac{p_n}{q_n} - Q_n \right| \leq \frac{1}{n}, \quad \left| \frac{p_n}{q_n} - Q \right| \leq \frac{2}{n}. \quad q_n \neq 0.$$

Proof. The second inequality can easily be proved from the first using the Cauchy convergence of Q_n to Q in $C(\Omega)$. To prove the first, we simply compute

$$\frac{r_n}{s_n} - Q_n = \frac{r_n}{s_n} - \frac{r_n + p_n}{s_n + q_n} = \frac{r_n s_n + r_n q_n - r_n s_n - s_n p_n}{s_n(s_n + q_n)}$$

We get

$$\frac{r_n}{s_n} - Q_n = \frac{r_n q_n - p_n s_n}{s_n(s_n + q_n)} = \frac{1}{s_n(s_n + q_n)}$$

where we used that the determinant of the of the matrix associated to x_1^n is 1 as a product of Farey matrices. Now, as we discussed earlier, $s_n + q_n \geq n$. Also, since

$s_n \neq 0, s_n \in \mathbb{Z}, s_n \geq 1$, and we have

$$\left| \frac{r_n}{s_n} - Q_n \right| \leq \frac{1}{n}.$$

The inequalities for $\frac{p_n}{q_n}$ follow via a similar argument. □

4.2 Entropy Functionals Z_n

In this section, we now compute the entropy functionals Z_n for our measure. We have

$$Z_n(x_1^n) := \frac{\mu_{n-1}(x_2^n)}{\mu_n(x_1^n)}$$

With the expressions for the marginals used earlier, we have

$$Z_n(x_1^n) = \frac{\frac{1}{2} \frac{1}{3^{n-1}} \begin{pmatrix} 1 & 1 \end{pmatrix} F_{x_2} F_{x_3} \dots F_{x_n} \begin{pmatrix} 1 \\ 1 \end{pmatrix}}{\frac{1}{2} \frac{1}{3^n} \begin{pmatrix} 1 & 1 \end{pmatrix} F_{x_1} F_{x_2} \dots F_{x_n} \begin{pmatrix} 1 \\ 1 \end{pmatrix}}$$

We can simplify by introducing factors of F_i in the numerator and denominator. We get

$$Z_n(x_1^n) = 3 \frac{\begin{pmatrix} 1 & 1 \end{pmatrix} F_{x_1}^{-1} F_0^{-1} F_0 F_{x_1} F_{x_2} F_{x_3} \dots F_{x_n} \begin{pmatrix} 1 \\ 1 \end{pmatrix}}{\begin{pmatrix} 1 & 1 \end{pmatrix} F_0^{-1} F_0 F_{x_1} F_{x_2} \dots F_{x_n} \begin{pmatrix} 1 \\ 1 \end{pmatrix}}$$

We can now use our expression for $Q_n(x_1^n : 1, 1)$ to simplify the entropy functionals.

We have $Q_n(x_1^n : 1, 1) = \frac{r_n + p_n}{s_n + q_n}$ and

$$F_0 F_{x_1} F_{x_2} \dots F_{x_n} = \begin{pmatrix} r_n & p_n \\ s_n & q_n \end{pmatrix}$$

so

$$F_0 F_{x_1} F_{x_2} \dots F_{x_n} \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} r_n & p_n \\ s_n & q_n \end{pmatrix} \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

So our numerator becomes,

$$\begin{pmatrix} 1 & 1 \end{pmatrix} F_{x_1}^{-1} F_0^{-1} \begin{pmatrix} r_n & p_n \\ s_n & q_n \end{pmatrix} \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

We see that the numerator can have two different expressions, one corresponding to $x_1 = 0$ and the other to $x_1 = 1$. When $x_1 = 0$, we have

$$\begin{pmatrix} 1 & 1 \end{pmatrix} F_0^{-1} F_0^{-1} \begin{pmatrix} r_n & p_n \\ s_n & q_n \end{pmatrix} \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 1 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ -1 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ -1 & 1 \end{pmatrix} \begin{pmatrix} r_n & p_n \\ s_n & q_n \end{pmatrix} \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

which gives us

$$s + q - (r + p).$$

When $x_1 = 1$, we have

$$\begin{pmatrix} 1 & 1 \end{pmatrix} F_1^{-1} F_0^{-1} \begin{pmatrix} r_n & p_n \\ s_n & q_n \end{pmatrix} \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 1 & 1 \end{pmatrix} \begin{pmatrix} 1 & -1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ -1 & 1 \end{pmatrix} \begin{pmatrix} r_n & p_n \\ s_n & q_n \end{pmatrix} \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

and we get

$$r + p$$

for the numerator. The denominator is the same in both situations. We compute

$$\begin{pmatrix} 1 & 1 \end{pmatrix} F_0^{-1} F_0 F_{x_1} F_{x_2} \dots F_{x_n} \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 1 & 1 \end{pmatrix} F_0^{-1} \begin{pmatrix} r_n & p_n \\ s_n & q_n \end{pmatrix} \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

and we have

$$\begin{pmatrix} 1 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ -1 & 1 \end{pmatrix} \begin{pmatrix} r_n & p_n \\ s_n & q_n \end{pmatrix} \begin{pmatrix} 1 \\ 1 \end{pmatrix} = s_n + q_n$$

So our entropy functionals Z_n are given by

$$Z_n = 3 \frac{s_n + q_n - (r_n + p_n)}{s_n + q_n} = 3(1 - \frac{r_n + p_n}{s_n + q_n}) = 3(1 - Q_n)$$

when $x_1 = 0$, and

$$Z_n = 3Q_n$$

when $x_1 = 1$.

Theorem 4.1. For the Farey matrix product measure, the entropy functionals $Z_n(x_1^n)$ satisfy

$$Z_n = \begin{cases} 3(1 - Q_n) & x_1 = 0 \\ 3Q_n & x_1 = 1 \end{cases}$$

where Q_n here represents the Q_n -map acting on the vector $(1, 1)$.

The calculation above shows why despite the $Q_n(x)$ function first being in a ad-hoc manner, is an important function to consider when investigating the Farey matrix product measure as the entropy functionals Z_n are given in terms of this function. In the next section, we take a brief detour and study continued fractions, and then finally use continued fractions to give a more explicit description of our function the limit of $Q_n(x)$, and hence our entropy functional Z_n for this statistical model.

4.3 Continued Fractions

Let $\{a_n\}_{n \geq 0}$, with $a_n \in \mathbb{Z}$ and $a_n \geq 1$. We denote by $\mathcal{F}$ the collection of all such sequences. We associate to each sequence a continued fraction via

$$[a_0 : a_1, a_2, a_3, \dots] := a_0 + \frac{1}{a_1 + \frac{1}{a_2 + \frac{1}{a_3 + \dots}}}$$

The limit $\epsilon := \lim_{n \rightarrow \infty} [a_0 : a_1, a_2, a_3, \dots, a_n]$ exists and ϵ is an irrational number if the sequence does not terminate [18]. Hence, the map defined by

$$\{a_n\}_{n \geq 0} \in \mathcal{F} \rightarrow \epsilon \in \mathbb{R} \setminus \mathbb{Q}$$

is a bijection. For $n \geq 1$, we define

$$[a_0 : a_1, a_2, a_3, \dots, a_n] = \frac{p'_n}{q'_n}$$

where the fraction p'_n/q'_n is called the n -th convergent. If we take $p'_{-1} = q'_{-1} = 0, p'_0 = a_0, q'_0 = 1$, then we have a recurrence relation relating the n -th convergent to the $(n-1)$ -th and $(n-2)$ -nd convergent. We have

$$\frac{p'_n}{q'_n} = \frac{a_n p'_{n-1} + p'_{n-2}}{a_n q'_{n-1} + q'_{n-2}}$$

This can be expressed in a matrix via

$$\begin{pmatrix} p'_n & p'_{n-1} \\ q'_n & q'_{n-1} \end{pmatrix} = \begin{pmatrix} p'_{n-1} & p'_{n-2} \\ q'_{n-1} & q'_{n-2} \end{pmatrix} \begin{pmatrix} a_n & 1 \\ 1 & 0 \end{pmatrix}.$$

and is proven using a simple induction argument [18]. We now define matrices A_k which will help us define the convergents $\frac{p'_n}{q'_n}$ in a different manner. We let

$$A_{a_k} = \begin{pmatrix} a_k & 1 \\ 1 & 0 \end{pmatrix}$$

which implies that

$$\begin{pmatrix} p'_n \\ q'_n \end{pmatrix} = A_{a_1} A_{a_2} \dots A_{a_n} \begin{pmatrix} 1 \\ 0 \end{pmatrix}$$

