UNIVERSITÄT LEIPZIG

FAKULTÄT FÜR MATHEMATIK UND INFORMATIK

DIPLOMARBEIT

Titel der Diplomarbeit

Sarnak's Conjecture about Möbius Function Randomness in Deterministic Dynamical Systems

Verfasser
Paul Wabnitz

angestrebter akademischer Grad
Diplom Mathematiker

Leipzig, März 2016

Studiengang: Mathematik
Betreuer: Prof. Dr. Tatjana Eisner

Zusammenfassung

Die vorliegende Arbeit befasst sich mit einer Vermutung von Sarnak aus dem Jahre 2010 über die Orthogonalität von durch deterministische dynamische Systeme induzierte Folgen zur Möbiusschen μ-Funktion. Ihre Hauptresultate sind zum einen der Ergodensatz mit Möbiusgewichten, welcher eine maßtheoretische (schwächere) Version von Sarnaks Vermutung darstellt, und zum anderen die bereits gesicherte Gültigkeit der genannten Vermutung in Spezialfällen, wobei hier exemplarisch unter anderem der Thue–Morse Shift und Schiefprodukterweiterungen von rationalen Rotationen auf dem Kreis gewählt worden sind.

Zum Zwecke der Motivation zeigen wir, dass eine gewisse Wachstumsabschätzung für $\sum_{n=1}^{N} \mu(n)$ äquivalent ist zum Primzahlsatz und skizzieren ein Resultat, welches die Äquivalenz einer weiteren solchen Abschätzung zur Riemannschen Vermutung liefert, um auf diese Weise die Bedeutung der Möbiusfunktion für die Zahlentheorie herauszustellen. Da sie für das Verständnis von Sarnaks Vermutung unerlässlich ist, geben wir eine Einführung in die Theorie der Entropie dynamischer Systeme auf Grundlage der Definitionen von Adler–Konheim–McAndrew, Bowen–Dinaburg und Kolmogorov–Sinai. Ferner berechnen wir die topologische Entropie des Thue–Morse Shifts und von Schiefprodukterweiterungen von Rotationen auf dem Kreis. Wir studieren die ergodische Zerlegung T-invarianter Maße auf kompakten metrischen Räumen mit stetiger Transformation T, welche wir für den Beweis des Ergodensatzes mit Möbiusgewichten benötigen.

Sodann beweisen wir den genannten gewichteten Ergodensatz. Wir geben eine hinreichende Bedingung an für das Erfülltsein von Sarnaks Vermutung in einem gegebenen dynamischen System, welche im anschließenden Kapitel Anwendung findet. So wird nachgewiesen, dass Sarnaks Vermutung im Falle des Thue–Morse Shifts und von Schiefprodukterweiterungen von rationalen Rotationen auf dem Kreis erfüllt ist. Abschließend wird gezeigt, dass Sarnaks Vermutung sich als Konsequenz aus einer Vermutung von Chowla ergibt.

Abstract

The thesis in hand deals with a conjecture of Sarnak from 2010 about the orthogonality of sequences induced by deterministic dynamical systems to the Möbius μ-function. Its main results are the ergodic theorem with Möbius weights, which is a measure theoretic (weaker) version of Sarnak's conjecture, and the already assured validity of Sarnak's conjecture in special cases, where we have exemplarily chosen the Thue–Morse shift and skew product extensions of rational rotations on the circle et al.

For the purpose of motivation, we show that a certain growth rate estimation for $\sum_{n=1}^{N} \mu(n)$ is equivalent to the prime number theorem and outline a result about another such estimation being equivalent to the Riemann hypothesis to underline the significance of the Möbius function for number theory. Since it is essential for the understanding of Sarnak's conjecture we give an introduction to the theory of entropy of dynamical systems based on the definitions of Adler–Konheim–McAndrew, Bowen–Dinaburg and Kolmogorov–Sinai. Furthermore, we calculate the topological entropy of the Thue–Morse shift and of skew product extensions of rotations on the circle. We study the ergodic decomposition for T-invariant measures on compact metric spaces with continuous transformations T, which we will need for the proof of the ergodic theorem with Möbius weights.

Thereafter, we prove the namely weighted ergodic theorem. We give a sufficient condition for Sarnak's conjecture to hold for a given dynamical system, which we make use of in the following chapter. Thereupon, it is varified that Sarnak's conjecture holds for the Thue–Morse shift and for skew product extensions of rational rotations on the circle. Lastly, it is shown that Sarnak's conjecture follows from one of Chowla.

Contents

1. Introduction

Many fundamental results of number theory deal with the phenomenon that is the prime numbers and for many still open questions to solve getting a better understandig of their true nature and distribution among the integers is essential. One way to approach this is the study of the well-known MÖBIUS function μ, which is the unique multiplicative arithmetical function taking -1 at each prime number and 0 at every higher power of a prime. So the values (that is the zero pattern or the sign) of μ correspond to the prime numbers and the question arises if they act predictably in any sense. From all we know about the nature of prime numbers, it seems reasonable to assume that this is not the case. But we can try to quantify this MÖBIUS *function randomness* and make it mathematically ascertainable. A first such attempt was made by CHOWLA in [10] in 1965 expressed in a (still unproven) conjecture of him (see Conjecture 8.1) which implies that the sign of μ fluctuates like random noise (cf. Remark 1 in [58]). But we want to focus more on another related hypothesis about the MÖBIUS function randomness heuristic.

We call two sequences $(a_n)_{n\in\mathbb{N}}, (b_n)_{n\in\mathbb{N}} \subset \mathbb{C}$[1] *mutually orthogonal* if

$$\frac{1}{N}\sum_{n=1}^{N} a_n b_n \xrightarrow[N\to\infty]{} 0.$$

It is known that not every sequence is orthogonal to the MÖBIUS function - e.g.

$$\frac{1}{N}\sum_{n=1}^{N} (\mu(n))^2 \nrightarrow 0$$

as $N \to \infty$ (see e.g. [53]; see also [21] (Proposition 5) for an example of a sequence $(z_n)_{n\in\mathbb{N}} \neq (\mu(n))_{n\in\mathbb{N}}$ not orthogonal to μ) - and according to the MÖBIUS function randomness the question arises if $(F_n)_{n\in\mathbb{N}} \subset \mathbb{C}$ is orthogonal to $(\mu(n))_{n\in\mathbb{N}}$ whenever $(F_n)_{n\in\mathbb{N}}$ acts - in some sense - predictably. This consideration culminates in a resent conjecture of SARNAK:

Conjecture 1.1 (SARNAK, [53])**.** *Let $T : X \to X$ be a deterministic continuous transformation on a compact metric space X. Then for each $x \in X$ and every $f \in C(X)$ we have*

$$\lim_{N\to\infty} \frac{1}{N}\sum_{n=1}^{N} f\left(T^n x\right)\mu(n) = 0,$$

where $C(X) \coloneqq \{f : X \to \mathbb{C} \,|\, f \text{ continuous}\}$.

[1] Note that we write $(x_n)_{n\in\mathbb{N}} \subset M$ whenever a sequence $(x_n)_{n\in\mathbb{N}}$ takes values in a set M while we reserve "$\subseteq$" for describing set inclusions (where we write $N \subsetneq M$ for $N \subseteq M$, $N \neq M$). Note furthermore, that we identify sequences $(x_n)_{n\in\mathbb{N}} \subset \mathbb{C}$, which are of the index set $\mathbb{N}$ and take values in $\mathbb{C}$, with arithmetical functions $x : \mathbb{N} \ni n \mapsto x(n) \in \mathbb{C}$ and use both terms synonymously.

So we consider sequences $(F_n)_{n\in\mathbb{N}}$ given by $F_n \coloneqq f(T^n x)$ with f and T as in Conjecture 1.1 and the predictability of $(F_n)_{n\in\mathbb{N}}$ encoded in the assumption about T being deterministic, which we will explain in more detail in Chapter 3.

While in general Sarnak's conjecture is still open, several special cases are already known, including

- constant or periodic sequences, see Chapter 7;
- rotations on the circle, which follows from the work of Davenport [14] (see also Chapter 7);
- nilsystems, which is a result of Green and Tao [28];
- horocycle flows, which was shown by Bourgain, Sarnak and Ziegler [9];
- the classical Thue–Morse shift, for which several proofs are known (the first has been found by Dartyge and Tenenbaum [13]). In Chapter 7 we will see this by following the argumentation of El Abdalaoui, Kasjan and Lemanczyk done in [21];
- a large class of rank one maps, as varified by Bourgain [8] and by El Abdalaoui, Lemanczyk and de la Rue [23];
- the dynamical system generated by the Rudin–Shapiro sequence, which is a result of Mauduit and Rivat [42] (see also [24]);
- subshifts given by bijective substitutions, as shown recently by Ferenczi, Kulaga-Przymus, Lemanczyk and Mauduit [24];

Conversely, it is possible to construct a sequence $(F_n)_{n\in\mathbb{N}}$ induced by a non-deterministic transformation (of arbitrarily small topological entropy) such that $\frac{1}{N}\sum_{n=1}^{N} F_n\mu(n) \nrightarrow 0$ as $N \to \infty$. Choose e.g.

$$F_n = \mathbf{1}_{k|n} \coloneqq \begin{cases} 1 & \text{if } k|n \\ 0 & \text{otherwise} \end{cases}$$

for some $k \in \mathbb{N}$ sufficiently large and such that $\nexists p \in \mathbb{P}\colon p^2|k$ (consult [58] for further information). Also the examples of sequences not orthogonal to μ given above come from non-deterministic transformations (see [53] or. [21]).

So Sarnak's conjecture appears to be a promising attempt to ascertainably describe the Möbius function randomness, despite it being far from being proven up until the present day, and the thesis in hand is intended to give a brief introduction to the study of this field of recent mathematical research. To do so we choose an approach based on instruments of ergodic theory by considering topological as well as measure-preserving dynamical systems.

The thesis is organized as follows: First of all, we want to substantiate the significance of the Möbius function for number theory by proving a fundamental equivalence to the prime number theorem and outlining another to the prominent Riemann hypothesis. Subsequently, we will give an introduction to the concept of topological entropy of a dynamical system to fully explain the meaning of the assumption about T being deterministic. Chapter 4 will be dedicated to a technical tool from ergodic

theory, namely the ergodic decomposition of an invariant measure for a given dynamical system. This will be applied (et al) in Chapter 5 to prove the ergodic theorem with MÖBIUS weights stating that SARNAK's conjecture holds true for ν-a.e. point of an arbitrary dynamical system with ν an invariant measure on it. In Chapter 6 the so called KBSZ-criterion will be discussed, which represents a sufficient condition for the claimed convergence in Conjecture 1.1 in a given system. By bringing that into usage we will collect some examples for which we show that SARNAK's conjecture holds. Finally, we will show that Conjecture 1.1 is a consequence of the conjecture of CHOWLA mentioned above.

2. The Möbius Function

Denote by $\omega : \mathbb{N} \to \mathbb{N}$ the arithmetical function which maps $n \in \mathbb{N}$ to the number of distinct prime factors of n, i.e., $\omega(n) \coloneqq \#\{p \in \mathbb{P} \mid p|n\}$. We call n *square-free* if there is no $p \in \mathbb{P}$ such that $p^2|n$. That means that, in the unique prime factorization $\prod_{k=1}^{m} p_k^{a_k}$ of n all the exponents a_k are equal to 1. Equivalently, an $n \in \mathbb{N}$ is square-free if and only if for all $m \in \mathbb{N}$ we have $m^2 \nmid n$.

On this basis, define the Möbius *function* $\mu : \mathbb{N} \to \{-1, 0, 1\}$ by

$$\mu(n) \coloneqq \begin{cases} 1 & \text{if } n = 1 \\ (-1)^{\omega(n)} & \text{if } n \text{ is square-free} \\ 0 & \text{otherwise} \end{cases}.$$

This function was introduced in 1832 by the German mathematician and theoretical astronomer August Ferdinand Möbius (17 November 1790 – 26 September 1868), who learned and lectured at the University of Leipzig (appointed as Extraordinary Professor to the "chair of astronomy and higher mechanics" in 1816, Full Professorship in astronomy by 1844).[1] Actually, this function was first considered by Gauss more than 30 years before Möbius in his work *Disquisitiones Arithmeticae* of 1801 (see [27], §81). The notation $\mu(n)$ was first used by Mertens in 1874.

So, why is this function of concern for present mathematicial research? At least the conjectures of Sarnak and Chowla (and thus the studies of several mathematicians working towards a proof of either of these) are essentially related to it.

In this chapter, which is to be understood as a motivation for the thesis in hand, we want to assess the significance of the Möbius function for number theory. For

$$M : [1, \infty) \ni x \mapsto \sum_{\substack{n \in \mathbb{N} \\ n \leq x}} \mu(n) \in \mathbb{Z}$$

(M is called the Mertens *function*) we will show that the asymptotical growth estimation

$$\lim_{x \to \infty} \frac{M(x)}{x} = 0 \qquad (\bigstar)$$

is equivalent to the prime number theorem, which describes the asymptotic distribution of the prime numbers among the positive integers:

Theorem 2.1 (Prime number theorem)**.** *We have*

$$\lim_{x \to \infty} \frac{\#\{p \in \mathbb{P} \mid p \leq x\}}{x/\log x} = 1.$$

[1] See e.g. http://www-history.mcs.st-andrews.ac.uk/Biographies/Mobius.html for further information.

Several different proofs for Theorem 2.1 are known; the first have been found independently by HADAMARD and DE LA VALLÉE-POUSSIN in 1896. One by NEWMAN, that is arguably the simplest known, can be found in [45].

Furthermore, following the work of Titchmarsh done in [59], we will give a brief sketch of proof that the much stronger (and still open) improvement

$$\limsup_{x\to\infty} \frac{M(x)}{x^{\frac{1}{2}+\varepsilon}} < \infty \qquad (\bigstar\bigstar)$$

is equivalent to the RIEMANN hypothesis:

For $s \in \mathbb{C}$ with $\Re(s) > 0$ let $\zeta(s) \coloneqq \sum_{n=1}^{\infty} \frac{1}{n^s}$. Then ζ can be uniquely analytically continued to a meromorphic[2] function on the whole complex plane (except $s = 1$), which we also denote by ζ. It can be shown that $\zeta(s) = 0$ for any $s \in \{-2t \,|\, t \in \mathbb{N}\}$ (this follows from the functional equation for ζ, see Remark 2.22 below). But ζ do have further zeros, called the *non-trival* zeros of the RIEMANN Zeta-function ζ.

Conjecture 2.2 (RIEMANN)**.** *All non-trival zeros of the* RIEMANN *Zeta-function lie on the line* $\left\{\frac{1}{2} + it \,\middle|\, t \in \mathbb{R}\right\}$*, i.e., whenever* $\zeta(s) = 0$ *for* $\Re(s) \geq 0$ *we have* $\Re(s) = \frac{1}{2}$*.*

We want to start our studies with a rather easy property of the MÖBIUS function we will need in several argumentations.

Definition 2.3. Let f be an arithmetical function. Then we call f

- *multiplicative*, if for each $m, n \in \mathbb{N}$ with $(m, n) = 1$ we have $f(m \cdot n) = f(m)f(n)$.
- *totally multiplicative*, if for each $m, n \in \mathbb{N}$ we have $f(m \cdot n) = f(m)f(n)$.

Proposition 2.4. *The* MÖBIUS *function is multiplicative but not totally multiplicative.*

Proof. First we show that μ is multiplicative. This is obvious in the case $n = 1$ or $m = 1$. So let $m, n \in \mathbb{N} \setminus \{1\}$ with $(m, n) = 1$. Then we have

$$\mu(m) = 0 \vee \mu(n) = 0 \Leftrightarrow \exists p \in \mathbb{P}\colon \left(p^2 | m \vee p^2 | n\right) \Leftrightarrow \exists p \in \mathbb{P}\colon p^2 | m \cdot n \Leftrightarrow \mu(m \cdot n) = 0.$$

On the other hand, for $\mu(m), \mu(n) \neq 0$, we have $m = \prod_{j=1}^{k} p_j$ and $n = \prod_{i=1}^{l} q_i$ for some $k, l \in \mathbb{N}$, $p_1, \ldots, p_k, q_1, \ldots, q_l \in \mathbb{P}$, $p_{j_1} \neq p_{j_2}$ and $q_{i_1} \neq q_{i_2}$ for each $j_1, j_2 \in [1, k] \cap \mathbb{Z}$ and each $i_1, i_2 \in [1, l] \cap \mathbb{Z}$. Moreover, since $(m \cdot n) = 1$, we have $p_j \neq q_i$ for any $j \in [1, k] \cap \mathbb{Z}$, $i \in [1, l] \cap \mathbb{Z}$, because otherwise $(m \cdot n) > r$, for $r \coloneqq p_j = q_i$. Hence $m \cdot n = \left(\prod_{j=1}^{k+l} p_j\right)$, for $p_{k+i} \coloneqq q_i$ for each $i \in [1, l] \cap \mathbb{Z}$, and thus

$$\mu(m \cdot n) = (-1)^{k+l} = (-1)^k (-1)^l = \mu(m)\mu(n).$$

To see that μ is not totally multiplicative, consider e.g. $m = 3$ and $n = 6$. Then $\mu(3) = -1$ and $\mu(6) = \mu(2 \cdot 3) = 1$, but

$$\mu(3 \cdot 6) = \mu(3^2 \cdot 2) = 0 \neq -1 = \mu(3)\mu(6). \qquad \square$$

By the fundamental theorem of arithmetic, which states that the prime factorization of any positive integer is unique up to the order of the factors, each multiplicative arithmetical function is uniquely determined by its values at prime powers. Thus μ is the unique multiplicative function that takes the value -1 at each prime and the value 0 at every higher power of a prime.[3]

[2] i.e., holomorphic on all $\mathbb{C}$ except a set of countably many isolated points

[3] Note that, since $(m, 1) = 1$ for each $m \in \mathbb{N}$, we have $f(1) = 1$ whenever $f : \mathbb{N} \to \mathbb{C}$ is multiplicative.

2.1. The Prime Number Theorem is equivalent to (★)

In this section we will consider sums of the form $\sum_{\substack{d\in\mathbb{N}\\ d|n}} f(d)$, for $n \in \mathbb{N}$ and f an arithmetical function, for which we shortly write $\sum_{d|n} f(n)$. We will content ourselves to real-valued functions, although each result holds for complex-valued arithmetical functions by dividing it into its real and imaginary part. All results of this section are taken from [4].

Let $\pi : [1,\infty) \to \mathbb{N}$ be given by $\pi(x) \coloneqq \#\{p \in \mathbb{P} \mid p \leq x\}$ for every $x \in [1,\infty)$. For $f, g : \mathbb{R} \to \mathbb{R}$, $g(x) \neq 0$ for each $x \in \mathbb{R}$, write $f(x) \sim g(x)$ if $\lim_{x\to\infty} \frac{f(x)}{g(x)} = 1$. We want to prove the following result:

Theorem 2.5. *The following statements are equivalent:*

i) $\pi(x) \sim \frac{x}{\log x}$.

ii) $\lim_{x\to\infty} \frac{M(x)}{x} = 0$.

To do so we need some preparation concerning Dirichlet products, the Möbius inversion and the Chebyshev functions.

2.1.1. Preliminaries

Definition 2.6 (Dirichlet multiplication)**.** For arithmetical functions $f, g : \mathbb{N} \to \mathbb{R}$ we call

$$f * g : \mathbb{N} \ni n \mapsto \sum_{d|n} f(d) g\left(\frac{n}{d}\right) \in \mathbb{R}$$

the Dirichlet *product* of f and g.

One can show that the Dirichlet multiplication is associative and commutative (see [4]).

For $x \in \mathbb{R}$ denote by $[x]$ the largest integer not greater than x, i.e.,

$$[x] \coloneqq \max\{n \in \mathbb{Z} \mid n \leq x\}.$$

Define $I : \mathbb{N} \to \{0,1\}$ by

$$I(n) \coloneqq \left[\frac{1}{n}\right] = \begin{cases} 1 & \text{if } n = 1 \\ 0 & \text{if } n > 1 \end{cases}.$$

Then I is the neutral element for the Dirichlet multiplication.

Proposition 2.7. *For $n \in \mathbb{N}$ we have $I(n) = \sum_{d|n} \mu(d)$.*

Proof. For $n = 1$ the assertion holds. Let $n > 1$ and write $n = \prod_{j=1}^{k} p_j^{a_j}$ with $p_1, \ldots, p_k \in \mathbb{P}$, $a_1, \ldots, a_k \in \mathbb{N}$. In $\sum_{d|n} \mu(d)$ only the terms for $d = 1$ and for those divisors of n which are products of distinct primes do contribute. Hence

$$\begin{aligned} \sum_{d|n} \mu(d) &= \mu(1) + \mu(p_1) + \ldots + \mu(p_k) + \mu(p_1p_2) + \ldots + \mu(p_{k-1}p_k) \\ &\quad + \ldots + \mu(p_1p_2\ldots p_k) \\ &= 1 + \binom{k}{1}(-1) + \binom{k}{2}(-1)^2 + \ldots + \binom{k}{k}(-1)^k = (1-1)^k = 0 \end{aligned}$$

by the binomial sum. □

From Proposition 2.7 we obtain $I(n) = \sum_{d|n} \mu(d) = \sum_{d|n} \mathbf{1}_{\mathbb{N}}(d)\mu\left(\frac{n}{d}\right) = \mathbf{1}_{\mathbb{N}} * \mu$.

Theorem 2.8 (MÖBIUS inversion)**.** *For arithmetical functions f, g the following conditions are equivalent:*

i) $f(n) = \sum_{d|n} g(d)$.

ii) $g(n) = \sum_{d|n} f(d)\mu\left(\frac{n}{d}\right)$.

Proof. From *i)* follows that $f = g * \mathbf{1}_{\mathbb{N}}$. Hence

$$f * \mu = (g * \mathbf{1}_{\mathbb{N}}) * \mu = g * (\mathbf{1}_{\mathbb{N}} * \mu) = g * I = g,$$

which implies *ii)*. The converse implication follows analogously by DIRICHLET multiplication of $f * \mu = g$ with $\mathbf{1}_{\mathbb{N}}$. □

Definition 2.9. The function $\Lambda : \mathbb{N} \to [0, \infty)$ given by

$$\Lambda(n) \coloneqq \begin{cases} \log p & \text{if } n = p^m \text{ for some } p \in \mathbb{P} \text{ and } m \in \mathbb{N} \\ 0 & \text{otherwise} \end{cases}$$

is called the VON MANGOLDT *function*.

Proposition 2.10. *For $n \in \mathbb{N}$ we have $\sum_{d|n} \Lambda(d) = \log n$.*

Proof. For $n = 1$ both sides are equal to zero. For $n > 1$ write $n = \prod_{j=1}^{k} p_j^{a_j}$ with $p_1, \ldots, p_k \in \mathbb{P}$, $a_1, \ldots, a_k \in \mathbb{N}$. Then

$$\log n = \log\left(\prod_{j=1}^{k} p_j^{a_j}\right) = \sum_{j=1}^{k} a_j \log p_j.$$

Now the only non-zero terms in the sum $\sum_{d|n} \Lambda(d)$ come from those divisors of d which are of the form p_k^m for $m \in [1, a_k] \cap \mathbb{Z}$, $k \in [1, r] \cap \mathbb{Z}$. Thus

$$\sum_{d|n} \Lambda(d) = \sum_{k=1}^{r} \sum_{m=1}^{a_k} \Lambda(p_k^m) = \sum_{k=1}^{r} a_k \log p_k = \log n. \qquad \square$$

Proposition 2.11. *For $n \in \mathbb{N}$ we have $\Lambda(n) = \sum_{d|n} \mu(d) \log\left(\frac{n}{d}\right) = -\sum_{d|n} \mu(d) \log d$.*

Proof. From Proposition 2.10 we know that $\sum_{d|n} \Lambda(d) = \log n$. MÖBIUS inversion of this equation yields

$$\begin{aligned} \Lambda(n) &= \sum_{d|n} \mu(d) \log\left(\frac{n}{d}\right) = (\log n) \sum_{d|n} \mu(d) - \sum_{d|n} \mu(d) \log d \\ &= I(n) \log n - \sum_{d|n} \mu(d) \log d = -\sum_{d|n} \mu(d) \log d, \end{aligned}$$

since $I(n) \log n = 0$ for each $n \in \mathbb{N}$ (because $I(n) \neq 0$ iff $n = 1$, which is the root of the logarithm). □

For $F : (0,\infty) \to \mathbb{R}$, $F(x) = 0$ for $x \in (0,1)$, and $f : \mathbb{N} \to \mathbb{R}$ define $f \star F : (0,\infty) \to \mathbb{R}$ by

$$(f \star F)(x) \coloneqq \sum_{\substack{n\in\mathbb{N}\\ n\le x}} f(n) F\left(\frac{x}{n}\right).$$

Proposition 2.12. *For all arithmetical functions f, g and F given as above we have*

$$f \star (g \star F) = (f * g) \star F.$$

Proof. For each $x \in (0,\infty)$ we have

$$\begin{aligned}(f \star (g \star F))(x) &= \sum_{\substack{n\in\mathbb{N}\\ n\le x}} f(n) \sum_{\substack{m\in\mathbb{N}\\ m\le \frac{x}{n}}} g(m) F\left(\frac{x}{m\cdot n}\right) = \sum_{\substack{m,n\in\mathbb{N}\\ m\cdot n\le x}} f(n)g(m) F\left(\frac{x}{m\cdot n}\right)\\ &= \sum_{\substack{k\in\mathbb{N}\\ k\le x}} \left(\sum_{n\mid k} f(n) g\left(\frac{k}{n}\right)\right) F\left(\frac{x}{k}\right) = \sum_{\substack{k\in\mathbb{N}\\ k\le x}} (f*g)(k) F\left(\frac{x}{k}\right)\\ &= ((f*g)\star F)(x).\end{aligned}$$

□

Proposition 2.13. *For $f, g : \mathbb{N} \to \mathbb{R}$ let $H(x) \coloneqq \sum_{\substack{n\in\mathbb{N}\\ n\le x}} (f*g)(n)$, $F(x) \coloneqq \sum_{\substack{n\in\mathbb{N}\\ n\le x}} f(n)$ and $G(x) \coloneqq \sum_{\substack{n\in\mathbb{N}\\ n\le x}} g(n)$. Then*

$$H(x) \coloneqq \sum_{\substack{n\in\mathbb{N}\\ n\le x}} f(n) G\left(\frac{x}{n}\right) = \sum_{\substack{n\in\mathbb{N}\\ n\le x}} g(n) F\left(\frac{x}{n}\right).$$

Proof. Define $U : (0,\infty) \to \{0,1\}$ by

$$U(x) \coloneqq \begin{cases} 0 & \text{if } x \in (0,1)\\ 1 & \text{otherwise}\end{cases}.$$

Then $F = f \star U$, $G = g \star U$ and $H = (f*g) \star U$ and from Proposition 2.12 we obtain

$$f \star G = f \star (g \star U) = (f*g) \star U = H$$

as well as

$$g \star F = g \star (f \star U) = (g*f) \star U = H.$$

□

Proposition 2.14. *For $x \in [1,\infty)$, $a, b \in [1,\infty)$ such that $a \cdot b = x$ and F, G as in Proposition 2.13 we have*

$$\sum_{\substack{q,d\in\mathbb{N}\\ q\cdot d\le x}} f(d)g(q) = \sum_{\substack{n\in\mathbb{N}\\ n\le a}} f(n) G\left(\frac{x}{n}\right) + \sum_{\substack{n\in\mathbb{N}\\ n\le b}} g(n) F\left(\frac{x}{n}\right) - F(a)G(b).$$

Sketch of proof. The sum $\sum_{\substack{q,d\in\mathbb{N}\\ q\cdot d\le x}} f(d)g(q)$ is extended over the lattice points underneath the graph of a certain hyperbolic function φ. Let $a > 1$ and $b \coloneqq \varphi(a)$. Set $B \coloneqq [1,a] \times [1,b]$ as well as $A \coloneqq \{(x,y) \in \mathbb{R}^2 \,|\, x \in (1,a),\, y \in (b, \varphi(x))\}$ and $C \coloneqq \{(x,y) \in \mathbb{R}^2 \,|\, x > a,\, y \in [1,\varphi(x)]\}$. Split the sum into to parts, one over the

lattice points in $A \cup B$ and the other over those in $B \cup C$. The lattice points in B are covered twice, so we have

$$\sum_{\substack{q,d\in\mathbb{N}\\ q\cdot d\le x}} f(d)g(q) = \sum_{\substack{d\in\mathbb{N}\\ d\le a}}\sum_{\substack{q\in\mathbb{N}\\ q\le \frac{x}{d}}} f(d)g(q) + \sum_{\substack{q\in\mathbb{N}\\ q\le b}}\sum_{\substack{d\in\mathbb{N}\\ d\le \frac{x}{q}}} f(d)g(q) - \sum_{\substack{d\in\mathbb{N}\\ d\le a}}\sum_{\substack{q\in\mathbb{N}\\ q\le b}} f(d)g(q)$$

which is the same as the asserted equality. □

Lemma 2.15 (Abel's identity)**.** *Let a be an arbitrary arithmetical function. Define $A : \mathbb{R} \to \mathbb{R}$ by*

$$A(x) \coloneqq \sum_{\substack{n\in\mathbb{N}\\ n\le x}} a(n)$$

(note that $A(x) = 0$ for each $x < 1$). For $0 < y < x$ let $f : \mathbb{R} \to \mathbb{R}$ be continuously differentiable in $[y, x]$. Then

$$\sum_{\substack{n\in\mathbb{N}\\ y<n\le x}} a(n)f(n) = A(x)f(x) - A(y)f(y) - \int_y^x A(t)f'(t)\,\mathrm{d}\,t.$$

Proof. Let $k \coloneqq [x]$ and $m \coloneqq [y]$. So $A(x) = A(k)$, $A(y) = A(m)$ and

$$\begin{aligned}
\sum_{\substack{n\in\mathbb{N}\\ y<n\le x}} a(n)f(n) &= \sum_{n=m+1}^{k} a(n)f(n) = \sum_{n=m+1}^{k} \left(A(n) - A(n-1)\right) f(n)\\
&= \sum_{n=m+1}^{k} A(n)f(n) - \sum_{n=m}^{k-1} A(n)f(n+1)\\
&= \sum_{n=m+1}^{k-1} A(n)\left(f(n) - f(n+1)\right) + A(k)f(k) - A(m)f(m+1)\\
&= -\sum_{n=m+1}^{k-1} A(n)\int_n^{n+1} f'(t)\,\mathrm{d}\,t + A(k)f(k) - A(m)f(m+1)\\
&= -\int_{m+1}^{k} A(t)f'(t)\,\mathrm{d}\,t + A(x)f(x)\\
&\quad -\int_k^x A(t)f'(t)\,\mathrm{d}\,t - A(y)f(y) - \int_y^{m+1} A(t)f'(t)\,\mathrm{d}\,t\\
&= A(x)f(x) - A(y)f(y) - \int_y^x A(t)f'(t)\,\mathrm{d}\,t.
\end{aligned}$$

Thus the assertion follows. □

It will be convenient to reformulate the prime number theorem. More specifically, we will show that Theorem 2.1 is equivalent to

$$\sum_{\substack{n\in\mathbb{N}\\ n\le x}} \Lambda(n) \sim x \tag{2.1}$$

as $x \to \infty$, where Λ denotes the van Mangoldt function as defined in Definition 2.9. The partial sums of Λ define a function introduced by Chebyshev.

Definition 2.16 (Chebyshev)**.** For $x \in (0,\infty)$ define $\vartheta, \psi : (0,\infty) \to \mathbb{R}$ by

$$\vartheta(x) \coloneqq \sum_{\substack{p\in\mathbb{P}\\ p\le x}} \log p \quad \text{and}$$
$$\psi(x) \coloneqq \sum_{\substack{n\in\mathbb{N}\\ n\le x}} \Lambda(n),$$

Then we call ϑ the *first* and ψ the *second* Chebyshev *function.*

Thus (2.1) takes the form

$$\lim_{x\to\infty} \frac{\psi(x)}{x} = 1. \tag{2.2}$$

To show that this is equivalent to the prime number theorem, we need to study the two Chebyshev functions a little further. Since $\Lambda(n) = 0$ unless n is a prime power, we can write

$$\psi(x) = \sum_{\substack{n\in\mathbb{N}\\ n\le x}} \Lambda(n) = \sum_{p\in\mathbb{P}} \sum_{\substack{m\in\mathbb{N}\\ p^m\le x}} \Lambda(p^m) = \sum_{m=1}^{\infty} \sum_{\substack{p\in\mathbb{P}\\ p\le \sqrt[m]{x}}} \log p.$$

The sum on m is actually finite. In fact, the sum on p is empty, if $\sqrt[m]{x} < 2$, that is, if $\frac{1}{m}\log x < \log 2$, which we rewrite as

$$m > \frac{\log x}{\log 2} = \log_2 x.$$

Hence

$$\psi(x) = \sum_{\substack{m\in\mathbb{N}\\ m\le\log_2 x}} \sum_{\substack{p\in\mathbb{P}\\ p\le \sqrt[m]{x}}} \log p = \sum_{\substack{m\in\mathbb{N}\\ m\le\log_2 x}} \vartheta(\sqrt[m]{x}). \tag{2.3}$$

Proposition 2.17. *For each $x \in (0,\infty)$ we have*

$$0 \le \frac{\psi(x)}{x} - \frac{\vartheta(x)}{x} \le \frac{(\log x)^2}{2\sqrt{x}\log 2}.$$

Proposition 2.17 shows that $\lim_{x\to\infty}\left(\frac{\psi(x)}{x} - \frac{\vartheta(x)}{x}\right) = 0$, i.e., if one of the functions $\frac{\psi(x)}{x}$, $\frac{\vartheta(x)}{x}$ converges then so does the other and the two limits coincide.

Proof of Proposition 2.17. From (2.3) we find

$$0 \le \psi(x) - \vartheta(x) = \sum_{\substack{m\in\mathbb{N}\\ m\le\log_2 x}} \vartheta(\sqrt[m]{x}) - \vartheta(x^{\frac{1}{1}}) = \sum_{\substack{m\in\mathbb{N}\\ 2\le m\le\log_2 x}} \vartheta(\sqrt[m]{x}).$$

But from the definition of ϑ we have

$$\vartheta(x) \le \sum_{\substack{p\in\mathbb{P}\\ p\le x}} \log x \le x\log x.$$

So

$$\begin{aligned} 0 \le \psi(x) - \vartheta(x) &= \sum_{\substack{m\in\mathbb{N}\\ 2\le m\le \log_2 x}} \vartheta(\sqrt[m]{x}) \le \sum_{\substack{m\in\mathbb{N}\\ 2\le m\le \log_2 x}} (\sqrt[m]{x}) \log (\sqrt[m]{x}) \\ &\le (\log_2 x)\sqrt{x}\log(\sqrt{x}) = \frac{\log x}{\log 2}\cdot\frac{\sqrt{x}}{2}\log x \\ &= \frac{\sqrt{x}\,(\log x)^2}{2\log 2}. \end{aligned}$$

Hence

$$0 \le \frac{\psi(x)}{x} - \frac{\vartheta(x)}{x} \le \frac{1}{x}\cdot\frac{\sqrt{x}\,(\log x)^2}{2\log 2} = \frac{(\log x)^2}{2\sqrt{x}\log 2}. \qquad \square$$

2.1.2. Proof of Theorem 2.5

As mentioned before, we want to deal with the prime number theorem in the form given by equation (2.2). Thus, we have to verify this equivalence first. Recall that for $x \in [1,\infty)$ we set $\pi(x) \coloneqq \#\{p \in \mathbb{P} \mid p \le x\} = \sum_{\substack{p\in\mathbb{P}\\ p\le x}} 1$. Furthermore, for two functions $f, g : \mathbb{R} \to \mathbb{R}$ we write $f(x) = \mathrm{O}(g(x))$ as $x \to \infty$, if there is a constant $C > 0$ such that for all sufficiently large $x \in \mathbb{R}$ (i.e., for all x greater than some $x_0 \in \mathbb{R}$) we have $|f(x)| \le C\,|g(x)|$. Note that we have $M(x) = \sum_{\substack{n\in\mathbb{N}\\ n\le x}} \mu(n) \le \sum_{\substack{n\in\mathbb{N}\\ n\le x}} 1 = [x] = \mathrm{O}(x)$ as $x \to \infty$.

Lemma 2.18. *For $x \in [2,\infty)$ we have*

a) $\vartheta(x) = \pi(x)\log x - \int_2^x \frac{\pi(t)}{t}\,\mathrm{d}\,t$.

b) $\pi(x) = \frac{\vartheta(x)}{\log x} + \int_2^x \frac{\vartheta(x)}{t(\log t)^2}\,\mathrm{d}\,t$.

Proof. We want to make use of Abel's identity (Lemma 2.15). Note that

$$\begin{aligned} \pi(x) &= \sum_{\substack{p\in\mathbb{P}\\ p\le x}} 1 = \sum_{\substack{n\in\mathbb{N}\\ 1<n\le x}} \mathbf{1}_{\mathbb{P}}(n) \quad \text{and} \\ \vartheta(x) &= \sum_{\substack{p\in\mathbb{P}\\ p\le x}} \log p = \sum_{\substack{n\in\mathbb{N}\\ 1<n\le x}} \mathbf{1}_{\mathbb{P}}(n)\log p, \end{aligned}$$

where

$$\mathbf{1}_A(x) \coloneqq \begin{cases} 1 & \text{for } x \in A \\ 0 & \text{otherwise} \end{cases},$$

for arbitrary sets A and Ω with $A \subseteq \Omega$ and $x \in \Omega$. So from Lemma 2.15 with $a(n) \coloneqq \mathbf{1}_{\mathbb{P}}(n)$, $f(x) \coloneqq \log x$ and $y \coloneqq 1$ we obtain

$$\vartheta(x) = \sum_{\substack{n\in\mathbb{N}\\ 1<n\le x}} \mathbf{1}_{\mathbb{P}}(n)\log p = \pi(x)\log x - \underbrace{\pi(1)\log 1}_{=0} - \int_1^x \frac{\pi(t)}{t}\,\mathrm{d}\,t,$$

which implies *a)* since $\pi(t) = 0$ for $t < 2$.

Now let $a(n) \coloneqq \mathbf{1}_{\mathbb{P}}(n) \log n$ and write

$$\pi(x) = \sum_{\substack{n \in \mathbb{N} \\ \frac{3}{2} < n \leq x}} a(n) \frac{1}{\log n} \quad \text{and}$$
$$\vartheta(x) = \sum_{\substack{n \in \mathbb{N} \\ n \leq x}} a(n).$$

Then Lemma 2.15 with $f(x) \coloneqq (\log x)^{-1}$ and $y \coloneqq \frac{3}{2}$ yields

$$\pi(x) = \frac{\vartheta(x)}{\log x} - \frac{\vartheta\left(\frac{3}{2}\right)}{\log \frac{3}{2}} + \int_{\frac{3}{2}}^{x} \frac{\vartheta(t)}{t \left(\log t\right)^2} \, \mathrm{d}\, t,$$

which implies *b*) since $\vartheta(t) = 0$ for $t < 2$. □

Theorem 2.19. *The following relations are equivalent:*

i) $\lim_{x \to \infty} \frac{\pi(x) \log x}{x} = 1.$

ii) $\lim_{x \to \infty} \frac{\vartheta(x)}{x} = 1.$

iii) $\lim_{x \to \infty} \frac{\psi(x)}{x} = 1.$

Proof. The equivalence of *ii*) and *iii*) follows from Proposition 2.17. So it remains to show that *i*) and *ii*) are equivalent.