This follows from the equation above, as we can express the matrix giving the $(n-1)$ -st convergent in terms of $(n-2)$ -nd and $(n-3)$ rd and A_{n-1} (or a_{n-1})

i.e.

$$\begin{pmatrix} p'_{n-1} & p'_{n-2} \\ q'_{n-1} & q'_{n-2} \end{pmatrix} = \begin{pmatrix} p'_{n-2} & p'_{n-3} \\ q'_{n-2} & q'_{n-3} \end{pmatrix} \begin{pmatrix} a_{n-1} & 1 \\ 1 & 0 \end{pmatrix}$$

We let $M_n = \begin{pmatrix} p'_n & p'_{n-1} \\ q'_n & q'_{n-1} \end{pmatrix}$, and hence the first matrix on the right can be further expressed in terms of $A_{a_{n-2}}$ and M_{n-2} . We have

$$M_{n-1} = M_{n-2} A_{a_{n-2}} A_{a_{n-1}}.$$

If we continue this procedure until we reach A_{a_1} , we get

$$\begin{pmatrix} p'_n & p'_{n-1} \\ q'_n & q'_{n-1} \end{pmatrix} = A_{a_1} A_{a_2} A_{a_3} \dots A_{a_n}$$

Now, to get the vector $(p'_n \ q'_n)^T$, we can multiply the matrices on the right and the right to $(1 \ 0)^T$, and we have

$$\begin{pmatrix} p'_n \\ q'_n \end{pmatrix} = A_{a_1} A_{a_2} A_{a_3} \dots A_{a_n} \begin{pmatrix} 1 \\ 0 \end{pmatrix}$$

Now, we define map x from I , the set of irrational numbers in $[0, 1]$, to the space of sequences on $\Omega = \{0, 1\}^N$. For each $\alpha \in I$, we let

$$\alpha = [a_0 : a_1, a_2, a_3, \dots, a_n, \dots] = [0 : a_1, a_2, a_3, \dots, a_n, \dots]$$

where we have $a_0 = 0$ since $\alpha \in [0, 1]$ and set

$$x_j = 0 \quad 1 \leq j \leq a_1 - 1$$

$$x_j = 1 \quad a_1 \leq j \leq a_2 + a_1 - 1$$

$$x_j = 0 \quad a_2 \leq j \leq a_1 + a_2 + a_3 - 1$$

and so on, *i.e.* the first $a_1 - 1$ terms in the sequence are 0, the next a_2 terms are 1, the next a_3 are again 0, and the next a_4 are 1 and so on. We have

$$x(\alpha) = (\underbrace{0, 0, 0, 0, \dots, 0, 0}_{a_1-1 \text{ times}}, \underbrace{1, 1, 1, \dots, 1, 1}_{a_2 \text{ times}}, \underbrace{0, 0, 0, 0, \dots, 0, 0}_{a_3 \text{ times}}, 1, 1, 1, \dots).$$

Given all this formulation, we are now in a position to prove a property of our Q_n function, that is $Q_n(x(\alpha)) = \alpha$ for all $\alpha \in I$. This property was first established in [2], and we adapt the proof discussed there. We first define a matrix J

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

We obviously have $J^2 = I$, where I is the identity. We consider the expression

$$JF_0F_{x_1}F_{x_2}\dots F_{x_n} \begin{pmatrix} 0 \\ 1 \end{pmatrix}.$$

We first write this matrix product in a different manner using $J^2 = I$,

$$JF_0F_{x_1}F_{x_2}\dots F_{x_n} = (JF_0J)(JF_{x_1}J)(JF_{x_2}J)\dots(JF_{x_n}J)J$$

Our expression above then becomes

$$(JF_0J)(JF_{x_1}J)\dots(JF_{x_n}J)J \begin{pmatrix} 0 \\ 1 \end{pmatrix} = (JF_0J)(JF_{x_1}J)\dots(JF_{x_n}J) \begin{pmatrix} 1 \\ 0 \end{pmatrix}$$

If we now look at the expressions $JF_0JF_{x_1}J\dots JF_{x_{a_0-1}}J = JF_0F_{x_1}\dots F_{x_{a_1-1}}J$, we see we have

$$\left(J \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix} J\right)^{a_1} = \left(J \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix} J\right)^{a_1} = \begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix}^{a_1} = \begin{pmatrix} 1 & 1 \\ a_1 & 1 \end{pmatrix} = A_{a_1}J$$

We hence see that $JF_0JF_{x_1}J\dots JF_{x_{a_1-1}}J = A_{a_1}J$, and similarly we have for the next a_2 terms, we have $JF_{x_{a_1}}F_{x_{a_1+1}}\dots F_{x_{a_2}}J = JA_{a_2}$. In general, we have

$$(JF_0J)(JF_{x_1}J)(JF_{x_2}J)\dots(JF_{x_n}J)J = J(A_{a_1}J)(JA_{a_2})(A_{a_3}J)\dots J(JA_{a_n})J.$$

We again use $J^2 = I$ to get

$$(JF_0J)(JF_{x_1}J)(JF_{x_2}J)\dots(JF_{x_n}J)J = JA_{a_1}A_{a_2}A_{a_3}\dots A_{a_n}J$$

We hence have

$$(JF_0J)(JF_{x_1}J)\dots(JF_{x_n}J)J \begin{pmatrix} 0 \\ 1 \end{pmatrix} = JA_{a_1}A_{a_2}A_{a_3}\dots A_{a_n} \begin{pmatrix} 1 \\ 0 \end{pmatrix} = J \begin{pmatrix} p'_n \\ q'_n \end{pmatrix}$$

We hence finally have

$$J \begin{pmatrix} p'_n \\ q'_n \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} p'_n \\ q'_n \end{pmatrix} = \begin{pmatrix} q'_n \\ p'_n \end{pmatrix}.$$

Now, if we let $t_n = a_1 + a_2 + a_3\dots a_n$, on the other hand, we also have

$$JF_0F_{x_1}F_{x_2}\dots F_{x_n} \begin{pmatrix} 0 \\ 1 \end{pmatrix} = J \begin{pmatrix} r_{t_n} & p_{t_n} \\ s_{t_n} & q_{t_n} \end{pmatrix} \begin{pmatrix} 0 \\ 1 \end{pmatrix} = \begin{pmatrix} s_{t_n} & q_{t_n} \\ r_{t_n} & p_{t_n} \end{pmatrix} \begin{pmatrix} 0 \\ 1 \end{pmatrix} = \begin{pmatrix} q_{t_n} \\ p_{t_n} \end{pmatrix}$$

using the expressions for $F_0F_{x_1}F_{x_2}\dots F_{x_n}$. Since we have two different expressions for the same result, they must be the same, and we have

$$\begin{pmatrix} q_{t_n} \\ p_{t_n} \end{pmatrix} = JF_0F_{x_1}F_{x_2}\dots F_{x_n} \begin{pmatrix} 1 \\ 0 \end{pmatrix} = A_{a_1}A_{a_2}A_{a_3}\dots A_{a_n} \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} q'_n \\ p'_n \end{pmatrix}.$$

Here, the p', q' represents the fraction $\frac{p'_n}{q'_n}$ which gives the $n - th$ convergent of α . The expression on the left, however, gives entries in the matrix expression of the

sequence $x(\alpha)$ via the Farey matrix product. Since we have

$$\frac{q_{t_n}}{p_{t_n}} = \frac{q'_n}{p'_n} \implies \frac{1}{p_{t_n}/q_{t_n}} = \frac{1}{p'_n/q'_n}$$

we can now take the limit to infinity on both sides. The right hand side simply converges to $\frac{1}{\alpha}$, while the left hand side gives us

$$\frac{1}{\lim_{n \rightarrow \infty} p_{t_n}/q_{t_n}} = \frac{1}{Q(x(\alpha))}.$$

We have the important result that

Theorem 4.2. The limit of the Q_n functions, $Q = \lim_{n \rightarrow \infty} Q_n$ satisfies

$$Q(x(\alpha)) = \alpha.$$

In the next section, we define the Minkowski question mark function first defined by Hermann Minkowski in 1904, and then finally express the function Q in relation to this function.

4.4 Minkowski Question Mark Function $?(x)$

The Minkowski question mark function [19] [20] [21] [22], defined on the unit interval, maps *quadratic surds* to irrational numbers in the unit interval. For any number $x \in [0, 1]$, one first takes its continued fraction expansion

$$x = c_0 + \frac{1}{c_1 + \frac{1}{c_2 + \frac{1}{c_3 + \dots}}} := .[c_0 : c_1, c_2, c_3, \dots]$$

Definition 4.2. The Minkowski question mark function is defined in terms of the continued fraction expansion of x . We have

$$?(x) = c_0 + \sum_k (-1)^{k-1} \frac{1}{2^{c_1+c_2+c_3+\dots+c_k-1}}.$$

where $[c_0 : c_1, c_2, c_3, \dots]$ is the continued fraction expansion of x . We first discuss some of its properties.

Firstly $?(0) = 0$ and $?(1) = 1$. This follows immediately from the continued fraction expansions of 0 and 1. It is to be noted that for $x \in (0, 1)$, $c_0 = 0$, and hence

$$?(x) = \sum_k (-1)^{k-1} \frac{1}{2^{c_1+c_2+c_3+\dots+c_k-1}}.$$

We will be using the form above for our calculations since all other numbers are in the open interval $(0, 1)$.

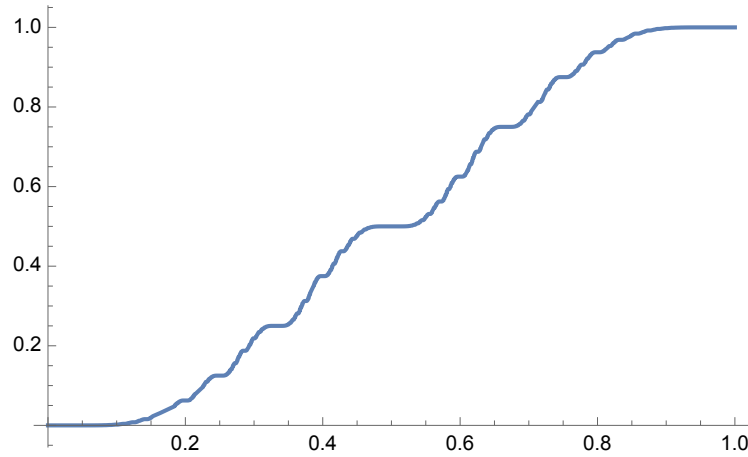


Figure 4.1: Minkowski Question Mark function

Lemma 4.5. If $x \in \mathbb{Q}$ we have that $?(x) = \frac{k}{2^s}$ for some s, k .

Proof. If x is a rational number, then its continued fraction must terminate. This follows from Euclid's algorithms, since we want to write $x = \frac{p}{q} = \frac{1}{q/p} = \frac{1}{1+r/p}$, where r is the remainder of q divided by p . Finding the next terms in the continued fraction expansion then corresponds to determining the next divisor in Euclid's algorithm. The continued fraction expansion of x hence terminates since Euclid's algorithm

terminates at 1. Now, we can write

$$x = [0 : c_1, c_2, \dots, c_l]$$

and hence

$$\begin{aligned} ?(x) &= \sum_{k=0}^{k=l} (-1)^{k-1} \frac{1}{2^{c_1+c_2+c_3+\dots+c_k-1}} \\ &= \frac{(-1)}{2^{c_1}} + \frac{(-1)^2}{2^{c_1+c_2-1}} + \frac{(-1)^3}{2^{c_1+c_2+c_3-1}} + \dots \frac{(-1)^l}{2^{c_1+c_2+c_3+\dots+c_l-1}} \end{aligned}$$

Now, we let $c_1 + c_2 + \dots c_l - 1 = m$, and take the LCM. Certainly, all the terms in the denominator are divisible by 2^m , and hence we have

$$?(x) = \frac{k}{2^m}$$

for some value of k . □

We make a note here that if x is irrational, then its continued fraction expansion is infinite, and we do not always get a rational or an irrational number for $?(x)$. Nevertheless, if x is a quadratic irrational, *i.e.* $x = \frac{a+b\sqrt{d}}{c}$, where d is squarefree, then $?(x)$ is a rational number. To prove this result, we first prove a theorem due to Lagrange that the continued fraction expansion of a quadratic irrational is eventually periodic.

Theorem 4.3. Continued fraction for a quadratic irrational is (eventually periodic)

i.e. $\exists n_0 \geq 0$ and $k \geq 1$ s.t $a_{n+k} = a_k$ where

$$\alpha = [a_0 : a_1, \dots, a_k, \dots]$$

is the continued fractional expansion of x .

Proof. Define $a_n = \lfloor \alpha_n \rfloor (n \geq 0, \alpha_0 = \alpha)$, and

$$\alpha_n = a_n + \frac{1}{\alpha_{n+1}}.$$

Now, since x is a quadratic irrational, there exists a polynomial equation with A, B, C integers such that $Ax^2 + Bx + C = 0$. We now defined for each $n \geq 0$, a polynomial

$$f_n(x) = A_nx^2 + B_nx + C_n$$

with A_n, B_n, C_n integers, $B_n^2 - 4A_nC_n > 0$ squarefree, such that $f_n(\alpha_n) = 0$. We know such a f_n exists as we started with a quadratic surd, and every iteration of the continued fraction expansion gives α_n as a surd. We have

$$f_n(\alpha_n) = 0 \implies A_n\alpha_n^2 + B_n\alpha_n + C_n = 0,$$

We can now multiply everything by α_{n+1}^2 to get

$$\alpha_{n+1}^2(A_n\alpha_n^2 + B_n\alpha_n + C_n) = 0$$

and we use $\alpha_n = a_n + \frac{1}{\alpha_{n+1}}$ to obtain

$$\alpha_{n+1}^2(A_n(a_n + \frac{1}{\alpha_{n+1}})^2 + B_n(a_n + \frac{1}{\alpha_{n+1}}) + C_n) = 0$$

After multiplying throughout by α_{n+1}^2 , we have

$$A_{n+1}\alpha_{n+1}^2 + B_{n+1}\alpha_{n+1} + C_{n+1} = 0$$

with

$$A_{n+1} = \alpha_n^2 A_n + \alpha_n B_n + C_n$$

$$B_{n+1} = 2\alpha_n A_n + B_n$$

and

$$C_{n+1} = A_n.$$

We see that $B_{n+1}^2 - 4A_{n+1}C_{n+1} = B_n^2 - 4A_nC_n := D = B_0^2 - 4A_0C_0$. We now make the remark that the A_n s change sign infinitely. Now, for the sake of contradiction, let $A_n > 0$ for $n \geq N$. Hence, we will have some polynomial $A_n\alpha_n^2 + B_n\alpha_n + C_n$ with all terms positive, and we cannot get a root. Similarly, all the A_n s cannot all be negative eventually. Now, let

$$E = \{n \in \mathbb{Z}^+ : A_nA_{n-1} < 0\}.$$

Since $C_{n-1} = A_n$, we have $A_nC_n < 0$ on E , and hence

$$|B_n| < \sqrt{D}$$

which implies $|A_n| \leq \frac{1}{4}D, |C_n| \leq \frac{1}{4}D$. Now, for $n \in E$, we thus see that there are finitely many distinct polynomials f_n since all the A_n, B_n, C_n s have to be integers and also satisfy the bounds. Hence, two of the α_n s have to be equal i.e. the continued fraction has period n . $\square$

We hence have the result below.

Theorem 4.4. If x is a quadratic irrational, then $?(x)$ is rational.