From Lemma 2.18 *a*) and *b*) we obtain respectively

$$\frac{\vartheta(x)}{x} = \frac{\pi(x) \log x}{x} - \frac{1}{x} \int_2^x \frac{\pi(t)}{t} \, \mathrm{d}\, t \quad \text{and}$$
$$\frac{\pi(x) \log x}{x} = \frac{\vartheta(x)}{x} + \frac{\log x}{x} \int_2^x \frac{\vartheta(t)}{t \left(\log t\right)^2} \, \mathrm{d}\, t.$$

Hence, to show that *i*) implies *ii*), it suffices to show that *i*) implies

$$\lim_{x \to \infty} \frac{1}{x} \int_2^x \frac{\pi(t)}{t} \, \mathrm{d}\, t = 0.$$

But from *i*) it follows that $\frac{\pi(t)}{t} = \mathrm{O}\left(\frac{1}{\log t}\right)$ as $t \to \infty$. Hence

$$\frac{1}{x} \int_2^x \frac{\pi(t)}{t} \, \mathrm{d}\, t = \mathrm{O}\left(\frac{1}{x} \int_2^x \frac{1}{\log t} \, \mathrm{d}\, t\right)$$

as $x \to \infty$. Now

$$\int_2^x \frac{1}{\log t} \, \mathrm{d}\, t = \int_2^{\sqrt{x}} \frac{1}{\log t} \, \mathrm{d}\, t + \int_{\sqrt{x}}^x \frac{1}{\log t} \, \mathrm{d}\, t \leq \frac{\sqrt{x}}{\log 2} + \frac{x - \sqrt{x}}{\log \sqrt{x}} = \left(\frac{2 - \frac{2}{\sqrt{x}}}{\log x} - \frac{1}{\sqrt{x} \log 2}\right) x$$

and thus

$$\frac{1}{x} \int_2^x \frac{1}{\log t} \, \mathrm{d}\, t \leq \frac{2 - \frac{2}{\sqrt{x}}}{\log x} - \frac{1}{\sqrt{x} \log 2} \xrightarrow[x \to \infty]{} 0.$$

On the other hand, to show that *ii*) implies *i*), it suffices to show that *ii*) implies

$$\lim_{x\to\infty} \frac{\log x}{x} \int_2^x \frac{\vartheta(t)}{t\,(\log t)^2}\,\mathrm{d}\,t = 0.$$

But from *ii*) it follows that $\vartheta(t) = \mathrm{O}(t)$ as $t \to \infty$. Hence

$$\frac{\log x}{x} \int_2^x \frac{\vartheta(t)}{t\,(\log t)^2}\,\mathrm{d}\,t = \mathrm{O}\left(\frac{\log x}{x} \int_2^x \frac{1}{(\log t)^2}\,\mathrm{d}\,t\right).$$

Like above we conclude

$$\int_2^x \frac{1}{(\log t)^2}\,\mathrm{d}\,t \le \frac{\sqrt{x}}{(\log 2)^2} + \frac{x-\sqrt{x}}{(\log\sqrt{x})^2} = \frac{x}{(\log x)^2}\left(-\frac{4}{\sqrt{x}} + \frac{(\log x)^2}{\sqrt{x}\,(\log 2)^2} + 4\right)$$

and thus

$$\frac{\log x}{x} \int_2^x \frac{1}{(\log t)^2}\,\mathrm{d}\,t \le \frac{1}{\log x}\left(-\frac{4}{\sqrt{x}} + \frac{(\log x)^2}{\sqrt{x}\,(\log 2)^2} + 4\right) \xrightarrow[x\to\infty]{} 0. \qquad \square$$

Denote by $\mathfrak{d} : \mathbb{N} \to \mathbb{N}$ the arithmetical function that counts the divisors of an integer n, i.e., $\mathfrak{d}(n) \coloneqq \sum_{d|n} 1$. Furthermore, denote $\mathfrak{C} \coloneqq \lim_{n\to\infty}\left(\sum_{k=1}^n \frac{1}{k} - \log n\right)$ (one can show that this is a real number, see e.g. [4]). We call $\mathfrak{C}$ Euler*'s constant.*

Lemma 2.20 (Dirichlet's Formula)**.** *We have*

$$\sum_{\substack{n\in\mathbb{N}\\ n\le x}} \mathfrak{d}(n) - x\log x - (2\mathfrak{C}-1)\,x = \mathrm{O}\left(\sqrt{x}\right)$$

as $x \to \infty$, where $\mathfrak{C}$ denotes Euler*'s constant.*

A proof of Lemma 2.20 can be found in [4] (Theorem 3.3).

Lemma 2.21. *Define $H : [1,\infty) \to \mathbb{R}$ by $H(x) \coloneqq \sum_{\substack{n\in\mathbb{N}\\ n\le x}} \mu(n)\log n$. Then for M the* Mertens *function we have*

$$\lim_{x\to\infty}\left(\frac{M(x)}{x} - \frac{H(x)}{x\log x}\right) = 0.$$

Analogous to Proposition 2.17, Lemma 2.21 implies that if one of the functions $\frac{M(x)}{x}$, $\frac{H(x)}{x\log x}$ converges then so does the other and the two limits coincide.

Proof of Lemma 2.21. Form Lemma 2.15 with $a(n) \coloneqq \mu(n)$, $f(x) \coloneqq \log x$ and $y \coloneqq 1$ we obtain

$$H(x) = \sum_{\substack{n\in\mathbb{N}\\ n\le x}} \mu(n)\log n = M(x)\log x - \int_1^x \frac{M(t)}{t}\,\mathrm{d}\,t.$$

Since $x > 1$ this implies

$$\frac{M(x)}{x} - \frac{H(x)}{x\log x} = \frac{1}{\log x}\int_1^x \frac{M(t)}{t}\,\mathrm{d}\,t.$$

Hence it remains to show that $\frac{1}{\log x}\int_1^x \frac{M(t)}{t}\,\mathrm{d}\,t \xrightarrow[x\to\infty]{} 0$. But this is immediate from

$$\int_1^x \frac{M(t)}{t}\,\mathrm{d}\,t = \mathrm{O}\left(\int_1^x \mathrm{d}\,t\right) = \mathrm{O}(x)$$

as $x \to \infty$, which is a consequence of $M(x) = \mathrm{O}(x)$ as $x \to \infty$. $\square$

Now we are ready to prove the claimed equivalence.

Proof of Theorem 2.5.
i) implies ii):

From Theorem 2.19 we know that *i)* is equivalent to $\psi(x) \sim x$. We aim to show that $\frac{H(x)}{x \log x} \xrightarrow[x\to\infty]{} 0$, with H as in Lemma 2.21, to obtain *ii)* using Lemma 2.21. In Proposition 2.11 we found

$$\Lambda(n) = -\sum_{d|n} \mu(d) \log d.$$

By applying MÖBIUS inversion (Theorem 2.8) to that we obtain

$$-\mu(n) \log n = \sum_{d|n} \mu(d) \Lambda\left(\frac{n}{d}\right).$$

Summing over all $n \in \mathbb{N}$ with $n \leq x$ and using Proposition 2.13 with $f = \mu$ and $g = \Lambda$ we find

$$-H(x) = -\sum_{\substack{n\in\mathbb{N}\\ n\leq x}} \mu(n) \log n = \sum_{\substack{n\in\mathbb{N}\\ n\leq x}} \mu(n) \psi\left(\frac{x}{n}\right). \tag{2.4}$$

Now fix $\varepsilon > 0$. Since $\psi(x) \sim x$, there is a constant $A = A(\varepsilon) > 0$ just depending on ε such that $\left|\frac{\psi(x)}{x} - 1\right| < \varepsilon$ for each $x \geq A$. In other words,

$$|\psi(x) - x| < \varepsilon x, \tag{2.5}$$

whenever $x \geq A$. Choose $x > A$ and write

$$\sum_{\substack{n\in\mathbb{N}\\ n\leq x}} \mu(n) \psi\left(\frac{x}{n}\right) = \sum_{\substack{n\in\mathbb{N}\\ n\leq y}} \mu(n) \psi\left(\frac{x}{n}\right) + \sum_{\substack{n\in\mathbb{N}\\ y<n\leq x}} \mu(n) \psi\left(\frac{x}{n}\right),$$

where $y \coloneqq \left[\frac{x}{A}\right]$. In the first sum, because of $n \leq y \leq \frac{x}{A}$, we have $\frac{x}{n} \geq A$ and thus obtain from (2.5)

$$\left|\psi\left(\frac{x}{n}\right) - \frac{x}{n}\right| < \varepsilon \frac{x}{n},$$

whenever $n \leq y$. Hence

$$\begin{aligned}\sum_{\substack{n\in\mathbb{N}\\ n\leq y}} \mu(n) \psi\left(\frac{x}{n}\right) &= \sum_{\substack{n\in\mathbb{N}\\ n\leq y}} \mu(n) \left(\frac{x}{n} + \psi\left(\frac{x}{n}\right) - \frac{x}{n}\right) \\ &= x \sum_{\substack{n\in\mathbb{N}\\ n\leq y}} \frac{\mu(n)}{n} + \sum_{\substack{n\in\mathbb{N}\\ n\leq y}} \mu(n) \left(\psi\left(\frac{x}{n}\right) - \frac{x}{n}\right)\end{aligned}$$

and therefore

$$\begin{aligned}\left|\sum_{\substack{n\in\mathbb{N}\\ n\leq y}} \mu(n) \psi\left(\frac{x}{n}\right)\right| &\leq x \left|\sum_{\substack{n\in\mathbb{N}\\ n\leq y}} \frac{\mu(n)}{n}\right| + \sum_{\substack{n\in\mathbb{N}\\ n\leq y}} \underbrace{|\mu(n)|}_{\leq 1} \left|\psi\left(\frac{x}{n}\right) - \frac{x}{n}\right| < x + \varepsilon \sum_{\substack{n\in\mathbb{N}\\ n\leq y}} \frac{x}{n} \\ &< x + \varepsilon x \left(1 + \log y\right) < x + \varepsilon x + \varepsilon x \log x.\end{aligned} \tag{2.6}$$

In the second sum, because of $y < n \leq x$, we have $n \geq y+1$ and since $y \leq \frac{x}{A} < y+1$ also $\frac{x}{n} \leq \frac{x}{y+1} < A$. The inequality $\frac{x}{n} < A$ implies $\psi\left(\frac{x}{n}\right) < \psi(A)$. Hence, the sum is dominated by $x\psi(A)$. Together with (2.6) we conclude

$$|H(x)| = \left| \sum_{\substack{n \in \mathbb{N} \\ n \leq x}} \mu(n) \psi\left(\frac{x}{n}\right) \right| < x + \varepsilon x + \varepsilon x \log x + \psi(A) < (2 + \psi(A))\, x + \varepsilon x \log x,$$

if $\varepsilon < 1$. Thus, for $\varepsilon \in (0,1)$ we have

$$\frac{|H(x)|}{x \log x} < \frac{2 + \psi(A)}{\log x} + \varepsilon.$$

Now we can find a $B > A$ so that $x > B$ implies $\frac{2+\psi(A)}{\log x} < \varepsilon$. Hence, for $x > B$,

$$\frac{|H(x)|}{x \log x} < 2\varepsilon.$$

Thus $\frac{H(x)}{x \log x} \xrightarrow[x \to \infty]{} 0$, which because of Lemma 2.21 implies *ii)*.

ii) implies i):

First recall that for each $x \in [1, \infty)$ we have

- $[x] = \sum_{\substack{n \in \mathbb{N} \\ n \leq x}} 1$,
- $\psi(x) = \sum_{\substack{n \in \mathbb{N} \\ n \leq x}} \Lambda(n)$,
- $1 = \sum_{\substack{n \in \mathbb{N} \\ n \leq x}} \left[\frac{1}{n}\right]$.

Using Möbius inversion on these we obtain

- $1 = \sum_{d|n} \mu(n) \mathfrak{d}\left(\frac{n}{d}\right)$,
- $\Lambda(n) = \sum_{d|n} \mu(d) \log\left(\frac{n}{d}\right)$,
- $\left[\frac{1}{n}\right] = \sum_{d|n} \mu(d)$,

where $\mathfrak{d}(n)$ denotes the number of divisors of n. Define $f : \mathbb{N} \to \mathbb{R}$ by

$$f(n) \coloneqq \mathfrak{d}(n) - \log n - 2\mathfrak{C},$$

where $\mathfrak{C}$ denotes Euler's constant. Then

$$\begin{aligned}
[x] - \psi(x) - 2\mathfrak{C} &= \sum_{\substack{n \in \mathbb{N} \\ n \leq x}} \left(1 - \Lambda(n) - 2\mathfrak{C}\left[\frac{1}{n}\right]\right) \\
&= \sum_{\substack{n \in \mathbb{N} \\ n \leq x}} \sum_{d|n} \mu(d) \left(\mathfrak{d}\left(\frac{n}{d}\right) - \log\left(\frac{n}{d}\right) - 2\mathfrak{C}\right) \\
&= \sum_{\substack{q,d \in \mathbb{N} \\ qd \leq x}} \mu(d) \left(\mathfrak{d}(q) - \log q - 2\mathfrak{C}\right) \\
&= \sum_{\substack{q,d \in \mathbb{N} \\ qd \leq x}} \mu(d) f(q).
\end{aligned}$$

This implies

$$\psi(x) - x + \sum_{\substack{q,d\in\mathbb{N}\\ qd\leq x}} \mu(d)f(q) = \mathrm{O}\left(1\right)$$

as $x \to \infty$. Hence it remains to show that $\frac{1}{x}\sum_{\substack{q,d\in\mathbb{N}\\ qd\leq x}} \mu(d)f(q) \xrightarrow[x\to\infty]{} 0$. To do so we make use of Proposition 2.14 and write

$$\sum_{\substack{q,d\in\mathbb{N}\\ qd\leq x}} \mu(d)f(q) = \sum_{\substack{n\in\mathbb{N}\\ n\leq b}} \mu(d)F\left(\frac{x}{n}\right) + \sum_{\substack{n\in\mathbb{N}\\ n\leq a}} f(n)M\left(\frac{x}{n}\right) - F(a)M(b), \tag{2.7}$$

where $a, b \in (0,\infty)$ such that $ab = x$ and $F(x) \coloneqq \sum_{\substack{n\in\mathbb{N}\\ n\leq x}} f(n)$.

Now, from Lemma 2.20 we know that

$$\sum_{\substack{n\in\mathbb{N}\\ n\leq x}} \mathfrak{d}(n) - x\log x - (2\mathfrak{C}-1)\,x = \mathrm{O}\left(\sqrt{x}\right).$$

Together with

$$\sum_{\substack{n\in\mathbb{N}\\ n\leq x}} \log n = \log \prod_{\substack{n\in\mathbb{N}\\ n\leq x}} n = \log\left([x]!\right) = x\log x - x + \mathrm{O}\left(\log x\right)$$

this yields

$$\begin{aligned} F(x) = \sum_{\substack{n\in\mathbb{N}\\ n\leq x}} f(n) &= \sum_{\substack{n\in\mathbb{N}\\ n\leq x}} \mathfrak{d}(n) - \sum_{\substack{n\in\mathbb{N}\\ n\leq x}} \log n - 2\mathfrak{C}\sum_{\substack{n\in\mathbb{N}\\ n\leq x}} 1 \\ &= x\log x + (2\mathfrak{C}-1)\,x + \mathrm{O}\left(\sqrt{x}\right) - \left(x\log x - x + \mathrm{O}\left(\log x\right)\right) - 2\mathfrak{C}x + \mathrm{O}(1) \\ &= \mathrm{O}\left(\sqrt{x}\right) + \mathrm{O}\left(\log x\right) + \mathrm{O}\left(1\right) = \mathrm{O}\left(\sqrt{x}\right). \end{aligned}$$

Hence there exists a constant $B > 0$ such that

$$|F(x)| \leq B\sqrt{x},$$

whenever $x \geq 1$. Applying this to the first sum on the right of (2.7) implies

$$\left|\sum_{\substack{n\in\mathbb{N}\\ n\leq b}} \mu(d)F\left(\frac{x}{n}\right)\right| \leq B\sum_{\substack{n\in\mathbb{N}\\ n\leq b}} \sqrt{\frac{x}{n}} \leq A\sqrt{xb} = \frac{Ax}{\sqrt{a}}$$

for some constant $A > B$. Now fix $\varepsilon > 0$ and choose $a > 1$ such that $\frac{A}{\sqrt{a}} < \varepsilon$. Then

$$\left|\sum_{\substack{n\in\mathbb{N}\\ n\leq b}} \mu(d)F\left(\frac{x}{n}\right)\right| < \varepsilon x, \tag{2.8}$$

for $x \geq 1$. Note that a depends on ε but not on x.

From *ii*) we deduce, that there exists a constant $D = D(\varepsilon) > 0$ such that for any $K > 0$ we have

$$x > D \Longrightarrow \frac{|M(x)|}{x} < \frac{\varepsilon}{K}.$$

The second sum on the right of (2.7) satisfies

$$\left|\sum_{\substack{n\in\mathbb{N}\\ n\le a}} f(n)M\left(\frac{x}{n}\right)\right| \le \sum_{\substack{n\in\mathbb{N}\\ n\le a}} |f(n)|\,\frac{\varepsilon x}{Kn} = \frac{\varepsilon x}{K}\sum_{\substack{n\in\mathbb{N}\\ n\le a}} \frac{|f(n)|}{n}$$

provided $\frac{x}{n} > D$ for any $n \le a$, thus for each $x > aD$. By choosing $K \coloneqq \sum_{\substack{n\in\mathbb{N}\\ n\le a}} \frac{|f(n)|}{n}$ we obtain

$$\left|\sum_{\substack{n\in\mathbb{N}\\ n\le a}} f(n)M\left(\frac{x}{n}\right)\right| \le \varepsilon x, \tag{2.9}$$

whenever $x > aD$.

Finally, we have

$$|F(a)M(b)| \le A\sqrt{a}\,|M(b)| < A\sqrt{a}b < \varepsilon\sqrt{a}\sqrt{a}b = \varepsilon x, \tag{2.10}$$

if $x > a^2$, since $ab = x$. Combining (2.7), (2.8), (2.9) and (2.10) yields

$$\left|\sum_{\substack{q,d\in\mathbb{N}\\ qd\le x}} \mu(d)f(q)\right| \le 3\varepsilon x,$$

whenever $x > \max\{a^2, aC\}$. Since a and D depend only on ε, we obtain

$$\lim_{x\to\infty} \frac{1}{x}\sum_{\substack{q,d\in\mathbb{N}\\ qd\le x}} \mu(d)f(q) = 0,$$

which, as explained above, shows that $i)$ holds.

□

2.2. The Riemann Hypothesis is equivalent to (★★)

We write (★★) in the form

$$M(x) = \mathrm{O}\left(x^{\frac{1}{2}+\varepsilon}\right)$$

uniformly in $\varepsilon > 0$ as $x \to \infty$.

Remark 2.22. Recall the well-known functional equation for the Riemann Zeta-function: For each $s \in \mathbb{C}$, $0 < \Re(s) < 1$, we have

$$\zeta(1-s) = \frac{2}{(2\pi)^s}\cos\left(\frac{\pi s}{2}\right)\Gamma(s)\zeta(s),$$

where $\Gamma(s) \coloneqq \int_0^\infty t^{s-1}e^{-t}\,\mathrm{d}\,t$ is defined for all $s \in \mathbb{C}\setminus(\mathbb{Z}\setminus\mathbb{N})$ (Γ generalizes the factorial; note that $\Gamma(s+1) = s\Gamma(s)$). A proof of this equation can be found e.g. in [37].

Using this relation Riemann has shown that all non-trival zeros of ζ lie in the *critical stripe* $\{z \in \mathbb{C}\,|\,0 < \Re(z) < 1\}$. Moreover, from the functional equation we

see that whenever $\zeta(s) = 0$ then also $\zeta(1-s) = 0$. Hence all zeros in the critical stripe are symmetric with respect to the *critical line* $\left\{z \in \mathbb{C} \,\middle|\, \Re(z) = \frac{1}{2}\right\}$. Thus we can reword Conjecture 2.2 as the condition

$$(\zeta(s) = 0 \wedge \Re(s) \in (0,1)) \Longrightarrow \Re(s) = \frac{1}{2}.$$

Furthermore, because of the mentioned symmetry of the zeros, it would suffice to show that either $\left\{z \in \mathbb{C} \,\middle|\, 0 < \Re(z) < \frac{1}{2}\right\}$ or $\left\{z \in \mathbb{C} \,\middle|\, \frac{1}{2} < \Re(z) < 1\right\}$ are free of zeros of ζ.

2.2.1. A series representation for ζ^{-1}

Lemma 2.23. *Let $f : \mathbb{N} \to \mathbb{C}$ be multiplicative and so that $\sum_{n=1}^{\infty} f(n)$ converges absolutely. Then*

$$S(f) \coloneqq \sum_{n=1}^{\infty} f(n) = \prod_{p \in \mathbb{P}} \sum_{k=0}^{\infty} f(p^k).$$

If f is totally multiplicative, then

$$S(f) = \prod_{p \in \mathbb{P}} \frac{1}{1 - f(p)}.$$

Proof. Since $\sum_{n=1}^{\infty} f(n)$ converges absolutely so does $\sum_{k=0}^{\infty} f(p^k)$ for each $p \in \mathbb{P}$. Moreover, for p sufficiently large,

$$0 < \sum_{k=0}^{\infty} f(p^k) \leq \sum_{k=0}^{\infty} \left|f(p^k)\right| \leq \sum_{\substack{n \in \mathbb{N} \\ n \geq p}} |f(n)| < 1.$$

Hence, for $x \in (0, \infty)$,

$$P(x) \coloneqq \sum_{\substack{p \in \mathbb{P} \\ p \leq x}} \sum_{k=0}^{\infty} f(p^k) = \sum_{n' \in \mathcal{N}_1} f(n'),$$

where $\mathcal{N}_1 \coloneqq \{n \in \mathbb{N} \,|\, \mathbb{P} \ni p | n \Rightarrow p \leq x\}$, and therefore

$$S(f) - P(x) = \sum_{n'' \in \mathcal{N}_2} f(n''),$$

where $\mathcal{N}_2 \coloneqq \mathbb{N} \setminus \mathcal{N}_1 = \{n \in \mathbb{N} \,|\, \exists p \in \mathbb{P} \colon\ (p | n \wedge p > x)\}$. Note that for each $n'' \in \mathcal{N}_2$ we have $n'' > x$. Thus for each $\varepsilon > 0$ there is an $x_0 = x_0(\varepsilon) \in (0, \infty)$ such that

$$|S(f) - P(x)| \leq \sum_{\substack{n \in \mathbb{N} \\ n > x}} |f(x)| \leq \varepsilon$$

for each $x \geq x_0$. Since $\sum_{p \in \mathbb{P}} \sum_{k=0}^{\infty} f(p^k)$ converges absolutely so does $\prod_{p \in \mathbb{P}} \sum_{k=1}^{\infty} f(p^k)$ and we conclude

$$\prod_{p \in \mathbb{P}} \sum_{k=1}^{\infty} f(p^k) = \lim_{x \to \infty} P(x) = S(f).$$

Now, if f is totally multiplicative then $\sum_{k=0}^{\infty} f(p^k) = \sum_{k=0}^{\infty} (f(p))^k$ and the assertion follows, since by absolute convergence we obtain $|f(p)| < 1$ for each $p \in \mathbb{P}$. $\square$

Theorem 2.24 (EULER's Formula)**.** *For $s \in \mathbb{C}$, $\Re(s) > 1$ we have*

$$\zeta(s) = \prod_{p \in \mathbb{P}} \frac{1}{1 - p^{-s}}.$$

Proof. Set $f(n) \coloneqq \frac{1}{n^s}$. Then f is totally multiplicative and $\sum_{n=1}^{\infty} f(n)$ converges absolutely for $\Re(s) > 1$. Hence Lemma 2.23 implies the assertion. □

Corollary 2.25. *For $s \in \mathbb{C}$, $\Re(s) > 1$ we have*

$$\sum_{n=1}^{\infty} \frac{\mu(n)}{n^s} = \frac{1}{\zeta(s)}.$$

Proof. Set $f(n) \coloneqq \frac{1}{n^s}\mu(n)$. Then f is multiplicative (cf. Proposition 2.4) and $\sum_{n=1}^{\infty} f(n)$ converges absolutely for $\Re(s) > 1$. Hence Lemma 2.23 implies

$$\sum_{n=1}^{\infty} f(n) = \prod_{p \in \mathbb{P}} \sum_{k=0}^{\infty} \frac{\mu(p^k)}{p^{ks}}.$$

We have

$$\frac{\mu(p^k)}{p^{ks}} = \begin{cases} 1 & \text{for } k = 0 \\ -\frac{1}{p^s} & \text{for } k = 1 \\ 0 & \text{for } k > 1 \end{cases}.$$

Thus

$$\sum_{n=1}^{\infty} \frac{\mu(n)}{n^s} = \prod_{p \in \mathbb{P}} \sum_{k=0}^{\infty} \frac{\mu(p^k)}{p^{ks}} = \prod_{p \in \mathbb{P}} \left(1 - \frac{1}{p^s}\right) = \left(\prod_{p \in \mathbb{P}} \left(1 - \frac{1}{p^{-s}}\right)\right)^{-1}.$$

Hence Theorem 2.24 yields the assertion. □

2.2.2. Outlining the Proof

From Corollary 2.25 we see that there cannot be any zeros of ζ in the half-plane $\{s \in \mathbb{C} \,|\, \Re(s) > 1\}$. If we could continue this series representation of ζ^{-1} onto $\left\{s \in \mathbb{C} \,\middle|\, \Re(s) > \frac{1}{2}\right\}$ the RIEMANN hypothesis would follow (because of the symmetry of the zeros of ζ in the critical stripe). This is the basic idea of the proof for the claimed equivalence.

Lemma 2.26 (LITTLEWOOD)**.** *We have*

$$\log \zeta(s) = \mathrm{O}\left((\log \Im(\mathrm{s}))^{2-2\Re(\mathrm{s})+\delta}\right)$$

uniformly in $\delta > 0$, whenever $\frac{1}{2} + \eta \leq \Re(s) \leq 1$ for some $\eta > 0$.

For a proof of Lemma 2.26 see e.g. Theorem 14.2 in [59].

Lemma 2.26 implies that for each $\varepsilon > 0$ there is a $t = t(\varepsilon) > 0$ such that for all $s \in \mathbb{C}$ with $\Im(s) > t$ we have

$$-\varepsilon \log \Im(s) < \log |\zeta(s)| < \varepsilon \log \Im(s).$$

Hence

$$\zeta(s) = \mathrm{O}\left(\Im(s)^{\varepsilon}\right) \tag{2.11}$$

$$\frac{1}{\zeta(s)} = \mathrm{O}\left(\Im(s)^{\varepsilon}\right) \tag{2.12}$$

uniformly in ε.

Lemma 2.27. *Let $s \in \mathbb{C}$, $\Re(s) > 1$ and $f(s) \coloneqq \sum_{n=1}^{\infty} \frac{a_n}{n^s}$, where $a_n = \mathrm{O}\left(g(n)\right)$ as $n \to \infty$ with $g : \mathbb{N} \to \mathbb{R}$ non-decreasing, as well as $\sum_{n=1}^{\infty} \left|\frac{a_n}{n^{\Re(s)}}\right| = \mathrm{O}\left(\left(\Re(s)-1\right)^{-\alpha}\right)$ as $\Re(s) \to 1$ with $\alpha > 0$. Then, for each $c > 0$, $x \in \mathbb{R} \setminus \mathbb{Z}$ and every $t \in \mathbb{R}$ we have*

$$\begin{aligned}\sum_{\substack{n \in \mathbb{N} \\ n < x}} \frac{a_n}{n^s} &= \frac{1}{2\pi i} \int_{c-it}^{c+it} f(s+w) \frac{x^w}{w} \,\mathrm{d}\, w + \mathrm{O}\left(\frac{x^c}{t\left(\Re(s)+c-1\right)^{\alpha}}\right) \\ &\quad + \mathrm{O}\left(\frac{1}{t} g(2x) x^{1-\Re(s)} \log x\right) + \mathrm{O}\left(\frac{g(m) x^{1-\Re(s)}}{t\left|x-m\right|}\right)\end{aligned}$$

as $x \to \infty$, where , $m \coloneqq \begin{cases} [x] & \text{if } x - [x] < \frac{1}{2} \\ [x]+1 & \text{otherwise} \end{cases}$.

For a proof of Lemma 2.27 see [59], Lemma 3.12.

Theorem 2.28. *The condition $M(x) = \mathrm{O}\left(x^{\frac{1}{2}+\varepsilon}\right)$ is equivalent to the* Riemann *hypothesis.*

Sketch of proof. First, assume that Riemann's hypothesis holds. Then by applying Lemma 2.27 with $a_n \coloneqq \mu(n)$, $f(s) \coloneqq \frac{1}{\zeta(s)}$, $c = 2$, $s = 0$ and $x = \frac{m}{2}$, for some odd $m \in \mathbb{Z}$, one deduces

$$\begin{aligned} M(x) = \sum_{\substack{n \in \mathbb{N} \\ n \leq x}} \frac{\mu(n)}{n^0} &= \frac{1}{2\pi i} \int_{2-it}^{2+it} \frac{1}{\zeta(w)} \frac{x^w}{w} \,\mathrm{d}\, w + \mathrm{O}\left(\frac{x^2}{t}\right) \\ &= \frac{1}{2\pi i} \int_{2-it}^{\frac{1}{2}+\delta-it} \frac{x^w}{w\zeta(w)} \,\mathrm{d}\, w + \frac{1}{2\pi i} \int_{\frac{1}{2}+\delta-it}^{\frac{1}{2}+\delta+it} \frac{x^w}{w\zeta(w)} \,\mathrm{d}\, w \\ &\quad + \frac{1}{2\pi i} \int_{\frac{1}{2}+\delta+it}^{2+it} \frac{x^w}{w\zeta(w)} \,\mathrm{d}\, w + \mathrm{O}\left(\frac{x^2}{t}\right) \\ &\overset{(2.12)}{=} \mathrm{O}\left(\int_{-t}^{t} (1+|w|)^{-1+\varepsilon} x^{\frac{1}{2}+\delta} \,\mathrm{d}\, w\right) + \mathrm{O}\left(t^{\varepsilon-1}x^2\right) + \mathrm{O}\left(\frac{x^2}{t}\right) \\ &= \mathrm{O}\left(t^{\varepsilon} x^{\frac{1}{2}+\delta}\right) + \mathrm{O}\left(x^2 t^{\varepsilon-1}\right), \end{aligned}$$

where $t \in \mathbb{R}$ and $\delta > 0$. Choosing $t \coloneqq x^2$ we obtain $M(x) = \mathrm{O}\left(x^{\frac{1}{2}+\varepsilon}\right)$ for $x = \frac{m}{2}$ and so generally.

Now assume that $M(x) = \mathrm{O}\left(x^{\frac{1}{2}+\varepsilon}\right)$ as $x \to \infty$. Then one shows by partial summation that $\sum_{n=1}^{\infty} \frac{\mu(n)}{n^s}$ converges for $\Re(s) > \frac{1}{2}$. By the symmetry of the zeros of ζ in the critical stripe, this implies Riemann's hypothesis. □

3. Entropy of Dynamical Systems

To understand Sarnak's conjecture we first have to understand the concept of entropy of a dynamical system. In short, entropy measures the amount of chaos in a given system. So a system with zero entropy appears to be deterministic in an a-priori sense and μ being orthogonal to any sequence realised in a deterministic dynamical system means, that μ does not act deterministically (or predictably) in any way.

There are various definitions of entropy requiring various presuppositions to the dynamical system in regard. We will consider the notion of topological entropy introduced by Adler–Konheim–McAndrew, where X just has to be a compact topological space, as well as the so called metrical entropy by Bowen–Dinaburg, that additionally requires a metric d on X. We will see that these two definitions are equivalent in case of a metric space. Finally, we take a look at the initial notion of entropy by Kolmogorov–Sinai - coming from the study of stochastic processes - which needs an invariant probability measure ν on X.

All results of this chapter, for which it is not indicated otherwise, are taken either from [12], [60], [2], [20] or [47].

3.1. Topological Entropy by Adler–Konheim–McAndrews

Let X be a compact topological space and $T : X \to X$ a continuous transformation on X. Then we call (X, T) a *topological dynamical system* (in short: TDS).

Since X is compact each of its open covers has a finite subcover. Let $\mathcal{U}$ be an open cover of X, denote by $N(\mathcal{U})$ the smallest possible (finite) number of sets of $\mathcal{U}$ sufficient to cover X and let $H(\mathcal{U}) := \log_2 N(\mathcal{U})$.[1] Furthermore, for two covers $\mathcal{U}$ and $\mathcal{V}$ denote by

$$\mathcal{U} \vee \mathcal{V} := \{U \cap V \,|\, U \in \mathcal{U}, V \in \mathcal{V}, U \cap V \neq \varnothing\}$$

their *common refinement*. In the same way define $\bigvee_{k=n_1}^{n_2} \mathcal{U}_k$ for finitely many open covers $\mathcal{U}_{n_1}, \ldots, \mathcal{U}_{n_2}$ of X, with $n_1, n_2 \in \mathbb{Z}$, $n_1 < n_2$.

To define the topological entropy of X we need some proper preparations.

Proposition 3.1. *Let $\mathcal{U}, \mathcal{V}$ be open covers of X. Then*

a) $H(\mathcal{U}) \geq 0$ *and* $H(\mathcal{U}) = 0$ *iff* $X \in \mathcal{U}$.

b) $H(\mathcal{U} \vee \mathcal{V}) \leq H(\mathcal{U}) + H(\mathcal{V})$.

c) $H(T^{-1}\mathcal{U}) \leq H(\mathcal{U})$ *for* $T \in C(X)$ *(with* $T^{-1}\mathcal{U} = \{T^{-1}U \,|\, U \in \mathcal{U}\}$*).*

[1] The choice of the base of logarithm is not essential, because its change only results in a constant scaling factor. One may think of the base 2 in view of storing information digitally.

Proof. *a*) This follows directly from the fact that $N(\mathcal{U})$ has to be a positive integer and $\log_2 x = 0 \Leftrightarrow x = 1$.

b) Let $\left\{U_1, \dots, U_{N(\mathcal{U})}\right\}$ be a minimal subcover of $\mathcal{U}$ and $\left\{V_1, \dots, V_{N(\mathcal{V})}\right\}$ be a minimal subcover of $\mathcal{V}$. Then $\{U_i \cap V_j \,|\, i \in [1, N(\mathcal{U})] \cap \mathbb{Z}, j \in [1, N(\mathcal{V})] \cap \mathbb{Z}\}$ is a (not necessarily minimal) subcover of $\mathcal{U} \vee \mathcal{V}$. Therefore, $N(\mathcal{U} \vee \mathcal{V}) \leq N(\mathcal{U})N(\mathcal{V})$ and the assertion follows.

c) Let $\left\{U_1, \dots, U_{N(\mathcal{U})}\right\}$ be a minimal subcover of $\mathcal{U}$. Then $\left\{T^{-1}U_1, \dots, T^{-1}U_{N(\mathcal{U})}\right\}$ is a cover of X, but possibly not minimal. Therefore, $N(T^{-1}\mathcal{U}) \leq N(\mathcal{U})$. □

Lemma 3.2. *Let $\mathcal{U}$ be an open cover of X. Then $H\left(\bigvee_{k=0}^{n-1} T^{-k}\mathcal{U}\right) \leq n \cdot H(\mathcal{U})$ for every $n \in \mathbb{N}$, and the limit $h(T,\mathcal{U}) \coloneqq \lim_{n\to\infty} \frac{1}{n} H\left(\bigvee_{k=0}^{n-1} T^{-k}\mathcal{U}\right)$ exists.*

Proof. Set $u_n \coloneqq H\left(\bigvee_{k=0}^{n-1} T^{-k}\mathcal{U}\right)$, for $n \in \mathbb{N}$. Then, by (b) and (c) of Proposition 3.1, $u_n \leq n \cdot H(\mathcal{U})$ and $u_{m+n} \leq u_m + u_n$, for all $m, n \in \mathbb{N}$. Fix m. Then, for each $n \in \mathbb{N}$, there are $l \in \mathbb{Z}, p \in [0, m) \cap \mathbb{Z}$ with $n = l \cdot m + p$. Therefore,

$$\frac{u_n}{n} = \frac{u_{lm+p}}{lm+p} \leq \frac{u_p}{lm} + \frac{u_{lm}}{lm} \leq \frac{u_p}{lm} + \frac{lu_m}{lm} \leq \frac{p}{lm} H(\mathcal{U}) + \frac{u_m}{m}.$$

For $n \to \infty$ also $l \to \infty$ and so $\limsup_{n\to\infty} \frac{u_n}{n} \leq \frac{u_m}{m}$. Because this is true for all $m \in \mathbb{N}$, it follows that

$$\limsup_{n\to\infty} \frac{u_n}{n} \leq \inf_{m\in\mathbb{N}} \frac{u_m}{m} \leq \liminf_{m\to\infty} \frac{u_m}{m}$$

which implies the convergence of $\left(\frac{u_n}{n}\right)_{n\in\mathbb{N}}$. □

Because of Lemma 3.2 the following definition is plausible.

Definition 3.3 (Adler–Konheim–McAndrews)**.** Let (X, T) be a TDS and let $\mathscr{X}$ be the set of all open covers of X. Then we call

$$h_{top}(T) \coloneqq \sup_{\mathcal{U}\in\mathscr{X}} h(T,\mathcal{U}) = \sup_{\mathcal{U}\in\mathscr{X}} \lim_{n\to\infty} \frac{1}{n} H\left(\bigvee_{k=0}^{n-1} T^{-k}\mathcal{U}\right)$$

the *(topological) entropy* of (X, T).

We call (X, T) *deterministic* if $h_{top}(T) = 0$.

3.2. Metric Entropy by Bowen–Dinaburg

Now let X be a compact metric space with metric $d : X \times X \to [0, +\infty)$ and $T : X \to X$ a continuous transformation on X.

Definition 3.4. For $n \in \mathbb{N}_0$ and $x, y \in X$ the map

$$d_n : X \times X \to [0, +\infty)\,,\ (x, y) \mapsto \max_{0\leq j<n} d(T^j x, T^j y)$$

is called the *n-th* Bowen *distance* between x and y.

Definition 3.5. For $\varepsilon > 0$ a set $M \subseteq X$ is called *(n,ε)-separated* if each pair of distinct $x, y \in M$ is more than ε apart in the metric d_n, i.e., $d_n(x, y) > \varepsilon$.

Lemma 3.6. *For X, T, d and d_n as before as well as for $n \in \mathbb{N}$ and $\varepsilon > 0$ we have*

a) Each (n,ε)-separated subset of X is finite.

b) Let $\varepsilon_1 > \varepsilon_2 > 0$ and let $n \in \mathbb{N}_0$. If $M \subseteq X$ is (n,ε_1)-separated then M is also (n,ε_2)-separated.

Proof. *a)* Let $\mathcal{U} = \{U_x \,|\, x \in X\}$ be an open cover of X with

$$U_x = B^{(n)}_{\frac{\varepsilon}{2}}(x) \coloneqq \left\{ y \in X \,\middle|\, d_n(x,y) < \frac{\varepsilon}{2} \right\}$$

for each $x \in X$. Now let M be an arbitrary (n,ε)-separated subset of X. Then a set $U_x \in \mathcal{U}$ contains at most one point of M. Since X is compact $\mathcal{U}$ must have a finite subcover, which retains the property, that each of its sets can not contain more than one point of M. Therefore M has to be finite itself.

b) Since M is (n,ε_1)-separated we have

$$\forall x, y \in M : d_n(x,y) > \varepsilon_1 > \varepsilon_2$$

and so M is also (n,ε_2)-separated. □

Because of part *a)* of Lemma 3.6 we can define:

$$h(T,\varepsilon) \coloneqq \limsup_{n\to\infty} \frac{1}{n} \log_2 s(T,n,\varepsilon)$$

for $\varepsilon > 0$ and $s(T,n,\varepsilon)$ the maximum cardinality of all (n,ε)-separated subsets of X (for fixed n and ε). Because of part *b)* of Lemma 3.6 $h(T,\varepsilon)$ is a monotonically decreasing function in ε. This allows the following definition.

Definition 3.7 (Bowen–Dinaburg)**.** Let (X,T) be a metric TDS (i.e., X a compact metric space and $T : X \to X$ continuous). Then we call

$$h_{met}(T) \coloneqq \lim_{\varepsilon\to 0+} h(T,\varepsilon) = \lim_{\varepsilon\to 0+} \limsup_{n\to\infty} \frac{1}{n} \log_2 s(T,n,\varepsilon)$$

the *(metric) entropy* of (X,T).

To ensure that $h_{met}(T)$ is well defined we have to examine whether $\lim_{\varepsilon\to 0+} h(T,\varepsilon)$ is independent from the chosen metric d.