Proof.

$$?(x) = \sum_k (-1)^{k-1} \frac{1}{2^{c_1+c_2+\dots+c_k-1}} = 2 \sum_k (-1)^{k-1} \frac{1}{2^{c_1+c_2+\dots+c_k}}$$

Now we have

$$?(x) = 2 \left[\frac{1}{2^{c_1}} - \frac{1}{2^{c_1+c_2}} - \frac{1}{2^{c_1+c_2+c_3}} + \dots \frac{1}{2^{c_1+c_2+c_3+\dots+l}} + \dots \right]$$

Now, by Lagrange's theorem, the continued fraction expansion of x is eventually

periodic. Let the period be T . We have

$$\begin{aligned} ?(x) = & 2 \left[\frac{1}{2^{c_1}} \left[1 + \frac{1}{2^{c_1+c_2+\dots+c_T}} + \left(\frac{1}{2^{c_1+c_2+\dots+c_T}} \right)^2 + \left(\frac{1}{2^{c_1+c_2+\dots+c_T}} \right)^3 + \dots \right] \right. \\ & - 2 \left[\frac{1}{2^{c_1+c_2}} \left[1 + \frac{1}{2^{c_1+c_2+\dots+c_T}} + \left(\frac{1}{2^{c_1+c_2+\dots+c_T}} \right)^2 + \left(\frac{1}{2^{c_1+c_2+\dots+c_T}} \right)^3 + \dots \right] \right. \\ & \quad \left. + 2 \left[\frac{1}{2^{c_1+c_2+c_3}} \left[1 + \frac{1}{2^{c_1+c_2+\dots+c_T}} + \left(\frac{1}{2^{c_1+c_2+\dots+c_T}} \right)^2 + \dots \right] \right] \right. \\ & \quad \left. + \dots 2 \left[\frac{1}{2^{c_1+c_2+c_3+\dots+c_k}} \left[1 + \frac{1}{2^{c_1+c_2+\dots+c_T}} + \left(\frac{1}{2^{c_1+c_2+\dots+c_T}} \right)^2 + \dots \right] \right] + \dots \right] \end{aligned}$$

We see that the terms with $1/2^{c_1}$ factor form a geometric series, and so do the terms with $1/2^{c_1+c_2+\dots+c_k}$ factor for any k , and hence they converge to a rational number.

We have a sum of rational numbers, and hence $?(x)$ is rational. $\square$

Infact, this was one of Minkowski's motivations, to define a function that takes quadratic surds to rational numbers, via an expression that relates the continued fraction expansions to binary expansions. Raphael Salem was one of the first people to do a detailed study of the Minkowski question mark function. He proved two interesting results about the function that we mention below, and refer to the original article of Salem[22] for a reader interested in the proofs.

Theorem 4.5. $?(x)$ is strictly increasing.

Theorem 4.6. $?(x)$ is purely singular.

There is another method of defining the Minkowski question mark function in terms of the Farey sequence. We first define the Farey sequence recursively. We have

$$F_1 = \left\{ \frac{0}{1}, \frac{1}{1} \right\}$$

and define an operation on the set of fractions in F_1 via

$$\frac{a}{b} \oplus \frac{c}{d} = \frac{a+c}{b+d},$$

We keep the initial fractions, and add the new fractions obtained via this operation.

We have

$$F_2 = \left\{ \frac{0}{1}, \frac{1}{2}, \frac{1}{1} \right\}.$$

We carry out the same procedure as we go along, and we omit fractions with either $a + c > n$ or $b + d > n$ when determining F_n . We have

$$F_3 = \left\{ \frac{0}{1}, \frac{1}{3}, \frac{1}{2}, \frac{2}{3}, \frac{1}{1} \right\}$$

$$F_4 = \left\{ \frac{0}{1}, \frac{1}{4}, \frac{1}{3}, \frac{1}{2}, \frac{2}{3}, \frac{3}{4}, \frac{1}{1} \right\}$$

$$F_5 = \left\{ \frac{0}{1}, \frac{1}{5}, \frac{1}{4}, \frac{1}{3}, \frac{2}{5}, \frac{1}{2}, \frac{3}{5}, \frac{2}{3}, \frac{3}{4}, \frac{4}{5}, \frac{1}{1} \right\}$$

and so on. This information is expressed beautifully in the Farey tree, where each row gives the n -th Farey fractions F_n . The Farey fractions have very interesting arithmetic and geometric properties, some of which are explored in Allen Hatcher's Topology of Numbers [23].

Now, given the Farey fractions, we can define the Minkowski question mark function as the function that satisfies

$$? \left(\frac{p+p'}{q+q'} \right) = \frac{?(p/q) + ?(p'/q')}{2}.$$

Salem gives a proof of the equivalence of the two definitions in his seminal 1943 paper on the Minkowski question mark function [22].

4.5 Stern Diatomic Sequence and Conway Box Function

In this section we define the Stern diatomic sequence, which we will later use to define the Conway Box function [24].

Definition 4.3. The stern diatomic sequence is a sequence defined recursively via the relations

$$a_{2n} = a_n$$

$$a_{2n+1} = a_n + a_{n+1}$$

with $a_0 = 0$ and $a_1 = 1$.

We can determine a_2 using $a_2 = a_1 = 1$. Similarly, plugging $n = 1$ into the second relation, we get $a_3 = a_1 + a_2 = 2$. a_4 again is given by $a_4 = a_2 = 1$, and finally we obtain $a_5 = a_2 + a_3 = 2 + 1 = 3$. We have the following terms in the sequence hence,

$$0, 1, 1, 2, 1, 2, 3, \dots$$

Again, similar to the Farey tree, we can represent it diagrammatically as a collection of rows, with each row containing $2^{n-1} + 1$ elements. See [25] for details.

Given the Stern diatomic sequence, we define the Conway box function from the dyadic rationals $\mathcal{D}$ to the rationals, $f : \mathcal{D} \rightarrow \mathbb{Q}$. We have

$$\mathcal{D} = \left\{ \frac{k}{2^n} : k, n \in \mathbb{Z}^+, k \leq 2^n \right\}$$

Now the dyadic rationals are dense in $[0, 1]$, and so any strictly increasing function on the dyadic rationals can be extended to a continuous function the whole unit interval.

Definition 4.4. The Conway box function f is a function defined on the dyadic rationals via the equation

$$f\left(\frac{k}{2^n}\right) := \frac{a_k}{a_{2^n+k}}$$

where a_i refers to the i th term in the Stern diatomic sequence.

We first check that f is well-defined. We have

$$f\left(\frac{2k}{2^{n+1}}\right) = \frac{a_{2k}}{a_{2^{n+1}+2k}} = \frac{a_k}{a_{2^n+k}} = f\left(\frac{k}{2^n}\right)$$

where we use the relations for the Stern diatomic tree for the last equality.

Lemma 4.6. f is an increasing function.

Proof. To prove this, we defines matrices

$$M(m, n) := \begin{pmatrix} a_{m+1} & a_{n+1} \\ a_m & a_n \end{pmatrix}$$

We define

$$A_0 := M(0, 1) = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$$

and

$$A_1 = M(1, 0) = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}.$$

We have that

$$A_i M(m, n) = M(2m + i, 2n + i)$$

which can be shown by implying computing the matrix products. i.e.

$$A_0 M(m, n) = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} a_{m+1} & a_{n+1} \\ a_m & a_n \end{pmatrix} = \begin{pmatrix} a_{m+1} + a_m & a_{n+1} + a_n \\ a_m & a_n \end{pmatrix}$$

and we have

$$A_0 M(m, n) = \begin{pmatrix} a_{2m+1} & a_{2n+1} \\ a_m & a_n \end{pmatrix}$$

using the first relation, and

$$A_0 M(m, n) = \begin{pmatrix} a_{2m+1} & a_{2n+1} \\ a_{2m} & a_{2n} \end{pmatrix} = M(2m, 2n)$$

via the second relation. Similarly,

$$A_1 M(m, n) = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} a_{m+1} & a_{n+1} \\ a_m & a_n \end{pmatrix} = \begin{pmatrix} a_{2m+1} & a_{2n+1} \\ a_{2m} & a_{2n} \end{pmatrix}$$

$$A_1 M(m, n) = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} a_{m+1} & a_{n+1} \\ a_m & a_n \end{pmatrix} = \begin{pmatrix} a_{2m+1} & a_{2n+1} \\ a_{2m} & a_{2n} \end{pmatrix} = M(2m+1, 2n+1).$$

Now, if we let $m = \sum i_k 2^k$, we have

$$A_{i_0} A_{i_1} \dots A_{i_n} M(0, 1) = A_{i_0} A_{i_1} \dots A_{i_{n-1}} M(i_n, 2 + i_n)$$

and

$$A_{i_0} A_{i_1} \dots A_{i_n} M(0, 1) = A_{i_0} A_{i_1} \dots A_{i_{n-2}} M(2i_n + i_{n-1}, 4 + 2i_n + i_{n-1}).$$

If we do this further we get

$$A_{i_0} A_{i_1} \dots A_{i_n} M(0, 1) = A_{i_0} A_{i_1} \dots A_{i_{n-3}} M(4i_n + 2i_{n-1} + i_{n-2}, 8 + 4i_n + 2i_{n-1} + i_{n-2}).$$

and eventually

$$M(\sum i_k 2^k, 2^{n+1} + \sum i_k 2^k) = M(m, 2^{n+1} + m)$$

which can be proved by induction. Now, $M(0, 1)$ has determinant 1, and since the A_i s have determinant 1, $M(m, n)$ is a product of matrices of determinant 1, and therefore we also have

$$\det M(k, 2^n + k) = 1$$

$$\begin{vmatrix} a_{k+1} & a_{2^n+k+1} \\ a_k & a_{2^n+k} \end{vmatrix} = 1$$

$$a_{k+1}a_{2^n+k} - a_k a_{2^n+k+1} = 1$$

Now, we have $a_{2^n+k}a_{2^n+k+1} \geq n+1$ from the n -th row of the Stern diatomic sequence. Hence, we have

$$\frac{a_{k+1}}{a_{2^n+k+1}} - \frac{a_k}{a_{2^n+k}} = \frac{1}{a_{2^n+k}a_{2^n+k+1}} \leq \frac{1}{n+1}$$

and we finally have

$$0 < \frac{a_{k+1}}{a_{2^n+k+1}} - \frac{a_k}{a_{2^n+k}} \leq \frac{1}{n+1}$$

and so

$$0 < f\left(\frac{k+1}{2^n}\right) - f\left(\frac{k}{2^n}\right) \leq \frac{1}{n+1}$$

and so f is a strictly increasing function on the dyadic rationals. $\square$

As stated earlier, since $\mathcal{D}$ is dense in $[0, 1]$, f extends to a continuous function from the unit interval to itself.

Now, we call our extended function on the whole unit interval the Conway box function. Since it is a strictly increasing continuous function on the unit interval, it must have an inverse. The important result we prove here following [25] is that the inverse function is the Minkowski question mark function $?(x)$ defined earlier.

Theorem 4.7. The Conway box function and the Minkowski function are inverse functions of each other. i.e.

$$f^{-1}(x) = ?(x).$$

Proof. We need to only prove

$$f\left(\frac{2-x}{2^n}\right) = \frac{1}{n+f(x)}$$

and $f(1) = 1$. This is because

$$?(x) = \frac{2-?(x)}{2^n}$$

as $\frac{1}{n+x}$ has continued fraction expansion $[0 : n, c_1, c_2, \dots]$ if the continued fraction expansion of x is $[0, c_1, c_2, \dots]$. The symmetry of the question mark function mentioned earlier then follows immediately.

We first state two additional results related to the Stern diatomic sequence.

$$(i) \quad a_{2^n+k} = a_{2^{n+1}-k} \quad (\text{pallindrome})$$

and

$$(ii) \quad a_{2^n+k} = a_k + a_{2^n-k},$$

both of which are easy to prove via induction using the defining relations. Now, we determine $a_{2^{m+1}+2^{m+1}-k}$, as we use it later. We get

$$a_{2^{m+1}+2^{m+1}-k} = a_{2^{m+2}-k} = a_{2^{m+1}-k}$$

using the pallindrome property for the last equality. We have

$$a_{2^{m+1}+2^{m+1}-k} = a_k + na_{2^{m+1}-k}$$

which again follows via an induction argument. Finally we use the pallindrome property again to conclude

$$a_{2^{m+1}+2^{m+1}-k} = a_k + na_{2^m+k}.$$

Now, given $x = \frac{k}{2^m}$, we compute $f(\frac{2-x}{2^n})$. We get

$$f(\frac{2-x}{2^n}) = f(\frac{2-\frac{k}{2^m}}{2^n}) = f(\frac{2^{m+1}-k}{2^{m+n}}) = \frac{a_{2^{m+1}-k}}{a_{2^{m+n}+2^{m+1}-k}}$$

We have

$$a_{2^{m+1}-k} = a_{2^{m+1}+k}$$

and also

$$\frac{a_{2^{m+1}-k}}{a_{2^{m+n}+2^{m+1}-k}} = a_k + na_{2^m+k},$$

so

$$f(\frac{2-x}{2^n}) = \frac{a_{2^{m+1}+k}}{a_k + na_{2^m+k}}$$

where if we divide by the $a_{2^{m+1}+k}$ and use the definition of f , we get

$$f(\frac{2-x}{2^n}) = \frac{1}{n + f(\frac{k}{2^m})} = \frac{1}{n + f(x)}.$$

Since we have proved the symmetry property for the dyadic rationals which are dense in the unit interval, and it is obvious that $f(1) = 1$, it follows that f is the inverse function of the Minkowski question function. $\square$

4.6 Relation between Q_n and $?(x)$

In this section, we would first like to prove a theorem from [1].

Theorem 4.8.

$$?(\alpha) = \eta^{-1} \circ x(\alpha)$$

where x associates to any real number $\alpha = [a_1, a_2, a_3, ..a_n...]$, the sequence

$$\underbrace{(0, 0, 0, 0, ...0, 0)}_{a_1-1 \text{ times}}, \underbrace{(1, 1, 1, ...1, 1)}_{a_2 \text{ times}}, \underbrace{(0, 0, 0, 0, ...0, 0)}_{a_3 \text{ times}}, (1, 1, 1, ...) \in \Omega = \{0, 1\}^{\mathcal{N}},$$

$[a_1, a_2, a_3, ..a_n...]$ gives the continued fraction expansion of α , i.e.

$$[a_0 : a_1, a_2, a_3, ...] := a_0 + \frac{1}{a_1 + \frac{1}{a_2 + \frac{1}{a_3 + \dots}}}$$

and $\eta^{-1} : \Omega \rightarrow (0, 1)$ takes a sequence $b_1 b_2 b_3 b_4 \in \Omega$ to $\sum_n b_n 2^{-n}$. Here, $?(x)$ is Minkowski question mark function, given by

$$?(x) = \sum_{k \geq 1} \frac{(-1)^{k-1}}{2^{a_1 + a_2 + \dots + a_k - 1}}$$

where again $[a_1, a_2, a_3, ...]$ is the continued fraction expansion of α .

We first prove a simple result involving geometric series.

Lemma 4.7. For $a, b \in \mathbb{Z}$, we have

$$2^{-a} + 2^{-a-1} + 2^{-a-2} + 2^{-a+3} + ... + 2^{-a-b+1} = \frac{1}{2^{a-1}} - \frac{1}{2^{a+b-1}}.$$

Proof. To prove this, we first factor out 2^{-a+1} from the LHS, and then use the formula for the sum of the geometric series up to n -terms, $S_n = \frac{a(1-r^n)}{1-r}$. We have

$$2^{-a+1} \left(\frac{1}{2} + \frac{1}{2^2} + \frac{1}{2^3} + ... \frac{1}{2^b} \right) = 2^{-a+1} \left(\frac{\frac{1}{2}(1 - (\frac{1}{2})^b)}{1 - \frac{1}{2}} \right) = 2^{-a+1} \left(1 - (\frac{1}{2})^b \right) = 2^{-a+1} \left(1 - \frac{1}{2^b} \right)$$

If we multiply out, we get

$$\frac{1}{2^{a-1}} \left(1 - \frac{1}{2^b} \right) = \frac{1}{2^{a-1}} - \frac{1}{2^{a+b-1}}.$$

□

This simply identity would be useful to prove the theorem we are interested in.