Proposition 3.8. *Let d_1 and d_2 be two metrics on X, inducing the same topology. For $\varepsilon > 0$ define $h_{d_1}(T,\varepsilon)$ and $h_{d_2}(T,\varepsilon)$ as before, respectively corresponding to d_1 and d_2. Then*

$$\lim_{\varepsilon\to 0+} h_{d_1}(T,\varepsilon) = \lim_{\varepsilon\to 0+} h_{d_2}(T,\varepsilon).$$

Proof. Fix $\varepsilon > 0$ and consider the set $D_\varepsilon \coloneqq \{(x_1,x_2) \in X \times X \,|\, d_1(x_1,x_2) \geq \varepsilon\}$, which is closed and therefore compact in $X \times X$. Since $d_2 : X \times X \to [0,+\infty)$ is continuous there is a $(x_1^0, x_2^0) \in D_\varepsilon$ with

$$\inf_{(x_1,x_2)\in D_\varepsilon} d_2(x_1,x_2) = \min_{(x_1,x_2)\in D_\varepsilon} d_2(x_1,x_2) = d_2(x_1^0,x_2^0) =: \delta > 0$$

(since $d_1(x_1^0, x_2^0) \geq \varepsilon > 0$ and therefore $x_1^0 \neq x_2^0$). Hence

$$d_2(x_1, x_2) < \delta \implies d_1(x_1, x_2) < \varepsilon.$$

This holds for the BOWEN distances in accordance, too, which implies (in compliance with the monotonically behavior of $\varepsilon \mapsto h(T, \varepsilon)$)

$$\lim_{\varepsilon \to 0+} h_{d_2}(T, \varepsilon) \geq \lim_{\varepsilon \to 0+} h_{d_1}(T, \varepsilon).$$

Swapping the roles of d_1 and d_2 yields the assertion. □

Remark 3.9. Because of the monotony of the map $\varepsilon \mapsto h(T, \varepsilon)$ the identity

$$h_{met}(T) = \lim_{k \to \infty} h(T, \varepsilon_k)$$

holds for every null sequence $(\varepsilon_k)_{k \in \mathbb{N}}$. In some cases this may simplify the computing of metric entropy.

To show that the two definitions of entropy are equivalent in case of a compact metric space, we have to consider an alternative characterization of the metric entropy first.

Definition 3.10. For $\varepsilon > 0$ a set $N \subseteq X$ is called *(n,ε)-spanning* if for every $x \in X$ there is a $y \in N$ such that $d_n(x, y) \leq \varepsilon$.

Lemma 3.11. *For $\varepsilon > 0$ and $n \in \mathbb{N}$ denote by $r(T, n, \varepsilon)$ the minimum cardinality of all (n,ε)-spanning subset of the compact metric space X. Then $0 < r(T, n, \varepsilon) < +\infty$.*

Proof. Let N be an (n,ε)-spanning subset of X. Then $\mathcal{V} = \{V_y \,|\, y \in X\}$ with

$$V_y \coloneqq B_\varepsilon^{(n)}(y) = \{x \in X \,|\, d_n(x, y) < \varepsilon\}$$

is an open cover of X. Since X is compact $\mathcal{V}$ has a finite subcover. The set of the centers of the balls of this subcover is still an (n,ε)-spanning subset of X. Hence, each (n,ε)-spanning subset of X has a finite subset, which is (n,ε)-spanning itself. Therefore, $r(T, n, \varepsilon) < +\infty$. □

Theorem 3.12. *Let (X, T) be a metric TDS. Then*

$$h_{met}(T) = \lim_{\varepsilon \to 0+} \limsup_{n \to \infty} \frac{1}{n} \log_2 r(T, n, \varepsilon).$$

Proof. We show that for all $n \in \mathbb{N}$ and all $\varepsilon > 0$ we have

$$r(T, n, \varepsilon) \leq s(T, n, \varepsilon) \leq r(T, n, \frac{\varepsilon}{2}).$$

Let M be an (n,ε)-separated subset of X of maximum cardinality. Then M is also (n,ε)-spanning. This implies the first inequality. Now let $P \subseteq X$ be (n,ε)-separated and $Q \subseteq X$ (n,$\frac{\varepsilon}{2}$)-spanning. Define $\rho : P \ni x \mapsto \rho(x) \in Q$ with $d_n(x, \rho(x)) \leq \frac{\varepsilon}{2}$. Then ρ is injective, hence $\#P \leq \#Q$. □

Theorem 3.13. *Let X be a compact metric space with metric d and T a continuous transformation on X. Then*

$$h_{top}(T) = h_{met}(T).$$

Proof. Fix $\varepsilon > 0$ and let $\mathcal{U}$ be an open cover of X with

$$\operatorname{diam}(\mathcal{U}) \coloneqq \sup_{U\in\mathcal{U}} \operatorname{diam}(U) = \sup_{U\in\mathcal{U}} \sup\{d(x,y) \mid x,y \in U\} < \varepsilon.$$

Since two points of an (n,ε)-separated subset of X cannot be included in the same $U \in \mathcal{U}$, we have $s(T,n,\varepsilon) \leq N\left(\bigvee_{k=0}^{n-1} T^{-k}\mathcal{U}\right)$. Hence

$$h_{met}(T) = \lim_{\varepsilon\to 0+} \limsup_{n\to\infty} \frac{1}{n}\log_2 s(T,n,\varepsilon) \leq \sup_{\mathcal{U}\in\mathscr{X}} \lim_{n\to\infty} \frac{1}{n}\log_2\left(N\bigvee_{k=0}^{n-1} T^{-k}\mathcal{U}\right) = h_{top}(T).$$

Now let $\mathcal{V}$ be an open cover of X with Lebesgue number δ (see Theorem A.1 in the Appendix). For $n \in \mathbb{N}$ let S be an (n,ε)-spanning subset of X with $\#S = r(T,n,\varepsilon)$ (i.e., S is minimal). Because δ is a Lebesgue number of $\mathcal{V}$ for each $k \in [0,n) \cap \mathbb{Z}$ and each $y \in S$ there is a $V_{k,y} \in \mathcal{V}$ with $B_\delta(T^k y) \subseteq V_{k,y}$. For every $x \in X$ there is a $y \in S$ with $d(T^k x, y) \leq \delta$ and therefore $x \in T^{-k}B_\delta(y)$. Hence

$$x \in \bigcap_{k=0}^{n-1} T^{-k}V_{k,y}.$$

Therefore, $\left\{\bigcap_{k=0}^{n-1} T^{-k}V_{k,y} \,\middle|\, y \in S\right\}$ is a subcover of $\bigvee_{k=0}^{n-1} T^{-k}\mathcal{V}$ with cardinality not larger than $\#S$. Hence $N(\bigvee_{k=0}^{n-1} T^{-k}\mathcal{V}) \leq r(T,n,\varepsilon)$, which implies

$$h_{top}(T) = \sup_{\mathcal{U}\in\mathscr{X}} \lim_{n\to\infty} \frac{1}{n}\log_2 N\left(\bigvee_{k=0}^{n-1} T^{-k}\mathcal{U}\right) \leq \lim_{\varepsilon\to 0+} \limsup_{n\to\infty} \frac{1}{n}\log_2 r(T,n,\varepsilon) = h_{met}(T).$$

□

Remark 3.14. Because of Theorem 3.13 it is reasonable to set

$$h(T) \coloneqq h_{top}(T) = h_{met}(T)$$

and just speak of the *(topological) entropy* of a TDS (X,T) (referring also to the term given by Definition 3.7).

Now we want to prove some basic properties of topological entropy.

Proposition 3.15. *For (X,T) a TDS we have $h(T) \in [0,\infty]$.*

Proof. This follows immediately from $s(T,n,\varepsilon)$, $r(T,n,\varepsilon)$ or $N(\mathcal{U})$ being positive integers, for all $\varepsilon > 0$ and all open covers $\mathcal{U}$ of X. □

Proposition 3.16. *Let (Y,S) be a factor of (X,T), i.e., we have a continuous surjection $\phi : X \to Y$ with $\phi \circ T = S \circ \phi$. Then $h(S) \leq h(T)$.*

Proof. Let d be a metric on X and d' a metric on Y. For each $\varepsilon > 0$ we find a $\delta > 0$ with $\lim_{\varepsilon \to o} \delta = 0$ and $d(x_1, x_2) > \delta$ for $d'(\phi(x_1), \phi(x_2)) > \varepsilon$. The same holds for the corresponding BOWEN distances, for every $n \in \mathbb{N}$.

Now, for $n \in \mathbb{N}$, let $Q \subseteq Y$ be a (n,ε)-separated set with maximum cardinality, i.e., $\#Q = s(S, n, \varepsilon)$. Then $\phi^{-1}(Q) = \{x \in X \mid \phi(x) \in Q\}$ is (n,δ)-separated in X with $\#\phi^{-1}(Q) \leq s(T, n, \delta)$. Therefore

$$h(S) = \lim_{\varepsilon \to o} \limsup_{n \to \infty} \frac{1}{n} \log_2 s(S, n, \varepsilon) \leq \lim_{\delta \to 0} \limsup_{n \to \infty} \frac{1}{n} \log_2 s(T, n, \delta) = h(T). \qquad \square$$

Corollary 3.17. *Let (X, T) be a deterministic TDS. Then every factor of (X, T) is deterministic, too.*

Proof. Let (Y, S) be a factor of (X, T). Then, by the Propositions 3.15 and 3.16 we have

$$0 \leq h(S) \leq h(T) = 0. \qquad \square$$

Proposition 3.18. *Let (X, T) and (Y, S) be isomorphic TDS, i.e., we have a continuous bijection $\eta : Y \to X$ with $\eta \circ S = T \circ \eta$. Then $h(T) = h(S)$.*

Proof. Since η is a continuous bijection, so is η^{-1}. Therefore, (X, T) is a factor of (Y, S) and (Y, S) is a factor of (X, T). So, the assertion follows from Proposition 3.16. $\square$

Proposition 3.19. *Let $((X_k, T_k))_{k=1}^m$ be a finite sequence of TDS. Then*

$$h(T_1 \times \ldots \times T_m) = \sum_{k=1}^{m} h(T_k).$$

Proof. First consider the case $m = 2$. For d_k a metric on X_k, $k \in \{1, 2\}$, choose the metric $d((x_1, y_1), (x_2, y_2)) := \max\{d_1(x_1, x_2), d_2(y_1, y_2)\}$ on $X_1 \times X_2$. For $n \in \mathbb{N}$, $\varepsilon > 0$ and $k \in \{1, 2\}$ let $Q_k \subseteq X_k$ be a (n,ε)-separated set of maximum cardinality. Then $Q_1 \times Q_2$ is (n,ε)-separated in $X_1 \times X_2$. This implies

$$h(T_1) + h(T_2) \leq h(T_1 \times T_2).$$

Assume that $Q_1 \times Q_2$ is not of maximum cardinality. Then $Q_1 \times Q_2$ is contained in another (n,ε)-separated subset M of $X_1 \times X_2$ and we can find a $(x, y) \in M$ with $(x, y) \notin Q_1 \times Q_2$. Without loss of generality, let $y \notin Q_2$. Then, for each $(u, v) \in Q_1 \times Q_2$, we find a $j \in [0, n) \cap \mathbb{Z}$ with $d_1(T_1^j x, T_1^j u) > \varepsilon$ or $d_2(T_2^j y, T_2^j v) > \varepsilon$. For a fixed u the second inequality can not hold for every v, because otherwise $Q_2 \cup \{v\}$ would be (n,ε)-separated in X_2, which contradicts the maximum cardinality of Q_2. Therefore the first inequality holds for every u, which implies $x \in Q_1$ (because of the maximum cardinality of Q_1). For $u = x$ we obtain a contradiction. Hence, $Q_1 \times Q_2$ is of maximum cardinality and we conclude

$$h(T_1) + h(T_2) = h(T_1 \times T_2).$$

For $m > 2$ the assertion follows by induction. $\square$

To compute the topological entropy of a given dynamical system it is often useful to search for so called topological generators. The reason for that is provided by Lemma 3.21 below, which is a variant of SINAI's theorem about a similar correlation in terms of the KOLMOGOROV–SINAI entropy we consider in the next section.

Definition 3.20. Let (X,T) be a metric TDS and let $\mathcal{G}$ be a finite open cover of X. Then $\mathcal{G}$ is called a *topological generator* for T, if for every map $\phi : \mathbb{Z} \to \mathcal{G}$ the set $\bigcap_{n\in\mathbb{Z}} T^{-n}\overline{\phi(n)}$ contains not more than one point of X (i.e. $\#\left(\bigcap_{n\in\mathbb{Z}} T^{-n}\overline{\phi(n)}\right) \leq 1$).

Lemma 3.21. *Let (X,T) be a TDS with metric d and $\mathcal{G}$ a topological generator for T. Then*

$$h(T) = h(T,\mathcal{G}).$$

Proof. First, let $\mathcal{V}$ be a finite open cover of X with LEBESGUE number δ. Then there has to be an $N \in \mathbb{N}$ with $\operatorname{diam}(\bigvee_{n=-N}^{N} T^{-n}\mathcal{G}) < \delta$, because otherwise, for every $j \in \mathbb{N}$, there would be $x_j, y_j \in X$ with $d(x_j,y_j) > \delta$ and a $\phi_j : [-j,j]\cap\mathbb{Z} \to \mathcal{G} = \{G_l\}$ with $x_j, y_j \in \bigcap_{i=-j}^{j} T^{-i}\phi_j(i)$. We could choose subsequences x_{j_k}, y_{j_k} with

$$x \coloneqq \lim_{k\to\infty} x_{j_k} \neq \lim_{k\to\infty} y_{j_k} =: y$$

since $d(x_j,y_j) > \delta$ for each $j \in \mathbb{N}$. Since $\mathcal{G}$ is finite, infinitely many of the sets $\phi_j(0)$ would have to coincide, and therefore, e.g., $x_{j_k}, y_{j_k} \in G_0$, for infinitely many k, which implies $x, y \in \overline{G_0}$. In the same way, for every $n \in [-j,j]$, infinitely many $\phi_{j_k}(n)$ would have to coincide and we obtain $x, y \in T^{-n}\overline{G_n}$, for some $G_n \in \mathcal{G}$. This would imply

$$\#\left(\bigcap_{n\in\mathbb{Z}} T^{-n}\overline{G_n}\right) \geq 2$$

in contradiction to the choise of $\mathcal{G}$ as a topological generator for T.

Now choose such an N. Then, since δ is a LEBESGUE number of $\mathcal{V}$, it follows from $\operatorname{diam}(\bigvee_{n=-N}^{N} T^{-n}\mathcal{G}) < \delta$ that

$$\begin{aligned}
h(T,\mathcal{V}) &\leq h\left(T, \bigvee_{k=-N}^{N} T^{-k}\mathcal{G}\right) \\
&= \lim_{n\to\infty} \frac{1}{n} H\left(\bigvee_{i=0}^{n-1} T^{-i}\left(\bigvee_{k=-N}^{N} T^{-k}\mathcal{G}\right)\right) \\
&= \lim_{n\to\infty} \frac{1}{n} H\left(\bigvee_{k=-N}^{N+n-1} T^{-k}\mathcal{G}\right) \\
&= \lim_{n\to\infty} \frac{1}{n} H\left(\bigvee_{k=0}^{2N+n-1} T^{-k}\mathcal{G}\right) \\
&= \lim_{n\to\infty} \frac{2N+n-1}{n} \cdot \frac{1}{2N+n-1} H\left(\bigvee_{k=0}^{2N+n-1} T^{-k}\mathcal{G}\right) \\
&= h(T,\mathcal{G}).
\end{aligned}$$

Therefore $h(T,\mathcal{V}) \leq h(T,\mathcal{G})$ for all open covers $\mathcal{V}$ of X. Since $\mathcal{G}$ itself is an open cover of X, we obtain

$$h(T,\mathcal{G}) = \sup_{\mathcal{U}\in\mathscr{X}} h(T,\mathcal{U}) = h(T). \qquad \square$$

3.3. Measure-Theoretic Entropy by Kolmogorov–Sinai

The first attempt to generalize the notion of entropy, known from probability theory and (originally) thermodynamics, dates back to the work of Kolmogorov. But it was not until Sinai's investigation that it was certain that this term is nontrivial.[2] We want to give a brief introduction to Kolmogorov's concept and demonstrate the correlation to the other notions of entropy.

In what follows let (X, Σ_X, ν) be a probability space, with $\nu : \Sigma_X \to [0,1]$ a complete measure (i.e., $\forall A \in \mathcal{N}_\nu \forall B \subseteq A : B \in \Sigma_X$, where $\mathcal{N}_\nu$ denotes the set of all ν-nullsets) and Σ_X countably generated, and let $T : X \to X$ be a measurable transformation which preserves the measure ν, i.e., $\forall A \in \Sigma_X : \nu(T^{-1}A) = \nu(A)$ (in that case we also call ν invariant under T). Then we call (X, Σ_X, ν, T) a *measure-preserving dynamical system* (in short: MDS).

Definition 3.22. Let $\mathcal{Q} \subseteq \Sigma_X$ be a finite partition of X, i.e., $\mathcal{Q} = \{Q_1, \dots, Q_r\}$ with $Q_i \cap Q_j = \varnothing$, for all $i, j \in [1, r] \cap \mathbb{Z}$, $i \neq j$, and $X = \bigcup_{k=1}^{r} Q_k$. Then we call

$$H_\nu(\mathcal{Q}) \coloneqq -\sum_{k=1}^{r} \nu(Q_k) \log_2 \nu(Q_k)$$

the *entropy of the partition* $\mathcal{Q}$.

Furthermore, let $\mathcal{P}, \mathcal{Q} \subseteq \Sigma_X$ be finite partitions of X and denote by $\Sigma(\mathcal{Q})$ the smallest σ-algebra which contains $\mathcal{Q}$. Then we call

$$H_\nu(\mathcal{P}|\mathcal{Q}) \equiv H_\nu(\mathcal{P}|\Sigma(\mathcal{Q})) \coloneqq -\sum_{Q \in \mathcal{Q}} \nu(Q) \sum_{P \in \mathcal{P}} \left(\frac{\nu(P \cap Q)}{\nu(Q)} \right) \log_2 \left(\frac{\nu(P \cap Q)}{\nu(Q)} \right)$$

the *conditional entropy of the partition* $\mathcal{P}$ *given* $\mathcal{Q}$.

Proposition 3.23. *Let $\mathcal{P}, \mathcal{Q} \subseteq \Sigma_X$ be finite partitions of X and denote by*

$$\mathcal{P} \vee \mathcal{Q} \coloneqq \{P \cap Q \mid P \in \mathcal{P}, Q \in \mathcal{Q}, \nu(P \cap Q) > 0\}$$

the common refinement of $\mathcal{P}$ and $\mathcal{Q}$. Then we have

$$H_\nu(\mathcal{P} \vee \mathcal{Q}) = H_\nu(\mathcal{P}) + H_\nu(\mathcal{P}|\mathcal{Q}).$$

For a proof of that see [52] (Theorem 4.1.).

Proposition 3.24. *For all finite partitions $\mathcal{P}, \mathcal{Q} \subseteq \Sigma_X$ we have*

$$H_\nu(\mathcal{P} \vee \mathcal{Q}) \leq H_\nu(P) + H_\nu(Q).$$

Proof. By Proposition 3.23 we have

$$H_\nu(\mathcal{P} \vee \mathcal{Q}) \leq H_\nu(P) + H_\nu(Q) \iff H_\nu(P|Q) \leq H_\nu(Q).$$

So we have to show that

$$\sum_{Q \in \mathcal{Q}} \nu(Q) \log_2 \nu(Q) \geq \sum_{Q \in \mathcal{Q}} \nu(Q) \sum_{P \in \mathcal{P}} \left(\frac{\nu(P \cap Q)}{\nu(Q)} \right) \log_2 \left(\frac{\nu(P \cap Q)}{\nu(Q)} \right).$$

[2] Sinai proved that the entropy of an automorphism of the two-dimensional torus is positive.

Consider the map $\varphi : t \mapsto -t \log_2 t$. Then the above inequality takes the form

$$\sum_{Q\in\mathcal{Q}} \varphi(\nu(Q)) \leq \sum_{Q\in\mathcal{Q}} \nu(Q) \sum_{P\in\mathcal{P}} \varphi\left(\frac{\nu(P\cap Q)}{\nu(Q)}\right),$$

which holds since φ is stricly concave. □

Definition 3.25 (Kolmogorov–Sinai)**.** Let $\mathscr{P}$ be the set of all measurable partitions of X. Then we call

$$h_\nu(T) \coloneqq \sup_{\mathcal{Q}\in\mathscr{P}} \lim_{N\to\infty} \frac{1}{N} H_\nu\left(\bigvee_{n=0}^{N} T^{-n}\mathcal{Q}\right),$$

with

$$\bigvee_{n=0}^{N} T^{-n}\mathcal{Q} \coloneqq \left\{ \bigcap_{n=0}^{N} T^{-n}Q_{i_n} \,\middle|\, Q_{i_j}\in\mathcal{Q},\, j\in[0,N]\cap\mathbb{Z} \text{ with } \nu\left(\bigcap_{n=0}^{N} T^{-n}Q_{i_n}\right) > 0 \right\},$$

the Kolmogorov–Sinai *entropy* or *measure-theoretic entropy* of (X, Σ_X, ν, T).

Remark 3.26. *a*) Since ν is a probability measure, $\nu(Q) \leq 1$ for all $Q \in \Sigma_X$. Hence, $\log_2 \nu(Q) \leq 0$. Therefore, the minus causes H_ν to be non-negative. Also note the convention $0 \cdot (\pm\infty) = 0$.

b) The convergence of the sequence $(\frac{1}{N} H_\nu(\bigvee_{n=0}^{N} T^{-n}\mathcal{Q}))_{N\in\mathbb{N}}$, which is indispensable for the above definition, was first shown by Shannon–McMillan (see Proposition 4.4 of [52] for a shorter proof).

c) Let (X, T) be a TDS. Then we can always find a probability measure ν on (X, Σ_X), for which (X, Σ_X, ν, T) is an MDS (i.e., for every continuous transformation on a compact metric space there is a Borel probability measure invariant under this transformation). This is the statement of the Krylov–Bogolyubov theorem (see Theorem A.6 in the Appendix). Note that the measure in consideration does not have to be unique (e.g. every probability measure on X is id_X-invariant). For a given system (X, T) we denote by $\mathfrak{M}_T$ the set of all Borel probability measures on X invariant under T.

The following statement establishes the interrelation between the notions of entropy.

Theorem 3.27. *For every TDS (X, T) we have*

$$h(T) = \sup_{\nu\in\mathfrak{M}_T} h_\nu(T).$$

For a proof of Theorem 3.27 see e.g. [52] (Theorem 4.7. and Theorem 4.9.).

Note that, according to Theorem 3.27, for all deterministic TDS (X, T) and all $\nu \in \mathfrak{M}_T$ we have

$$0 \leq h_\nu(T) \leq \sup_{\xi\in\mathfrak{M}_T} h_\xi(T) = h(T) = 0.$$

3.4. Examples

In this section we show that for selected dynamical systems, for which we will show in Chapter 6 that Sarnak's conjecture holds, the assumption about being deterministic is satisfied.

3.4.1. The Thue–Morse Shift is Deterministic

In this subsection we mainly follow the work of FORYS done in [25].

Consider $X = \{0,1\}^{\mathbb{N}_0}$ and the (one-sided) shift $S : X \ni (x_0 x_1 \ldots) \mapsto (x_1 x_2 \ldots) \in X$. Then (X, S) is a TDS (X is compact in the product topology by TYCHONOFF's theorem (see Theorem A.2 in the Appendix)). We call $\{0,1\}$ an *alphabet* and each $x \in X$ a *word*. For $x \in X$ denote by $K_x := \overline{\{S^n x \,|\, n \in \mathbb{N}_0\}}$ the closed orbit under S of x. Then (K_x, S) is again a TDS, since K_x is closed and S-invariant (i.e., $S(K_x) \subseteq K_x$). We call (K_x, S) a *subshift* of (X, T). If x is almost periodic (i.e., for all open subsets U of X with $x \in U$ there is an $N \in \mathbb{N}$ so that for all $n \in \mathbb{N}$ we have $\{m \in \mathbb{N} \,|\, S^m \in U\} \cap [n, n+N] \neq \varnothing$), then the subshift is nontrivial, i.e., $K_x \neq X$.

For $X \ni x = (x_0 x_1 \ldots)$ and $k, l \in \mathbb{N}$ we call $(x_k x_{k+1} \ldots x_{k+l})$ a finite *subword* of length $l+1$ of x. Denote by $\mathcal{B}_n(x)$ the set of all subwords of length n of x. Then we can simplify the notion of topological entropy for such systems in the following way.

Proposition 3.28 ([17])**.** *For each $x \in X$ we have*

$$h(S|_{K_x}) = \lim_{n\to\infty} \frac{1}{n} \log_2 \#\mathcal{B}_n(x).$$

Proof. Let $\mathcal{P}$ be the partition over the *0*th coordinate. Then $\mathcal{P}$ is open in the product topology and thus a cover of K_x. The same holds for $\bigvee_{k=0}^{n-1} S^{-k}\mathcal{P}$. Since this cover consists of disjoint sets, subcovers can only be obtained by removing empty sets. Hence,

$$H\left(\bigvee_{k=0}^{n-1} S^{-k}\mathcal{P}\right) = \log_2 \#\left(\bigvee_{k=0}^{n-1} S^{-k}\mathcal{P}\right)$$

(by not counting empty sets). On the other hand, $\#\bigvee_{k=0}^{n-1} S^{-k}\mathcal{P}$ equals the number of subwords of length n appearing in an arbitrary $y \in K_x$. Because of the definition of K_x, these are the subwords of length n appearing in x. Therefore, $\#\bigvee_{k=0}^{n-1} S^{-k}\mathcal{P} = \#\mathcal{B}_n(x)$ and thus

$$h(S, \mathcal{P}) = \lim_{n\to\infty} \frac{1}{n} \log_2 \#\mathcal{B}_n(x).$$

Now let $y, y' \in (\bigcap_{n\in\mathbb{Z}} S^{-n}P_n) \cap K_x$, for $P_n \in \mathcal{P}$. This means $y, y' \in S^{-n}P_n$ for every $n \in \mathbb{Z}$ and thus for every $n \in \mathbb{Z} \setminus \mathbb{N}$. This implies $y_n = y'_n$ for each such n and therefore, $y = y'$. Hence, $\mathcal{P}$ is a topological generator for S and the assertion follows by Lemma 3.21. □

Definition 3.29. Let $t \in X$ be the sequence defined by

$$\begin{aligned} t_0 &= 0 \\ t_{2n} &= t_n \\ t_{2n-1} &= 1 - t_n \end{aligned}$$

for all $n \in \mathbb{N}$. Then we call t the THUE–MORSE *sequence.*

Remark 3.30. *a*) Equivalently, t can be defined as the (unique) fixed point of the transformation $\rho : 0 \mapsto 01, 1 \mapsto 10$ with $t_0 = 0$. This implies, that none of the subwords 000 and 111 can be found in t.

b) Another possible procedure yielding t is given as follows: Set $\overline{b}$ to be the complementary word of b, which one receives by switching all 0's of b into 1's and

vice versa. Let $(f_n)_{n\in\mathbb{N}_0}$ be the sequence of finite words recursively defined by $f_0 = 0$, $f_{n+1} = f_n\overline{f_n}$. Then $t = \lim_{n\to\infty} f_n$ (where the limit is understood pointwise).

For both representations see e.g. [33].

We intend to show, that $h(S|_{K_t}) = 0$. For that purpose we need the following little preparation.

For finite subwords $a = (a_0 \dots a_n)$ and $b = (b_0 \dots b_m)$ denote by ab the word $(a_0 \dots a_n b_0 \dots b_m)$. Furthermore, denote by $|w|$ the length of a finite word w. Finally, let ϵ be the empty word. Then the following holds.

Lemma 3.31. *Let w be a subword of t with $|w| \geq 7$. Then there are $l, r \in \{\epsilon, 0, 1\}$, $k \in \mathbb{N}$ and $u \in \{01, 10\}^k$ so that*

$$w = lur$$

and this decomposition is unique.

Proof. The sequence t can be devided into subwords $01, 10$. Starting at the 0th position, every such subword appears at an even position $t_{2n}t_{2n+1}$, for an $n \in \mathbb{N}$. Pairs 00 and 11 can only occur between such subwords. Therefore, starting at the beginning of the sequence, t can also be devided into the subwords $0110, 1001$ of length 4.

Now, if a subword w contains only one of the blocks $00, 11$, than w is placed in the middle of one of the subwords of length 4. Therefore, w can be uniquely decomposed into blocks $01, 10$. Does w contain more than one of the blocks $00, 11$, the decomposition of t into subwords $01, 10$ can be used for w as well. If there are any leftovers, then they are of length at most 1 and have to be at the beginning or the end of w. This yields the assertion. □

The middle subword u in Lemma 3.31 consists entirely of blocks $01, 10$, so there has to be a subword v with $|u| = 2|v|$ and $\rho(v) = u$. This observation implies the following lemma.

Lemma 3.32. *Let w be a subword of t with $|w| \geq 7$. Then we have the unique decomposition*

$$w = l_0 \dots l_{k-1}\rho^k(u)r_{k-1} \dots r_0,$$

with $k \in \mathbb{N}$, $l_i, r_i \in \{\epsilon, \rho^i(0), \rho^i(1)\}$ and $u \in \{0, 1\}^h$ with $3 \leq h \leq 6$.

Proof. Lemma 3.31 yields the decomposition $w = l_0\rho(u_0)r_0$. We can find an $n \in \mathbb{N}$ such that w is a subword of $f_n = \rho(f_{n-1})$, where f_n is a member of the sequence given in Remark 3.30 *b*). Hence, u_0 is a subword of f_{n-1} and therefore a subword of t. Is $|u_0| \geq 7$ we can again apply Lemma 3.31 and get

$$w = l_0\rho(u_0)r_0 = l_0\rho(l_1'\rho(u_1)r_1')r_0 = l_0l_1\rho^2(u_1)r_1r_0.$$

This procedure can be repeated recursively as long as $|u_k| \geq 7$. □

Theorem 3.33. *The* Thuer–Morse *shift is deterministic, i.e., $h(S|_{K_t}) = 0$.*

Proof. We show that there is a $C > 0$ so that for all $n \in \mathbb{N}$ we have

$$\#\mathcal{B}_n(t) \leq C \cdot n^{2\log_2 3}.$$

Then, with Proposition 3.28 we are able to conclude

$$h(S|_{K_t}) = \lim_{n\to\infty} \frac{1}{n} \log_2 \#\mathcal{B}_n(t) \leq \lim_{n\to\infty} \frac{1}{n} \log_2 Cn^{2\log_2 3} = 0.$$

To find such a C fix an $n \in \mathbb{N}_0$. Then Lemma 3.32 yields

$$\#\mathcal{B}_n(t) \leq \# \left\{ w = l_0 \dots l_{k-1} \rho^k(u) r_{k-1} \dots r_0 \,\middle|\, l_i, r_i \in \left\{ \epsilon, \rho^i(0), \rho^i(1) \right\}, 3 \leq |u| \leq 6 \right\}.$$

The blocks l_i and r_i can take one out of three possible values, while also u can take just a finite number of values. Let $\frac{C}{2}$ denote this number.

Note that for a subword w of length h the exponent k is always smaller than $\log_2 h$ and for $i \in [0, k-1] \cap \mathbb{Z}$ we have

$$0 \leq |l_i| \leq 2^i \quad \text{and} \quad 0 \leq |r_i| \leq 2^i,$$

while $3 \leq |u| \leq 6$ implies

$$3 \cdot 2^k \leq |\rho^k(u)| \leq 6 \cdot 2^k.$$

Hence for a subword $w = l_0 \dots l_{k-1} \rho^k(u) r_{k-1} \dots r_0$ of length h we have

$$2^{k+1} < 3 \cdot 2^k \leq n \leq 2 \cdot \sum_{i=0}^{k-1} 2^i + 6 \cdot 2^k = 2 \cdot (2^k - 1) + 6 \cdot 2^k = 8 \cdot 2^k - 2 < 2^{k+3}.$$

Logarithmizing these inequalities yields

$$\log_2 n - 3 < k < \log_2 n - 1.$$

Therefore, since k is a positive integer, it can take at most

$$\#\left((\log_2 n - 3, \log_2 n - 1) \cap \mathbb{N} \right) \leq 2$$

values. Thus we can estimate

$$\#\mathcal{B}_n(t) \leq 2\frac{C}{2} \cdot 3^{2\log_2 n} = C \cdot n^{2\log_2 3}$$

and the assertion follows as shown above. □

3.4.2. Each Rotation on the Circle is Deterministic

First, consider the following statement.

Proposition 3.34. *Let X be a compact metric space and let $T \in C(X)$ be an isometry, i.e., for all $x, y \in X$ we have $d(Tx, Ty) = d(x, y)$. Then (X, T) is deterministic.*

Proof. Since T is an isometry on X, for each $n \in \mathbb{N}$ and each $x, y \in X$ we have

$$d_n(x, y) = \max_{0 \leq j < n} d(T^j x, T^j y) = \max_{0 \leq j < n} d(x, y) = d(x, y).$$

Therefore, the value $s(T, n, \varepsilon)$ does not depend on n and hence

$$h(T) = \lim_{\varepsilon \to 0+} \limsup_{n\to\infty} \frac{1}{n} \log_2 s(T, n, \varepsilon) = \lim_{\varepsilon \to 0+} 0 = 0.$$ □

Now, denote by $\mathbb{S}_1 \coloneqq \{z \in \mathbb{C} \mid |z| = 1\}$ the additive unit circle. Recall that $(\mathbb{S}_1, \cdot)$ is isomorphic to $(\mathbb{T}, +) = (\mathbb{R} \setminus \mathbb{Z}, +)$ and consider the map

$$R_\alpha : \mathbb{T} \to \mathbb{T},\ x \mapsto x + \alpha \mod 1,$$

where $\alpha \in \mathbb{R}$.[3] We call R_α a *rational* or an *irrational rotation* on $\mathbb{T}$ (through the angle α), depending on α respectively being rational or irrational. Since R_α is continuous, $(\mathbb{T}, R_\alpha)$ is a TDS.

Lemma 3.35. *For each $\alpha \in \mathbb{R}$ we have*

$$h(R_\alpha) = 0.$$

Proof. Consider $d : \mathbb{T} \times \mathbb{T} \to [0, \infty)$, $(x, y) \mapsto \min\{|x - y|, 1 - |x - y|\}$. Then d is a metric on $\mathbb{T}$, inducing the circle topology.[4] Since

$$\begin{aligned} d(R_\alpha x, R_\alpha y) &= \min\{|R_\alpha x - R_\alpha y|, 1 - |R_\alpha x - R_\alpha y|\} \\ &= \min\{|x - y|, 1 - |x - y|\} \\ &= d(x, y), \end{aligned}$$

R_α is an isometry on $\mathbb{T}$ and the assertion follows from Proposition 3.34. □

3.4.3. Each Skew Product Extension of a Rotation is Deterministic

Consider an arbitrary TDS (X, T) as well as a continuous map $\phi : X \to \mathbb{T}$. Then we call (Y, S), where $Y = X \times \mathbb{T}$ and

$$S(x, u) \coloneqq (Tx, u + \phi(x)) \,(\mathrm{mod}\, 1),$$

the skew product extension of T by ϕ.

This notion was introduced by Anzai in [3]. The following identity was shown by Abramov in [1] (et al).

Theorem 3.36 ([1])**.** *Let (X, T), ϕ and (Y, S) be as before. Then*

$$h(S) = h(T).$$

Proof. If $\phi \equiv c \in \mathbb{T}$ we have $S = T \times R_c$, with R_c the rotation on $\mathbb{T}$ through the angle c. Hence, by Proposition 3.19 and Lemma 3.35, we obtain

$$h(S) = h(T) + h(R_c) = h(T)$$

and furthermore - also in the non-constant case - $h(T) \leq h(S)$.

Now let ϕ be non-constant. To show the reverse inequality we make use of the Kolmogorov–Sinai entropy. Therefore, let ν be a T-invariant Borel measure on X, λ the ordinary Lebesgue measure on $\mathbb{T}$ and ρ an S-invariant Borel measure on Y. Since ϕ is continuous it is measurable.

We need several partitions on the involved sets. Let $\mathcal{D} \coloneqq \{D_1, D_2, \ldots\}$ be a base[5] for the topology of X and, for $m \in \mathbb{N}$, let $\mathcal{D}_m$ denote the partition of X

[3] We speak of $\mathbb{S}_1$ and $\mathbb{T}$ synonymously.

[4] Note that, because of Proposition 3.8, every metric inducing the same topology is suitable.

[5] i.e., an open cover of X such that for any $D_i, D_j \in \mathcal{D}$ and any $x \in D_i \cap D_j$ there is a $D_k \in \mathcal{D}$ such that $D_i \cap D_j \supseteq D_k \ni x$.

generated by the sets $D_1, \dots, D_m$ as well as $\mathcal{F}_m$ the partition of Y generated by the sets $D_1 \times \mathbb{T}, \dots, D_m \times \mathbb{T}$ and $\mathcal{F}$ the partition of Y into intervals $x \times \mathbb{T}$, i.e., $\mathcal{F} = \{x \times \mathbb{T} | x \in X\}$. Furthermore, denote by $\mathcal{P}_r = \left\{\triangle_1^{(r)}, \dots, \triangle_r^{(r)}\right\}$ the partition of $\mathbb{T}$ into r equal parts and let $\mathcal{Q}_r$ be the partition of Y into the sets $X \times \triangle_j^{(r)}$, $j \in [1, r] \cap \mathbb{Z}$. Since $\mathcal{D}_m \leq \mathcal{D}_{m+1}$, for all $m \in \mathbb{N}$, and $\bigvee_{m\in\mathbb{N}} \mathcal{F}_m = \mathcal{F}$ (modulo nullsets), it follows that for any n and r we have

$$\lim_{m\to\infty} H_\rho\left(\bigvee_{j=0}^{n} S^{-j}\mathcal{Q}_r \middle| \mathcal{F}_m\right) = H_\rho\left(\bigvee_{j=0}^{n} S^{-j}\mathcal{Q}_r \middle| \mathcal{F}\right)$$

(See [50], §1). The partition $\bigvee_{j=0}^{n} S^{-j}\mathcal{Q}_r$ induces, in each element $x \times \mathbb{T}$ of $\mathcal{F}$, a partition into not more than $n \cdot r$ intervals. Hence

$$\begin{aligned} H_\rho\left(\bigvee_{j=0}^{n} S^{-j}\mathcal{Q}_r \middle| \mathcal{F}\right) &= -\sum_{F\in\mathcal{F}} \rho(F) \sum_{E\in\bigvee_{j=0}^{n} S^{-j}\mathcal{Q}_r} \left(\frac{\rho(E\cap F)}{\rho(F)}\right) \log_2\left(\frac{\rho(E\cap F)}{\rho(F)}\right) \\ &\leq \log_2(nr). \end{aligned}$$

Let $\varepsilon > 0$. For each $r \in \mathbb{N}$ choose n_r such that $\frac{1}{n_r}\log_2(n_r r) < \frac{\varepsilon}{2}$ and denote by $\widetilde{m}_r$ the smallest of all $m \in \mathbb{N}$ for which

$$H_\rho\left(\bigvee_{j=0}^{n_r} S^{-j}\mathcal{Q}_r \middle| \mathcal{F}_m\right) < \log_2(n_r r) + \frac{\varepsilon}{2}.$$

Now define the sequence $(m_r)_{r\in\mathbb{N}_0}$ inductively by $m_0 = 1$, $m_r = \max\{m_{r-1}, \widetilde{m}_r, r\}$, for $r \in \mathbb{N}$. Then, for any $r, k \in \mathbb{N}$, one can show that (see [1])

$$\frac{1}{kn_r} H_\rho\left(\bigvee_{j=0}^{kn_r} S^{-j}\mathcal{Q}_r \middle| \bigvee_{j=0}^{kn_r} S^{-j}\mathcal{F}_{m_r}\right) \leq \varepsilon. \tag{3.1}$$

Furthermore, Proposition 3.23 yields

$$\begin{aligned} H_\rho\left(\bigvee_{j=0}^{kn_r} S^{-j}(\mathcal{F}_{m_r} \vee \mathcal{Q}_r)\right) &= H_\rho\left(\left(\bigvee_{j=0}^{kn_r} S^{-j}\mathcal{F}_{m_r}\right) \vee \left(\bigvee_{j=0}^{kn_r} S^{-j}\mathcal{Q}_r\right)\right) \\ &= H_\rho\left(\bigvee_{j=0}^{kn_r} S^{-j}\mathcal{F}_{m_r}\right) + H_\rho\left(\bigvee_{j=0}^{kn_r} S^{-j}\mathcal{Q}_r \middle| \bigvee_{j=0}^{kn_r} S^{-j}\mathcal{F}_{m_r}\right). \end{aligned}$$

Now divide this equality by kn_r and pass to the limit $k \to \infty$. Then, as a consequence of the identity $H_\rho(\bigvee_{j=0}^{n} S^{-j}\mathcal{F}_m) = H_\nu(\bigvee_{j=0}^{n} T^{-j}\mathcal{D}_m)$ and (3.1), we obtain

$$\limsup_{N\to\infty} \frac{1}{N} H_\rho\left(\bigvee_{n=0}^{N} S^{-n}\mathcal{Q}_{m_r}\right) \leq \limsup_{N\to\infty} \frac{1}{N} H_\nu\left(\bigvee_{n=0}^{N} T^{-n}\mathcal{D}_{m_r}\right) + \varepsilon.$$

Since ε has been chosen arbitrarily, passing to the limit $r \to \infty$ yields

$$h_\rho(S) \leq h_\nu(T)$$

(See [50], §4), which also holds for the suprema, too, since ρ and ν have been chosen arbitrarily. Therefore,

$$h(S) = \sup_{\rho\in\mathfrak{M}_S} h_\rho(S) \leq \sup_{\nu\in\mathfrak{M}_T} h_\nu(T) = h(T). \qquad \square$$

Corollary 3.37. *For $\phi : \mathbb{T} \to \mathbb{T}$ continuous and $\alpha \in \mathbb{R}$ consider the transformation*

$$\mathbb{T}^2 \ni (x, y) \mapsto T_\phi(x, y) \coloneqq (x + \alpha, y + \phi(x)) \in \mathbb{T}^2.$$

Then

$$h(T_\phi) = 0.$$

Proof. Since T_ϕ is a skew product extension of the rotation R_α, by Theorem 3.36 and Lemma 3.35 we obtain

$$h(T_\phi) = h(R_\alpha) = 0. \qquad \square$$

4. Ergodic Decomposition

This chapter provides a tool we will need mainly for proving the ergodic theorem with MÖBIUS weights. Almost all of the results presented here are taken from [31].