Proof of Theorem 4.8. For any real number $\alpha \in (0, 1)$ we compute $\eta^{-1} \circ x(\alpha)$. First, if the continued fraction expansion of α is given by $[a_1, a_2, a_3, \dots, a_n, \dots]$, we have the sequence $x(\alpha) \in \Omega$ is given by

$$\underbrace{(0, 0, 0, 0, \dots, 0, 0)}_{a_1-1 \text{ times}}, \underbrace{(1, 1, 1, \dots, 1, 1)}_{a_2 \text{ times}}, \underbrace{(0, 0, 0, 0, \dots, 0, 0)}_{a_3 \text{ times}}, (1, 1, 1, \dots) \in \Omega = \{0, 1\}^{\mathbb{N}}$$

Since the first $a_1 - 1$ terms are 0, when we compute $\eta^{-1}(x(\alpha))$, the first term in the sum would be $= 2^{-a_1} = \frac{1}{2^a}$. However, the next $a_2 - 1$ terms in the sequence are all 1, so from the first $a_1 + a_2 - 1$ terms in the sequence $x(\alpha)$, we have the contribution

$$2^{-a_1} + 2^{-a_1-1} + 2^{-a_1-2} + 2^{-a_1-3} + \dots + 2^{-a_1-a_2+1} = \frac{1}{2^{a_1-1}} - \frac{1}{2^{a_1+a_2-1}}.$$

Now, the next a_3 terms in the sequence are all 0, so they contribute nothing to the sum $\eta^{-1}(x(\alpha))$. However, we again have a_4 1s now, so their contribution to the sum is given by

$$\begin{aligned} & 2^{-(a_1+a_2+a_3)} + 2^{-(a_1+a_2+a_3+1)} + \dots 2^{-(a_1+a_2+a_3+\dots+a_4-1)} \\ &= 2^{-a_1-a_2} \left(2^{-a_3} + 2^{-a_3-1} + 2^{-a_3-2} + 2^{-a_3-3} + \dots + 2^{-a_3-a_4+1} \right) \end{aligned}$$

Again, we use our identity to get

$$2^{-a_1-a_2} \left(\frac{1}{2^{a_3-1}} - \frac{1}{2^{a_3+a_4-1}} \right) = \frac{1}{2^{a_1+a_2+a_3-1}} - \frac{1}{2^{a_1+a_2+a_3+a_4-1}}$$

The next a_5 terms are again all 0, and the a_6 terms after that are all 1s. So, these

terms contribute

$$2^{-(a_1+a_2+a_3+a_4+a_5)} + 2^{-(a_1+a_2+a_3+a_4+a_5+1)} + 2^{-(a_1+a_2+a_3+a_4+a_5+2)} + \dots + 2^{-(a_1+a_2+a_3+a_4+a_5+a_6+1)}$$

Again, we can factor out $2^{-a_1-a_2-a_3-a_4}$, and use an argument similar to the one above to conclude that the sum above is equal to

$$\frac{1}{2^{a_1+a_2+a_3+a_4+a_5-1}} - \frac{1}{2^{a_1+a_2+a_3+a_4+a_5+a_6-1}}.$$

Now, the pattern is somewhat clear as to how the terms in the expansion of $\eta^{-1}(x(\alpha))$ look. Since for our sequence we have $a_1 = 1$ 0s, followed by $a_2 = 1$ 1s, followed by $a_3 = 1$ 0s, followed by $a_4 = 1$ 1s, followed by $a_5 = 1$ 0s and so on we have $\eta^{-1} \circ x(\alpha)$ is given by

$$\begin{aligned} & \frac{1}{2^{a_1-1}} - \frac{1}{2^{a_1+a_2-1}} + \frac{1}{2^{a_1+a_2+a_3-1}} - \frac{1}{2^{a_1+a_2+a_3+a_4-1}} + \frac{1}{2^{a_1+a_2+a_3+a_4+a_5-1}} - \frac{1}{2^{a_1+a_2+a_3+a_4+a_5+a_6-1}} \\ & + \frac{1}{2^{a_1+a_2+a_3+a_4+a_5+a_6+a_7-1}} - \frac{1}{2^{a_1+a_2+a_3+a_4+a_5+a_6+a_7+a_8-1}} + \dots \\ & + \dots \dots \frac{1}{2^{a_1+a_2+a_3+a_4+a_5+a_6+\dots+a_k-1}} - \frac{1}{2^{a_1+a_2+a_3+a_4+a_5+a_6+\dots+a_k+a_{k+1}-1}} + \dots \end{aligned}$$

We can write this sum more compactly as

$$\sum_{k \geq 1} \frac{(-1)^{k-1}}{2^{a_1+a_2+a_3+\dots+a_k-1}}$$

However, this is exactly $?(\alpha)$, and since α was arbitrary, we just showed that $?(\alpha) = \eta^{-1} \circ x(\alpha)$ for all $\alpha \in I$ (where I the set of irrational numbers in $(0, 1)$ is a dense subset of $(0, 1)$). We therefore have that

$$\eta^{-1} \circ x(\alpha) = ?(\alpha)$$

for all $\alpha \in (0, 1)$, and since both sides are equal to 0 when $\alpha = 0$, and 1 when $\alpha = 1$, we also have they agree on $[0, 1]$. $\square$

As discussed earlier, since our $?(\alpha)$ is an increasing continuous function, it has a continuous inverse (often called the Conway box function) and we have

$$(?^{-1} \circ \eta^{-1}) \circ x(\alpha) = \alpha \implies (\eta \circ ?)^{-1} x(\alpha) = \alpha.$$

However, $Q(x(\alpha)) = \alpha$ too, and so we have that $Q = (\eta \circ ?)^{-1}$, i.e. our Q function is the inverse of the Minkowski question mark function after the identification of the space of sequences in $\Omega = \{0, 1\}^{\mathbb{N}}$ with the open interval $(0, 1)$ via the η function. Given our results in the last chapter involving $Z = \lim_{n \rightarrow \infty} Z_n$, we now have

Theorem 4.9. The function $Z = \lim_{n \rightarrow \infty} Z_n$ satisfies

$$Z : \Omega \rightarrow \mathbb{R}, \quad Z = \begin{cases} 3(1 - (\eta \circ ?)^{-1}) & x_1 = 0 \\ 3(\eta \circ ?)^{-1} & x_1 = 1. \end{cases}$$

where $?(x)$ is the Minkowski question mark function.

5. The Number Theoretic Spin Chain via Matrix Products

In this chapter, we give a brief introduction to the number theoretic spin chain discussed in [1],[26],[27],[28],[29],[30], and discuss how the Farey matrix product measure can be used to give a physically equivalent description of the spin chain.

We consider a one-dimensional lattice, and assign a *spin* to each point of this lattice. The individual spins ω_i are allowed to have values either 0 or 1 i.e. spin down or spin up, and the configuration of the spin chain is given via the sequence $\omega = (\omega_1, \omega_2, \omega_3, \dots, \omega_i, \omega_{i+1}, \dots)$, where we allow the length of our spin chain to be infinite if required. We make contact with our previous notation by defining $\Omega_N = X^N = \{0, 1\}^N$, where $\sigma_i \in X$, for a spin chain of length N .

thermodynamic quantity of interest is the partition function $Z_N(\beta)$ of the system. We only define the canonical partition function for in this thesis. We first define a sequence of functions $h_N : \Omega_N \rightarrow \mathcal{N}$ inductively via

$$h_1(0) = 0, h_1(1) = 2$$

and

$$h_{N+1}(\omega, 0) = h_N(\omega)$$

$$h_{N+1}(\omega, 1) = h_N(\omega) + h_N(\bar{\omega})$$

where $\bar{\omega} = (1 - \omega_1, 1 - \omega_2, 1 - \omega_3, \dots, 1 - \omega_N) \bmod 2$ i.e. all spins are reversed. These spin chains were first defined by Andreas Knauf in a 1993 paper [26] [27] [28] [29] [30]. In his paper, he first associated to each spin configuration ω an element in the cyclic group of order 2^N via the expansion $f : \Omega \rightarrow G_{2^N}$

$$f(\omega_1, \omega_2, \dots, \omega_N) = \sum_{i=1}^N \omega_i 2^{N-i}.$$

The functions h_N were then defined on the elements of the group G_{2^N} , in which case, h_k simply turns out to be the elements in Stern-Diatomic sequence at the k -th row i.e. $h_3(0) = 1, h_3(1) = 4, h_3(2) = 3, h_3(3) = 5, h_3(4) = 2, \dots$. With our definition of h_N , we are now in a position to define the partition function $\mathcal{Z}_N(\beta)$. We first define the Hamiltonian H_N that represents the energy of each spin configuration.

$$H_N(\omega) = \log h_N(\omega)$$

Definition 5.1. The partition function $\mathcal{Z}(\beta)$ in statistical mechanics is given by summing $\sum_s e^{-\beta H(s)}$ where s represents the possible states of a system.