Let (X, Σ_X, ν, T) be a measure-preserving dynamical system (MDS). We call ν *ergodic* if for each $A \in \Sigma_X$ the following implication holds

$$\nu(T^{-1}(A) \setminus A) = 0 \implies \nu(A) \in \{0, 1\}\,.$$

Sets for which $\nu(T^{-1}(A) \setminus A) = 0$ holds are called *invariant*. For an invariant set A its complement $X \setminus A$ is also invariant. Hence, for $0 < \nu(A) < 1$ the natural decomposition of (X, Σ_X, ν, T) into the two systems

$$(A, \Sigma_A, \nu_A, T\big|_A) \quad \text{and} \quad (X \setminus A, \Sigma_{X\setminus A}, \nu_{X\setminus A}, T\big|_{X\setminus A})$$

(with $\Sigma_A = A \cap \Sigma_X$ and $\nu_A : \Sigma_A \ni B \mapsto \frac{\nu(B)}{\nu(A)} \in [0, 1]$, where $A \in \Sigma_X$) arises. In this sense ergodic systems are "indecomposable" (since one of the two systems above would have measure zero). For a non-ergodic measure ν and an invariant set A the measures ν_A and $\nu_{X\setminus A}$ do not have to be ergodic either, but are - in some sense - closer to be ergodic than ν has been, since we eliminated potential invariant sets of positive measure $\neq 1$. Note that ν_A and $\nu_{X\setminus A}$ are supported on disjoint invariant sets (i.e., $\operatorname{supp}(\nu_A) \cap \operatorname{supp}(\nu_{X\setminus A}) = \varnothing$ for $\operatorname{supp}(\nu)$ defined as the set of all points x in X for which every open neighbourhood of x has positive measure[1]) and are mutually singular (i.e., $\exists B \in \Sigma_X : (\nu_A(X \setminus B) = 0 \wedge \nu_{X\setminus A}(B) = 0)$). So iterating this procedure yields a representation of ν as a combination of mutually singular measures supported on increasingly small disjoint invariant sets. So the question arises, if - by a somehow natured process of passing to a limit - we can hope to obtain a representation of ν as a combination of measures actually being ergodic.

Denote by $\mathfrak{M}_T$ the set of all probability measures on X invariant under the transformation T. Then one can characterize ergodic measures as the extreme points of $\mathfrak{M}_T$. In finite-dimensional spaces each point of a compact convex set M can be represented as a convex combination of the extreme points of M. One can also formulate infinite-dimensional versions of this fact (See CHOQUET theory). So what we are looking for is a representation of ν as a convex combination of the extreme points of the - in a certain sense - compact convex set $\mathfrak{M}_T$ (satisfying some other mild conditions). But here we choose a more measure-theoretic approach by studying measure integration and disintegration.

4.1. Measure Integration

Let (X, Σ_X) an (Y, Σ_Y) be measurable spaces. A family $\{\nu_x\}_{x\in X}$ of probability measures on (Y, Σ_Y) is called *measurable*, if for every $A \in \Sigma_Y$ the map $X \ni x \mapsto$

[1] It is not necessary to take the closure of this set, since the support of a measure is already closed in X as its complement is the union of the open sets of ν-measure 0.

$\nu_x(A) \in [0,1]$ is measurable with respect to Σ_X, or - equivalently - if for each bounded measurable function $f : Y \to \mathbb{R}$ the map $X \ni x \mapsto \int_Y f(y) \,\mathrm{d}\, \nu_x(y)$ is measurable. Denote by $\mathfrak{M}(X)$ the set of all probability measures on X.

Definition 4.1. For $\rho \in \mathfrak{M}(X)$, we define the *measure integration* ν of $\{\nu_x\}_{x \in X}$, which is a probability measure on Y, by

$$\nu(A) \coloneqq \int_X \nu_x(A) \,\mathrm{d}\, \rho(x),$$

where $A \in \Sigma_Y$, and we also write $\int_X \nu_x \,\mathrm{d}\, \rho(x)$ for ν.

For a bounded measurable function $f : Y \to \mathbb{R}$ Definition 4.1 yields

$$\int_Y f d\nu = \int_X \left(\int_Y f \,\mathrm{d}\, \nu_x \right) \mathrm{d}\, \rho(x).$$

The same holds, by approximation, for $f \in L^1(Y, \nu)$. Note that, although f is defined only on a set $E \in \Sigma_Y$ of full ν-measure, we have $\nu_x(E) = 1$ for ρ-a.e. x, so the integral $\int_Y f \,\mathrm{d}\, \nu_x$ is well defined ρ-a.e.

The following example accounts for the above definition to generalize convex combinations.

Example 4.2. Let X be a finite set and $\Sigma_X = 2^X$. Then we have

$$\int_X \nu_x \,\mathrm{d}\, \rho(x) = \sum_{x \in X} \rho(x) \cdot \nu_x$$

and any convex combination of measures on Y can be represented this way.

4.2. Measure Disintegration

We intend to reverse the above procedure to find a representation of any measure as such an integral. Of particular interest will be the decomposition of a measure with respect to a partition.

Example 4.3. Let (X, Σ_X, ν) be a probability space and let $\mathcal{P} = \{P_1, \ldots, P_n\} \subseteq \Sigma_X \setminus \mathcal{N}_\nu$ be a finite partition of X. For $x \in X$ denote by $\mathcal{P}(x)$ the unique $P_i \ni x$ and set

$$\nu_x \coloneqq \frac{1}{\nu(\mathcal{P}(x))} \nu\big|_{\mathcal{P}(x)}.$$

Then, for $A \in \Sigma_X$, we have

$$\begin{aligned} \int_X \nu_x(A) \,\mathrm{d}\, \nu(x) &= \int_X \frac{1}{\nu(\mathcal{P}(x))} \nu\big|_{\mathcal{P}(x)}(A) \,\mathrm{d}\, \nu(x) \\ &= \frac{1}{\nu(\mathcal{P}(x))} \nu(\mathcal{P}(x)) \nu(A) = \nu(A). \end{aligned}$$

We want to find a comparable decomposition with respect to an infinite (in most cases uncountable) partition $\mathcal{E}$ of X. But in this case most of the sets $E \in \mathcal{E}$ will have measure 0 and the formula $\frac{1}{\nu(E)} \nu\big|_E$ will no longer make sense. So we pass to the conditional probability of an event $E \ni x$: Define

$$\nu_x(E) \coloneqq \mathbb{E}_\nu(\mathbf{1}_E | \mathcal{E})(x)$$

for any countably generated algebra $\mathcal{E}$ and $E \in \mathcal{E}$. This yields a countably additive measure defined for ν-a.e. x. But we want to define $\nu_x(E)$ for all measurable sets, which will occupy the rest of this section.

In what follows let X be a compact metric space with BOREL σ-algebra Σ_X and let $\mathcal{E}$ be a countably generated sub-σ-algebra of Σ_X. For (Y, Σ_Y) from the above section we also take (X, Σ_X).

Proposition 4.4. *The map $X \ni x \mapsto \nu_x \in \mathfrak{M}(X)$ is $\mathcal{E}$-measurable and we have*

$$\mathbb{E}_\nu(\mathbf{1}_A|\mathcal{E})(x) = \nu_x(A) \ \ \nu\text{-a.e.},$$

for any $A \in \Sigma_X$ and ν the measure integration of $\{\nu_x\}_{x\in X}$.

Proof. Denote by $\mathcal{A} \subseteq \Sigma_X$ the family of all sets $A \subseteq X$ for which the assertion holds. We show that $\mathcal{A} = \Sigma_X$.

Denote by $\mathcal{A}_0 \subseteq \Sigma_X$ the family of all sets $A \subseteq X$ such that $\mathbf{1}_A$ has a representation as a pointwise limit of a uniformly bounded sequence $(f_n)_{n\in\mathbb{N}}$ of continuous functions. Then we have

- $\{X, \varnothing\} \in \mathcal{A}_0$,
- if $\lim_{n\to\infty} f_n = \mathbf{1}_A$ then $\lim_{n\to\infty}(1 - f_n) = \mathbf{1}_{X\setminus A}$,
- if $\lim_{n\to\infty} f_n = \mathbf{1}_A$ and $\lim_{n\to\infty} g_n = \mathbf{1}_B$ then $\lim_{n\to\infty} f_n g_n = \mathbf{1}_A\mathbf{1}_B = \mathbf{1}_{A\cap B}$.

Therefore, $\mathcal{A}_0$ is an algebra.

Now, for $\lim_{n\to\infty} f_n = \mathbf{1}_A$ and $\|f_n\|_\infty \leq C \in [0, \infty)$ we obtain

$$\lim_{n\to\infty} \int_X f_n \,\mathrm{d}\,\nu_x = \int_X \mathbf{1}_A \,\mathrm{d}\,\nu_x = \nu_x(A)$$

by dominated convergence. Thus, $x \mapsto \nu_x(A)$ is the pointwise limit of the sequence $(x \mapsto \int_X f_n \,\mathrm{d}\,\nu_x)_{n\in\mathbb{N}}$, which is a.e. identical with the sequence $(\mathbb{E}_\nu(f_n|\mathcal{E}))_{n\in\mathbb{N}}$. Therefore, $x \mapsto \nu_x(A)$ is measurable and a.e. identical to $\mathbb{E}_\nu(\mathbf{1}_A|\mathcal{E})$, since $\mathbb{E}(\cdot|\mathcal{E})$ is continuous in $L^1(X, \nu)$ and $(f_n)_{n\in\mathbb{N}}$ is uniformly bounded. Hence,

$$\mathcal{A}_0 \subseteq \mathcal{A}.$$

For $A \subseteq X$ closed and $n \in \mathbb{N}$ set $f_n(x) \coloneqq \exp(-n \cdot d(x, A))$, where $d(x, A) = \inf_{y\in A} d(x, y)$ and d the metric on X. Then $\lim_{n\to\infty} f_n = \mathbf{1}_A$ and therefore $A \in \mathcal{A}_0$. Hence $\mathcal{A}_0$ generates the BOREL σ-algebra Σ_X.

Now let $(A_j)_{j\in\mathbb{N}}$ with $A_j \subseteq A_{j+1}$ for all $j \in \mathbb{N}$ and $A \coloneqq \bigcup_{j\in\mathbb{N}} A_j$. Then $\nu_x(A) = \lim_{j\to\infty} \nu_x(A_j)$ and thus $x \mapsto \nu_x(A)$ is the pointwise limit of the measurable sequence $(x \mapsto \nu_x(A_j))_{j\in\mathbb{N}}$, which is the same as the sequence $(\mathbb{E}_\nu(\mathbf{1}_{A_j}|\mathcal{E}))_{j\in\mathbb{N}}$ and therefore, since $\lim_{j\to\infty} \left\|\mathbf{1}_{A_j} - \mathbf{1}_A\right\|_{L^1} = 0$, by continuity of the conditional expectation, we obtain $\lim_{j\to\infty} \left\|\mathbb{E}_\nu(\mathbf{1}_{A_j}|\mathcal{E}) - \mathbb{E}_\nu(\mathbf{1}_A|\mathcal{E})\right\|_{L^1} = 0$. Hence we have

$$\nu_x(A) = \mathbb{E}_\nu(\mathbf{1}_A|\mathcal{E}) \ \ \nu - \text{a.e.}$$

and thus $\mathcal{A}$ is a monotone class containing the algebra $\mathcal{A}_0$, which for its part generates Σ_X. By the monotone class theorem (see Appendix) we can conclude $\Sigma_X \subseteq \mathcal{A}$ and thus $\Sigma_X = \mathcal{A}$, which yields the assertion. □

Proposition 4.5. *For every $f \in L^1(X,\nu)$ we have $\mathbb{E}_\nu(f|\mathcal{E})(x) = \int_X f d\nu_x \;\; \nu - a.e.$*

Proof. By Propositon 4.4 the assertion holds for indicator functions. Since both sides of the equation are linear and continuous under monotone increasing sequences, approximation by simple functions yields the claim for positive functions and thus, by taking differences, for all f in $L^1(X,\nu)$. □

For $x, y \in X$ write $x \sim_\mathcal{E} y$ if $\mathbf{1}_E(x) = \mathbf{1}_E(y)$ for every $E \in \mathcal{E}$. Since $\mathcal{E}$ has been chosen to be generated by a countable family $\{E_n\}_{n\in\mathbb{N}}$, we have $x \sim_\mathcal{E} y$ if and only if $\mathbf{1}_{E_n}(x) = \mathbf{1}_{E_n}(y)$ for each $n \in \mathbb{N}$. Then $\sim_\mathcal{E}$ is an equivalence relation and its equivalence classes are measurable, being intersections of sequences F_n of the form $F_n \in \{E_n, X \setminus E_n\}$ respectively. We call the equivalence classes of $\sim_\mathcal{E}$ the *atoms* of $\mathcal{E}$ (not to be mistaken as the atoms of a measure).

For $\mathcal{E}$ as above and $x \in X$ we denote by $\mathcal{E}(x)$ the atom containing x, i.e., $\mathcal{E}(x) = [x]_{\sim_\mathcal{E}}$.

Proposition 4.6. *ν_x is $\nu - a.s.$ supported on $\mathcal{E}(x)$, i.e., $\nu_x(\mathcal{E}(x)) = 1 \;\; \nu - a.e.$*

Proof. For $E \in \mathcal{E}$ we have

$$\mathbf{1}_E(x) = \mathbb{E}_\nu(\mathbf{1}_E|\mathcal{E})(x) = \int_X \mathbf{1}_E \,\mathrm{d}\,\nu_x = \nu_x(E)$$

and therefore $\nu_x(E) = \mathbf{1}_E(x)$ a.e. Due to the choice of $\mathcal{E}$ there is a family $\{E_n\}_{n\in\mathbb{N}}$ which generates $\mathcal{E}$. Let $M \subseteq X$ be a set of full measure such that the above holds for all $x \in M$ and all E_n. For $x \in M$ and $n \in \mathbb{N}$ choose $F_n \in \{E_n, X \setminus E_n\}$ so that $\mathcal{E}(x) = \bigcap_{n\in\mathbb{N}} F_n$. The above implies $\nu_x(F_n) = 1$ for every $n \in \mathbb{N}$, and so we obtain $\nu_x(\mathcal{E}(x)) = 1$ for all $x \in M$, which yields the assertion. □

Theorem 4.7. *Let X be a compact metric space with* Borel *σ-algebra Σ_X and let $\mathcal{E}$ be a countably generated sub-σ-algebra of Σ_X. Then there is an $\mathcal{E}$-measurable family $\{\nu_x\}_{x\in X}$ in $\mathfrak{M}(X)$ such that ν_x is supported on $\mathcal{E}(x)$ and*

$$\nu = \int_X \nu_x \,\mathrm{d}\,\nu(x).$$

Proof. Let V be a countable dense $\mathbb{Q}$-linear subspace of $C(X)$ with $\mathbf{1}_X \in V$. For $f \in V$ let $\widetilde{f} \coloneqq \mathbb{E}_\nu(f|\mathcal{E})$. Since V is countable, there is a subset $X_0 \subseteq X$ of full ν-measure such that $\widetilde{f}$ is defined for each $f \in V$ and every $x \in X_0$. Furthermore, $f \mapsto \widetilde{f}$ is $\mathbb{Q}$-linear and positive on X_0 as well as $\tilde{\mathbf{1}}_X = \mathbf{1}_X$. Hence, for each $x \in X_0$ the function

$$\tilde{\Lambda}_x : V \to \mathbb{R},\; f \mapsto \widetilde{f}(x)$$

is a positive continuous $\mathbb{Q}$-linear functional on the normed space $(V, \|\cdot\|_\infty)$ (continuous, since by positivity of the conditional expectation we have $\left\|\widetilde{f}\right\|_\infty \leq \|f\|_\infty$). Therefore, for each $x \in X_0$, $\tilde{\Lambda}_x$ extends to a positive $\mathbb{R}$-linear functional $\Lambda_x : C(X) \to \mathbb{R}$ (note that $\Lambda_x \mathbf{1}_X = \tilde{\mathbf{1}}_X(x) = 1$). So, by the representation theorem of Riesz–Markov–Kakutani (see Theorem A.9 in the Appendix), for each $x \in X_0$ there is a $\nu_x \in \mathfrak{M}(X)$ such that

$$\Lambda_x f = \int_X f(x) \,\mathrm{d}\,\nu_x(x).$$

To ensure measurability, for $x \in X \setminus X_0$ set ν_x to be some fixed measure in $\mathfrak{M}(X)$. Then, by Propositons 4.5 and 4.6 the assertion follows. □

Remark 4.8. The $\mathcal{E}$-measurability of the family $\{\nu_x\}_{x\in X}$ implies that for each $x' \in \mathcal{E}(x)$ we have $\nu_{x'} = \nu_x$ for ν-a.e. $x \in X$. And since $\nu_x(\mathcal{E}(x)) = 1$, we obtain $\nu_{x'} = \nu_x$ for ν_x-a.e. x'.

The representation $\nu = \int_X \nu_x d\nu(x)$ is called the *disintegration* of ν over $\mathcal{E}$. The following statement assures that this representation is unique (in the measure-theoretic sense).

Lemma 4.9. *If $\{\nu'_x\}_{x\in X}$ is another family with the same properties, then for ν-a.e. $x \in X$ we have $\nu'_x = \nu_x$.*

Proof. For $f \in L^1(X,\nu)$ define $f' : x \mapsto \int_X f d\nu'_x$. Then f' is a bounded linear operator on $L^1(X,\Sigma_X,\nu)$ with $\mathrm{im}(f) \subseteq L^1(X,\mathcal{E},\nu)$, since, by Definition 4.1 and the choice of $\{\nu'_x\}_{x\in X}$,

$$\int_X |f'| \,\mathrm{d}\,\nu \le \int_X \left(\int_X |f| \,\mathrm{d}\,\nu'_x\right) \mathrm{d}\,\nu(x) = \int_X |f| \,\mathrm{d}\,\nu = \|f\|_{L^1}.$$

Furthermore,

$$\int_X f' \,\mathrm{d}\,\nu = \int_X \left(\int_X f d\nu'_x\right) \mathrm{d}\,\nu(x) = \int_X \left(\int_X f d\nu_x\right) \mathrm{d}\,\nu(x) = \int_X f \,\mathrm{d}\,\nu.$$

Due to the fact, that, for $E \in \mathcal{E}$, ν_x is supported on E for ν-a.e. $x \in E$ and on $X \setminus E$ for ν-a.e. $x \in X \setminus E$, for ν-a.e. $x \in X$ we obtain

$$(\mathbf{1}_E f)'(x) = \int_X \mathbf{1}_E f \,\mathrm{d}\,\nu'_x = \mathbf{1}_E(x) \int_X f \,\mathrm{d}\,\nu'_x = (\mathbf{1}_E f')(x).$$

Therefore, $f' = \mathbb{E}_\nu(f|\mathcal{E}) = \widetilde{f}$, which $\widetilde{f}$ as in the proof of Theorem 4.7, and the assertion follows. □

4.3. Ergodic Decomposition

Let (X,Σ_X,ν,T) be a metric MDS (with Σ_X the Borel σ-algebra on X) and let $\mathcal{T} \subseteq \Sigma_X$ be the family of measurable sets invariant under the transformation T. Then $\mathcal{T}$ is a σ-algebra but in general not countably generated (e.g. consider $T = \mathrm{id}_X$. Then $\mathcal{T} = \Sigma_X$ and thus not countably generated).

Therefore it is inevitable to pass over to a fixed countably generated ν-dense sub-σ-algebra $\mathcal{T}_0$ of $\mathcal{T}$ in the following way: Choose a dense sequence $(f_n)_{n\in\mathbb{N}} \subset L^1(X,\mathcal{T},\nu)$ by choosing representatives of the functions that are genuinely $\mathcal{T}$-measurable, not just modulo a set $C \in \mathcal{N}_\nu$ (note that $L^1(X,\mathcal{T},\nu)$ is a closed subspace of $L^1(X,\Sigma_X,\nu)$, and since $L^1(X,\Sigma_X,\nu)$ is separable, so is $L^1(X,\mathcal{T},\nu)$). Now for $p,q \in \mathbb{Q}$ consider the (countable) family $\mathcal{A}$ of the sets $A_{n,p,q} = \{p < f_n < q\}$ and let $\mathcal{T}_0 \coloneqq \sigma(\mathcal{A})$. Then $\mathcal{T}_0 \subseteq \mathcal{T}$ and each f_n is $\mathcal{T}_0$-measurable, hence $L^1(X,\mathcal{T}_0,\nu) = L^1(X,\mathcal{T},\nu)$. In particular, $\mathcal{T}$ is contained in the ν-completion of $\mathcal{T}_0$.

The following theorem yields the decomposition we are looking for.

Theorem 4.10 ([44])**.** *Let (X,Σ_X,ν,T) be a MDS on a* Borel *space and let $\mathcal{T}_0 \subseteq \mathcal{T} \subseteq \Sigma_X$ be as above. Then there is a disintegration $\nu = \int_X \nu_x d\nu(x)$ of ν over $\mathcal{T}_0$ (and therefore over $\mathcal{T}$) such that a.e. ν_x is T-invariant, ergodic and supported on $\mathcal{T}_0(x)$, and the disintegration is unique in the measure-theoretic sense, i.e., for any other family $\{\nu'_x\}_{x\in X}$ with the same properties we have $\nu_x = \nu'_x$ for ν-a.e. x.*

Proof. Let $\{\nu_x\}_{x\in X}$ be the (because of Lemma 4.9 unique) family yielding the disintegration of ν relative to $\mathcal{T}_0$ given by Theorem 4.7. It remains to show that for ν-a.e. $x \in X$ the measure ν_x is T-invariant and ergodic.

1. For ν-a.e. $x \in X$ the measure ν_x is T-invariant.
For $A \in \Sigma_X$ define $g_1 : X \to \mathbb{R}$ by

$$g_1(x) \coloneqq \nu_x(T^{-1}A) - \nu_x(A).$$

Then, for each $x \in X$, $g_1(x) = \int_X (\mathbf{1}_{T^{-1}A} - \mathbf{1}_A)\,\mathrm{d}\,\nu_x$ and hence g_1 is $\mathcal{T}$-measurable. Moreover, by Definition 4.1, for each $I \in \mathcal{T}$,

$$\begin{aligned}\int_I g_1(x)\,\mathrm{d}\,\nu(x) &= \int_I \left(\int_X (\mathbf{1}_{T^{-1}A} - \mathbf{1}_A)\,\mathrm{d}\,\nu_x\right)\mathrm{d}\,\nu(x) = \int_I (\mathbf{1}_{T^{-1}A}(x) - \mathbf{1}_A(x))\,\mathrm{d}\,\nu(x)\\ &= \nu\left(I \cap T^{-1}A\right) - \nu\left(I \cap A\right).\end{aligned}$$

Since I is T-invariant and T preserves the measure ν, we obtain

$$\nu\left(I \cap T^{-1}A\right) = \nu\left(T^{-1}I \cap T^{-1}A\right) = \nu\left(T^{-1}\left(I \cap A\right)\right) = \nu\left(I \cap A\right)$$

and therefore $\int_I g_1\,\mathrm{d}\,\nu = \nu\left(I \cap T^{-1}A\right) - \nu\left(I \cap A\right) = 0$, for every $I \in \mathcal{T}$. Since g_1 is $\mathcal{T}$-measurable (and thus $I \cap \mathcal{T}$-measurable, for each $I \in \mathcal{T}$) we conclude $g_1 = 0$ a.e., which implies $\nu_x(T^{-1}A) = \nu_x(A)$ for ν-a.e. $x \in X$, which yields assertion *1.*

2. For ν-a.e. x the measure ν_x is ergodic.
For $I \in \mathcal{T}$ define $g_2 : X \to \mathbb{R}$ by

$$g_2(x) \coloneqq \nu_x(I).$$

Then, for each $x \in X$, $g_2(x) = \int_X \mathbf{1}_I\,\mathrm{d}\,\nu_x$ and hence g_2 is $\mathcal{T}$-measurable. Furthermore, for each $J \in \mathcal{T}$,

$$\int_J g_2(x)\,\mathrm{d}\,\nu(x) = \int_J \left(\int_X \mathbf{1}_I\,\mathrm{d}\,\nu_x\right)\mathrm{d}\,\nu(x) = \int_J \mathbf{1}_I d\nu(x)$$

and thus $g_2 = \mathbf{1}_I$. Since $\mathcal{T}$ is the set of all T-invariant sets $I \in \Sigma_X$, we obtain that, for ν-a.e. $x \in X$ and each T-invariant set I,

$$\nu_x(I) = \begin{cases} 1 & \text{if } x \in I \\ 0 & \text{otherwise} \end{cases},$$

i.e., $\nu_x(I) \in \{0, 1\}$, which yields assertion *2.*

□

5. The Ergodic Theorem with Möbius Weights

In ergodic theory there are always two sides to a result about dynamical systems: a measure-theoretic version (stated for ν-a.e. point, for some measure ν) and a topological one (stated for all points). Most of the time the measure-theoretic formulation is easier to prove, but at the cost of loosing validity for ν-nullsets. Nevertheless proofing the measure-theoretical version of such a claim might be the first step in understanding the entire context.

In case of SARNAK's conjecture, which itself can be considered as a statement about certain sequences being orthogonal to $(\mu(n))_{n\in\mathbb{N}}$, its measure-theoretic version appears to be in form of a weighted ergodic thereom, whose weights are given by the MÖBIUS function. Surprisingly, the latter version does not demand the given system to be deterministic.

This chapter is dedicated to the proof of the mentioned weighted ergodic theorem. For that purpose some preliminaries are required, namely BIRKHOFF's pointwise ergodic theorem, DAVENPORT's estimation and the spectral theorem for bounded unitary operators on a separable HILBERT space.

Recall that for X a compact metric space and $T : X \to X$ a continuous transformation we also denote by T the KOOPMAN operator on $C(X)$ given by $(Tf)(x) \coloneqq f(Tx)$ for $x \in X$. Note that T is linear, positive, multiplicative, contractive and preserves conjugation.

5.1. The Pointwise Ergodic Theorem by Birkhoff

The statements of this section are taken from [20]. We start with an important consequence of the BOREL-CANTELLI lemma (see Theorem A.10 in the Appendix), which states that convergence in the L^p-norm ($1 \leq p \leq \infty$) forces pointwise convergence along a subsequence.

Proposition 5.1. *Let (X, Σ_X, ν) be a probability space and let $(f_n)_{n\in\mathbb{N}} \subset L^p(X, \nu)$, where $1 \leq p \leq \infty$, be convergent to an f in the L^p-norm. Then there is a sequence $(n_k)_{k\in\mathbb{N}} \subset \mathbb{N}$ such that $(f_{n_k})_{k\in\mathbb{N}}$ converges pointwise ν-a.e. to f.*

Proof. Since $(f_n)_{n\in\mathbb{N}}$ converges to f in the L^p-norm, we can choose $(n_k)_{k\in\mathbb{N}} \subset \mathbb{N}$ such that $\|f_{n_k} - f\|_{L^p}^p < \frac{1}{k^{2+p}}$ for each $k \in \mathbb{N}$. Then

$$\nu\left(\left\{x \in X \;\middle|\; |f_{n_k}(x) - f(x)| > \frac{1}{k}\right\}\right) < \frac{1}{k^2}.$$

By the BOREL-CANTELLI lemma we obtain for ν-a.e. x, that $|f_{n_k}(x) - f(x)| > \frac{1}{k}$ holds for only finitely many k. So $\lim_{k\to\infty} f_{n_k}(x) = f(x)$ for ν-a.e. $x \in X$. □

To obtain the desired statement of this section we need to consider two significant results of ergodic theory, namely the mean and the maximal ergodic theorem, which are of outstanding significance for the study of dynamical systems.

Theorem 5.2 (Mean Ergodic Theorem; VON NEUMANN)**.** *Let (X, Σ_X, ν, T) be an MDS and denote by π_T the orthogonal projection onto the closed subspace*

$$\operatorname{Fix} T \coloneqq \left\{ g \in L^2(X,\nu) \,\middle|\, Tg = g \right\} \subseteq L^2(X,\nu).$$

Then, for any $f \in L^2(X,\nu)$ the sequence $(\frac{1}{N}\sum_{n=0}^{N-1} T^n f)_{N\in\mathbb{N}}$ converges to $\pi_T(f)$ in the L^2-norm.

Proof. Let $B \coloneqq \{Tg - g \,|\, g \in L^2(X,\nu)\}$. For $f \in \operatorname{Fix} T$ we have

$$\langle f, Tg - g\rangle = \langle Tf, Tg\rangle - \langle f, g\rangle = 0,$$

so $f \in B^\perp$. For $f \in B^\perp$ we find

$$\langle Tg, f\rangle = \langle g, f\rangle$$

for all $g \in L^2(X,\nu)$. Therefore, $T^*f = f$, and thus (by the parallelogram identity)

$$\begin{aligned} \|Tf - f\|_{L^2} &= \langle Tf - f, Tf - f\rangle \\ &= \|Tf\|_{L^2}^2 - \langle f, Tf\rangle - \langle Tf, f\rangle + \|f\|_{L^2}^2 \\ &= 2\,\|f\|_{L^2}^2 - \langle T^*f, f\rangle - \langle f, T^*f\rangle \\ &= 0, \end{aligned}$$

which implies $f \in \operatorname{Fix} T$. Altogether we obtain $B^\perp = \operatorname{Fix} T$, which implies

$$L^2(X,\nu) = \operatorname{Fix} T \oplus \overline{B}.$$

So each $f \in L^2(X,\nu)$ can be decomposed as

$$f = \pi_T f + h,$$

with a unique $h \in \overline{B}$. Hence, it remains to show that $\frac{1}{N}\sum_{n=0}^{N-1} T^n h \xrightarrow{L^2} 0$ as $N \to \infty$, for each $h \in \overline{B}$. For $h = Tg - g \in B$ we obtain

$$\left\| \frac{1}{N} \sum_{n=0}^{N-1} T^n (Tg - g) \right\|_{L^2} = \frac{1}{N} \left\| T^N g - g \right\|_{L^2} \to 0 \ \text{ as } N \to \infty, \tag{5.1}$$

since $\sum_{n=0}^{N-1} T^n(Tg - g)$ is a telescoping sum. Now, for an arbitrary $h \in \overline{B}$, choose $(g_k)_{k\in\mathbb{N}} \subset L^2(X,\nu)$ with $h_k \coloneqq Tg_k - g_k \to h$ as $k \to \infty$. Then, for each $k \in \mathbb{N}$,

$$\left\| \frac{1}{N} \sum_{n=0}^{N-1} T^n h \right\|_{L^2} \le \left\| \frac{1}{N} \sum_{n=0}^{N-1} T^n (h - h_k) \right\|_{L^2} + \left\| \frac{1}{N} \sum_{n=0}^{N-1} T^n h_k \right\|_{L^2}. \tag{5.2}$$

Because of (5.1), for any fixed $\varepsilon > 0$, we can find l and N sufficiently large such that

$$\|h - h_l\|_{L^2} < \frac{\varepsilon}{2}$$

and

$$\left\| \frac{1}{N} \sum_{n=0}^{N-1} T^n h_l \right\|_{L^2} < \frac{\varepsilon}{2}.$$

Together with (5.2) these imply

$$\left\| \frac{1}{N} \sum_{n=0}^{N-1} T^n h \right\|_{L^2} \leq \varepsilon,$$

which yields the assertion, since ε has been chosen arbitrarily. □

Corollary 5.3. *Let (X, Σ_X, ν, T) be an MDS and let $f \in L^1(X, \nu)$. Then there exists an $\widetilde{f} \in L^1(X, \nu)$ such that*

$$\frac{1}{N} \sum_{n=0}^{N-1} f \circ T^n \xrightarrow[N\to\infty]{L^1} \widetilde{f}.$$

Proof. For $f \in L^1(X, \nu)$, and $N \in \mathbb{N}$ set $C_N(f) \coloneqq \frac{1}{N} \sum_{n=0}^{N-1} f \circ T^n$. By Theorem 5.2 for any $g \in L^\infty(X, \nu) \subseteq L^2(X, \nu)$ its averages $C_N(g)$ converge in $L^2(X, \nu)$ to some $\widetilde{g} \in L^2(X, \nu)$. Since $\|\cdot\|_{L^1} \leq \|\cdot\|_{L^2}$, we also have

$$C_N(g) \xrightarrow[N\to\infty]{L^1} \widetilde{g}. \tag{5.3}$$

Now let $f \in L^1(X, \nu)$, fix an $\varepsilon > 0$ and choose $g \in L^\infty(X, \nu)$ with $\|g - f\|_{L^1} < \frac{\varepsilon}{4}$ (which is possible for any $\varepsilon > 0$ since $L^\infty(X, \nu)$ is dense in $L^1(X, \nu)$). By taking averages for any $N \in \mathbb{N}$ we obtain

$$\|C_N(f) - C_N(g)\|_{L^1} < \frac{\varepsilon}{4}$$

and by (5.3) there is an $N_0 \in \mathbb{N}$ such that

$$\|C_N(g) - \widetilde{g}\|_{L^1} < \frac{\varepsilon}{4}$$

for all $N \geq N_0$. Hence, for all $N, N' \geq N_0$,

$$\begin{aligned}
\left\| \frac{1}{N} \sum_{n=0}^{N-1} f \circ T^n - \frac{1}{N'} \sum_{n=0}^{N'-1} f \circ T^n \right\|_{L^1} &= \|C_N(f) - C_{N'}(f)\|_{L^1} \\
&= \|C_N(f) - C_N(g) + C_N(g) - \widetilde{g} \\
&\quad + \widetilde{g} - C_{N'}(g) + C_{N'}(g) - C_{N'}(f)\|_{L^1} \\
&\leq \|C_N(f) - C_N(g)\|_{L^1} + \|C_N(g) - \widetilde{g}\|_{L^1} \\
&\quad + \|\widetilde{g} - C_{N'}(g)\|_{L^1} + \|C_{N'}(g) - C_{N'}(f)\|_{L^1} \\
&\leq \frac{\varepsilon}{4} + \frac{\varepsilon}{4} + \frac{\varepsilon}{4} + \frac{\varepsilon}{4} \\
&= \varepsilon.
\end{aligned}$$

So $(\frac{1}{N} \sum_{n=0}^{N-1} f \circ T^n)_{n\in\mathbb{N}}$ is a CAUCHY sequence in $L^1(X, \nu)$ and therefore, since each L^p-space is complete, converges to an $\widetilde{f} \in L^1(X, \nu)$. □

Lemma 5.4 (Maximal Inequality)**.** *Let (X, Σ_X, ν) be a probability space and let $U : L^1(X,\nu) \to L^1(X,\nu)$ be a positive linear operator with $\|U\| \leq 1$. For $f \in L^1(X,\nu)$ real-valued define recursively*

$$\begin{aligned} f_0 &= 0 \\ f_1 &= f \\ f_{n+1} &= \sum_{k=0}^{n} U^k f \end{aligned}$$

for all $n \in \mathbb{N}$, as well as $F_N \coloneqq \max\{f_n \mid n \in [0,N] \cap \mathbb{Z}\}$ for $N \in \mathbb{N}$ (all functions defined pointwise). Then, for $\mathcal{F}_X \coloneqq \{x \in X \mid \inf_{N\in\mathbb{N}} F_N(x) > 0\}$,

$$\int_{\mathcal{F}_X} f \,\mathrm{d}\,\nu \geq 0.$$

Proof. For each $N \in \mathbb{N}$ we have $F_N \in L^1(X,\nu)$ and $F_N \geq f_n$ for all $n \in [0,N] \cap \mathbb{Z}$. By U being positive and linear, we obtain

$$UF_N + f \geq Uf_n + f = f_{n+1}$$

for all $n \in [0,N] \cap \mathbb{Z}$. Hence

$$UF_N + f \geq \max_{n\in[1,N]\cap\mathbb{Z}} f_n.$$

For $x \in \mathcal{F}_X$ we have

$$F_N(x) = \max_{n\in[0,N]\cap\mathbb{Z}} f_n(x) = \max_{n\in[1,N]\cap\mathbb{Z}} f_n(x) \leq UF_N(x) + f(x),$$

since $f_0 = 0$. Therefore, for each $x \in \mathcal{F}_X$,

$$f(x) \geq F_N(x) - UF_N(x). \tag{5.4}$$

Since U is positive we have $UF_N(x) \geq F_N(x) > 0$ for all $x \in \mathcal{F}_X$. This implies, together with $\|U\| \leq 1$, (5.4) and the fact, that $F_N(x) = 0$ for $x \notin \mathcal{F}_X$,

$$\begin{aligned} \int_{\mathcal{F}_X} f \,\mathrm{d}\,\nu &\geq \int_{\mathcal{F}_X} F_N \,\mathrm{d}\,\nu - \int_{\mathcal{F}_X} UF_N \,\mathrm{d}\,\nu \\ &= \int_X F_N \,\mathrm{d}\,\nu - \int_{\mathcal{F}_X} UF_N \,\mathrm{d}\,\nu \\ &\geq \int_X F_N \,\mathrm{d}\,\nu - \int_X UF_N \,\mathrm{d}\,\nu \\ &= \|F_N\|_{L^1} - \|UF_N\|_{L^1} \\ &\geq 0 \end{aligned}$$

for all $N \in \mathbb{N}$. □

Theorem 5.5 (Maximal Ergodic Theorem)**.** *Let (X, Σ_X, ν, T) be an MDS and let $g \in L^1(X,\nu)$ be real-valued. For $c \in \mathbb{R}$ define*

$$E_c \coloneqq \left\{ x \in X \;\middle|\; \sup_{N\in\mathbb{N}} \frac{1}{N} \sum_{n=0}^{N-1} g(T^n x) > c \right\}.$$

Then

$$c\nu(E_c) \leq \int_{E_c} g \,\mathrm{d}\,\nu \leq \|g\|_{L^1}.$$

Moreover, for $A \in \Sigma_X$ such that $T^{-1}A = A$,

$$c\nu(E_c \cap A) \leq \int_{E_c \cap A} f \,\mathrm{d}\,\nu.$$

Proof. Let $f \coloneqq g - c$ and $Uf \coloneqq f \circ T$. Then, in the notation of Lemma 5.4,

$$E_c = \left\{ x \in X \;\middle|\; \sup_{N \in \mathbb{N}} \frac{1}{N} \sum_{n=0}^{N-1} g(T^n x) > c \right\} = \bigcup_{N \in \mathbb{N}_0} \{x \in X \mid F_N(x) > 0\}.$$

From Lemma 5.4 it follows that $\int_{E_c} f \,\mathrm{d}\,\nu \geq 0$ and therefore $\int_{E_c} g \,\mathrm{d}\,\nu \geq c\nu(E_c)$.