In this case, we have

$$\mathcal{Z}_N(\beta) = \sum_{\omega \in \Omega_N} e^{-\beta H_N(\omega)} = \sum_{\omega \in \Omega_N} [h_N(\omega)]^{-\beta}$$

It is easy to see, that if we let the length of our spin chain to be infinite, the expression we obtain for $\mathcal{Z}(\beta)$ resembles that of a zeta function in number theory. In fact, for this model, Knauf computed the partition function and found that

$$\lim_{N \rightarrow \infty} \mathcal{Z}_N(\beta) = \sum_{n=1}^{\infty} \frac{\phi(n)}{n^\beta}$$

where $\phi(n)$ is the Euler totient function. The sum above is a common Dirichlet series encountered in number theory, and one has

$$\sum_{n=1}^{\infty} \frac{\phi(n)}{n^\beta} = \frac{\zeta(\beta-1)}{\zeta(\beta)}.$$

Theorem 5.1. The partition function of the infinite number theoretic spin chain is given as the equal to the ratio of Riemann zeta functions evaluated at $\beta-1$ and β [28].

$$\mathcal{Z}_\infty(\beta) = \lim_{N \rightarrow \infty} \mathcal{Z}_N(\beta) = \frac{\zeta(\beta-1)}{\zeta(\beta)}.$$

The result above probably justifies the name 'number theoretic spin chain' for

this statistical model. A reason why there was a lot of initial excitement around this result was because it was hoped that the Lee-Yang theorem [31] [32] [33] [34] could be used to investigate the zeroes of the partition function $\mathcal{Z}_N(\beta)$, and that could lead to new base on the zeroes of the Riemann zeta function.

Now, we change gears, and see how the Farey matrices can be used to study the spin chain. For a spin configuration $\omega = (\omega_1, \omega_2, \dots, \omega_n) \in \Omega_N$, we define the associated matrix

$$M(\omega) = \begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix} F_{\omega_n} F_{\omega_{n-1}} \dots F_{\omega_1}$$

and then $h_N(\omega)$ is simply given by the trace of this matrix $M(\omega)$. We hence have

$$\mathcal{Z}_N(\beta) = \sum_{\omega \in \Omega_N} e^{-\beta(\log \text{tr}(M(\omega)))} = \sum_{\omega \in \Omega_N} (\text{tr}(M(\omega)))^{-\beta}.$$

To prove this equivalence, we first show $h_1(M(0)) = 1$ and $h_1(M(1)) = 2$. This is easy since

$$M(0) = \begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix} F_0 = \begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix}$$

and hence $\text{tr}(M(0)) = 1$. Similarly,

$$M(1) = \begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix} F_1 = \begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}$$

and $\text{tr}(M(1)) = 2$. In general, if

$$F_{\omega_n} F_{\omega_{n-1}} \dots F_{\omega_1} = \begin{pmatrix} a & b \\ c & d \end{pmatrix},$$

we have $\text{tr}(M(\omega)) = a + b$. We use this expression prove the other two recurrence relations for the Hamiltonian generating functions h_N . For the first one,

$$F_0 F_{\omega_n} F_{\omega_{n-1}} \dots F_{\omega_1} = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} a & b \\ a+c & b+d \end{pmatrix},$$

and therefore $\text{tr}(M(\omega, 0)) = a + b = \text{tr}(M(\omega))$. For the second relation, we first

compute $\text{tr}(M(\bar{\omega}))$. Now,

$$M(\bar{\omega}) = \begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix} F_{\omega_n}^T F_{\omega_{n-1}}^T \dots F_{\omega_1}^T = \begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix} (F_{\omega_1} F_{\omega_2} \dots F_{\omega_n})^T$$

We want to show

$$F_{\omega_1} F_{\omega_2} \dots F_{\omega_n} := \begin{pmatrix} a' & b' \\ c' & d' \end{pmatrix} = \begin{pmatrix} d & b \\ c & a \end{pmatrix}.$$

We prove this by induction. Now, since both Farey matrices F_0 and F_1 have 1 on the diagonals, the formula above is true $n = 1$. Now, let for us assume for a product of n Farey matrices

$$F_{\omega_n} F_{\omega_{n-1}} \dots F_{\omega_1} = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \implies F_{\omega_1} F_{\omega_2} \dots F_{\omega_n} = \begin{pmatrix} d & b \\ c & a \end{pmatrix}$$

Now, we take a product of $n + 1$ Farey matrices such that

$$F_{\omega_{n+1}} F_{\omega_n} F_{\omega_{n-1}} \dots F_{\omega_1} = \begin{pmatrix} p & q \\ r & s \end{pmatrix}$$

Depending on whether $F_{\omega_{n+1}} = F_0$ or F_1 , we can have expressions for p, q, r, s in terms of a, b, c, d . If $F_{\omega_{n+1}} = F_0$, we have

$$\begin{pmatrix} p & q \\ r & s \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} a & b \\ a + c & b + d \end{pmatrix}.$$

Now, we want to compute $F_{\omega_1} F_{\omega_2} \dots F_{\omega_n} F_{\omega_{n+1}}$. We have

$$F_{\omega_1} F_{\omega_2} \dots F_{\omega_n} F_{\omega_{n+1}} = \begin{pmatrix} d & b \\ c & a \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix} = \begin{pmatrix} d + b & b \\ c + a & a \end{pmatrix} = \begin{pmatrix} s & q \\ r & p \end{pmatrix}.$$

where we use the induction hypothesis to write the product in terms of a, b, c, d .

Similarly, if $F_{\omega_{n+1}} = F_1$,

$$\begin{pmatrix} p & q \\ r & s \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} a+c & b+d \\ c & d \end{pmatrix}.$$

and we again have

$$F_{\omega_1} F_{\omega_2} \dots F_{\omega_n} F_{\omega_{n+1}} = \begin{pmatrix} d & b \\ c & a \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} d & d+b \\ c & c+a \end{pmatrix} = \begin{pmatrix} s & q \\ r & p \end{pmatrix}.$$

So,

$$M(\bar{\omega}) = \begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} d & b \\ c & a \end{pmatrix}^T = \begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} d & c \\ b & a \end{pmatrix}$$

and $tr(M(\bar{\omega})) = c + d$. Finally,

$$F_1 F_{\omega_n} F_{\omega_{n-1}} \dots F_{\omega_1} = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} a+c & b+d \\ c & d \end{pmatrix}$$

and so $tr(M(\omega, 1)) = a + c + b + d = (a + b) + (c + d) = tr(M(\omega)) + tr(M(\bar{\omega}))$. Since these matrices now satisfy the same relations as the $h_N(\omega)$, we have a *physically* equivalent description of the number theoretic spin chain via matrix products of Farey matrices.

Given this new description of the number theoretic spin chain, it would be interesting to ask what new properties of the spin chain may be discovered via this description, and if some of the properties discovered earlier have a more natural explanation.

6. Conclusion

In this thesis, we studied extensively a particular example of matrix product measure (the Farey matrix product measure), and used it to give an equivalent description of the number theoretic spin chain via matrix products, largely following the exposition in [1]. The number theoretic spin chain is an interesting object on the crossroad of statistical mechanics and number theory, and there has been a lot of research on this topic in the last 30 years studying it from both perspectives. The initial work on the subject focused on studying many of its thermodynamic properties such as its ferromagnetism, magnetization, and phase transitions, and computing various thermodynamic quantities of interest such as its partition function, pressure and free energy. Given the explicit description of the limit of the entropy functionals Z_n of the Farey matrix product measure in terms of the Minkowski question mark function $?(x)$ studied in chapter 2 of this thesis, it would be interesting to see if the thermodynamic properties mentioned above can be derived from this explicit formula. It is believed that the unusual fractal properties of the Minkowski question mark function would also give interesting base into the fractal and chaotic properties of the spin chain defined in this thesis and perhaps for a more general class of spin chains defined similarly. As for the thermodynamic quantities calculated for this system, there seems to be a shadow of modularity in some of the calculations. For case in point, the free energy $\mathcal{F}(\beta)$ of the spin chain was calculated to be given by

$$\mathcal{F}(\beta) = \lim_{N \rightarrow \infty} \ln(\mathcal{Z}_N(\beta)) = \frac{-\ln(\lambda(\beta))}{\beta}$$

for $0 \leq \beta < 2$ (there exists a phase transition at $\beta = 2$) where λ is given by the largest eigenvalue of the Lewis-three term functional equation

$$\lambda\psi(x) = \psi(x+1) + x^{-\beta}\psi(1 + \frac{1}{x})$$

in the theory of modular forms. In an interesting paper [35], J Lewis and Don Zagier and defined a one-to-one *period correspondence* between eigenvalues of the Lewis three-term equation and *Maass forms* for the full modular group $\Gamma = PSL_2(\mathbb{Z})$. Maass forms are Γ -invariant functions on the complex upper half-plane $\mathbb{H}$ which

are eigenvalues of the hyperbolic Laplacian $\nabla = -y^2(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2})$ and satisfy a suitable growth conditions. They often carry interesting arithmetic information, and hence are studied extensively in number theory. Since the Farey matrices generate a subgroup of the modular group Γ , it would be interesting to investigate if modularity phenomenon has a more natural explanation in this matrix product description, perhaps in relation to some congruence subgroup of the full modular group Γ .