For the last statement apply the same argument to $f \coloneqq g - c$ on the measure-preserving system $(A, \Sigma_A, \frac{1}{\nu(A)}\nu|_A, T|_A)$, with $\Sigma_A \coloneqq \Sigma(\{B \cap A \mid B \in \Sigma_X\})$. □

Now we have everything together we need to prove Birkhoff's pointwise ergodic theorem. It describes the relationship between the space average of a function and its time average along the orbit of a typical point, i.e., except for those contained in a certain nullset.

Theorem 5.6 (Birkhoff)**.** *Let (X, Σ_X, ν, T) be an MDS and $f \in L^1(X, \nu)$ then*

$$\frac{1}{N} \sum_{n=0}^{N-1} f \circ T^n$$

converges ν-a.e. to a function $g \in L^1(X, \nu)$ with

$$\int_X g \,\mathrm{d}\,\nu = \int_X f \,\mathrm{d}\,\nu.$$

If (X, Σ_X, ν, T) is ergodic, then

$$g(x) = \int_X f \,\mathrm{d}\,\nu$$

for ν-a.e. $x \in X$.

Proof. Let f be real-valued (for a complex-valued function the claim then follows by deviding it into its real and imaginary part). For each $x \in X$ define

$$f^*(x) \coloneqq \limsup_{N \to \infty} \frac{1}{N} \sum_{n=0}^{N-1} f(T^n x),$$
$$f_*(x) \coloneqq \liminf_{N \to \infty} \frac{1}{N} \sum_{n=0}^{N-1} f(T^n x).$$

Then

$$\frac{1}{N} \sum_{n=0}^{N-1} f(T^n(Tx)) + \frac{1}{N} f(x) = \frac{1}{N} \sum_{n=0}^{N} f(T^n x) = \frac{N+1}{N} \left(\frac{1}{N+1} \sum_{n=0}^{N} f(T^n x) \right). \quad (5.5)$$

By taking the limit along a subsequence for which the left-hand side of (5.5) converges to its limit superior, this implies $f^* \geq f^* \circ T$. A limit along a subsequence for which the right-hand side of (5.5) converges to its limit superior shows $f^* \leq f^* \circ T$. Altogether we obtain $f^* = f^* \circ T$. A similar argument for f_* (by considering limit inferiors) yields $f_* = f_* \circ T$.

Now fix $a, b \in \mathbb{Q}$, $a > b$, and define

$$E_a^b \coloneqq \{x \in X \mid f_*(x) < b,\, f^*(x) > a\}\,.$$

Then $T^{-1}E_a^b = E_a^b$, since $h = h \circ T$ for $h \in \{f_*, f^*\}$. Moreover, $E_a^b \subseteq E_a$ with E_a defined as in Theorem 5.5 (with $c = a$ and $g = f$). Hence $E_a^b = E_a^b \cap E_a$ and by Theorem 5.5 we obtain

$$\int_{E_a^b} f \,\mathrm{d}\,\nu \geq a\nu(E_a^b). \tag{5.6}$$

Analogously, by replacing f by $-f$, we obtain

$$\int_{E_a^b} f \,\mathrm{d}\,\nu \leq b\nu(E_a^b). \tag{5.7}$$

Now

$$\bigcup_{\substack{a,b\in\mathbb{Q}\\ a>b}} E_a^b = \bigcup_{\substack{a,b\in\mathbb{Q}\\ a>b}} \{x \in X \mid f_*(x) < b,\, f^*(x) > a\} = \{x \in X \mid f_*(x) < f^*(x)\}\,,$$

while (5.6) and (5.7) show that $\nu(E_a^b) = 0$ for $a > b$. Therefore,

$$\nu\left(\bigcup_{\substack{a,b\in\mathbb{Q}\\ a>b}} E_a^b\right) = 0,$$

so $f_*(x) = f^*(x)$ ν-a.e. Thus, for $g \coloneqq f^*$,

$$g_N(x) \coloneqq \frac{1}{N}\sum_{n=0}^{N-1} f(T^n x) \to g(x) \;\; \nu\text{-a.e.} \tag{5.8}$$

By Corollary 5.3 we also know that

$$\lim_{n\to\infty} \left\| g_n - \widetilde{f} \right\|_{L^1} = 0 \tag{5.9}$$

for a certain $\widetilde{f} \in L^1(X, \Sigma_X, \nu)$. By Proposition 5.1 this implies the existence of a sequence $(n_k)_{k\in\mathbb{N}} \subset \mathbb{N}$ with $\lim_{k\to\infty} n_k = \infty$ for which

$$g_{n_k}(x) \to \widetilde{f}(x) \;\; \nu\text{-a.e.} \tag{5.10}$$

So by (5.8) and (5.10) we obtain $g = \widetilde{f}$ and hence, by (5.9), that the convergence in (5.8) does also happen in $L^1(X, \nu)$. Finally we also get

$$\int_X f \,\mathrm{d}\,\nu = \frac{1}{N}\int_X \sum_{n=0}^{N-1} f \circ T^n \,\mathrm{d}\,\nu = \int_X g \,\mathrm{d}\,\nu.$$

The last claim follows from the above by taking in consideration, that $\operatorname{Fix} T = \mathbb{C} \cdot \mathbf{1}_X$ (i.e., $Tf = f$ iff f is constant) whenever (X, Σ_X, ν, T) is ergodic. □

5.2. Davenport's Estimation

We want to study the behavior correlation of the MÖBIUS function with functions on the unit circle, for what we need is an estimation for the growth of sums $\sum_{n\leq x}\mu(n)e^{2\pi in\theta}$ for any angle θ and a real value x going to infinity. This is what this section is dedicated to and the results below were shown by DAVENPORT in [14] in 1937.

We will bountifully make use of the big O notation by BACHMANN–LANDAU: for any real-valued functions f and g and any $a\in\mathbb{R}$ write $f(x)=\mathrm{O}(g(x))$ as $x\to a$ if there are a constant $C>0$ just depending on a and a $\delta>0$ so that for any $x\in(a-\delta,a+\delta)$ we have

$$|f(x)|\leq C\,|g(x)|\,.$$

Analogously, we write $f(x)=\mathrm{O}(g(x))$ as $x\to\infty$ if there are a constant $C>0$ and an $x_0\in\mathbb{R}$ such that $|f(x)|\leq C\,|g(x)|$ whenever $x\geq x_0$. Furthermore, we denote by $[x]$ the largest integer not greater than $x\in\mathbb{R}$, by $\mathbb{P}$ the set of all prime numbers and by (p,q) the greatest common divisor of $p,q\in\mathbb{N}$.

We aim to prove the following statement.

Theorem 5.7 (DAVENPORT)**.** *For each $r>0$ and every $\theta\in[0,1)$ we have*

$$\sum_{\substack{n\in\mathbb{N}\\ n\leq x}}\mu(n)e^{2\pi in\theta}=\mathrm{O}(x(\log x)^{-r})$$

as $x\to\infty$, uniformly in θ.

Note that for each $z\in\mathbb{T}$ there exists exactly one $\theta\in[0,1)$ such that $z\simeq e^{2\pi i\theta}$. Furthermore, since $|\mu(n)|\leq 1$ for all $n\in\mathbb{N}$, replacing $x\in\mathbb{R}$ by $N\in\mathbb{N}$ in Theorem 5.7 does not change the growth rate of the sum. So, following the definition of the big O notation, we can reword the above relation as

$$\max_{z\in\mathbb{T}}\left|\sum_{n=1}^{N}\mu(n)z^n\right|\leq\frac{CN}{\log^r N}\tag{5.11}$$

for each $r>0$, every $N\in\mathbb{N}$ and a constant $C=C(r)>0$ just depending on r.

To prove Theorem 5.7 we need a little preparation. We start with three technical lemmas for which we omit the proofs; they can be found in [14].

Lemma 5.8. *Let $x\in(0,\infty)$ and $l,q,H\in\mathbb{N}$ with $q\leq(\log x)^H$. Then there is a constant $C=C(H)>0$ just depending on H such that*

$$\sum_{\substack{n\in\mathbb{N}\\ n\leq x\\ n\equiv l \bmod q}}\mu(n)=\mathrm{O}\left(xe^{-C(H)\sqrt{\log x}}\right)$$

as $x\to\infty$.

Remark 5.9. The condition $n\equiv l \bmod q$ is equivalent to $n\in\{l+qk\,|\,k\in\mathbb{N}_0\}$. Furthermore, for $n\in\mathbb{N}$ and $a\in\mathbb{N}$ with $(a,q)=1$,

$$\sum_{m=1}^{n}\mu(m)e^{2\pi im\frac{a}{q}}=\sum_{r=1}^{q}e^{2\pi ia\frac{r}{q}}\sum_{\substack{m=1\\ m\equiv r \bmod q}}^{n}\mu(m),$$

so, for $n \coloneqq [x]$, Lemma 5.8 implies

$$\sum_{m=1}^{n} \mu(m) e^{2\pi i m \frac{a}{q}} = \mathrm{O}\left(qxe^{-C(H)\sqrt{\log x}}\right) = \mathrm{O}\left(xe^{-C(H)\sqrt{\log x}}\right)$$

as $x \to \infty$.

Lemma 5.10. *Let $N, u_0, u_1, q, a \in \mathbb{N}$ such that $1 < u_0 < u_1 < N$, $1 \leq q \leq N$ and $(a, q) = 1$. Let $\theta, \gamma : \mathbb{R} \times \mathbb{N} \to \mathbb{R}$ be bounded functions and $\psi : \mathbb{R} \times \mathbb{N} \to \mathbb{R}$ (not necessarily bounded). Then*

$$\sum_{u_0 < x \leq u_1} \theta(x, N) \sum_{\substack{1 \leq y \leq \frac{N}{x} \\ \psi(y,N) < x}} \gamma(y, N) e^{2\pi i a \frac{xy}{q}} = \mathrm{O}\left(N(\log N)^2 \sqrt{\frac{1}{u_0} + \frac{u_1}{N} + \frac{1}{q} + \frac{q}{N}}\right)$$

as $N \to \infty$.

Lemma 5.11. *For $h_1 > 3$ and $N_1 \in \mathbb{N}$ choose $q_1, b \in \mathbb{N}$ such that $(\log N_1)^{3h_1} < q_1 \leq N_1 (\log N_1)^{-3h_1}$ and $(b, q_1) = 1$. Then*

$$\sum_{\substack{p \in \mathbb{P} \\ p \leq N_1}} e^{2\pi i b \frac{p}{q_1}} = \mathrm{O}\left(N_1 (\log N_1)^{2-h_1}\right)$$

as $N_1 \to \infty$.

Now we split the statement of Theorem 5.7 into two parts (Lemma 5.12 and Lemma 5.14 below) and prove them separately.

Lemma 5.12. *Let $x \in (0, \infty)$, $H \in \mathbb{N}$ and set $\tau \coloneqq \left[x(\log x)^{-H}\right]$. For all $\theta \in \left(\frac{a}{q} - \frac{1}{\tau q}, \frac{a}{q} + \frac{1}{\tau q}\right)$, for some $a, q \in \mathbb{N}$ with $q \leq (\log x)^H$ and $(a, q) = 1$, there is a constant $C(H) > 0$ just depending on H such that*

$$\sum_{\substack{n \in \mathbb{N} \\ n \leq x}} \mu(n) e^{2\pi i n \theta} = \mathrm{O}\left(xe^{-C(H)\sqrt{\log x}}\right)$$

as $x \to \infty$, uniformly in θ.

Proof. For each $n \in \mathbb{N}$, $1 \leq n \leq x$ define

$$S_0 \coloneqq 0$$
$$S_n \coloneqq \sum_{m=1}^{n} \mu(m) e^{2\pi i m \frac{a}{q}}$$

and

$$S_x \coloneqq \sum_{\substack{m \in \mathbb{N} \\ m \leq x}} \mu(m) e^{2\pi i m \frac{a}{q}}.$$

From Lemma 5.8 and Remark 5.9, for $n \coloneqq [x]$ we know that

$$S_n = \mathrm{O}\left(xe^{-C(H)\sqrt{\log x}}\right)$$

as $x \to \infty$, with a constant $C(H) > 0$ just depending on H. Write $\theta = \frac{a}{q} + \beta$ with $\beta \in \left(-\frac{1}{\tau q}, \frac{1}{\tau q}\right)$. Then, by the choice of τ,

$$x \left|1 - e^{2\pi i\beta}\right| = \mathrm{O}\left(\frac{x}{q\tau}\right) = \mathrm{O}(1)$$

as $x \to \infty$, and thus we obtain

$$\begin{aligned}
\sum_{\substack{n\in\mathbb{N}\\ n\le x}} \mu(n) e^{2\pi i n\left(\frac{a}{q}+\beta\right)} &= \sum_{\substack{n\in\mathbb{N}\\ n\le x}} (S_n - S_{n-1})\, e^{2\pi i n\beta} \\
&= \sum_{\substack{n\in\mathbb{N}\\ n\le x}} S_n e^{2\pi i n\beta} - \sum_{\substack{n\in\mathbb{N}_0\\ n\le x}} S_n e^{2\pi i(n+1)\beta} \\
&= \sum_{\substack{n\in\mathbb{N}\\ n\le x}} S_n e^{2\pi i n\beta}(1 - e^{2\pi i\beta}) + \mathrm{O}\left(x e^{-C(H)\sqrt{\log x}}\right) \\
&= \mathrm{O}\left(\left(x\left|1 - e^{2\pi i\beta}\right| + 1\right) x e^{-C(H)\sqrt{\log x}}\right) \\
&= \mathrm{O}\left(\left(\frac{x}{q\tau} + 1\right) x e^{-C(H)\sqrt{\log x}}\right) \\
&= \mathrm{O}\left(x e^{-C(H)\sqrt{\log x}}\right). \qquad \square
\end{aligned}$$

Lemma 5.13. *For $h > 1$ and $N \in \mathbb{N}$ choose $q, a \in \mathbb{N}$ such that $(\log N)^{12h} < q \le N(\log N)^{-12h}$ and $(a, q) = 1$. Then*

$$\sum_{n=1}^{N} \mu(n) e^{2\pi i n\frac{a}{q}} = \mathrm{O}\left(N(\log N)^{2-h}\right)$$

as $N \to \infty$.

Sketch of proof. For $n \in \mathbb{N}$ denote by $\Psi(n)$ the largest prime factor of n and by $d(n)$ the sum of all prime factors of n. If n is square-free (i.e. for every two prime factors p_1, p_2 of n we have $p_1 \neq p_2$) with $\sqrt{N} \le n \le N$ and $\Psi(n) \le (\log N)^{2h}$ then n has not less than $\frac{\log N}{4h\log(\log N)}$ prime factors, which implies

$$d(n) \ge 2^{\frac{\log N}{4h\log(\log N)}} > (\log N)^h,$$

for N sufficiently large. Using this together with $\left|\sum_{n=1}^{N} \mu(n)\right| \le \left|\sum_{n=1}^{N} d(n)\right|$ one shows that (see [14])

$$\sum_{\substack{n=2\\ \Psi(n)\le(\log N)^{2h}}}^{N} \mu(n) e^{2\pi i n\frac{a}{q}} = \mathrm{O}\left(N(\log N)^{1-h}\right). \tag{5.12}$$

On the other hand, since $\mu(p) = -1$, for each $p \in \mathbb{P}$, and $\mu(m \cdot n) = \mu(m)\mu(n)$, for

$m, n \in \mathbb{N}$, $(m,n) = 1$, we have

$$\begin{aligned}
\sum_{\substack{n=2 \\ \Psi(n)>(\log N)^{2h}}}^{N} \mu(n) e^{2\pi i n \frac{a}{q}} &= \sum_{\substack{p \in \mathbb{P} \\ (\log N)^{2h} < p \le N}} \sum_{\substack{1 \le m \le \frac{N}{p} \\ \Psi(m) < p}} \mu(pm) e^{2\pi i a m \frac{p}{q}} \\
&= - \sum_{\substack{p \in \mathbb{P} \\ (\log N)^{2h} < p \le N}} \sum_{\substack{1 \le m \le \frac{N}{p} \\ \Psi(m) < p}} \mu(m) e^{2\pi i a m \frac{p}{q}} \\
&= \underbrace{- \sum_{m \le (\log N)^{2h}} \mu(m) \sum_{(\log N)^{2h} < p \le \frac{N}{m}} e^{2\pi i a m \frac{p}{q}}}_{=:P_1} \\
&\quad \underbrace{- \sum_{(\log N)^{2h} < p < N(\log N)^{-2h}} \sum_{\substack{(\log N)^{2h} < m \le \frac{N}{p} \\ \Psi(m) < p}} \mu(m) e^{2\pi i a m \frac{p}{q}}}_{=:P_2} \\
&= -P_1 - P_2.
\end{aligned}$$

The inner sum in P_1 satisfies the conditions of Lemma 5.11 with

$$q_1 \coloneqq \frac{q}{(m,q)},\ b \coloneqq \frac{am}{(m,q)},\ N_1 \coloneqq \frac{N}{m} \text{ and } h_1 \coloneqq 3h.$$

Hence, by Lemma 5.11,

$$P_1 = \mathrm{O}\left((\log N)^{2h} N (\log N)^{2-3h}\right) = \mathrm{O}\left(N(\log N)^{2-h}\right) \tag{5.13}$$

as $N \to \infty$. Now, set

$$\theta(x,N) \coloneqq \begin{cases} 1 & \text{for } x \in \mathbb{P} \\ 0 & \text{otherwise} \end{cases} \quad \text{and} \quad \gamma(y,N) \coloneqq \begin{cases} \mu([y]) & \text{for } y > (\log N)^{2h} \\ 0 & \text{otherwise} \end{cases}$$

as well as $\psi(y,N) \coloneqq \Psi([y])$. Thus, by using Lemma 5.10, one shows that (see [14])

$$P_2 = \mathrm{O}\left(N(\log N)^{2-h}\right) \tag{5.14}$$

as $N \to \infty$. So, by (5.12), (5.13) and (5.14) we obtain

$$\sum_{n=1}^{N} \mu(n) e^{2\pi i n \frac{a}{q}} = \mathrm{O}\left(N(\log N)^{1-h}\right) + \mathrm{O}\left(N(\log N)^{2-h}\right) = \mathrm{O}\left(N(\log N)^{2-h}\right)$$

as $N \to \infty$. □

Lemma 5.14. *Let $x \in (0,\infty)$, $H \in \mathbb{N}$, $H > 14$ and set $\tau \coloneqq \left\lfloor x(\log x)^{-H} \right\rfloor$. Then for each $\theta \in [0,1)$ for which there are $a, q \in \mathbb{N}$ with $(a,q) = 1$, $\left|\theta - \frac{a}{q}\right| \le \frac{1}{q\tau}$ and $(\log x)^H < q \le \tau$, we have*

$$\sum_{\substack{n \in \mathbb{N} \\ n \le x}} \mu(n) e^{2\pi i n \theta} = \mathrm{O}\left(x(\log x)^{2-\frac{1}{14}H}\right)$$

as $x \to \infty$.

Proof. Let $h \coloneqq \frac{1}{14}H$, write $\theta = \frac{a}{q} + \beta$ and consider the same partial summation as that used in the proof of Lemma 5.12. Then it suffices to show that for $N \leq x$

$$\left| \sum_{n=1}^{N} \mu(n) e^{2\pi i n \frac{a}{q}} \right| \leq C x (\log x)^{2-h}$$

for x sufficiently large and $C = C(H) > 0$ a constant just depending on H. For $N \leq x(\log x)^{-h}$ we have

$$\left| \sum_{n=1}^{N} \mu(n) e^{2\pi i n \frac{a}{q}} \right| \leq \sum_{n=1}^{N} |\mu(n)| \left| e^{2\pi i n \frac{a}{q}} \right| \leq N \leq x \left(\log x\right)^{-h} \leq x \left(\log x\right)^{2-h},$$

since $x \left(\log x\right)^2 \leq x \left(\log x\right)^{2h}$ whenever $x > 1$ (since $h > 1$ by the choice of H). For $x(\log x)^{-h} < N \leq x$ the assertion follows from Lemma 5.13. □

Now we can put everything together to obtain the desired result.

Proof of Theorem 5.7. Fix $x \in [0, \infty)$ and choose $H \in \mathbb{N}$ such that $2 - \frac{1}{14}H < -r$. Set $\tau \coloneqq \left\lfloor x(\log x)^{-H} \right\rfloor$. Then, by the DIRICHLET drawer principle, there are $a, q \in \mathbb{N}$ such that $(a, q) = 1$, $1 \leq q \leq \tau$ and $\left|\theta - \frac{a}{q}\right| \leq \frac{1}{q\tau}$. For $q \leq (\log x)^H$ the assertion follows from Lemma 5.12 and for $(\log x)^H < q \leq \tau$ it follows from Lemma 5.14. □

5.3. Spectral Theorem for Bounded Unitary Operators

The results of this section are taken from [30],[46] and [41]. Throughout this section let H be a separable HILBERT space over $\mathbb{C}$. Denote by $\mathfrak{L}(H)$ the set of all bounded linear operators $T : H \to H$.

Definition 5.15. Let $T \in \mathfrak{L}(H)$. Then we call T

- *normal*, if $TT^* = T^*T$,
- *unitary*, if $TT^* = T^*T = \mathrm{id}_H$,
- *self-adjoint*, if $T^* = T$,

where T^* denotes the adjoint operator of T.

Obviously, every unitary operator on H is normal and bijective with $T^{-1} = T^*$, and T^* is also unitary. Furthermore, for any $x, y \in H$,

$$\langle Tx, Ty \rangle_H = \langle x, T^*Ty \rangle_H = \langle x, y \rangle_H$$

and therefore

$$\|Tx\|_H = \sqrt{\langle Tx, Tx \rangle_H} = \sqrt{\langle x, x \rangle_H} = \|x\|_H,$$

i.e., T is isometric (thus $\|T\| = 1$). The converse is also true (see e.g. [30]). So the unitary operators on H are exactly the isometric automorphisms on H.

Denote by $\sigma(T) \coloneqq \{\lambda \in \mathbb{C} \mid \lambda \mathrm{id}_H - T \text{ is not invertible}\}$ the *spectrum* of an arbitrary operator $T \in \mathfrak{L}(H)$ and by $r(T) \coloneqq \sup\{|\lambda| \mid \lambda \in \sigma(T)\}$ the *spectral radius* of

T. One can show (see [30], Satz 58.6) that $r(T) = \lim_{n\to\infty} \sqrt[n]{\|T^n\|}$. Furthermore, for normal operators and all $n \in \mathbb{N}$ we have $\|T^n\| = \|T\|^n$. Therefore,

$$r(T) = \lim_{n\to\infty} \sqrt[n]{\|T^n\|} = \|T\| \tag{5.15}$$

and, if T is unitary,

$$r(T) = \|T\| = 1. \tag{5.16}$$

Proposition 5.16. *Let T be a unitary operator on H. Then $\sigma(T) \subseteq \mathbb{T}$.*

Proof. By (5.16) and since T is bijective, we have $\sigma(T) \subseteq \{z \in \mathbb{C} \,|\, |z| \leq 1\} \setminus \{0\}$. Let $\lambda \in \mathbb{C}$ with $0 < |\lambda| < 1$. Then $\left|\frac{1}{\lambda}\right| > 1$ and therefore $\frac{1}{\lambda} \in \mathbb{C} \setminus \sigma(T^*)$ (since T^* is unitary, too, and thus $r(T^*) = 1$). Hence $\frac{1}{\lambda}\mathrm{id}_H - T^*$ is an invertible operator and so is $-\lambda T$. Consequently, $\left(\frac{1}{\lambda}\mathrm{id}_H - T^*\right)(-\lambda T) = \lambda \mathrm{id}_H - T$ is also invertible and thus $\lambda \notin \sigma(T)$. This yields the assertion. □

Remark 5.17. One can also show, that for T self-adjoint we have $\sigma(T) \subseteq \mathbb{R}$, see [46] for a proof. For T normal $\sigma(T)$ is an arbitrary (non-empty) compact subset of $\mathbb{C}$.

We will prove the required spectral theorem for all normal operators on separable HILBERT spaces, because a limitation on unitary operators beforehand would not make the task any easier. But we will be content with the case of a *cyclic space*, which is to say that there is a vector $h \in H$ such that the linear span of the T-orbit of h is dense in H, i.e., we have $H = \overline{\mathrm{lin}\,\{T^n h \,|\, n \in \mathbb{N}\}}$.

Let $\phi : H_1 \to H_2$ be a homomorphism between two HILBERT spaces over $\mathbb{C}$. Then we call ϕ a **-homomorphism*, if ϕ preserves the involution, i.e., we have $\phi(\overline{x}) = \overline{\phi(x)}$, for each $x \in H_1$. By a $\mathbb{C}$*-algebra* we mean a vector space over $\mathbb{C}$ with a bilinear multiplication on it. A BANACH *algebra* $\mathcal{A}$ is an associative $\mathbb{C}$-algebra with a sub-multiplicative norm $\|\cdot\|_{\mathcal{A}}$ such that $(\mathcal{A}, +, \|\cdot\|_{\mathcal{A}})$ is a BANACH space (the sub-multiplicativity ensures the multiplication operation to be continuous). If we equip a commutative BANACH algebra $\mathcal{A}$ with an involution * such that $\|x^*x\|_{\mathcal{A}} = \|x\|^2_{\mathcal{A}}$ for all $x \in \mathcal{A}$, we obtain a C^**-algebra*. Finally, for $T \in \mathfrak{L}(H)$ normal, denote by

$$C^*(T) \coloneqq \bigcap \{\mathcal{A} \subseteq \mathfrak{L}(H) \,|\, \mathcal{A} \text{ is a } C^*\text{-algebra with } \{T, \mathrm{id}_H\} \subseteq \mathcal{A}\}$$

the smallest C^*-algebra which contains T and id_H (note that $C^*(T)$ also contains T^*). One can show (see [41]) that $C^*(T) = \overline{\{P(T, T^*) \,|\, P \text{ a polynomial}\}}$. We say $C^*(T)$ has a *cyclic vector* h, if

$$\overline{\{Bh \,|\, B \in C^*(T)\}} = \overline{\{P(T, T^*)h \,|\, P \text{ a polynomial}\}} = H.$$

Note that, if h is a cyclic vector for T, then h is also a cyclic vector for $C^*(T)$ (for T self-adjoint the inversion holds, too; see e.g. [41]).

Definition 5.18. We call a bounded $T \in \mathfrak{L}(H)$ *unitary equivalent to a multiplicator*, if there exist a σ-finite measure space (X, Σ_X, ν), a function $\phi \in L^\infty(X, \nu)$ and a unitary operator $\Phi : L^2(X, \nu) \to H$ such that

$$\Phi^* T \Phi = M_\phi,$$

where $M_\phi : L^2(X, \nu) \to L^2(X, \nu)$ is given by $M_\phi f(z) = \phi(z) f(z)$, for each $f \in L^2(X, \nu)$ and every $z \in X$.

We intend to show that each normal operator on a separable HILBERT space H with a cyclic vector is unitary equivalent to the multiplicator $M_{\mathrm{id}_{\sigma(T)}}$. To do so we need some proper preparation.

Proposition 5.19. *Let $\mathcal{A}$ be a commutative unital*[1] BANACH *algebra and denote by Γ the* GELFAND *transformation on $\mathcal{A}$:*

$$\Gamma(A)(\gamma) \equiv \hat{A}(\gamma) \coloneqq \langle A, \gamma \rangle ,$$

where $A \in \mathcal{A}$ and $\gamma \in \widehat{\mathcal{A}}$ with $\widehat{\mathcal{A}}$ the set of all characters of $\mathcal{A}$ (i.e., of all surjective (multiplicative) homomorphisms $\phi : \mathcal{A} \to \mathbb{C}$). Then $\widehat{\mathcal{A}}$ possesses a compact HAUSDORFF *topology such that Γ is a norm-contractive homomorphism from $\mathcal{A}$ into a subalgebra of $C(\widehat{\mathcal{A}})$, which separates the points of $\widehat{\mathcal{A}}$. For each $A \in \mathcal{A}$ we have $\hat{A}(\widehat{\mathcal{A}}) = \sigma(A)$ and $\left\|\hat{A}\right\|_\infty = r(A)$.*

Proof. One can show (see [46], Proposition 4.2.2) that

$$\sigma(A) = \left\{ \langle A, \gamma \rangle \,\middle|\, \gamma \in \widehat{\mathcal{A}} \right\} \tag{5.17}$$

for each $A \in \mathcal{A}$. Therefore, for each $A \in \mathcal{A}$ and every $\gamma \in \widehat{\mathcal{A}}$,

$$|\langle A, \gamma \rangle| \leq r(A) \leq \|A\| .$$

Thus $\|\gamma\| \leq 1$, regarding γ as an element in the dual space $\mathcal{A}'$. Denote by $\mathcal{B}'$ the closed unit ball in $\mathcal{A}'$. Then $\widehat{\mathcal{A}} \subseteq \mathcal{B}'$ and, considering the w^*-topology on $\mathcal{A}'$, we have a HAUSDORFF topology on $\widehat{\mathcal{A}}$.

Now let J be a directed set and $(\gamma_j)_{j \in J}$ a net[2] which w^*-converges to a $\gamma \in \mathcal{B}'$. Then, for $A, B \in \mathcal{A}$,

$$\langle AB, \gamma \rangle = \lim_{j \in J} \langle AB, \gamma_j \rangle = \lim_{j \in J} \langle A, \gamma_j \rangle \langle B, \gamma_j \rangle = \langle A, \gamma \rangle \langle B, \gamma \rangle ,$$

with the limits in the w^*-sense. Hence, $\gamma \in \widehat{\mathcal{A}}$, which implies that $\widehat{\mathcal{A}}$ is a w^*-closed subset of the compact set $\mathcal{B}'$ and thus compact itself.

Since w^*-convergence is pointwise convergence, it follows that each function $\hat{A}$ on $\widehat{\mathcal{A}}$, with $A \in \mathcal{A}$, $\hat{A}(\gamma) = \langle A, \gamma \rangle$, is continuous. Furthermore, by (5.17), $\hat{A}(\widehat{\mathcal{A}}) = \sigma(A)$ and therefore $\left\|\hat{A}\right\|_\infty = r(A)$. Finally, since each $\gamma \in \widehat{\mathcal{A}}$ is multiplicative and for $\gamma_1, \gamma_2 \in \widehat{\mathcal{A}}$, $\gamma_1 \neq \gamma_2$ implies $\langle A, \gamma_1 \rangle \neq \langle A, \gamma_2 \rangle$ for at least one $A \in \mathcal{A}$, we conclude that $\Gamma : \mathcal{A} \to C(\widehat{\mathcal{A}})$, $A \mapsto \hat{A}$, is a homomorphism, which separates the points of $\widehat{\mathcal{A}}$. □

Proposition 5.20. *Each commutative unital C^*-algebra $\mathcal{A}$ is isometrically *-isomorphic to $C(\widehat{\mathcal{A}})$.*

Proof. Since $\mathcal{A}$ is commutative, each $A \in \mathcal{A}$ is normal. So by (5.15) and Proposition 5.19 the GELFAND transformation is isometric. If A is self-adjoint, then for each $\gamma \in \widehat{\mathcal{A}}$

$$\hat{A}(\gamma) = \langle A, \gamma \rangle \in \sigma(A) \subseteq \mathbb{R}$$

[1] i.e., $\mathcal{A}$ has a neutral element for the multiplication

[2] also called a MOORE–SMITH sequence; see [46] for definition and properties

(see Lemma 4.3.12 in [46]).

Let $T \in \mathcal{A}$. Then, since T is normal, there are self-adjoint $A, B \in \mathcal{A}$ such that $T = A + iB$; these are

$$A \coloneqq \frac{1}{2}\left(T + T^*\right) \quad \text{and} \quad B \coloneqq \frac{i}{2}\left(T - T^*\right).$$

Therefore,

$$\Gamma(T^*) = \Gamma(A - iB) = \Gamma(A) - i\Gamma(B) = \overline{\Gamma(A) + i\Gamma(B)} = \overline{\Gamma(A + iB)} = \overline{\Gamma(T)},$$

i.e., Γ preserves the involution *. In particular, $\Gamma(\mathcal{A}) \coloneqq \left\{\hat{A} \,\middle|\, A \in \mathcal{A}\right\}$ is a subalgebra of real-valued functions in $C(\widehat{\mathcal{A}})$ and thus $\Gamma(\mathcal{A}) = C(\widehat{\mathcal{A}})$ by Proposition 5.19 and the Stone–Weierstrass theorem (Theorem A.16 in the Appendix; see also Theorem 4.3.4 in [46]). This yields the assertion. □

Lemma 5.21. *Let T be a normal element in a C^*-algebra $\mathcal{A}$ with unit I and denote by $C^*(T)$ the smallest C^*-subalgebra of $\mathcal{A}$ which contains T and I. Then there is an isometric *-isomorphism $\widetilde{\phi} : C(\sigma(T)) \to C^*(T)$ which maps $\mathbf{1}_{\sigma(T)}$ to I and $\mathrm{id}_{\sigma(T)}$ to T.*

Proof. As mentioned above, we have $C^*(T) = \overline{\{P(T, T^*) \,|\, P \text{ a polynomial}\}}$. This implies that the C^*-algebra $C^*(T)$ is unital and commutative. Thus, by Proposition 5.20, $C^*(T)$ is isometrically *-isomorphic to $C(\widehat{C^*(T)})$ (by the Gelfand transformation Γ). Denote by $\sigma(T)$ the spectrum of T in $\mathcal{A}$ and by $\sigma^*(T)$ the spectrum of T in $C^*(T)$. Since $C^*(T) \subseteq \mathcal{A}$, we have

$$\sigma(T) \subseteq \sigma^*(T). \tag{5.18}$$

By Proposition 5.19 the map $\gamma \mapsto \langle T, \gamma\rangle$ is a surjection from $\widehat{C^*(T)}$ onto $\sigma^*(T)$ and continuous, since $\widehat{C^*(T)}$, as a subset of $(C^*(T))^\star$, has the w^*-topology. Note that for $\gamma_1, \gamma_2 \in \widehat{C^*(T)}$, with $\langle T, \gamma_1\rangle = \langle T, \gamma_2\rangle$, also

$$\langle T^*, \gamma_1\rangle = \overline{\langle T, \gamma_1\rangle} = \overline{\langle T, \gamma_2\rangle} = \langle T^*, \gamma_2\rangle,$$

and furthermore, $\langle I, \gamma_1\rangle = 1 = \langle I, \gamma_2\rangle$. Therefore, γ_1 and γ_2 match on the subset $\{P(T, T^*) \,|\, P \text{ a polynomial}\}$ and thus, because of the continuity, also on

$$\overline{\{P(T, T^*) \,|\, P \text{ a polynomial}\}} = C^*(T).$$

Hence, $\gamma_1 = \gamma_2$ and therefore $\gamma \mapsto \langle T, \gamma\rangle$ is injective, too.

Let $\Psi : C(\sigma^*(T)) \to C(\widehat{C^*(T)})$ be given by $\Psi(f)(\gamma) = f(\langle T, \gamma\rangle)$, where $f \in C(\sigma^*(T))$ and $\gamma \in \widehat{C^*(T)}$. Then, by the above, Ψ is an isometric *-isomorphism. Therefore, $\widetilde{\phi} \coloneqq \Gamma^{-1} \circ \Psi$ is an isometric *-isomorphism between $C(\sigma^*(T))$ and $C^*(T)$. For each $\gamma \in \widehat{C^*(T)}$ we have

$$\Gamma(T)(\gamma) = \langle T, \gamma\rangle = \mathrm{id}_{\sigma^*(T)}(\langle T, \gamma\rangle) = \Psi(\mathrm{id}_{\sigma^*(T)})(\gamma),$$

which implies $\Gamma(T) = \Psi(\mathrm{id}_{\sigma^*(T)})$ and thus $T = \widetilde{\phi}(\mathrm{id}_{\sigma^*(T)})$. Analogously we obtain $I = \widetilde{\phi}(\mathbf{1}_{\sigma^*(T)})$.

Now, showing that $\sigma^*(T) = \sigma(T)$ will finish the proof. Because of (5.20) it just remains to show that $\sigma^*(T) \subseteq \sigma(T)$. For this purpose choose $\lambda \in \sigma^*(T)$ arbitrarily. Fix $\varepsilon > 0$. Then there is an $f \in C(\sigma^*(T))$ such that $\|f\|_\infty = 1$ and $f(\lambda) = 1$ but $f(\rho) = 0$ for every $\rho \in \sigma^*(T)$ with $|\lambda - \rho| \geq \varepsilon$. Let $A \coloneqq \widetilde{\phi}(f)$. Then

$$\|(T - \lambda I)A\| = \left\|\widetilde{\phi}^{-1}\left((T - \lambda I)A\right)\right\|_\infty = \left\|(\mathrm{id}_{\sigma^*(T)} - \lambda)f\right\|_\infty \leq \varepsilon.$$

Thus, $T - \lambda I$ cannot be invertible in $\mathcal{A}$ (because the inverse would have to have norm greater than ε^{-1}). Hence, $\lambda \in \sigma(T)$. □

Now we are able to prove the desired spectral theorem.

Theorem 5.22. *Let H be a separable* Hilbert *space over $\mathbb{C}$ and $T \in \mathfrak{L}(H)$ a bounded normal operator such that $C^*(T)$ has a cyclic vector. Then T is unitary equivalent to the multiplicator $M_{\mathrm{id}_{\sigma(T)}} : L^2(\sigma(T), \nu) \to L^2(\sigma(T), \nu)$ given by $M_{\mathrm{id}_{\sigma(T)}} g(z) = zg(z)$, for each $g \in L^2(\sigma(T), \nu)$ and every $z \in \sigma(T)$, with ν a unique positive finite* Borel *measure on $\sigma(T)$.*

Proof. Let h be the cyclic vector of $C^*(T)$. By Lemma 5.21 there is an isometric *-isomorphism ϕ ($\coloneqq \widetilde{\phi}^{-1}$) between $C^*(T)$ and $C(\sigma(T))$ such that $\phi(T) = \mathrm{id}_{\sigma(T)}$. By Lemma 5.21 for $P = P(x, y)$ a polynomial and $f, g \in C(\sigma(T))$ we obtain the mappings

$$\begin{aligned}
C^*(T) &\leftrightarrow C(\sigma(T)) \\
\mathrm{id}_H &\leftrightarrow \mathbf{1}_{\sigma(T)} \\
T &\leftrightarrow \mathrm{id}_{\sigma(T)} \\
T^* &\leftrightarrow \overline{\mathrm{id}_{\sigma(T)}} \\
P(T, T^*) &\leftrightarrow P(id_{\sigma(T)}, \overline{\mathrm{id}_{\sigma(T)}}) \\
f(T) &\leftrightarrow f \\
f(T)^* &\leftrightarrow \overline{f} \\
f(T)g(T) &\leftrightarrow fg \\
f(T) + g(T) &\leftrightarrow f + g
\end{aligned}$$

where $f(T) \coloneqq \phi^{-1}(f)$. Define Λ on $C(\sigma(T))$ by $\Lambda(f) \coloneqq \langle f(T)h, h\rangle$. Then Λ is a bounded positive linear functional on $C(\sigma(T))$, because:

- Linearity follows from $f(T) + g(T) = (f + g)(T)$ and $(\alpha f)(T) = \alpha f(T)$ for each $\alpha \in \mathbb{C}$ (see the last two mappings in the above scheme).
- Since ϕ isometric, we have

  $$|\Lambda(f)| = |\langle f(T)h, h\rangle| \leq \|f(T)h\|_H \|h\|_H \leq \|f(T)\| \|h\|_H^2 = \|f\|_\infty \|h\|_H^2 .$$

  So Λ is bounded with $\|\Lambda\| \leq \|h\|_H^2$.
- Let $f \in C(\sigma(T))$ be real-valued with $f(x) \geq 0$ for each $x \in \sigma(T)$. Then $g \coloneqq \sqrt{f} \in C(\sigma(T))$ is also real-valued with $g(x) \geq 0$ for each $x \in \sigma(T)$. Therefore,

  $$\Lambda(f) = \Lambda(g^2) = \left\langle g^2(T)h, h\right\rangle = \langle g(T)h, g(T)h\rangle = \|g(T)h\|_H^2 \geq 0,$$

  since g is real-valued.

Hence, by the Riesz-Markov-Kakutani representation theorem (see Theorem A.9 in the Appendix) there is a unique positive finite Borel measure ν_h on $\sigma(T)$ such that

$$\Lambda(f) = \langle f(T)h, h\rangle = \int_{\sigma(T)} f \,\mathrm{d}\,\nu_h$$

for each $f \in C(\sigma(T))$, with $\nu_h(\sigma(T)) = \|\Lambda\|$.

Now, for $f \in C(\sigma(T))$, define $\Phi f \coloneqq f(T)h = \phi^{-1}(f)h$. By the linearity of ϕ, Φ has to be linear, too. Furthermore, for $f, g \in C(\sigma(T))$,

$$\begin{aligned}\langle \Phi f, \Phi g\rangle_H &= \langle f(T)h, g(T)h\rangle_H = \langle g(T)^* f(T)h, h\rangle_H = \langle (\overline{g}f)(T)h, h\rangle_H \\ &= \Lambda(\overline{g}f) = \int_{\sigma(T)} \overline{g}f \,\mathrm{d}\,\nu_h = \langle f, g\rangle_{L^2(\sigma(T),\nu_h)}\,.\end{aligned}$$

Hence, Φ is a linear isometry from $C(\sigma(T))$ (equipped with the L^2-norm) into H. Since $C(\sigma(T))$ is dense in $L^2(\sigma(T), \nu_h)$, we can, in a unique way, extend Φ to a linear isometry on $L^2(\sigma(T), \nu_h)$, which range is a closed subspace of H that includes $\{f(T)h \,|\, f \in C(\sigma(T))\} = \{Bh \,|\, B \in C^*(T)\}$ as a subset (we denote this extension by Φ, too). Since h is cyclic for T in H, $\{Bh \,|\, B \in C^*(T)\}$ is dense in H. Thus, Φ is a unitary map from $L^2(\sigma(T), \nu_h)$ onto H.

It remains to show that $\Phi^* T \Phi$ is the claimed multiplicator on $L^2(\sigma(T), \nu_h)$. For each $f \in C(\sigma(T))$ let $M_f : C(\sigma(T)) \to C(\sigma(T))$ be given by $M_f(g)(z) = f(z)g(z)$. Then, for $f, g \in C(\sigma(T))$, we have

$$\Phi M_f g = \Phi(fg) = (fg)(T)h = f(T)g(T)h = f(T)\Phi g,$$

thus $\Phi M_f = f(T)\Phi$ on the dense subset $C(\sigma(T))$ of $L^2(\sigma(T), \nu_h)$ and therefore on $L^2(\sigma(T), \nu_h)$. This implies

$$\Phi^{-1} f(T) \Phi = M_f$$

for each $f \in C(\sigma(T))$, and hence, in particular,

$$\Phi^{-1} T \Phi = M_{\mathrm{id}_{\sigma(T)}}. \qquad \square$$

The measure ν_h we obtained in the above proof by bringing the Riesz–Markov–Kakutani representation theorem into use, is a finite Borel measure on $\mathbb{T}$ and is called the *spectral measure* of T.

Remark 5.23. From the construction of the isomorphism Φ in the above proof we can conclude that

$$\begin{aligned} & \Phi(\mathbf{1}_{\sigma(T)}) = \phi^{-1}(\mathbf{1}_{\sigma(T)})h \\ \iff & \mathbf{1}_{\sigma(T)} \quad = \Phi^{-1} \underbrace{\phi^{-1}(\mathbf{1}_{\sigma(T)})}_{=\mathrm{id}_H} h \\ \iff & \mathbf{1}_{\sigma(T)} \quad = \Phi^{-1}(h). \end{aligned}$$

5.4. The Ergodic Theorem with Möbius Weights

Again we denote by $[x]$ the largest integer not greater than $x \in \mathbb{R}$.

Theorem 5.24 ([22]). *Let (X, Σ_X, ν, T) be an invertible MDS and $f \in L^1(X, \nu)$. Then, for ν-a.e. $x \in X$, we have*

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} f(T^n x)\mu(n) = 0.$$

Proof. If ν is non-ergodic, then, by the previous chapter, there is a disintegration $\nu = \int_X \nu_x d\nu(x)$ of ν such that a.e. ν_x is T-invariant, ergodic and supported on disjoint invariant sets, and we can pass over to these ν_x. Therefore, without loss of generality, we may assume that ν is ergodic.

First, let $f \in L^2(X, \nu)$. Define $H \coloneqq \overline{\operatorname{lin}\{T^n f \mid n \in \mathbb{Z}\}}$. Then $H \subseteq L^2(X, \nu)$ is a separable Hilbert space (since L^2 is separable itself), with the cyclic vector f, and $T|_H : H \to H$ is unitary (as an invertible isometry). Recall that f is also cyclic for $C^*(T)$ and, by Proposition 5.16, we have $\sigma(T) \subseteq \mathbb{T}$. Therefore, by the spectral theorem (Theorem 5.22), T is unitary equivalent to the multiplicator

$$M_{\mathrm{id}_{\mathbb{T}}} : L^2(\mathbb{T}, \nu_f) \to L^2(\mathbb{T}, \nu_f)$$

with ν_f as in Theorem 5.22. So, together with Remark 5.23, we obtain

$$\Phi^{-1}(f \circ T^n)\Phi = [\mathbb{T} \to \mathbb{T},\, z \mapsto z^n],$$

thus

$$\begin{aligned}
\left\| \frac{1}{N} \sum_{n=1}^{N} f(T^n x)\mu(n) \right\|^2_{L^2(X,\nu)} &= \int_X \left| \frac{1}{N} \sum_{n=1}^{N} f(T^n x)\mu(n) \right|^2 \mathrm{d}\,\nu(x) \\
&= \int_{\mathbb{T}} \left| \frac{1}{N} \sum_{n=1}^{N} z^n \mu(n) \right|^2 \mathrm{d}\,\nu_f(z) \\
&= \left\| \frac{1}{N} \sum_{n=1}^{N} z^n \mu(n) \right\|^2_{L^2(\mathbb{T},\nu_f)}
\end{aligned}$$

and therefore,

$$\left\| \frac{1}{N} \sum_{n=1}^{N} f(T^n x)\mu(n) \right\|_{L^2(X,\nu)} = \left\| \frac{1}{N} \sum_{n=1}^{N} z^n \mu(n) \right\|_{L^2(\mathbb{T},\nu_f)}.$$

Hence, by Davenport's estimation (Theorem 5.7) in the form (5.11), for each $r > 0$ there is a constant $C_1 = C_1(r) > 0$ which depends only on r, such that

$$\left\| \frac{1}{N} \sum_{n=1}^{N} f(T^n x)\mu(n) \right\|_{L^2} \leq \frac{C_1}{(\log N)^r}. \tag{5.19}$$

For $\rho \in (1, \infty)$ and $m \in \mathbb{N}$ (5.19) takes the form (for $N \coloneqq [\rho^m]$)

$$\left\| \frac{1}{[\rho^m]} \sum_{n=1}^{[\rho^m]} f(T^n x)\mu(n) \right\|_{L^2} \leq \frac{C_2}{(m \log \rho)^r},$$

with $C_2 = C_2(r) > 0$ only depending on r. In particular, by choosing $r = 2$, this implies $\sum_{m=1}^{\infty} \left\| \frac{1}{[\rho^m]} \sum_{n=1}^{[\rho^m]} f(T^n x)\mu(n) \right\|_{L^2} < \infty$. Hence, by the Borel–Cantelli

lemma (Theorem A.10 in the Appendix, see also Corollary A.11), for ν-a.e. $x \in X$ we obtain

$$\frac{1}{[\rho^m]} \sum_{n=1}^{[\rho^m]} f(T^n x)\mu(n) \xrightarrow[m\to\infty]{} 0. \tag{5.20}$$

Now, suppose additionally that $f \in L^\infty(X,\nu)$. Then, for $[\rho^m] \le N < \left[\rho^{m+1}\right] + 1$,

$$\begin{aligned}
\left|\frac{1}{N}\sum_{n=1}^{N} f(T^n x)\mu(n)\right| &= \left|\frac{1}{N}\sum_{n=1}^{[\rho^m]} f(T^n x)\mu(n) + \frac{1}{N}\sum_{n=[\rho^m]+1}^{N} f(T^n x)\mu(n)\right| \\
&\le \left|\frac{1}{[\rho^m]}\sum_{n=1}^{[\rho^m]} f(T^n x)\mu(n)\right| + \frac{\|f\|_\infty}{[\rho^m]}\left(N - [\rho^m]\right) \\
&\le \left|\frac{1}{[\rho^m]}\sum_{n=1}^{[\rho^m]} f(T^n x)\mu(n)\right| + \frac{\|f\|_\infty}{[\rho^m]}\left(\left[\rho^{m+1}\right] - [\rho^m]\right).
\end{aligned}$$

Because of $\frac{\|f\|_\infty}{[\rho^m]}\left(\left[\rho^{m+1}\right] - [\rho^m]\right) \xrightarrow[m\to\infty]{} \|f\|_\infty(\rho - 1)$ and (5.20), for $\rho \to 1$ we obtain

$$\lim_{N\to\infty} \frac{1}{N}\sum_{n=1}^{N} f(T^n x)\mu(n) = 0 \tag{5.21}$$

for ν-a.e. $x \in X$ and each $f \in L^\infty(X,\nu)$.

Now, let $f \in L^1(X,\nu)$. Then, for any $\varepsilon > 0$ there exists a $g \in L^\infty(X,\nu)$ such that $\|f - g\|_{L^1} < \varepsilon$. Applying Birkhoff's pointwise ergodic theorem (Theorem 5.6) to $|f - g|$ yields

$$\lim_{N\to\infty} \frac{1}{N}\sum_{n=1}^{N} |f-g|(T^n x) = \left(\int_X |f-g|\,\mathrm{d}\,\nu\right)\cdot \mathbf{1}_X(x) = \|f-g\|_{L^1} < \varepsilon, \tag{5.22}$$

for ν-a.e. $x \in X$. Therefore, for ν-a.e. $x \in X$, $\limsup_{N\to\infty}\left|\frac{1}{N}\sum_{n=1}^{N} f(T^n x)\mu(n)\right|$ equals

$$\begin{aligned}
&\limsup_{N\to\infty}\left|\frac{1}{N}\sum_{n=1}^{N}(f-g)(T^n x)\mu(n) + \frac{1}{N}\sum_{n=1}^{N} g(T^n x)\mu(n)\right| \\
&\quad\le \limsup_{N\to\infty}\frac{1}{N}\sum_{n=1}^{N}|f-g|(T^n x)\underbrace{|\mu(n)|}_{\le 1} + \limsup_{N\to\infty}\left|\frac{1}{N}\sum_{n=1}^{N} g(T^n x)\mu(n)\right| \\
&\quad\le \underbrace{\lim_{N\to\infty}\frac{1}{N}\sum_{n=1}^{N}|f-g|(T^n x)}_{\overset{(5.22)}{<}\varepsilon} + \underbrace{\lim_{N\to\infty}\left|\frac{1}{N}\sum_{n=1}^{N} g(T^n x)\mu(n)\right|}_{\overset{(5.21)}{=}0} \\
&\quad< \varepsilon.
\end{aligned}$$

So the limit exists and the assertion follows, since ε has been chosen arbitrarily close to 0. □

Whenever we obtain a statement for almost every $x \in X$, the question arises if there is a simple way to apply it to any x. One could be tempted to merge the

results for the various MDS (X, Σ_X, ν, T) for each $\nu \in \mathfrak{M}_T$, hoping to cover every nullset this way. This would imply that Sarnak's conjecture holds for any dynamical system regardless of its topological entropy, since in Theorem 5.24 we did not need the given system to be deterministic. But that is not true. Several counterexamples show that, in general, we cannot renounce the zero entropy assumption. One for certain Toeplitz sequences was given by El Abdalaoui, Kułaga-Przymus, Lemańczyk and de la Rue in [22].

So, despite the unquestionable significance of the above ergodic theorem, the conjecture we are primarily occupied with remains unproven.

6. A Sufficient Condition for Sarnak's Conjecture

In this chapter we want to prove the orthogonality criterion of Katai–Bourgain–Sarnak–Ziegler (in short KBSZ-criterion) which applies to any bounded multiplicative function and thus, in particular, yields a sufficient condition for Sarnak's conjecture to hold.

As before, we denote by $\mathbb{P}$ the set of all prime numbers and by $\#A$ the cardinality of a finite set A.

Theorem 6.1 (KBSZ-criterion, quantitative version)**.** *Let $F : \mathbb{N} \to \mathbb{C}$ be bounded by 1 and let $\varphi : \mathbb{N} \to \{-1, 0, 1\}$ be a multiplicative number-theoretic function. Let $\tau \in (0, 1)$ be a small parameter and assume that for all $p_1, p_2 \in \left[1, e^{\frac{1}{\tau}}\right] \cap \mathbb{P}$, $p_1 \neq p_2$, there is an $M_0 \in \mathbb{N}$ such that for all $M \geq M_0$ we have*

$$\frac{1}{M}\left|\sum_{m=1}^{M} F(p_1 m)\overline{F(p_2 m)}\right| \leq \tau. \tag{6.1}$$

Then there exists an $N_0 \in \mathbb{N}$ such that for all $N \geq N_0$ we have

$$\frac{1}{N}\left|\sum_{n=1}^{N} \varphi(n)F(n)\right| \leq 2\sqrt{-\tau \log \tau}.$$

Note that it is sufficient to assume F to be bounded by an arbitrary $C > 0$. Therefore, Theorem 6.1 implies the following useful criterion.

Theorem 6.2 (KBSZ-criterion, qualitative version)**.** *Let $(F(n))_{n\in\mathbb{N}}$ be a complex-valued sequence for which $(|F(n)|)_{n\in\mathbb{N}}$ is bounded and which is such that for any pair of sufficiently large distinct primes p_1, p_2,*

$$\sum_{n=1}^{N} F(p_1 n)\overline{F(p_2 n)} = \mathrm{o}(N) \tag{6.2}$$

for $N \to \infty$. Then

$$\sum_{n=1}^{N} F(n)\mu(n) = \mathrm{o}(N)$$

for $N \to \infty$.

To apply this for varifying Sarnak's conjecture for a given TDS (X, T), for each $f \in C(X)$ and every $x \in X$ consider the sequence $(F(n))_{n\in\mathbb{N}}$ given by $F(n) := f(T^n x)$. Then, because of the continuity of the involved functions, $(|F(n)|)_{n\in\mathbb{N}}$ is bounded and the task is to find an $n_0 \in \mathbb{N}$ such that for all distinct primes p_1, p_2 greater than n_0 we have $\frac{1}{N}\sum_{n=1}^{N} F(p_1 n)\overline{F(p_2 n)} \xrightarrow[N\to\infty]{} 0$.

We will look into two different proofs of the criterion, but in both cases we will content ourselves with just a sketch of the respective proof.

6.1. About a Proof for the KBSZ-Criterion

The proof we will consider in this section was given by Bourgain, Sarnak and Ziegler in [9]. It makes use of the Chinese remainder theorem (see Theorem A.12 in the Appendix) and the prime number theorem (see Theorem 2.1). The basic idea of it is to decompose $[1, N] \cap \mathbb{Z}$ into a fixed number of pieces depending on the small parameter τ and chosen in a way that they cover most of the inteval and so that the members of the pieces have unique prime factors in suitable dyadic intervals. Then we will be able to estimate the key sum by bringing the multiplicativity of φ into usage.

For $f, g : \mathbb{N} \to \mathbb{R}$ write $f \lesssim g$ if asymptotically as $N \to \infty$, we have $f \leq g$, i.e., there is an $N_0 \in \mathbb{N}$ such that $\sup \{f(N) \,|\, N \geq N_0\} \leq \inf \{g(N) \,|\, N \geq N_0\}$.

Sketch of proof of Theorem 6.1. Let $\alpha \in (0, 1)$ be such that

$$(\log \alpha)^4 + \alpha \log \alpha > 0 \tag{6.3}$$

(to be chosen later depending on the parameter τ) and set

$$j_0 \coloneqq \frac{1}{\alpha} \left(\log \frac{1}{\alpha} \right)^3 = -\frac{(\log \alpha)^3}{\alpha},$$
$$j_1 \coloneqq j_0^2 = \frac{(\log \alpha)^6}{\alpha^2}.$$

Then, since

$$0 < (\log \alpha)^4 + \alpha \log \alpha \overset{\alpha > 0}{\iff} 0 < \frac{(\log \alpha)^6}{\alpha^2} + \frac{(\log \alpha)^3}{\alpha},$$

we have $j_0 < j_1$. Furthermore, define

$$D_0 \coloneqq (1 + \alpha)^{j_0},$$
$$D_1 \coloneqq (1 + \alpha)^{j_1}.$$

In order to decompose $[1, N] \cap \mathbb{Z}$ suitably, consider first the set S given by

$$S \coloneqq \{n \in [1, N] \cap \mathbb{Z} \,|\, n \text{ has a prime factor in } (D_0, D_1)\}.$$

Then one can show, by using the Chinese remainder theorem, that

$$\# \left([1, N] \cap \mathbb{Z} \setminus S \right) \lesssim \prod_{p \in (D_0, D_1) \cap \mathbb{P}} \left(1 - \frac{1}{p} \right) N.$$

By the prime number theorem and the choice of α we obtain

$$\prod_{p \in (D_0, D_1) \cap \mathbb{P}} \left(1 - \frac{1}{p} \right) \sim \frac{\log D_0}{\log D_1} = \frac{1}{j_0}$$

which implies

$$\# \left([1, N) \cap \mathbb{Z} \setminus S \right) \lesssim \alpha N,$$

that is, up to a fraction of α, S covers $[1, N) \cap \mathbb{Z}$.

Now, for each $j \in [j_0, j_1] \cap \mathbb{Z}$, define $P_j \coloneqq \mathbb{P} \cap \left[(1+\alpha)^j, (1+\alpha)^{j+1}\right]$ and

$$S_j \coloneqq \left\{ n \in [1, N) \cap \mathbb{Z} \,\middle|\, n \text{ has exactly one divisor in } P_j \text{ and no divisor in } \bigcup_{i<j} P_i \right\}.$$

Then for $j, j' \in [j_0, j_1] \cap \mathbb{Z}$, $j \neq j'$ we have $S_j \cap S_{j'} = \varnothing$. As above, consider $[1, N) \cap \mathbb{Z} \setminus S_j$ and appeal it to the prime number theorem to obtain

$$\#P_j = \frac{(1+\alpha)^{j+1}}{(j+1)\log(1+\alpha)} - \frac{(1+\alpha)^j}{j\log(1+\alpha)} + \mathrm{O}\left((1+\alpha)^j e^{-\sqrt{\alpha j}}\right). \tag{6.4}$$

Hence, for α sufficiently small,

$$\#P_j \leq (1+\alpha)^j \left(\frac{1}{j} + \frac{1}{\alpha j^2} + \mathrm{O}\left(e^{-\sqrt{\alpha j}}\right)\right). \tag{6.5}$$

Now, from the definition of S we have

$$S \setminus \bigcup_{j=j_0}^{j_1} S_j \subseteq \bigcup_{j=j_0}^{j_1} \left\{ n \in [1, N) \cap \mathbb{Z} \,\middle|\, \substack{n \text{ has at least two distinct} \\ \text{prime factors in } P_j} \right\}.$$

Hence one can show that

$$\#\left(S \setminus \bigcup_{j=j_0}^{j_1} S_j\right) \lesssim \sum_{j\in\mathbb{N}} \sum_{p_1,p_2\in P_j} \frac{N}{p_1 p_2} \leq N \sum_{\substack{j\in\mathbb{N} \\ j_0\leq j\leq j_1}} \left(\frac{\#P_j}{(1+\alpha)^j}\right)^2$$

and for α sufficiently small one deduces from (6.5) that

$$\begin{aligned}
\#\left(S \setminus \bigcup_{j=j_0}^{j_1} S_j\right) &\lesssim N \sum_{\substack{j\in\mathbb{N} \\ j_0\leq j\leq j_1}} \left(\frac{1}{j} + \frac{1}{\alpha j^2} + \mathrm{O}\left(e^{-\sqrt{\alpha j}}\right)\right)^2 \\
&\leq N\left(\frac{1}{j_0} + \frac{1}{j_0^3\alpha^2} + \mathrm{O}\left(\frac{1}{\alpha}\left(1+\sqrt{\alpha j_0}\right)e^{-\sqrt{\alpha j_0}}\right)\right) \\
&\leq \alpha N.
\end{aligned}$$

So the disjoint union $\bigcup_{j=j_0}^{j_1} S_j$ covers $[1, N] \cap \mathbb{Z}$ up to a fraction of α. Now we decompose each S_j into a well factored set and its complement. For $j \in [j_0, j_1] \cap \mathbb{Z}$ let

$$Q_j \coloneqq \left\{ m \in \left[1, \frac{N}{(1+\alpha)^{j+1}}\right) \cap \mathbb{Z} \,\middle|\, m \text{ has no prime factor in } \bigcup_{i\leq j} P_i \right\}.$$

Then, for each $j \in [j_0, j_1] \cap \mathbb{Z}$, the product sets $P_j \cdot Q_j \coloneqq \{pq \,|\, p \in P_j, q \in Q_j\}$ satisfy

$$P_j \cdot Q_j \subseteq S_j.$$

Moreover, for each $j \in [j_0, j_1] \cap \mathbb{Z}$,

$$S_j \setminus (P_j \cdot Q_j) \subseteq P_j \cdot \left(\left[\frac{N}{(1+\alpha)^{j+1}}, \frac{N}{(1+\alpha)^j}\right] \cap \mathbb{Z}\right)$$

and hence using (6.4) one shows that

$$\begin{aligned}\sum_{\substack{j\in\mathbb{N}\\ j_0\le j\le j_1}} \#\left(S_j\setminus(P_j\cdot Q_j)\right) &\le \sum_{\substack{j\in\mathbb{N}\\ j_0\le j\le j_1}} (\#P_j)\,\frac{\alpha N}{(1+\alpha)^j}\\ &\le N\left(\alpha\log\frac{j_1}{j_0}+\frac{1}{j_0}+\mathrm{O}\left(\left(1+\sqrt{\alpha j_0}\right)e^{-\sqrt{\alpha j_0}}\right)\right),\end{aligned} \tag{6.6}$$

From which for α sufficiently small one obtains

$$\sum_{\substack{j\in\mathbb{N}\\ j_0\le j\le j_1}} \#\left(S_j\setminus(P_j\cdot Q_j)\right)\le 2\alpha N. \tag{6.7}$$

Now, by (6.7) and the definition of Q_j, one deduces

$$\#\left([1,N)\cap\mathbb{Z}\setminus\bigcup_{j=j_0}^{j_1}(P_j\cdot Q_j)\right)\lesssim 3\alpha N,$$

which yields a decomposition of $[1,N)\cap\mathbb{Z}$ into disjoint sets $P_j\cdot Q_j$, $j\in[j_0,j_1]\cap\mathbb{Z}$, with only a small proportion of points omitted.

Now, since the map $P_j\times Q_j\to P_j\cdot Q_j$, $(p,q)\mapsto pq$, is injective and because of $|F|\le 1$ and $|\varphi|\le 1$ we have

$$\left|\sum_{n=1}^{N}\varphi(n)F(n)\right|\lesssim\sum_{\substack{j\in\mathbb{N}\\ j_0\le j\le j_1}}\left|\sum_{p\in P_j,q\in Q_j}\varphi(pq)F(pq)\right|+3\alpha N. \tag{6.8}$$

By the choice of Q_j and P_j we have $(p,q)=1$ for each $p\in P_j$ and each $q\in Q_j$. Therefore, by the multiplicativity of φ, we have $\varphi(pq)=\varphi(p)\varphi(q)$ and hence

$$\begin{aligned}\left|\sum_{n=1}^{N}\varphi(n)F(n)\right| &\lesssim \sum_{\substack{j\in\mathbb{N}\\ j_0\le j\le j_1}}\sum_{q\in Q_j}\underbrace{|\varphi(q)|}_{\le 1}\left|\sum_{p\in P_j}\varphi(p)F(pq)\right|+3\alpha N\\ &\le \sum_{\substack{j\in\mathbb{N}\\ j_0\le j\le j_1}}\sum_{q\in Q_j}\left|\sum_{p\in P_j}\varphi(p)F(pq)\right|+3\alpha N.\end{aligned} \tag{6.9}$$

By estimating the inner sum using the Cauchy–Schwarz inequality we obtain

$$\begin{aligned}\sum_{q\in Q_j}\left|\sum_{p\in P_j}\varphi(p)F(pq)\right| &\le \sqrt{\sum_{q\in Q_j}1}\sqrt{\sum_{q\in Q_j}\left|\sum_{p\in P_j}\varphi(p)F(pq)\right|^2}\\ &\le \sqrt{\#Q_j}\sqrt{\sum_{\substack{q\in\mathbb{N}\\ q\le\frac{N}{(1+\alpha)^j}}}\left|\sum_{p\in P_j}\varphi(p)F(pq)\right|^2}\\ &= \sqrt{\#Q_j}\sqrt{\sum_{\substack{q\in\mathbb{N}\\ q\le\frac{N}{(1+\alpha)^j}}}\sum_{p_1,p_2\in P_j}\varphi(p_1)\overline{\varphi(p_2)}F(p_1q)\overline{F(p_2q)}} \qquad (6.10)\\ &\overset{|\varphi|\le 1}{\le} \sqrt{\#Q_j}\sqrt{\sum_{p_1,p_2\in P_j}\left|\sum_{\substack{q\in\mathbb{N}\\ q\le\frac{N}{(1+\alpha)^j}}}F(p_1q)\overline{F(p_2q)}\right|}.\end{aligned}$$

Note that here

$$p_1,p_2<(1+\alpha)^{j_1}<e^{\frac{1}{\alpha^2}}. \qquad (6.11)$$

The diagonal contribution in (6.10), that is $p_1=p_2\,(=:p)$ for each j, yields (by using that $|F|\le 1$ and the definition of Q_j)

$$\sum_{p\in P_j}\left|\sum_{\substack{q\in\mathbb{N}\\ q\le\frac{N}{(1+\alpha)^j}}}F(pq)\overline{F(pq)}\right|\le\sqrt{\frac{(\#Q_j)\,(\#P_j)\,N}{(1+\alpha)^j}}=\sqrt{\#Q_j}\sqrt{\#P_j}\frac{\sqrt{N}}{(1+\alpha)^{\frac{j}{2}}}$$

and thus, again with the Cauchy–Schwarz inequality,

$$\begin{aligned}\sum_{\substack{j\in\mathbb{N}\\ j_0\le j\le j_1}}\sum_{p\in P_j}\left|\sum_{\substack{q\in\mathbb{N}\\ q\le\frac{N}{(1+\alpha)^j}}}F(pq)\overline{F(pq)}\right| &\le \frac{\sqrt{\sum_{\substack{j\in\mathbb{N}\\ j_0\le j\le j_1}}(\#P_j)\,(\#Q_j)}\sqrt{\sum_{\substack{j\in\mathbb{N}\\ j_0\le j\le j_1}}\frac{1}{(1+\alpha)^j}}}{\sqrt{N}}\\ &\le \sqrt{N}\sqrt{N}\sqrt{\sum_{\substack{j\in\mathbb{N}\\ j_0\le j\le j_1}}\frac{1}{(1+\alpha)^j}}\\ &= N\sqrt{\sum_{\substack{j\in\mathbb{N}\\ j_0\le j\le j_1}}\frac{1}{(1+\alpha)^j}}\\ &\le \alpha N,\end{aligned}$$

(6.12)

since $(\#P_j)\,(\#Q_j)=\#\,(P_j\cdot Q_j)\le\#S_j$, for each $j\in[j_0,j_1]\cap\mathbb{Z}$, and

$$\sum_{\substack{j\in\mathbb{N}\\ j_0\le j\le j_1}}\#S_j\le N.$$

For $p_1 \neq p_2$, we apply the assumption (6.1) for p_1, p_2 sufficiently large in view of (6.11), that is

$$\frac{1}{N}\left|\sum_{\substack{q\in\mathbb{N}\\ q\leq\frac{N}{(1+\alpha)^j}}} F(p_1q)\overline{F(p_2q)}\right| \leq \frac{\tau}{(1+\alpha)^j}.$$

Hence, once again by the Cauchy–Schwarz inequality,

$$\begin{aligned}
&\sum_{\substack{j\in\mathbb{N}\\ j_0\leq j\leq j_1}} \sum_{\substack{p_1,p_2\in P_j\\ p_1\neq p_2}} \left|\sum_{\substack{q\in\mathbb{N}\\ q\leq\frac{N}{(1+\alpha)^j}}} F(p_1q)\overline{F(p_2q)}\right| \\
&\quad\leq \sum_{\substack{j\in\mathbb{N}\\ j_0\leq j\leq j_1}} \sqrt{\#Q_j}\sqrt{\sum_{\substack{p_1,p_2\in P_j\\ p_1\neq p_2}} \frac{\tau N}{(1+\alpha)^j}} \\
&\quad\leq \sqrt{\tau N}\sum_{\substack{j\in\mathbb{N}\\ j_0\leq j\leq j_1}} (\#P_j)\sqrt{\#Q_j}\,(1+\alpha)^{-\frac{j}{2}} \\
&\quad\leq \sqrt{\tau N}\sqrt{\sum_{\substack{j\in\mathbb{N}\\ j_0\leq j\leq j_1}} (\#P_j)\,(\#Q_j)}\sqrt{\sum_{\substack{j\in\mathbb{N}\\ j_0\leq j\leq j_1}} (\#P_j)\,(1+\alpha)^{-j}} \\
&\quad\overset{(6.6)}{\leq} \sqrt{\tau N}\sqrt{N}\sqrt{\log\frac{j_1}{j_0}+\frac{1}{j_0\alpha}+\frac{1}{\alpha}\left(1+\sqrt{\alpha j_0}\right)e^{-\sqrt{j_0}}} \\
&\quad\leq N\sqrt{\tau}\sqrt{\log\left(-\frac{(\log\alpha)^3}{\alpha}\right)-(\log\alpha)^{-3}+\frac{1}{\alpha}+\sqrt{-(\log\alpha)^3}} \\
&\quad\leq N\sqrt{\tau}\sqrt{\log\frac{1}{\alpha}},
\end{aligned} \tag{6.13}$$

for α sufficiently small.

Now, combining (6.9), (6.12) and (6.13) yields

$$\left|\sum_{n=1}^{N}\varphi(n)F(n)\right| \lesssim \alpha N + N\sqrt{\tau}\sqrt{\log\frac{1}{\alpha}} + 3\alpha N = N\left(4\alpha+\sqrt{\tau}\sqrt{\log\frac{1}{\alpha}}\right).$$

By choosing $\alpha = \sqrt{\tau}$ we obtain

$$\frac{1}{N}\left|\sum_{n=1}^{N}\varphi(n)F(n)\right| \lesssim \sqrt{\tau}\left(4+\sqrt{\log\frac{1}{\sqrt{\tau}}}\right) = 4\sqrt{\tau}+\frac{1}{\sqrt{2}}\sqrt{-\tau\log\tau},$$

which is not greater than $2\sqrt{-\tau\log\tau}$ for $\tau \in \left(0, e^{\frac{32}{4\sqrt{2}-9}}\right]$ (cf. Remark 6.3 below) and the assertion follows, since this α suffices the condition (6.3) for such a τ. □

Remark 6.3. We obtained the estimation for τ in the last step of the above proof by the following calculation (assume that $\tau \in (0,1)$):

$$\begin{aligned}
& 4\sqrt{\tau} + \frac{1}{\sqrt{2}}\sqrt{-\tau\log\tau} \leq 2\sqrt{-\tau\log\tau} \\
\iff & 4\sqrt{\tau} && \leq \left(2 - \frac{1}{\sqrt{2}}\right)\sqrt{-\tau\log\tau} \\
\iff & 16\tau && \leq \left(2 - \frac{1}{\sqrt{2}}\right)^2 (-\tau\log\tau) \\
\iff & \frac{16}{\left(2 - \frac{1}{\sqrt{2}}\right)^2} && \leq -\log\tau \\
\iff & -\frac{64}{\left(4 - \sqrt{2}\right)^2} && \geq \log\tau \\
\iff & \frac{32}{4\sqrt{2} - 9} && \geq \log\tau.
\end{aligned}$$

Note that $0.000069 < e^{\frac{32}{4\sqrt{2}-9}} < 0.00007$. This gives a good impression about just how small the parameter τ has to be. Moreover, condition (6.3) holds for $\alpha = \sqrt{\tau}$, whenever $\tau \in (0, \eta)$, where η denotes the root of $(\log x)^3 + 8x$ near $x = 0.273163$, which is obviously the case.

Furthermore, recall that we assumed (6.1) to hold for $p_1, p_2 \in \left[1, e^{\frac{1}{\tau}}\right] \cap \mathbb{P}$, $p_1 \neq p_2$. So, by the above proof, for the namely estimation to hold we need to have (6.1) for at least all distinct primes not greater than $\exp\left(e^{-\frac{32}{4\sqrt{2}-9}}\right)$, which is larger than $1.2814 \cdot 10^{6234}$.

(All values have been calculated using `Mathematica`.)

6.2. About another Proof for the KBSZ-Criterion

We want to give another proof for the desired criterion, which this time varifies the assertion of Theorem 6.2 directly. The namely proof was given by Tao in [57] and makes use of the Turan–Kubilius inequality (see Lemma 6.6 below) as well as of the following classical result.

Lemma 6.4 (Theorem of Euler)**.** *The series* $\sum_{p\in\mathbb{P}} \frac{1}{p}$ *diverges.*

Proof. We have $e = \sup_{n\in\mathbb{N}} \left(1 + \frac{1}{n}\right)^n$. Hence, for each prime number p we have $\left(1 + \frac{1}{p-1}\right)^{p-1} < e$ and therefore

$$1 + \frac{1}{p-1} < e^{\frac{1}{p-1}}. \tag{6.14}$$

Thus, we can conclude

$$\begin{aligned}
\log\left(\frac{1}{1-\frac{1}{p}}\right) = \log\left(\frac{p}{p-1}\right) = \log\left(1 + \frac{1}{p-1}\right) &\overset{(6.14)}{<} \frac{1}{p-1} = \frac{p}{p\,(p-1)} \\
= \frac{p-1}{p\,(p-1)} + \frac{1}{p(p-1)} &= \frac{1}{p} + \frac{1}{p\,(p-1)} \leq \frac{2}{p}.
\end{aligned}$$

Now let $N \in \mathbb{N}$ and $(p_k)_{k=1,\dots,k(N)}$ be the sequence of all primes not greater than N. Then, by the above,

$$\sum_{k=1}^{k(N)} \frac{1}{p_k} > \frac{1}{2} \sum_{k=1}^{k(N)} \log\left(\frac{1}{1-\frac{1}{p_k}}\right) = \frac{1}{2} \log\left(\prod_{k=1}^{k(N)} \frac{1}{1-\frac{1}{p_k}}\right). \tag{6.15}$$

Because of $\log x \xrightarrow[x\to\infty]{} \infty$ it sufficies to show that

$$\lim_{N\to\infty} \prod_{k=1}^{k(N)} \frac{1}{1-\frac{1}{p_k}} = \infty.$$

Note that we have $\prod_{k=1}^{k(N)} \frac{1}{1-\frac{1}{p_k}} = \prod_{k=1}^{k(N)} \left(\sum_{n=0}^{\infty} \left(\frac{1}{p_k}\right)^n\right)$ and the geometric series involved converge absolutely for any such k. So we can expand this product and obtain, by the fundamental theorem of arithmetic, the sum over the reciprocals of all positive integers of the form $\prod_{k=1}^{k(N)} p_k^{a_k}$ with $a_k \in \mathbb{N}_0$ for each $k \in [1, k(N)] \cap \mathbb{Z}$ (and each such integer exactly once). Hence, by setting $\mathcal{N} \coloneqq \{n \in \mathbb{N} \,|\, p \in \mathbb{P},\, p|n \Rightarrow p \leq N\}$, we can write

$$\prod_{k=1}^{k(N)} \frac{1}{1-\frac{1}{p_k}} = \sum_{n\in\mathcal{N}} \frac{1}{n}.$$

Because of $\mathcal{N} \setminus [1, N] \neq \varnothing$ (e.g. we have $\prod_{k=1}^{k(N)} p_k > N$) it follows that

$$\prod_{k=1}^{k(N)} \frac{1}{1-\frac{1}{p_k}} > \sum_{n=1}^{N} \frac{1}{n}.$$

Since the harmonic series diverges the assertion follows from (6.15). □

Remark 6.5. Lemma 6.4 dates back to 1737 and is one of the first results implying that there are infinitely many prime numbers.

Let $\eta = \eta(N)$ be a slowly growing function with $\eta(N) \to \infty$ as $N \to \infty$. By Lemma 6.4 we have

$$\sum_{\substack{p\in\mathbb{P}\\ p<\eta(N)}} \frac{1}{p} \xrightarrow[N\to\infty]{} \infty.$$

It will also be convenient to eliminate small primes. Note that we can find an even slower growing function $\omega = \omega(N)$, with $\omega(N) \to \infty$ as $x \to \infty$, such that

$$\sum_{\substack{p\in\mathbb{P}\\ \omega(N)\leq p<\eta(N)}} \frac{1}{p} \xrightarrow[N\to\infty]{} \infty.$$

Therefore, for $P(N) \coloneqq \mathbb{P} \cap [\omega(N), \eta(N))$ and $\beta = \beta(N)$ given by

$$\beta(N) \coloneqq \sum_{p\in P(N)} \frac{1}{p}$$

we have $\beta(N) \to \infty$ as $N \to \infty$. We will take ω and η to be powers of 2.

In what follows all Bachmann–Landau symbols are meant in the sense $N \to \infty$.

Lemma 6.6 (Turan–Kubilius inequality)**.** *We have*

$$\sum_{n=1}^{N} \left| \sum_{\substack{p \in P(N) \\ p \mid n}} 1 - \beta(N) \right|^2 \ll N\beta(N)$$

as $N \to \infty$.