The Farey matrix product measure studied in this thesis is one particular example of a more general class of multifractal measures. These multifractal measures are often studied in relation to Bernoulli convolutions [1], which has been an active area of study in statistical mechanics since the 1930s. It is believed that the formalism studied in this paper can perhaps be generalized to study at least some class of the multifractal measures, and hopefully shed some new light on some aspect of Bernoulli convolutions. In Chapter 3 of the thesis, we also discussed how Markov, hidden Markov, and multi-step Markov measures can be described in terms of general matrix product measures. As discussed in [10], this idea might have possible application to the study of quantum instruments and repeated quantum measurements.

Another interesting way in which the spin chain studied here can be generalized is by following Fiala and Klebun [36], and defining the function h_n as

$$h_n(\omega, x) = ax + b$$

where $F_{\omega_n} F_{\omega_{n-1}} \dots F_{\omega_1} = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$, and $\omega = (\omega_1, \omega_2, \dots, \omega_n)$ is the configuration of the spin chain. The case $x = 1$ recovers our spin chain, and the additional parameter x gives more freedom in doing performing various mathematical analysis.

The Lee-Yang theorem is one of the major results in statistical field theory, first proved by Lee and Yang for the Ising model, and then proved in a more general mathematical setting in the work of Asano, Newman, Simon and Griffiths and many others. The Lee-Yang theorem is a statement about the zeroes of a partition function of a Hamiltonian that is *ferromagnetic* and the measures in the partition function have the *Lee-Yang property*. We do not get into the details of the definitions here,

but emphasize that it allows one to look at partition functions in statistical mechanics as function defined on the complex plane, and states that for a particular class of such partition functions, all of its zeroes are imaginary. The work of Knauf on the number theoretic spin chain initially created a lot of excitement particularly because it was hoped that the expression for the partition function of the infinite spin chain

$$\lim_{N \rightarrow \infty} \mathcal{Z}_N(\beta) = \frac{\zeta(\beta - 1)}{\zeta(\beta)}$$

might lead to some results about the Riemann zeroes using the Lee-Yang theorem. While the work of Knauf and others have succeeded in showing that the interactions of the number theoretic spin chain are in fact ferromagnetic in the sense required for the Lee-Yang theorem, there has not been much progress in that direction. One can speculate if the number theoretic spin chain can in fact give new base into the Riemann zeroes, and it can be investigated if there are connections with the nearby field of quantum chaos.

In conclusion, we would like to state that while the number theoretic spin chain and its generalizations have been studied for a while, we are only starting to explore many of its fascinating properties. One is hopeful that the number theoretic spin chain will continue to inspire fruitful interaction among members of the number theory and statistical mechanics communities and lead to interesting new mathematical ideas in both disciplines in the coming decades.

Bibliography

- [1] T. Benoist, N. Cueno, D. Jakobson, V. Jaksic, and C.A. Pillet. Number Theoretic Spin Chains Revisited. *To appear*.
- [2] V. Jaksic. Entropic Information Theory. *Lecture notes*, 2020.
- [3] V. Jaksic and C.A. Pillet. Mathematical theory of non-equilibrium quantum statistical mechanics. *Journal of Statistical Physics*, 108, 2002.
- [4] W.H. Ashbacher, V. Jaksic, Y Pautrat, and C.A. Pillet. Topics in nonequilibrium quantum statistical mechanics. *Open Quantum Systems III Springer*, 2006.
- [5] V. Jaksic and C.A. Pillet. On entropy production in quantum statistical mechanics. *Communications in Mathematical Physics*, 217, 2001.
- [6] V. Jaksic, C.A. Pillet, and A. Shirikyan. Entropic Fluctuations in Gaussian Dynamical Systems. *Reports on Mathematical Physics*, 77, 2016.
- [7] V. Jaksic, C.A. Pillet, and A. Shirikyan. Entropic fluctuations in thermally driven harmonic networks. *Journal of Statistical Physics*, 166, 2017.
- [8] V. Jaksic, C.A. Pillet, and L. Rey-Bellet. Entropic fluctuations in statistical mechanics I. Classical dynamical systems. *Nonlinearity*, 24, 2016.
- [9] T. Benoist, V. Jaksic, Y Pautrat, and C.A. Pillet. On entropy production of repeated quantum measurements I. General theory. *Communications in Mathematical Physics*, 77-123, 2018.
- [10] T. Benoist, N. Cueno, V. Jaksic, and C.A. Pillet. On entropy production of repeated quantum measurements II. Examples. *Journal of Statistical Physics*, 182, 2021.
- [11] V. Jaksic. Lectures On Entropy I: Information-Theoretic Notions. *Open Quantum Systems: A mathematical Perspective*, 2019.
- [12] P. Walters. *An Introduction to Ergodic Theory*. Springer New York, NY, 1982.

- [13] D. Ruelle. *Thermodynamic formalism*. Cambridge University Press, 1978.
- [14] D. Ruelle. Is Our Mathematics Natural? The Case of Equilibrium Statistical Mechanics. *Bulletins of the AMS*, 19, 1988.
- [15] P. Walters. A variational principle for the pressure of continuous transformations. *American Journal of Mathematics*, 97, 1975.
- [16] M.R. Palmer, W. Parry, and P. Walters. Large sets of endomorphisms and of g -measures. *The Structure of Attractors in Dynamical Systems. Lecture Notes in Mathematics*, 2006.
- [17] F. R. Gantmacher. *The Theory of Matrices, Volume 1*. AMS Chelsea Publishing, 1959.
- [18] A. Y. Khinchin. *Continued Fractions*. University of Chicago Press, 1964.
- [19] H. Minkowski. Zur Geometrie der Zahlen. *Verhandlungen des III. internationalen Mathematiker-Kongresses in Heidelberg*, 1904.
- [20] A. Gierdrius. The moments of Minkowski question mark function: the dyadic period function. *Glasgow Mathematical Journal*, 52 (1), 2010.
- [21] A. Denjoy. Sur une fonction réelle de Minkowski. *J. Math. Pures Appl., Série IX*, 1938.
- [22] R. Salem. On some singular monotonic functions which are strictly increasing. *Transactions of the American Mathematical Society*, 53 (3), 1943.
- [23] A. Hatcher. *Topology of Numbers*. American Mathematical Society, 2022.
- [24] J. H. Conway. *On Numbers and Games*. Wellesley, MA: A K Peters, 2000.
- [25] S. Northshield. On Stern's Diatomic Sequence 0,1,1,2,1,3,2,3,1,4,... *American Mathematical Monthly*, 1117(7), 2010.
- [26] A. Knauf. On a Ferromagnetic Spin Chain. *Communications in Mathematical Physics*, 153, 1993.
- [27] A. Knauf. Phases of the Number-Theoretical Spin Chain. *Journal of Statistical Physics*, 73, 1993.

- [28] A. Knauf. On a Ferromagnetic Spin Chain. Part II: Thermodynamic Limit. *Journal of Mathematical Physics*, 35, 1994.
- [29] A. Knauf. Number theory, dynamical systems and statistical mechanics. *Reviews in Mathematical Physics*, 11(8), 1999.
- [30] A. Knauf. The Number-Theoretical Spin Chain and the Riemann Zeroes. *Communication in Mathematical Physics*, 196, 1998.
- [31] C. N. Yang and T. D. Lee. Statistical Theory of Equations of State and Phase Transitions. I. Theory of Condensation. *Physical Review*, 87 (3), 1952.
- [32] Itzykson, C. and Drouffe, J-M. *Statistical field theory. Vol. 1.* Cambridge Monographs on Mathematical Physics, 1989.
- [33] H.L. Lieb and A.D. Sokal. A general Lee-Yang theorem for one-component and multicomponent ferromagnets. *Communications in Mathematical Physics*, 80(2), 1981.
- [34] M.C. Newman. Zeros of the partition function for generalized Ising systems. *Communications on Pure and Applied Mathematics*, 27 (2), 1974.
- [35] J. Lewis and D. Zagier. Period Functions for Maass Wave Forms. I. *Annals of Mathematics*, 153(1), 2001.
- [36] J. Fiala and P. Klebun. Generalized Number Theoretic Spin Chain-Connections to Dynamical Systems and Expectation Values. *Journal of Statistical Physics*, 121, 2005.