Proof. We have

$$\sum_{n=1}^{N} \sum_{\substack{p \in P(N) \\ p \mid n}} 1 = \sum_{p \in P(N)} \sum_{\substack{n=1 \\ p \mid n}}^{N} 1.$$

Moreover,

$$\sum_{\substack{n=1 \\ p \mid n}}^{N} 1 = \frac{N}{p} + \mathrm{O}\left(1\right)$$

and therefore (for η sufficiently slowly growing)

$$\sum_{n=1}^{N} \sum_{\substack{p \in P(N) \\ p \mid n}} 1 = N \cdot \beta(N) + \mathrm{O}\left(N\right).$$

Analogously, we obtain

$$\sum_{n=1}^{N} \left(\sum_{\substack{p \in P(N) \\ p \mid n}} 1 \right)^2 = \sum_{p,q \in P(N)} \sum_{\substack{n=1 \\ p \mid n,\, q \mid n}}^{N} 1.$$

Note, that

$$\sum_{\substack{p,q \in P(N) \\ p \mid n, q \mid n}} 1 = \begin{cases} \frac{N}{p} + \mathrm{O}\left(1\right) & \text{for } p = q \\ \frac{N}{pq} + \mathrm{O}\left(1\right) & \text{otherwise.} \end{cases}$$

Putting everything together, we obtain

$$\sum_{n=1}^{N} \left(\sum_{\substack{p \in P(N) \\ p \mid n}} 1 \right)^2 = N\left(\beta(N)\right)^2 + \mathrm{O}\left(N \cdot \beta(N)\right),$$

for η sufficiently slowly growing, which yields the assertion. □

Sketch of proof of Theorem 6.2. From Lemma 6.6 and the Cauchy–Schwarz inequality we have

$$\sum_{n=1}^{N} \left(\sum_{\substack{p \in P(N) \\ p \mid n}} 1 - \beta(N) \right) \mu(n) F(n) = \mathrm{O}\left(N\sqrt{\beta(N)} \right),$$

which we can rearrange as

$$\sum_{n=1}^{N} \mu(n)F(n) = \frac{1}{\beta(N)} \sum_{p\in P(N)} \sum_{\substack{n=1\\ p|n}}^{N} \mu(n)F(n) + \mathrm{O}\left(N\sqrt{\beta(N)}\right).$$

Since $\beta(N) \xrightarrow[N\to\infty]{} \infty$ we have $\mathrm{O}\left(N\sqrt{\beta(N)}\right) = \mathrm{o}(N)$ and hence it sufficies to show that

$$\sum_{p\in P(N)} \sum_{\substack{n=1\\ p|n}}^{N} \mu(n)F(n) = \mathrm{o}\left(N\cdot\beta(N)\right). \tag{6.16}$$

For $p|n$ we have $\frac{n}{p} =: m \in \mathbb{N}$ and $\mu(n)F(n) = -\mu(m)F(pm)$ for all but $\mathrm{O}\left(\frac{N}{p^2}\right)$ values of n (see [57]). These exceptional values contribute at most

$$\sum_{p\in P(N)} \frac{N}{p^2} \leq \sum_{p\in P(N)} \frac{N}{p\cdot\omega(N)} = \mathrm{O}\left(\frac{N\cdot\beta(N)}{\omega(N)}\right) = \mathrm{o}\left(N\cdot\beta(N)\right),$$

which is acceptable. Taking this into account as well as (6.16), it sufficies to show that

$$\sum_{p\in P(N)} \sum_{m\leq\frac{N}{p}} \mu(m)F(pm) = \mathrm{o}\left(N\cdot\beta(N)\right). \tag{6.17}$$

Now, by splitting up $P(N)$ into dyadic blocks $P_k(N) \coloneqq \left\{p\in P(N)\,\middle|\, 2^k < p < 2^{k+1}\right\}$ and noting that $\beta(N) \geq \sum_k \frac{\#P_k(N)}{2^{k+1}}$, (6.17) follows from

$$\sum_{p\in P_k(N)} \sum_{m\leq\frac{N}{p}} \mu(m)F(pm) = \mathrm{o}\left(\frac{N}{2^k}\left(\#P_k(N)\right)\right) \tag{6.18}$$

uniformly in k whenever $\omega(N) \leq 2^k < \eta(N)$. So, it sufficies to show that (6.18) holds. To do so fix k. Then one can show that

$$\sum_{p\in P_k(N)} \sum_{m\leq\frac{N}{p}} \mu(m)F(pm) = \sum_{m\leq\frac{N}{2^k}} \mu(m) \sum_{p\in P_k(N)} F(pm)\cdot\mathbf{1}_{\left[1,\frac{N}{p}\right]\cap\mathbb{Z}}(p).$$

So, by the Cauchy–Schwarz inequality, the fact that $(\mu(m))^2 \in \{0,1\}$ for each $m\in\mathbb{N}$, and (6.18) it sufficies to show that

$$\sum_{m\leq\frac{N}{2^k}} \left|\sum_{p\in P_k(N)} F(pm)\cdot\mathbf{1}_{\left[1,\frac{N}{p}\right]\cap\mathbb{Z}}(p)\right|^2 = \mathrm{o}\left(\frac{N}{2^k}\left(\#P_k(N)\right)^2\right),$$

where one can rewrite the left-hand side as

$$\sum_{p,q\in P_k(N)} \sum_{m\leq\min\left\{\frac{N}{p},\frac{N}{q}\right\}} F(pm)\overline{F(qm)}$$

so that we have to show

$$\sum_{p,q\in P_k(N)} \sum_{m\leq\min\left\{\frac{N}{p},\frac{N}{q}\right\}} F(pm)\overline{F(qm)} = \mathrm{o}\left(\frac{N}{2^k}\left(\#P_k(N)\right)^2\right). \tag{6.19}$$

Now, if η grows sufficiently slowly in N, the assumption (6.2) implies that for any sufficiently large $p, q \in \mathbb{P}$, $p \neq q$, $p, q \leq \eta(N)$, we have

$$\sum_{m \leq \min\left\{\frac{N}{p}, \frac{N}{q}\right\}} F(pm)\overline{F(qm)} = \mathrm{o}\left(\frac{N}{2^k}\right)$$

uniformly in p and q, for any k such that $\omega(N) \leq 2^k < \eta(N)$, while for $p = q$ we find

$$\sum_{m \leq \frac{N}{p}} F(pm)\overline{F(pm)} = \mathrm{O}\left(\frac{N}{2^k}\right).$$

By taking into account that $\#P_k(N) = \mathrm{o}\left((\#P_k(N))^2\right)$ (which follows from Lemma 6.4), this implies (6.19) and the proof is complete. □

7. Some Examples of Systems for which Sarnak's Conjecture holds

We want to collect some examples of dynamical systems for which Sarnak's conjecture is known to hold. In this context the easiest systems imaginable are those providing $(f(T^n x))_{n\in\mathbb{N}}$ to be either constant or periodic. In both cases one easily checks that the underlying dynamical system is deterministic (in the first case consider $X = \{x\}$ for some $x \in \mathbb{C}$ and in the second case consider X to be the finite (and therefore compact) abelian group $\mathbb{Z}/q\mathbb{Z}$ with $T : m \mapsto m + 1 \mod q$, $q \in \mathbb{N}_0$ the period).

Proposition 7.1. SARNAK*'s conjecture holds for constant sequences.*

Proof. Since the value of the constant does not contribute in terms of convergence it suffices to show that

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} \mu(n) = 0.$$

But this is immediate from Theorem 2.5 and the prime number theorem (Theorem 2.1). □

Proposition 7.2. SARNAK*'s conjecture holds for periodic sequences.*

To prove Proposition 7.2, consider the following decomposition (taken from [43]) first:

Let $(b_n)_{n\in\mathbb{N}}$ be a periodic sequence, i.e., there is a $q \in \mathbb{N}_0$ such that $b_{n+q} = b_n$ for each $n \in \mathbb{N}$ (for the purpose of uniqueness, since $b_{n+q'} = b_n$ also holds for each multiple q' of q, let q be the least integer with this property). Then we can express $(b_n)_{n\in\mathbb{N}}$ as a linear combination

$$(b_n)_{n\in\mathbb{N}} = \sum_{a=1}^{q} b_a \left(\mathbf{1}_{a,q}(n)\right)_{n\in\mathbb{N}},$$

where $(\mathbf{1}_{a,q}(n))_{n\in\mathbb{N}}$ denotes the characteristic function of the arithmetic progression $\{a + lq \,|\, l \in \mathbb{N}_0\}$. For $q = 2$, we can express $\mathbf{1}_{a,2}$ as a linear combination $\frac{1}{2}1^n \pm \frac{1}{2}(-1)^n$ of 1^n and $(-1)^n$, the nth powers of the square-roots of 1. Similarly, we express the characteristic function of an arithmetic progression modulo q as a linear combination of the sequences ξ_k^n where ξ_k runs over the q different qth roots of unity:

$$\xi_k := \exp\left(2\pi i \frac{k}{q}\right)$$

for $k \in [1, q] \cap \mathbb{Z}$. From the formula for the finite geometric series

$$\sum_{n=0}^{q-1} \xi^n = \frac{1 - \xi^q}{1 - \xi}$$

we see that for ξ the qth root of unity

$$\sum_{n=0}^{q-1} \xi^n = 0,$$

unless $\xi = 1$, since for $(k, q) = 1$, q is the least integer n such that $\xi_k^n = 1$. Hence

$$\frac{1}{q} \sum_{k=1}^{q} \exp\left(2\pi i \frac{k(n-a)}{q}\right) = \begin{cases} 1 & \text{if } n \in \{a + lq \,|\, l \in \mathbb{N}_0\} \\ 0 & \text{otherwise} \end{cases}$$

and we can express the characteristic function $(\mathbf{1}_{a,q}(n))_{n\in\mathbb{N}}$ of an arbitrary arithmetic progression $\{a + lq \,|\, l \in \mathbb{N}_0\}$ as a linear combination of the sequences $(\chi_{a,k}(n))_{n\in\mathbb{N}}$ with $\chi_{a,k}(n) \coloneqq \exp(2\pi i \frac{k(n-a)}{q})$.

Proof of Proposition 7.2. Denote by $q \in \mathbb{N}_0$ the period of $(b_n)_{n\in\mathbb{N}}$. From Lemma 5.8 and Remark 5.9 we deduce

$$\sum_{n=1}^{N} \chi_{a,k}(n)\mu(n) = \mathrm{O}\left(N \exp\left(-c\sqrt{\log N}\right)\right)$$

for each $N \in \mathbb{N} \setminus \{1\}$ and every $a, k \in [1, q] \cap \mathbb{Z}$, and $c > 0$ a constant just depending on q. Hence there exists a $C_{a,k,q} > 0$ such that

$$\left|\sum_{n=1}^{N} \chi_{a,k}(n)\mu(n)\right| \leq C_{a,k,q} N \exp\left(-c\sqrt{\log N}\right).$$

Together with the above decomposition, for any periodic sequence $(b_n)_{n\in\mathbb{N}}$ with period $q \in \mathbb{N}_0$, we obtain

$$\begin{aligned}
0 \leq \left|\frac{1}{N}\sum_{n=1}^{N} b_n\mu(n)\right| &= \left|\frac{1}{N}\sum_{n=1}^{N}\left(\sum_{a=1}^{q} b_a \mathbf{1}_{a,q}(n)\right)\mu(n)\right| \\
&= \left|\frac{1}{N}\sum_{n=1}^{N}\left(\sum_{a=1}^{q}\frac{b_a}{q}\sum_{k=1}^{q}\chi_{a,k}(n)\right)\mu(n)\right| \\
&\leq \frac{1}{N}\sum_{a=1}^{q}\frac{|b_a|}{q}\sum_{k=1}^{q}\left|\sum_{n=1}^{N}\chi_{a,k}(n)\mu(n)\right| \\
&\leq \sum_{a=1}^{q}\frac{|b_a|}{q}\sum_{k=1}^{q} C_{a,k,q}\exp\left(-c\sqrt{\log N}\right) \\
&\leq q \max_{a\in[1,q]\cap\mathbb{Z}} |b_a| \cdot \max_{a,k\in[1,q]\cap\mathbb{Z}} C_{a,k,q} \cdot \exp\left(-c\sqrt{\log N}\right) \\
&\xrightarrow[N\longrightarrow\infty]{} 0.
\end{aligned}$$

□

Before we tend to two more complicated examples we want to record the following result stating that it suffices to show that Sarnak's conjecture holds for a linearly dense subset of $C(X)$. Recall that we call a subset N of a vector space M *linearly dense* in M, if the set $\mathrm{lin}(N)$ of all finite linear combinations of elements of N is dense in M.

Lemma 7.3. *Let (X,T) be a metric TDS with $h(T)=0$ and let $\mathcal{M}\subseteq C(X)$ be linearly dense such that for each $x\in X$ and every $g\in\mathcal{M}$ we have*

$$\lim_{N\to\infty}\frac{1}{N}\sum_{n=1}^{N} g(T^n x)\mu(n)=0. \tag{7.1}$$

Then Sarnak*'s conjecture holds for (X,T).*

Proof. Fix $f\in C(X)$. Since $\mathcal{M}$ is linearly dense in $C(X)$ for each $\varepsilon>0$ there are a $k\in\mathbb{N}$, $g_1,\dots,g_k\in\mathcal{M}$ and $a_1,\dots,a_k\in\mathbb{C}$ such that

$$\sup_{x\in X}\left|f(x)-\sum_{j=1}^{k}a_j g_j(x)\right|<\frac{\varepsilon}{2}.$$

Furthermore, because of (7.1), there is an $N_0\in\mathbb{N}$ such that for all $N\geq N_0$, each $j\in[1,k]\cap\mathbb{Z}$ and every $x\in X$ we have

$$\left|\frac{1}{N}\sum_{n=1}^{N} g_j(T^n x)\mu(n)\right|<\frac{\varepsilon}{2}\left(k\cdot\max_{j\in[1,k]\cap\mathbb{Z}}|a_j|\right)^{-1}.$$

Hence for each $N\geq N_0$ and every $x\in X$ we have

$$\begin{aligned}
\left|\frac{1}{N}\sum_{n=1}^{N} f(T^n x)\mu(n)\right| &= \left|\frac{1}{N}\sum_{n=1}^{N}\left(f-\sum_{j=1}^{k}a_j g_j+\sum_{j=1}^{k}a_j g_j\right)(T^n x)\mu(n)\right| \\
&\leq \frac{1}{N}\sum_{n=1}^{N}\underbrace{\left|f(T^n x)-\sum_{j=1}^{k}a_j g_j(T^n x)\right|}_{<\frac{\varepsilon}{2}}\underbrace{|\mu(n)|}_{\leq 1} \\
&\quad +\sum_{j=1}^{k}a_j\underbrace{\left|\frac{1}{N}\sum_{n=1}^{N} g_j(T^n x)\mu(n)\right|}_{<\frac{\varepsilon}{2}\left(k\cdot\max_{j\in[1,k]\cap\mathbb{Z}}a_j\right)^{-1}} \\
&< \frac{\varepsilon}{2}+\left(k\cdot\max_{j\in[1,k]\cap\mathbb{Z}}|a_j|\right)\cdot\frac{\varepsilon}{2}\left(k\cdot\max_{j\in[1,k]\cap\mathbb{Z}}|a_j|\right)^{-1}=\varepsilon. \quad \square
\end{aligned}$$

7.1. Möbius Function Randomness for the Thue–Morse Shift

The results of this section can be found in [21]. Denote by t the Thue–Morse sequence as defined in Subsection 3.4.1. For S the left shift on $\{0,1\}^{\mathbb{N}_0}$ consider $K_t:=\overline{\{S^n t\,|\,n\in\mathbb{N}_0\}}$ and let X_t be the set of all sequences $x\in\{0,1\}^{\mathbb{Z}}$ such that any finite subword of x also appears on some $y\in K_t$ (and hence on t). Denote also by t any extension of t to a two-sided member of X_t. Then, for any such extension t, we have $\overline{\{S^n t\,|\,n\in\mathbb{Z}\}}=X_t$ (see [21]), where S denotes the (invertible) shift on $\{0,1\}^{\mathbb{Z}}$ (therefore any such extension of t is suitable). Furthermore, X_t is closed and S-invariant.

Denote by ρ the map on X_t which interchanges 0s and 1s. Then ρ is a homeomorphism which ranges over X_t (i.e., ρ preserves X_t) and commutes with the shift S. For $f \in C(X_t)$ define

$$f_1 \coloneqq \frac{1}{2}\left(f + f \circ \rho\right) \quad \text{and} \quad f_2 \coloneqq \frac{1}{2}\left(f - f \circ \rho\right).$$

Then $f = f_1 + f_2$ and $f_1 = f_1 \circ \rho$, $f_2 = -\left(f_2 \circ \rho\right)$ (pointwise). Moreover, both f_1, f_2 are continuous. Hence, recalling Theorem 3.33, it suffices to verify the two statements

Proposition 7.4. *For each $x \in X_t$*

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} f_1(S^n x)\mu(n) = 0.$$

Proposition 7.5. *For each $x \in X_t$*

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} f_2(S^n x)\mu(n) = 0.$$

A proof for Proposition 7.5 can be found in [21]. It goes along the following lines demanding further investigations in spectral theory:

1. Denote by $\nu_t \in \mathfrak{M}(\mathbb{T})$ the spectral measure associated to the Thue–Morse sequence as well as by $\nu_t^{(k)}$ the image of ν_t via the map $z \mapsto z^k$. Then, for any odd $p, q \in \mathbb{N}$, $p \neq q$, one shows that $\nu_t^{(p)}$ and $\nu_t^{(q)}$ are mutually singular (write $\nu_t^{(p)} \perp \nu_t^{(q)}$), i.e., there is an $A \in \Sigma_{\mathbb{T}}$ such that $\nu_t^{(p)}(A) = \nu_t^{(q)}(X \setminus A) = 0$ (cf. Corollary 3 in [21]).

2. In [48] it is shown that ν_t is the sum of the discrete measure ν' concentrated on all roots of unity of degree 2^n, $n \geq 0$, and the continuous measure $\widetilde{\nu}_t$ which arises as the convolution of ν' with ν_t. Thus Remark 1 in [21] implies that 1. also holds for $\widetilde{\nu}_t$ instead of ν_t.

3. One proves that the spectral measure ν_{f_2} of f_2 given by Theorem 5.22 is absolutely continuous with respect to $\widetilde{\nu}_t$, i.e., $\mathcal{N}_{\widetilde{\nu}_t} \subseteq \mathcal{N}_{\nu_{f_2}}$, and concludes that 1. also holds for ν_{f_2} instead of ν_t. Hence we have $\nu_{f_2}^{(p)} \perp \nu_{f_2}^{(q)}$ for all odd $p, q \in \mathbb{N}$, $p \neq q$, and therefore, in particular, for all distinct $p, q \in \mathbb{P} \setminus \{2\}$.

4. In [23] it is shown that $\nu_{f_2}^{(p)}$ being mutually singular to $\nu_{f_2}^{(q)}$, for any distinct odd primes p, q, implies

$$\frac{1}{N} \sum_{n=1}^{N} f_2\left(S^{pn}(x)\right) \overline{f_2\left(S^{qn}(x)\right)} \xrightarrow[N\to\infty]{} 0 \tag{7.2}$$

   for any such p, q and all $x \in X_t$. Together with the KBSZ-criterion (Theorem 6.2) this yields the assertion.

To obtain Proposition 7.4 we will show that Sarnak's conjecture holds for the so called *associated* Toeplitz *dynamical system* (X_z, S), which arises as described above from the sequence z constructed as follows:

1. For each $m \in \mathbb{N}_0$ set $z(2m) = 1$ and leave odd positions undefined.

2. For each $m \in \mathbb{N}_0$ set $z(4m+1) = 0$, that is, we fill every second unfilled place by 0.

3. Set 1 at every second unfilled place.

At the nth step fill every second unfilled place with either 1 or 0 whether n is odd or even respectively. Then, for each $n \geq 0$, we have

$$z = B_n?B_n?B_n?\ldots, \tag{7.3}$$

where $\#B_n = 2^n - 1$ and "?" stands for an unfilled position (half of these unfilled positions will be filled at step $n+1$). This way we obtain the sequence z with

$$z(n) = t(n) + t(n+1) \mod 2 \tag{7.4}$$

(see [21]) for each $n \in \mathbb{N}_0$.

We will obtain the result for (X_z, S) by proving that SARNAK's conjecture holds for a linearly dense subset of $C(X_z)$. So the first step will be the construction of such a set.

Given a sequence $w = (w(i))_{i\in\mathbb{Z}}$ and $a \in \mathbb{Z}$, $l \in \mathbb{N}_0$, denote by $w[a, a+l)$ the finite subword $(w(a), w(a+1), \ldots, w(a+l-1))$. For fixed $l \in \mathbb{N}_0$, $a \in \mathbb{Z}$ consider continuous functions $f_l : \{0,1\}^l \to \mathbb{C}$. They extend to continuous maps $f : X_z \to \mathbb{C}$ taking only finitely many values by setting

$$f : X_z \ni w \mapsto f_l(w[a, a+l)) \in \mathbb{C}.$$

Conversely, for any continuous map f on X_z taking only finitely many values, there are $l \in \mathbb{N}_0$, $a \in \mathbb{Z}$ such that f is obtained this way (see [21]). Denote by $\mathcal{F}$ the set of all such functions $f \in C(X_z)$. Furthermore, under this notation, $f(S^n w) = f_l(w[a+n, a+n+l))$, for each $n \in \mathbb{N}$.

For $l \in \mathbb{N}$, $a \in \mathbb{Z}$ and $u \in \{0,1\}^l$ set $U_{u,a} \coloneqq \{w \in X_z \,|\, w[a, a+l) = u\}$. Then $U_{u,a}$ is open (and closed) in the product topology.

Lemma 7.6. *$\mathcal{F}$ is dense in $C(X_z)$.*

Proof. Fix $\varepsilon > 0$ and let $f \in C(X_z)$. Then f is uniformly continuous and hence there exist $l \in \mathbb{N}$ and $a \in \mathbb{Z}$ such that for any $u \in \{0,1\}^l$

$$\operatorname{diam} f(U_{u,a}) < \varepsilon. \tag{7.5}$$

Fix such an a. Define the relation $\sim$ by $w \sim w'$ if $w[a, a+l) = w'[a, a+l)$. Then $\sim$ is an equivalence relation on X_z. Let $\widetilde{w}_1, \widetilde{w}_2, \ldots \in X_z$ be representatives of the equivalence classes of $\sim$ and define $f' \in C(X_z)$ by $f'(w) \coloneqq f(\widetilde{w}_j)$ for $w \in [\widetilde{w}_j]_\sim$. Then $f' \in \mathcal{F}$ and from (7.5) we obtain that, for any $w \in X_z$,

$$|f(w) - f'(w)| < \varepsilon. \qquad \square$$

Proof of Proposition 7.4. Because of (7.4), instead of functions $f_1 \in C(X_t)$ we consider $f \in C(X_z)$ (see also [21]).

First, let $f \in \mathcal{F}$. Then there are $a \in \mathbb{Z}$, $l \in \mathbb{N}_0$ such that for $w \in X_z$ the value $f(w)$ depends only on $w[a, a+l)$. Since f is continuous, $|f|$ is bounded, say $|f(w)| \leq A$ for each $w \in X_z$.

Fix $\varepsilon > 0$. Then there is an $m \in \mathbb{N}$ such that, for each $n \geq m$,

$$\frac{A \cdot l}{2^n} < \frac{\varepsilon}{3}. \tag{7.6}$$

From Proposition 7.2 we know that $\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} b_n \mu(n) = 0$ for all periodic sequences $(b_n)_{n\in\mathbb{N}}$. Hence, for any periodic $(b_n)_{n\in\,\mathbb{N}}$ of period $< 2^m$ and bounded by C, there is an $M_1 \in \mathbb{N}$, $M_1 > \frac{3A}{\varepsilon}$, such that, for each $N \geq 2^m M_1$,

$$\frac{1}{N} \sum_{n=1}^{N} b_n \mu(n) < \frac{\varepsilon}{3 \cdot 2^m}. \tag{7.7}$$

Fix an $N > 2^m M_1$ and set $M \coloneqq \left[\frac{N}{2^m}\right]$. Then, by the construction of X_z, for any $w \in X_z$ there is an $i \in \mathbb{Z}$ such that $w\,[1, N+1) = z\,[i, i+N)$. Thus, from (7.3) we obtain

$$w\,[1, N+1) = C B_m y_1 B_m y_2 \dots B_m y_M D, \tag{7.8}$$

where $B_m \in \{0,1\}^{2^m-1}$, $y_0, \dots, y_M \in \{0,1\}$ and C a suffix of $B_m y_0$ as well as D a prefix of B_m. Let c, d denote the length of C, D, respectively, and set

$$\begin{aligned}
E'(w) &\coloneqq \sum_{k=1}^{c} f(S^k w)\mu(k), \\
E''(w) &\coloneqq \sum_{k=c+M\cdot 2^m+1}^{c+M\cdot 2^m+d} f(S^k w)\mu(k), \\
\Sigma_i(w) &\coloneqq \sum_{k=0}^{M-1} f(S^{c+k\cdot 2^m+i} w)\mu(c + k\cdot 2^m + i),
\end{aligned}$$

for $i \in [1, 2^m] \cap \mathbb{Z}$ and $w \in X_z$. Then, from (7.8) we see that

$$\sum_{n=1}^{N} f(S^n w)\mu(n) = E'(w) + \sum_{i=1}^{2^m} \Sigma_i(w) + E''(w). \tag{7.9}$$

Since $|f(w)| \leq C$, for each $w \in X_t$ we have

$$\left|E'(w) + E''(w)\right| \leq \sum_{k=1}^{c} A + \sum_{k=c+M\cdot 2^m+1}^{c+M\cdot 2^m+d} A = (c+d-1)\,A \leq (2\cdot 2^m + 2)\,A \tag{7.10}$$

by the choice of C and D.

Moreover, each Σ_i is an expression of the form $\sum_{n=1}^{N} b_n \mu(n)$, where $(b_n)_{n\in\mathbb{N}}$ is a periodic sequence of the form $(0, \dots, 0, \phi, 0, \dots, 0, \phi, 0, \dots)$ and period 2^m, where ϕ is a fixed value of f, provided the segment $w\,[c+i, c+i+l)$ does not meet any of the entries $y_1, \dots, y_M$. This certainly holds for $i \leq 2^m - l$ and it follows by (7.7) that, for any $i \in [1, \dots, 2^m - l] \cap \mathbb{Z}$,

$$|\Sigma_i(w)| \leq \frac{N\varepsilon}{3 \cdot 2^m}, \tag{7.11}$$

while for $i \in [2^m - l + 1, 2^m] \cap \mathbb{Z}$ we have

$$|\Sigma_i(w)| \leq \sum_{k=0}^{M-1} A = MA. \tag{7.12}$$

Combining (7.6), (7.9), (7.10), (7.11), (7.12) and $M_1 > \frac{3A}{\varepsilon}$ yields

$$\begin{aligned}\left|\frac{1}{N}\sum_{n=1}^{N} f(S^n w)\mu(n)\right| &\leq \frac{1}{N}\left(|E'(w)+E''(w)| + \sum_{i=1}^{2^m}|\Sigma_i(w)|\right)\\ &\leq \frac{1}{N}\left((2\cdot 2^m+2)\,A + (2^m-l)\,\frac{N\varepsilon}{3\cdot 2^m} + (l-1)\,MA\right)\\ &\leq \frac{(2\cdot 2^m+2)\,A}{N} + \frac{2^m\varepsilon}{3\cdot 2^m} - \frac{l\varepsilon}{3\cdot 2^m} + \frac{lMA}{N} - \frac{MA}{N}.\end{aligned} \tag{7.13}$$

Now

$$\begin{aligned}\left|\frac{lMA}{N} - \frac{l\varepsilon}{3\cdot 2^m}\right| &= \frac{lA}{2^m}\left|\frac{2^m M}{N} - \frac{\varepsilon}{3A}\right| \overset{(7.6)}{<} \frac{\varepsilon}{3}\left|\frac{2^m M}{N} - \frac{\varepsilon}{3A}\right|\\ &= \left|\frac{2^m M\varepsilon}{3N} - \frac{\varepsilon^2}{9A}\right| \overset{M=\left[\frac{N}{2^m}\right]}{\leq} \left|\frac{\varepsilon}{3} - \frac{\varepsilon^2}{9A}\right| < \frac{\varepsilon}{3}.\end{aligned} \tag{7.14}$$

From $N \geq 2^m M_1$ and $M_1 > \frac{3A}{\varepsilon}$ we obtain $N > \frac{3\cdot 2^m A}{\varepsilon}$, which implies

$$\frac{(2\cdot 2^m+2)\,A}{N} < \frac{(2\cdot 2^m+2)\,A}{3\cdot 2^m A} = \frac{\varepsilon}{3} + \left(\frac{1}{3} + \frac{2}{3\cdot 2^m}\right)\varepsilon < \frac{\varepsilon}{3} + \varepsilon. \tag{7.15}$$

Together with $\frac{MA}{N} > 0$, inserting (7.14) and (7.15) into (7.13) yields

$$\left|\frac{1}{N}\sum_{n=1}^{N} f(S^n w)\mu(n)\right| < \frac{\varepsilon}{3} + \varepsilon + \frac{2^m\varepsilon}{3\cdot 2^m} + \frac{\varepsilon}{3} = 2\varepsilon.$$

So $\frac{1}{N}\sum_{n=1}^{N} f(S^n w)\mu(n) \xrightarrow[N\to\infty]{} 0$ for $f \in \mathcal{F}$.

Now let $f \in C(X_z)$ be chosen arbitrarily. Then the assertion follows from the above together with Lemma 7.6 and Lemma 7.3. □

Theorem 7.7. Sarnak*'s conjecture holds for the* Thue–Morse *shift.*

Proof. In Theorem 3.33 we have seen that the Thue–Morse shift satisfies the zero entropy assumption.

For each $f \in C(X_t)$ we have $f = f_1 + f_2$ with $f_1, f_2 \in C(X_t)$ given as above. Thus, for any $x \in X_t$,

$$\frac{1}{N}\sum_{n=1}^{N} f(S^n x)\mu(n) = \underbrace{\frac{1}{N}\sum_{n=1}^{N} f_1(S^n x)\mu(n)}_{\xrightarrow[N\to\infty]{\text{Proposition 7.4}} 0} + \underbrace{\frac{1}{N}\sum_{n=1}^{N} f_2(S^n x)\mu(n)}_{\xrightarrow[N\to\infty]{\text{Proposition 7.5}} 0} \xrightarrow[N\to\infty]{} 0. \qquad \square$$

7.2. Möbius Function Randomness for Skew Product Extensions of Rational Rotations

Proposition 7.8. Sarnak*'s conjecture holds for rotations on the circle.*

For rotations through a rational angle this follows from Proposition 7.2. Thus Proposition 7.8 is a natural generalization of the result for periodic sequences.

Proof of Proposition 7.8. In Lemma 3.35 we have seen that each rotation on the circle is deterministic.

For $\alpha \in \mathbb{T}$ consider $R_\alpha : \mathbb{T} \ni x \mapsto \alpha x \in \mathbb{T}$ and let $a \in [0,1)$ such that $\alpha = e^{2\pi i a}$. Fix $x \in \mathbb{T}$. Then there is a $\theta \in [0,1)$ so that $x = e^{2\pi i \theta}$ and therefore $R_\alpha(x) = e^{2\pi i(a+\theta)}$ with $a+\theta \in [0,1)$ (modulo 1). Hence, from DAVENPORT's estimation (Theorem 5.7) we know that for each $r > 0$

$$\sum_{n=1}^{N} R_\alpha^n(x)\mu(n) = \sum_{n=1}^{N} \mu(n) e^{2\pi i n(a+\theta)} = \mathrm{O}\left(\frac{N}{(\log N)^r}\right)$$

as $N \to \infty$, i.e., for each $r > 0$ there is a constant $C = C(r) > 0$ such that

$$0 \le \frac{1}{N}\left|\sum_{n=1}^{N} \mu(n) e^{2\pi i n(a+\theta)}\right| \le C\left(\frac{1}{(\log N)^r}\right) \xrightarrow[N\to\infty]{} 0. \qquad \square$$

Now consider a skew product extension of a rational rotation as defined in Subsection 3.4.3:

$$\mathbb{T}^2 \ni (x,y) \mapsto R_{\alpha,\phi}(x,y) \coloneqq (R_\alpha(x), y + \phi(x)) \,(\mathrm{mod}\, 1) \in \mathbb{T}^2,$$

where $\phi : \mathbb{T} \to \mathbb{T}$ is continuous and $\alpha =: \frac{p}{q} \in \mathbb{Q}$. We want to show that SARNAK's conjecture holds for such systems by making use of Lemma 7.3. So we need to find a linearly dense subset of $C(\mathbb{T}^2)$ first.

Lemma 7.9. *For $a, b \in \mathbb{Z}$ let $\chi_{a,b} : \mathbb{T}^2 \ni (x,y) \mapsto e^{2\pi i(ax+by)} \in \mathbb{C}$ and denote by $\mathcal{K} \coloneqq \{\chi_{a,b} \,|\, a,b \in \mathbb{Z}\}$ the set of all such functions. Then $\mathcal{K}$ is linearly dense in $C(\mathbb{T}^2)$.*

Proof. First, note that for each $a, b \in \mathbb{Z}$ we have $e^{2\pi i(ax+by)} \in C(\mathbb{T}^2)$, thus $\mathcal{K} \subseteq C(\mathbb{T}^2)$.

Now, since

$$\chi_{a,b}(x,y) = e^{2\pi i(ax+by)} = \cos(2\pi(ax+by)) + i\sin(2\pi(ax+by))$$

for each $(x,y) \in \mathbb{T}^2$ and

$$\chi_{0,0} = \mathbf{1}_{\mathbb{T}^2},$$

the linear span of $\mathcal{K}$ equals the set of all complex-valued trigonometric polynomials on $\mathbb{T}^2$. Hence $\mathrm{lin}(\mathcal{K})$ is a sub-$\mathbb{C}$-algebra of $C(\mathbb{T}^2)$ such that

- $\mathrm{lin}(\mathcal{K})$ separates the points of $\mathbb{T}$, i.e., $\forall x, y \in \mathbb{T}\, \exists P \in \mathrm{lin}(\mathcal{K}) : P(x) \neq P(y)$,
- $\mathrm{lin}(\mathcal{K})$ vanishes nowhere on $\mathbb{T}$, i.e., $\forall x \in \mathbb{T}\, \exists P \in \mathrm{lin}(\mathcal{K}) : P(x) \neq 0$,
- $\mathrm{lin}(\mathcal{K})$ is invariant under conjugation, i.e., $\forall P \in \mathrm{lin}(\mathcal{K}) : \overline{P} \in \mathrm{lin}(\mathcal{K})$.

Thus by the STONE–WEIERSTRASS theorem (see Theorem A.14 in the Appendix) $\mathrm{lin}(\mathcal{K})$ is dense in $C(\mathbb{T}^2)$ and the assertion follows. $\square$

Lemma 7.10 ([38])**.** *Consider $R_{0,\phi} : \mathbb{T}^2 \ni (x,y) \mapsto (x, y+\phi(x)) \in \mathbb{T}^2$ where $\phi : \mathbb{T} \to \mathbb{T}$ is continuous. Then for each $(x_1,y_1), (x_2,y_2) \in \mathbb{T}^2$ and every $\chi_{a,b} \in \mathcal{K}$ with $b \neq 0$ we have*

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} \chi\left(R_{0,\phi}^{rn}(x_1,y_1)\right) \overline{\chi\left(R_{0,\phi}^{sn}(x_2,y_2)\right)} = 0 \tag{7.16}$$

for sufficiently large $r, s \in \mathbb{P}$, $r \neq s$, whenever $\phi(x_1) \notin \mathbb{Q}$ or $\phi(x_2) \notin \mathbb{Q}$.

Proof. For each $m \in \mathbb{N}$ and every $(x, y) \in \mathbb{T}^2$ we have

$$R^m_{0,\phi}(x, y) = (x, y + m\phi(x)).$$

Hence

$$\begin{aligned}\sum_{n=1}^{N} \chi\left(R^{rn}_{0,\phi}(x_1, y_1)\right)\overline{\chi\left(R^{sn}_{0,\phi}(x_2, y_2)\right)} &= \sum_{n=1}^{N} \chi\left(x_1, y_1 + rn\phi(x_1)\right)\overline{\chi\left(x_2, y_2 + sn\phi(x_2)\right)} \\ &= \sum_{n=1}^{N} e^{2\pi i(ax_1+by_1+brn\phi(x_1))}e^{-2\pi i(ax_2+by_2+bsn\phi(x_2))} \\ &= e^{2\pi i(a(x_1-x_2)+b(y_1-y_2))}\sum_{n=1}^{N} e^{2\pi in(r\phi(x_1)-s\phi(x_2))}.\end{aligned}$$

If exactly one of the numbers $\phi(x_1), \phi(x_2)$ is irrational then the result follows from Weyl's theorem (see Theorem A.16 in the Appendix) for all $r, s \in \mathbb{N}$. If both of these numbers are irrational then there is at most one pair $(r, s) \in \mathbb{P} \times (\mathbb{P} \setminus \{r\})$ such that $r\phi(x_1) - s\phi(x_2) \in \mathbb{Q}$. Hence the result follows again from Weyl's theorem, for r, s sufficiently large. □

Remark 7.11. Note that the above proof of Lemma 7.10 reveals that the convergence in (7.16) happens uniformly in y_1, y_2.

Theorem 7.12 ([38])**.** Sarnak*'s conjecture holds for skew product extensions of rational rotations.*

Proof. By Corollary 3.37 the zero topological entropy assumption is satisfied. Consider $R_{\alpha,\phi} : \mathbb{T}^2 \to \mathbb{T}^2$ given by $R_{\alpha,\phi}(x, y) \coloneqq (x + \alpha, y + \phi(x)) \pmod 1$, with $\alpha \in \mathbb{Q}$ and $\phi : \mathbb{T} \to \mathbb{T}$ continuous. Write $\alpha = \frac{p}{q}$ with $p \in \mathbb{Z}$, $q \in \mathbb{N}$.

Because of Lemma 7.3 and Lemma 7.9 it suffices to consider functions $\chi_{a,b} \in \mathcal{K} \subsetneq C(\mathbb{T}^2)$, with $\chi_{a,b}(x, y) \coloneqq e^{2\pi i(ax+by)}$. If $b = 0$, the assertion follows from Proposition 7.2. So let $b \neq 0$.

First, note that

$$R^q_{\alpha,\phi}(x, y) = (x + p, y + \phi_q(x)) = (x, y + \phi_q(x)), \tag{7.17}$$

where $\phi_q(x) \coloneqq \sum_{k=0}^{q-1} \phi(x + \frac{kp}{q})$. For $n \in \mathbb{N}$ we find $n' \in \mathbb{N}$ such that $n = qn' + j$ with $j \in [0, q) \cap \mathbb{Z}$. Then, for each $\chi_{a,b} \in \mathcal{K}$, every $r, s \in \mathbb{N}$ and all $(x, y) \in \mathbb{T}^2$ we have

$$\begin{aligned}&\chi_{a,b}\left(R^{rn}_{\alpha,\phi}(x, y)\right)\overline{\chi_{a,b}\left(R^{sn}_{\alpha,\phi}(x, y)\right)} \\ &\qquad = \chi_{a,b}\left(R^{qrn'}_{\alpha,\phi}\left(R^{rj}_{\alpha,\phi}(x, y)\right)\right)\overline{\chi_{a,b}\left(R^{qsn'}_{\alpha,\phi}\left(R^{sj}_{\alpha,\phi}(x, y)\right)\right)},\end{aligned}$$

where, because of (7.17), the first coordinates of the points $R^{rj}_{\alpha,\phi}(x, y)$, $R^{sj}_{\alpha,\phi}(x, y)$ belong to the finite set $M \coloneqq \left\{x + \frac{kp}{q} \,\middle|\, k \in [0, q) \cap \mathbb{Z}\right\}$ (and hence do not depend on r and s). Thus, to show that

$$\lim_{N\to\infty} \frac{1}{N}\sum_{n=1}^{N} \chi_{a,b}\left(R^{rn}_{\alpha,\phi}(x, y)\right)\overline{\chi_{a,b}\left(R^{sn}_{\alpha,\phi}(x, y)\right)} = 0$$

holds for sufficiently large $r, s \in \mathbb{P}$, $r \neq s$, considering Remark 7.11, we need to varify that

$$\lim_{N\to\infty} \frac{1}{N} \sum_{\substack{n'\in\mathbb{N} \\ n'\leq \frac{N}{q}}} \chi_{a,b}\left(R_{\alpha,\phi}^{qrn'}(x_1,*)\right) \overline{\chi_{a,b}\left(R_{\alpha,\phi}^{qsn'}(x_2,*)\right)} = 0, \tag{7.18}$$

whenever $x_1, x_2 \in M$. If $\phi_q(x_1) \notin \mathbb{Q}$ or $\phi_q(x_2) \notin \mathbb{Q}$, (7.18) follows from (7.17) together with Lemma 7.10 and Remark 7.11 (since $b \neq 0$). Hence the assertion follows from the KBSZ-criterion (Theorem 6.2).

Suppose now that $\phi_q(x + \frac{jrp}{q}), \phi_q(x + \frac{jsp}{q}) \in \mathbb{Q}$. This is only possible in the case $\phi_q(x) \in \mathbb{Q}$, since ϕ_q is constant on the R_α-orbit of x. Moreover, since $n = qn' + j$ with $j \in [0, q) \cap \mathbb{Z}$, we have

$$\begin{aligned}
\phi^{(n)}(x) \coloneqq \sum_{k=0}^{n-1} \phi\left(x + \frac{kp}{q}\right) &= \sum_{k=qn'}^{qn'+j-1} \phi\left(x + \frac{kp}{q}\right) + \sum_{k=0}^{qn'-1} \phi\left(x + \frac{kp}{q}\right) \\
&= \sum_{k=0}^{j-1} \phi\left(x + \frac{kp}{q}\right) + \sum_{k=0}^{qn'-1} \phi\left(x + \frac{kp}{q}\right) \\
&= \phi^{(j)}(x) + \phi^{(qn')}\left(x + \frac{jp}{q}\right) \\
&= \phi^{(j)}(x) + n'\phi_q(x).
\end{aligned}$$

It follows that

$$\begin{aligned}
&\frac{1}{N} \sum_{n=1}^{N} \chi_{a,b}\left(R_{\alpha,\phi}^{n}(x,y)\right) \mu(n) \\
&\quad = \frac{1}{N} \sum_{j=0}^{q-1} \sum_{\substack{n'\in\mathbb{N} \\ n'\leq \frac{N}{q}}} \chi_{a,b}\left(x + \frac{(qn'+j)\,p}{q}, \phi^{(n)}(x) + y\right) \mu\left(qn'+j\right) \\
&\quad = \sum_{j=0}^{q-1} \frac{1}{N} \sum_{\substack{n'\in\mathbb{N} \\ n'\leq \frac{N}{q}}} \chi_{a,b}\left(x + \frac{jp}{q}, \phi^{(j)}(x) + n'\phi_q(x) + y\right) \mu\left(qn'+j\right) \\
&\quad = \sum_{j=0}^{q-1} \frac{1}{N} \sum_{\substack{n'\in\mathbb{N} \\ n'\leq \frac{N}{q}}} e^{2\pi i\left(a\left(x+\frac{jp}{q}\right)+b\left(\phi^{(j)}(x)+n'\phi_q(x)+y\right)\right)} \mu\left(qn'+j\right) \\
&\quad = \sum_{j=0}^{q-1} e^{2\pi i\left(a\left(x+\frac{jp}{q}\right)+b\left(\phi^{(j)}(x)+y\right)\right)} \frac{1}{N} \sum_{\substack{n'\in\mathbb{N} \\ n'\leq \frac{N}{q}}} e^{2\pi i b n'\phi_q(x)} \mu\left(qn'+j\right).
\end{aligned}$$

By writing $\phi_q(x) = \frac{c}{d}$ with $c \in \mathbb{Z}$, $d \in \mathbb{N}$, and setting $n' = dn'' + k$ with $k \in [0, d) \cap \mathbb{Z}$, we obtain

$$n = qn' + j = qdn'' + (qk + j).$$

So the above yields the representation

$$\frac{1}{N} \sum_{n=1}^{N} \chi_{a,b}\left(R_{\alpha,\phi}^{n}(x,y)\right) \mu(n) = \sum_{j=0}^{q-1} \sum_{k=0}^{d-1} \frac{1}{N} \sum_{\substack{n''\in\mathbb{N} \\ n''\leq \frac{N}{dq}}} a_{j,k}(n'') \mu\left(qdn'' + qk + j\right),$$

and hence the result again follows from Proposition 7.2. □

Here we just considered the skew product extension of a rotation R_α on $\mathbb{T}$ through a rational α. So there is much room for generalization and, by the time this thesis has been written, no proof is known for the general case. However, in [39] it is shown that SARNAK's conjecture holds for any $\alpha \in \mathbb{R}$, provided $\phi(x) = cx + \psi(x)$ with $c \in \mathbb{Z}$ and an analytic $\psi : \mathbb{T} \to \mathbb{T}$ such that $\left|\widehat{\psi}(m)\right| \gg e^{-\tau|m|}$ for some $\tau > 0$, while in [38] the result is obtained without the strong assumption on ϕ, but at the cost of reducing validity to those $\alpha \in \mathbb{R}$ which are generic for some R_α-invariant measure on $\mathbb{T}$.

8. Chowla's Conjecture implies Sarnak's Conjecture

In this chapter we want to show that the conjecture of Sarnak is a consequence of the following (also unproven) classical conjecture given by Chowla in [10] in 1965:

Conjecture 8.1 (Chowla)**.** *Let $r \in \mathbb{N}$, $a_0, \ldots, a_r \in \mathbb{N}_0$, with $0 = a_0 < a_1 < \ldots < a_r$, and $i_0, \ldots, i_r \in \{1, 2\}$ not all equal to 2. Then*

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} \prod_{s=0}^{r} \mu^{i_s}(n + a_s) = 0.$$

The exponents $i_0, \ldots, i_r$ are to be chosen not all even to not completely destroy the sign cancellation, since for $i_0 = \ldots = i_r = 2$ the above limit equals $\frac{6}{\pi^2}$ (density of the set of square-free integers; see [35]). Furthermore, note that the case $r = 0$ follows from Proposition 7.1 while for $r \geq 1$ the conjecture is still open.

Actually, the correlation between the both conjectures lies much deeper, since it still holds for an arbitrary sequence z taking values in $\{-1, 0, 1\}$ instead of the Möbius function μ. Therefore, in this chapter, we will take the more abstract approach and deal with such an arbitrary sequence, following the work of El Abdalaoui, Kułaga-Przymus, Lemańczyk and de la Rue done in [22].

8.1. Definitions

Let (X, T) be a metric TDS, i.e., X a compact metric space and $T : X \to X$ continuous. As before, we denote by $\mathfrak{M}(X)$ the set of all probability measures on (X, Σ_X), where Σ_X stands for the Borel σ-algebra on X, and by $\mathfrak{M}_T(X) = \mathfrak{M}_T \subseteq \mathfrak{M}(X)$ the subset of those measures, which are invariant under T (recall that by the Krylov–Bogolyubov theorem we have $\mathfrak{M}_T \neq \varnothing$). $\mathfrak{M}(X)$ is endowed with the (metrizable) weak topology where $(\nu_n)_{n\in\mathbb{N}}$ converges to ν in $\mathfrak{M}(X)$ if for each $f \in C(X)$ the series $\left(\int_X f \,\mathrm{d}\,\nu_n\right)_{n\in\mathbb{N}}$ converges to $\int_X f \,\mathrm{d}\,\nu$.

For $x \in X$ we denote by δ_x the Dirac measure on (X, Σ_X) given by

$$\delta_x(A) \coloneqq \begin{cases} 1 & \text{for } x \in A \\ 0 & \text{otherwise} \end{cases} = \mathbf{1}_A(x)$$

for $A \in \Sigma_X$. Note that, since δ_x is a probability measure on (X, Σ_X), for each $x \in X$, each limit point ρ of the sequence $(\delta_{T,N,x})_{N\in\mathbb{N}}$ with $\delta_{T,N,x} \coloneqq \frac{1}{N}\sum_{n=1}^{N} \delta_{T^n x}$ is again a probability measure on (X, Σ_X). Moreover, ρ is invariant under T, i.e., we have $\rho \in \mathfrak{M}_T$ (cf. the proof of Theorem A.6 in the Appendix).

Definition 8.2. Let $\nu \in \mathfrak{M}_T$, i.e., (X, Σ_X, ν, T) be an MDS. For $x \in X$ set

$$\mathrm{Q} - \mathrm{gen}(x) \coloneqq \left\{ \rho \in \mathfrak{M}_T \;\middle|\; \lim_{k\to\infty} \frac{1}{N_k} \sum_{n=1}^{N_k} \delta_{T^n x} = \rho \text{ for } (N_k)_{k\in\mathbb{N}} \subset \mathbb{N} \text{ strictly increasing} \right\}.$$

We call x

- *quasi-generic* for ν, if $\nu \in \mathrm{Q} - \mathrm{gen}(x)$.
- *generic* for ν, if $\{\nu\} = \mathrm{Q} - \mathrm{gen}(x)$ (i.e., we have $\frac{1}{N}\sum_{n=1}^{N} \delta_{T^n x} \xrightarrow[N\to\infty]{} \nu$).

Definition 8.3. Let (X,T) be a TDS and $x \in X$. Then we call x *completely deterministic* if for each $\nu \in \mathrm{Q} - \mathrm{gen}(x)$ we have $h_\nu(T) = 0$. (i.e., the associated MDS (X, Σ_X, ν, T) is of zero Kolmogorov–Sinai entropy).

Remark 8.4. By Theorem 3.27 every point of a deterministic system is completely deterministic.

To obtain the following results for both, one-sided and two-sided sequences, let $\mathbb{I} \in \{\mathbb{N}_0, \mathbb{Z}\}$ for the rest of this section (minor changes would also provide validity for $\mathbb{I} = \mathbb{N}$). Note that $\{-1,0,1\}^{\mathbb{I}}$ endowed with the product topology is a compact metric space. Furthermore, to cut down on indices, we identify sequences $z \in \{-1,0,1\}^{\mathbb{I}}$ with functions $z : \mathbb{I} \to \{-1,0,1\}$ and write $z(n)$ for z_n.

Now we can formulate the central conditions of this chapter.

Definition 8.5. We say that $z \in \{-1,0,1\}^{\mathbb{I}}$ satisfies

- condition (**Ch**) if, for each $r \in \mathbb{N}$, $a_1, \ldots, a_r \in \mathbb{N}$ with $a_1 < \ldots < a_r$, $a_0 = 0$ and all $i_0, \ldots, i_r \in \{1,2\}$ not all equal to 2, we have

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} \prod_{s=0}^{r} z^{i_s}(n + a_s) = 0. \tag{Ch}$$

- condition (**S**) if, for any TDS (X,T) with $h(T) = 0$, all $f \in C(X)$ and every $x \in X$, we have

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} f(T^n x)\, z(n) = 0. \tag{S}$$

- condition ($\widehat{\mathbf{S}}$) if, for any TDS (X,T), all $f \in C(X)$ and every completely deterministic $x \in X$, we have

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} f(T^n x)\, z(n) = 0. \tag{$\widehat{\mathbf{S}}$}$$

The Möbius function μ satisfies the condition (**Ch**) if and only if Chowla's conjecture holds; and analogous for the condition (**S**) and Sarnak's conjecture. Furthermore, by Remark 8.4, the implication ($\widehat{\mathbf{S}}$) $\Longrightarrow$ (**S**) is obvious. We will show that (**Ch**) implies ($\widehat{\mathbf{S}}$) and thus obtain the desired implication (**Ch**) $\Longrightarrow$ (**S**). To do so it will come in handy to consider a further system besides (X,T). So we denote by S the left shift on $A^{\mathbb{I}}$ with $A \in \{\{0,1\}, \{-1,0,1\}\}$ as well as the left shift on any closed shift-invariant subset of $A^{\mathbb{I}}$ (called a subshift). Then $(A^{\mathbb{I}}, S)$ is a TDS, since $A^{\mathbb{I}}$ endowed with the product topology is a compact metric space and S is continuous for this topology.

Now, let $F : A^{\mathbb{I}} \to A$ (or any subset of $A^{\mathbb{I}}$) be the continuous map given by

$$F(w) \coloneqq w(0)$$

for each $w \in A^{\mathbb{I}}$. Therefore we can write any member of a sequence $z \in A^{\mathbb{I}}$ in the form $f(T^n x)$, since for each $n \in \mathbb{N}_0$ we have $z(n) = F(S^n z)$.

Since the cylinder sets

$$C_t(a_0, \dots, a_{k-1}) \coloneqq \left\{ w \in A^{\mathbb{I}} \,\middle|\, w_{t+j} = a_j \text{ for each } j \in [0, k-1] \cap \mathbb{Z} \right\},$$

where $k \in \mathbb{N}$ and $t \in \mathbb{I}$ (that are the sets of all sequences in which the block $(a_0, \dots, a_{k-1})$ appears on at position t), form a base for the product topology, any $\nu \in \mathfrak{M}_S(A^{\mathbb{N}_0})$ is determined by the values it takes on blocks. Hence it can be extended to a measure in $\mathfrak{M}_S(A^{\mathbb{Z}})$ taking the same value on each block as ν. This measure will be denoted by ν as well. Note that this extension preserves quasi-genericity, i.e., if $w \in A^{\mathbb{N}_0}$ is quasi-generic for $\nu \in \mathfrak{M}_S(A^{\mathbb{N}_0})$ along $(N_k)_{k\in\mathbb{N}}$, then any $\widetilde{w} \in A^{\mathbb{Z}}$ with $\widetilde{w}(j) = w(j)$, for each $j \in \mathbb{N}_0$, is quasi-generic for $\nu \in \mathfrak{M}_S(A^{\mathbb{Z}})$ along $(N_k)_{k\in\mathbb{N}}$.

Finally, we need a way to obtain measures in $\mathfrak{M}_S(\{-1,0,1\}^{\mathbb{I}})$ from given measures in $\mathfrak{M}_S(\{0,1\}^{\mathbb{I}})$. To do so, let $\chi : \{-1,0,1\}^{\mathbb{I}} \to \{0,1\}^{\mathbb{I}}$ be the coordinate square map, i.e.,

$$\chi : (w_n)_{n\in\mathbb{I}} \mapsto (w_n^2)_{n\in\mathbb{I}}$$

for each $w = (w_n)_{n\in\mathbb{I}} \in \{-1,0,1\}^{\mathbb{I}}$. We also let χ act on blocks in the same way.

Definition 8.6. Let $\nu \in \mathfrak{M}_S(\{0,1\}^{\mathbb{I}})$ and denote $\operatorname{supp}(b) \coloneqq \{i \,|\, b(i) \neq 0\}$ for any block b taking values in $\{-1,0,1\}$. Let $\widehat{\nu}$ be the measure on $\{-1,0,1\}^{\mathbb{I}}$ defined by

$$\widehat{\nu}(b) \coloneqq 2^{-\#\operatorname{supp}(b)} \nu(\chi(b)).$$

for each such block b. Then we call $\widehat{\nu}$ the *relatively independent extension* of ν.

Remark 8.7. For any $\nu \in \mathfrak{M}_S(\{0,1\}^{\mathbb{I}})$ we have $\widehat{\nu} \in \mathfrak{M}_S(\{-1,0,1\}^{\mathbb{I}})$ (see [22]).

In what follows we write z^2 for $\chi(z)$, i.e., we have $z^2(n) \coloneqq (z(n))^2$ for each $n \in \mathbb{I}$.

8.2. Preliminaries

As before, for $w \in \{-1,0,1\}^{\mathbb{N}_0}$ consider the subshift (X_w, S) of $(\{-1,0,1\}^{\mathbb{Z}}, S)$ with

$$X_w \coloneqq \left\{ u \in \{-1,0,1\}^{\mathbb{Z}} \,\middle|\, \text{all blocks that appear on } u \text{ also appear on } w \right\}.$$

Fix $z \in \{-1,0,1\}^{\mathbb{N}_0}$ such that z^2 is quasi-generic for some $\nu \in \mathfrak{M}_S(X_{z^2})$, i.e., there is $(N_k)_{k\in\mathbb{N}} \subset \mathbb{N}$ strictly increasing such that

$$\delta_{S,N_k,z^2} = \frac{1}{N_k} \sum_{n=1}^{N_k} \delta_{S^n z^2} \xrightarrow[k\to\infty]{} \nu. \tag{8.1}$$

Note that $X_{z^2} \subseteq \{0,1\}^{\mathbb{Z}}$ and therefore $\nu \in \mathfrak{M}_S(\{0,1\}^{\mathbb{Z}})$. Hence the relatively independent extension $\widehat{\nu}$ of ν is a measure in $\mathfrak{M}_S(\{-1,0,1\}^{\mathbb{Z}})$.

We want to give equivalent characterizations for the condition (**Ch**), at least along a certain subsequence $(N_k)_{k\in\mathbb{N}} \subset \mathbb{N}$. To do so, we need the following lemma, for which we omit the proof. It can be found in [22] (Lemma 4.5).

Lemma 8.8. *Consider the subshift (X_{z^2}, S) and $F : \{-1,0,1\}^{\mathbb{Z}} \ni w \mapsto w(0) \in \{-1,0,1\}$. Let $r \in \mathbb{N}$, $a_0, \ldots, a_r \in \mathbb{N}_0$, with $0 = a_0 < a_1 < \ldots < a_r$, and $i_0, \ldots, i_r \in \{1,2\}$ not all equal to 2. Then*

$$\int_{\{-1,0,1\}^{\mathbb{Z}}} \prod_{s=0}^{r} \left(F^{i_s} \circ S^{a_s}\right) \mathrm{d}\,\widehat{\nu} = 0,$$

where $\widehat{\nu}$ denotes the relatively independent extension of the above measure $\nu \in \mathfrak{M}_S(X_{z^2})$, and $F^i(z) \coloneqq (F(z))^i = (z(0))^i$ for each $i \in \mathbb{N}$, $z \in \{-1,0,1\}^{\mathbb{Z}}$.

Remark 8.9. Since, for each $u \in \{-1,0,1\}^{\mathbb{Z}}$, we have $F^2(u) = (u(0))^2 = F(u^2)$, it follows that

$$\int_{\{-1,0,1\}^{\mathbb{Z}}} \prod_{s=0}^{r} \left(F^2 \circ S^{a_s}\right) \mathrm{d}\,\widehat{\nu} = \int_{\{0,1\}^{\mathbb{Z}}} \prod_{s=0}^{r} \left(F \circ S^{a_s}\right) \mathrm{d}\,\nu.$$

Lemma 8.10. *Let $(N_k)_{k\in\mathbb{N}} \subset \mathbb{N}$ be such that (8.1) holds. Then the following conditions are equivalent:*

i) For ν and $\widehat{\nu}$ as before we have

$$\lim_{k\to\infty} \delta_{S,N_k,z} = \widehat{\nu}.$$

ii) For each choice of $r \in \mathbb{N}$, $a_0, \ldots, a_r \in \mathbb{N}_0$, with $0 = a_0 < a_1 < \ldots < a_r$, and $i_0, \ldots, i_r \in \{1,2\}$ not all equal to 2, we have

$$\lim_{k\to\infty} \frac{1}{N_k} \sum_{n=1}^{N_k} \prod_{s=0}^{r} z^{i_s}(n + a_s) = 0.$$

Proof. By the definition of the map F, for each $k \in \mathbb{N}$ we have

$$\frac{1}{N_k} \sum_{n=1}^{N_k} \prod_{s=0}^{r} z^{i_s}(n + a_s) = \frac{1}{N_k} \sum_{n=1}^{N_k} \prod_{s=0}^{r} \left(F^{i_s} \circ S^{a_s}\right)(S^n z). \tag{8.2}$$

Suppose that *i)* holds. Then (8.2) implies

$$\frac{1}{N_k} \sum_{n=1}^{N_k} \prod_{s=0}^{r} z^{i_s}(n + a_s) \xrightarrow[k\to\infty]{} \int_{\{-1,0,1\}^{\mathbb{Z}}} \prod_{s=0}^{r} \left(F^{i_s} \circ S^{a_s}\right) \mathrm{d}\,\widehat{\nu},$$

which, by Lemma 8.8, is equal to zero. Thus, *ii)* follows.

Suppose now that *ii)* holds. Without loss of generality (cf. [22]) we may assume that $\delta_{S,N_k,z} \xrightarrow[k\to\infty]{} \rho$. Then, by (8.2) we have

$$\frac{1}{N_k} \sum_{n=1}^{N_k} \prod_{s=0}^{r} z^{i_s}(n + a_s) \xrightarrow[k\to\infty]{} \int_{\{-1,0,1\}^{\mathbb{Z}}} \prod_{s=0}^{r} \left(F^{i_s} \circ S^{a_s}\right) \mathrm{d}\,\rho.$$

Since *ii)* holds, we obtain

$$\int_{\{-1,0,1\}^{\mathbb{Z}}} \prod_{s=0}^{r} \left(F^{i_s} \circ S^{a_s}\right) \mathrm{d}\,\rho = 0, \tag{8.3}$$

whenever not all i_s are equal to 2. Moreover, since $F^2(u) = F(u^2)$ for each $u \in \{-1,0,1\}^{\mathbb{Z}}$, we obtain from (8.1) that

$$\int_{\{-1,0,1\}^{\mathbb{Z}}} \prod_{s=0}^{r} \left(F^2 \circ S^{a_s}\right) \mathrm{d}\,\rho = \int_{\{0,1\}^{\mathbb{Z}}} \prod_{s=0}^{r} (F \circ S^{a_s})\,\mathrm{d}\,\nu. \tag{8.4}$$

Hence, from Remark 8.9, (8.3), (8.4) and Lemma 8.8 we deduce that

$$\int_{\{-1,0,1\}^{\mathbb{Z}}} G\,\mathrm{d}\,\widehat{\nu} = \int_{\{-1,0,1\}^{\mathbb{Z}}} G\,\mathrm{d}\,\rho$$

for any $G \in \mathcal{A} := \left\{\prod_{s=0}^{r} \left(F^{i_s} \circ S^{a_s}\right) \,\middle|\, 0 = a_0 < a_1 < \ldots < a_r,\ r \in \mathbb{N},\ i_s \in \mathbb{N}\right\}$. Since $\mathcal{A} \subseteq C(\{-1,0,1\}^{\mathbb{Z}})$ is closed under taking products and separates the points of $\{-1,0,1\}^{\mathbb{Z}}$, the Stone–Weierstrass theorem (see Theorem A.14 in the Appendix) implies $\rho = \widehat{\nu}$. □

Theorem 8.11. *For $z \in \{-1,0,1\}^{\mathbb{N}_0}$ and $(N_k)_{k\in\mathbb{N}} \subset \mathbb{N}$ strictly increasing the following conditions are equivalent:*

i) z satisfies (**Ch**) *along $(N_k)_{k\in\mathbb{N}}$, i.e., for each choice of $r \in \mathbb{N}$, $a_0, \ldots, a_r \in \mathbb{N}_0$, with $0 = a_0 < a_1 < \ldots < a_r$, and $i_0, \ldots, i_r \in \{1,2\}$ not all equal to 2, we have*

$$\lim_{k\to\infty} \frac{1}{N_k} \sum_{n=1}^{N_k} \prod_{s=0}^{r} z^{i_s}(n + a_s) = 0.$$

ii) $\lim_{k\to\infty} \delta_{S,N_k,z^2} = \nu$ *if and only if* $\lim_{k\to\infty} \delta_{S,N_k,z} = \widehat{\nu}$.

iii) $\mathrm{Q} - \mathrm{gen}(z) = \left\{\widehat{\nu} \,\middle|\, \nu \in \mathrm{Q} - \mathrm{gen}(z^2)\right\}$.

Sketch of proof. The equivalence of *ii)* and *iii)* follows directly from the definition of the set $\mathrm{Q} - \mathrm{gen}(z)$, while one shows the equivalence of *i)* and *ii)* by using Lemma 8.10 and Definition 8.6 (see [22], Remark 4.8). □

8.3. (Ch) implies ($\widehat{\mathrm{S}}$)

Let (X,T) be a TDS and $\nu \in \mathfrak{M}_S(\{0,1\}^{\mathbb{Z}})$, where S denotes the shift on $\{0,1\}^{\mathbb{Z}}$.

Lemma 8.12. *For $w \in \{-1,0,1\}^{\mathbb{Z}}$ we have $\mathbb{E}_{\widehat{\nu}}(F|\chi(w) = u) = 0$ for ν-a.e. $u \in \{0,1\}^{\mathbb{Z}}$.*

Proof. We have

$$\mathbb{E}_{\widehat{\nu}}(F \mid \chi(w) = u) = \mathbb{E}_{\widehat{\nu}}\left(F \middle| \{0,1\}^{\mathbb{Z}}\right)(u) = \int_{\chi^{-1}(u)} F\,\mathrm{d}\,\widehat{\nu}_u, \tag{8.5}$$

where $\{\widehat{\nu}_u\}_{u\in\{0,1\}^{\mathbb{Z}}}$ denotes the measure disintegration of $\widehat{\nu}$. Then, for $u \in \{0,1\}^{\mathbb{Z}}$, $\widehat{\nu}_u$ is the $\left(\frac{1}{2},\frac{1}{2}\right)$-product measure of all positions belonging to $\{i \in \mathbb{Z} \,|\, u(i) \neq 0\}$. If $u(0) = 0$, formula (8.5) holds. If $u(0) = 1$, then F takes the two values ± 1 on $\chi^{-1}(u)$ with the same probability, so the integral on the right hand side of (8.5) is still zero. □

Theorem 8.13. (**Ch**) *implies* ($\widehat{\mathbf{S}}$).

Proof. Let $z \in \{-1,0,1\}^{\mathbb{N}_0}$ satisfy the condition (**Ch**). Then, in particular, z satisfies (**Ch**) along any strictly increasing $(N_k)_{k\in\mathbb{N}} \subset \mathbb{N}$, i.e., for each choice of $r \in \mathbb{N}$, $a_0, \ldots, a_r \in \mathbb{N}_0$, with $0 = a_0 < a_1 < \ldots < a_r$, and $i_0, \ldots, i_r \in \{1,2\}$ not all equal to 2, we have

$$\lim_{k\to\infty} \frac{1}{N_k} \sum_{n=1}^{N_k} \prod_{s=0}^{r} z^{i_s}(n + a_s) = 0.$$

For (X,T) the metric TDS in accordance, fix a completely deterministic $x \in X$ and $(N_k)_{k\in\mathbb{N}}$ such that the limit

$$\lim_{k\to\infty} \delta_{T\times S, N_k, (x,z)} \in \mathfrak{M}_{T\times S} =: \rho. \tag{8.6}$$

exists (which is possible by the Banach–Alaoglu theorem; see Theorem A.3 and Corollary A.4 in the Appendix). Then, since x is completely deterministic, the projection of ρ onto the first coordinate yields a measure $\theta \in \mathfrak{M}_T$ such that $h_\theta(T) = 0$. Furthermore, by Theorem 8.11, the projection of ρ onto the second coordinate is of the form $\widehat{\nu}$ for some $\nu \in \mathrm{Q} - \mathrm{gen}(z^2)$. Hence, using Lemma 8.12 we obtain

$$\mathbb{E}_\rho\left(F \middle| \{0,1\}^{\mathbb{Z}}\right) = \mathbb{E}_{\widehat{\nu}}\left(F \middle| \{0,1\}^{\mathbb{Z}}\right) = 0. \tag{8.7}$$

From (8.6) it follows that

$$\frac{1}{N_k} \sum_{n=1}^{N_k} f(T^n x) z(n) = \frac{1}{N_k} \sum_{n=1}^{N_k} f(T^n x) F(S^n z) \xrightarrow[k\to\infty]{} \int_{X\times\{-1,0,1\}^{\mathbb{N}_0}} f \otimes F \,\mathrm{d}\,\rho, \tag{8.8}$$

where $(f \otimes F)(x,z) \coloneqq (f(x), F(z))$, for each $x \in X$, $z \in \{-1,0,1\}^{\mathbb{N}_0}$, and

$$\int_{X\times\{-1,0,1\}^{\mathbb{N}_0}} f \otimes F \,\mathrm{d}\,\rho = \mathbb{E}_\rho\left(f \otimes F \middle| \{0,1\}^{\mathbb{Z}}\right).$$

Now, since we have

$$\mathbb{E}_\rho\left(f \otimes F \middle| \{0,1\}^{\mathbb{Z}}\right) = \mathbb{E}_\rho\left(f \middle| \{0,1\}^{\mathbb{Z}}\right) \mathbb{E}_\rho\left(F \middle| \{0,1\}^{\mathbb{Z}}\right)$$

(see Lemma 4.16 in [22]), from (8.7) and (8.8) we conclude

$$\lim_{k\to\infty} \frac{1}{N_k} \sum_{n=1}^{N_k} f(T^n x) z(n) = \mathbb{E}_\rho\left(f \middle| \{0,1\}^{\mathbb{Z}}\right) \mathbb{E}_\rho\left(F \middle| \{0,1\}^{\mathbb{Z}}\right) = 0. \qquad \square$$

Remark 8.14. Another proof of Chowla's conjecture implying Sarnak's conjecture was given by Tao in [58]. It yields a more measure-theoretic approach by making use of a variant of the moment method used in the large deviation estimates such as Chernoff's bound or Hoeffding's inequality (cf. [56]) to achieve an exponentially high concentration of a certain random variable given by the union bound and the zero topological entropy of the considered dynamical system.

Lastly, for the sake of completeness, it should be mentioned that the condition (**Ch**) is indeed stronger than the condition (**S**), which is to say that the converse implication does not hold. A counterexample confirming that can be found in [22] (Example 5.1).

A. Appendix

A.1. Lebesgue Numbers of Open Covers

Theorem A.1 (Lebesgue's number lemma)**.** *Let (X, d) be a compact metric space and $\mathcal{U}$ be an open cover of X. Then there exists a $\delta > 0$ such that to each $A \subseteq X$ with $\operatorname{diam}(A) \coloneqq \sup\{d(x, y) \mid x, y \in A\} < \delta$ there is a $U \in \mathcal{U}$ so that $A \subseteq U$.*

Proof. If $X \in \mathcal{U}$, then every $\delta > 0$ is suitable for each $A \subseteq X$. So let $X \notin \mathcal{U}$. Since X is compact, $\mathcal{U}$ contains a finite subcover $\{U_1, \ldots, U_m\}$ of X, with some $m \in \mathbb{N}$. Define a map $f : X \to \mathbb{R}$ by

$$f(x) \coloneqq \frac{1}{m} \sum_{k=1}^{m} d(x, X \setminus U_k),$$

where $d(x, C) \coloneqq \inf\{d(x, y) \mid y \in C\}$ for each $C \subseteq X$. Then f is continuous on the compact X and therefore takes its minimum which we denote by δ. Clearly, $\delta \geq 0$. Fix an $x_0 \in X$ and choose $i \in [1, m] \cap \mathbb{Z}$ such that $x_0 \in U_i$. Since U_i is open, there is an $\varepsilon > 0$ such that $B_\varepsilon(x_0) \subseteq U_i$. Therefore, $d(x_0, X \setminus U_i) \geq \varepsilon$ and hence

$$f(x_0) \geq \frac{\varepsilon}{m}.$$

Since this is possible for any $x \in X$, f is positive on X and thus $\delta > 0$.

Now let $A \subseteq X$ with $\operatorname{diam}(A) < \delta$ and choose $x_1 \in A$ arbitrarily. Then $A \subseteq B_\delta(x_1)$. Furthermore, choose $j \in [1, m] \cap \mathbb{Z}$ such that $d(x_1, X \setminus U_k)$ takes its maximum for $k = j$. Then $\delta \leq f(x_1) \leq d(x_1, X \setminus U_j)$ and hence

$$A \subseteq B_\delta(x_1) \subseteq X \setminus (X \setminus U_j) = U_j \in \mathcal{U}. \qquad \square$$

A.2. The Krylov–Bogolyubov Theorem

Theorem A.2 (Tychonoff)**.** *Let $(X_i)_{i \in I}$ be a (countable or uncountable) family of compact spaces. Then $\prod_{i \in I} X_i$ is compact in the product topology.*

A proof of Theorem A.2 can be found e.g. in [34].

Theorem A.3 (Banach–Alaoglu, *sequential version*)**.** *Let V be a separable normed vector space. Then the closed unit ball $\overline{B_1}(V^\star)$ in the dual $V^\star$ of V is sequentially compact in the w^*-topology, i.e., each sequence $(x_n)_{n \in \mathbb{N}} \subset \overline{B_1}(V^\star)$ contains a subsequence which w^*-converges in $\overline{B_1}(V^\star)$.*

The proof of Theorem A.3 makes use of Theorem A.2. One version can be found e.g. in [55].

Corollary A.4. *Let X be a compact metric space. Then every sequence $(\nu_n)_{n \in \mathbb{N}}$ of* Borel *probability measures on (X, Σ_X) has a limit point in the w^*-topology which is again a* Borel *probability measure on (X, Σ_X).*

Proof. The assertion follows from Theorem A.3, recalling that $C(X)$ is separable and for each probability measure ν on X we have

$$\left| \int_X f \,\mathrm{d}\,\nu \right| \leq \|f\|_\infty \,\nu(X) = \|f\|_\infty$$

for each $f \in C(X)$, which implies $\nu \in \overline{B_1}(C(X)^\star)$. □

Lemma A.5. *Let X be a compact metric space, $T : X \to X$ continuous, and $\nu \in \mathfrak{M}(X)$, where $\mathfrak{M}(X)$ denotes the set of all probability measures on (X, Σ_X). Then the following statements are equivalent:*

i) ν is T-invariant.

ii) For each $f \in C(X)$ we have $\int_X f \,\mathrm{d}\,\nu = \int_X (f \circ T) \,\mathrm{d}\,\nu$.

Proof. *i)*$\Longrightarrow$*ii)*: This follows directly from $C(X) \subseteq L^\infty(X, \nu)$.

ii)$\Longrightarrow$*i)*: It sufficies to show this for open sets $A \subseteq X$. One can find a sequence $(f_n)_{n\in\mathbb{N}} \subset C(X)$ such that $0 \leq f_n \nearrow \mathbf{1}_A$ for $n \to \infty$. Hence

$$\int_X f_n \,\mathrm{d}\,\nu \xrightarrow[n\to\infty]{} \int_X \mathbf{1}_A \,\mathrm{d}\,\nu = \nu(A)$$

as well as

$$\int_X (f_n \circ T) \,\mathrm{d}\,\nu \xrightarrow[n\to\infty]{} \int_X (\mathbf{1}_A \circ T) \,\mathrm{d}\,\nu = \nu(T^{-1}A).$$

Because of *ii)* this implies *i)*. □

Theorem A.6 (Krylov–Bogolyubov)**.** *Let X be a compact metric space and $T : X \to X$ continuous. Then $\mathfrak{M}_T(X) \coloneqq \{\nu \in \mathfrak{M}(X) \,|\, \nu \text{ is } T\text{-invariant}\} \neq \varnothing$.*

Proof. For $p \in X$ denote by $\delta_p : \Sigma_X \to [0, 1]$ the Dirac measure in p, given by

$$\delta_p(A) \coloneqq \begin{cases} 1 & \text{if } p \in A \\ 0 & \text{otherwise} \end{cases}.$$

Fix $a \in X$ and consider the sequence $(\nu_n)_{n\in\mathbb{N}}$ where $\nu_n \coloneqq \frac{1}{n} \sum_{k=0}^{n-1} \delta_{T^k a}$. Then for each $n \in \mathbb{N}$ we have $\nu_n \in \mathfrak{M}(X)$.

For each $p \in X$ and every $f \in L^1(X, \delta_p)$ we have $\int_X f \,\mathrm{d}\,\delta_p = f(p)$. Hence

$$\int_X (f(Tx) - f(x)) \,\mathrm{d}\,\nu_n(x) = \frac{1}{n} \left(\sum_{k=1}^{n} f(T^k a) - \sum_{k=0}^{n-1} f(T^k a) \right) = \frac{1}{n} \left(f(T^n a) - f(Ta) \right)$$

and thus

$$\left| \int_X (f \circ T) \,\mathrm{d}\,\nu_n - \int_X f \,\mathrm{d}\,\nu_n \right| \leq \frac{2}{n} \|f\|_\infty \xrightarrow[n\to\infty]{} 0. \tag{A.1}$$

Now let ν be a limit point of $(\nu_n)_{n\in\mathbb{N}}$ in the w^*-topology (whose existence is assured by Corollary A.4). Then we have $\nu \in \mathfrak{M}(X)$, since

$$\nu(X) = \int_X \mathbf{1}_X \,\mathrm{d}\,\nu = \lim_{k\to\infty} \int_X \mathbf{1}_X d\nu_{n_k} = 1,$$

for $(\nu_{n_k})_{k\in\mathbb{N}}$ an appropriate subsequence of $(\nu_n)_{n\in\mathbb{N}}$. By (A.1) we find

$$\int_X (f \circ T) \,\mathrm{d}\,\nu - \int_X f \,\mathrm{d}\,\nu = 0$$

and conclude $\nu \in \mathfrak{M}_T(X)$ using Lemma A.5. □

A.3. The Monotone Class Theorem

Definition A.7. Let X be a set and $\mathcal{M} \subseteq \mathfrak{P}(X)$, where $\mathfrak{P}(X)$ denotes the power set of X. Then we call $\mathcal{M}$ a *monotone class* in X, if

- $X \in \mathcal{M}$.
- For each sequence $(A_n)_{n\in\mathbb{N}}$ we have
  - $\bigcup_{n\in\mathbb{N}} A_n \in \mathcal{M}$,
  - $\bigcap_{n\in\mathbb{N}} A_n \in \mathcal{M}$.

Theorem A.8 (Monotone class theorem)**.** *Let X be a set. For $\mathcal{C} \subseteq \mathfrak{P}(X)$ denote by $\Sigma(\mathcal{C})$ the σ-algebra generated by A and by $\mathscr{M}(\mathcal{C})$ the smallest monotone class in X containing $\mathcal{C}$. Then, for each algebra $\mathcal{A}$ on X we have*

$$\Sigma(\mathcal{A}) = \mathscr{M}(\mathcal{A}).$$

A proof of Theorem A.8 can be found e.g. in [18].

A.4. The Representation Theorem of Riesz–Markov–Kakutani

Consider a finite regular Borel measure ν on a compact Hausdorff space X. Then $C(X) \subseteq L^1(X, \nu)$ and the mapping $f \mapsto \int_X f \,\mathrm{d}\,\nu$ yields a positive linear functional on $C(X)$. Theorem A.9 states that any positive linear functional on $C(X)$ can be obtained that way and that this representation is unique.

Riesz discovered a full classification of the dual space for the vector space $C([a, b])$, consisting of the continuous functions on $[a, b]$, $a, b \in \mathbb{R}$, $a < b$, and equipped with the L^∞-norm (which is a Banach space), by proving that each bounded linear functional on that space can be described as $\int_a^b f \,\mathrm{d}\,\nu$ using a suitable finite Borel measure ν. Later, Markov extended this theorem to the compactly supported functions on whole $\mathbb{R}$. Finally, Kakutani generalized the statement to cover the vector space of all continuous functions on any compact Hausdorff space.[1]

Theorem A.9 (Riesz–Markov–Kakutani)**.** *Let X be a compact* Hausdorff *space and Λ be a positive linear functional on $C(X)$. Then there is a unique finite regular* Borel *measure ν on (X, Σ_X) such that*

$$\Lambda(f) = \int_X f \,\mathrm{d}\,\nu$$

for each $f \in C(X)$.

For a proof of Theorem A.9 see e.g. [29].

[1] See [49].

A.5. The Borel–Cantelli Lemma

Theorem A.10 (Borel–Cantelli)**.** *Let $(X_n)_{n\in\mathbb{N}}$ be a sequence of random variables in a probability space $(\Omega, \mathcal{A}, P)$.*

a) If $\sum_{n\in\mathbb{N}} P(X_n) < \infty$, then $P\left(\limsup_{n\to\infty} X_n\right) = 0$.

b) If $(X_n)_{n\in\mathbb{N}}$ are pairwise stochastically independent and $\sum_{n\in\mathbb{N}} P(X_n) = \infty$, then $P\left(\limsup_{n\to\infty} X_n\right) = 1$.

Proof. a) Let $\varepsilon > 0$. Since the series $\sum_{n\in\mathbb{N}} P(X_n)$ converges, there is an $n_0 \in \mathbb{N}$ such that

$$\sum_{n=n_0}^{\infty} P(X_n) \leq \varepsilon.$$

Therefore, since $\limsup_{n\to\infty} X_n = \bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} X_k \subseteq \bigcup_{k=n_0}^{\infty} X_k$, from the isotony and the σ-subadditivity of the measure P we obtain

$$0 \leq P\left(\limsup_{n\to\infty} X_n\right) \leq P\left(\bigcup_{k=n_0}^{\infty} X_k\right) \leq \sum_{n=n_0}^{\infty} P(X_n) \leq \varepsilon.$$

b) Let $X \coloneqq \limsup_{n\to\infty} X_n$. Then

$$\Omega \setminus X = \Omega \setminus \left(\bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} X_k\right) = \bigcup_{n=1}^{\infty} \bigcap_{k=n}^{\infty} (\Omega \setminus X_k).$$

For $n, l \in \mathbb{N}$, $n < l$, set $Y_{n,l} \coloneqq \bigcap_{k=n}^{l} (\Omega \setminus X_k)$. Then $Y_{n,l} \supseteq Y_{n,l+1}$ and therefore $\bigcap_{l=1}^{\infty} Y_{n,l} = \bigcap_{k=n}^{\infty} (\Omega \setminus X_k)$. Thus

$$P\left(\bigcap_{k=n}^{\infty} (\Omega \setminus X_k)\right) = P\left(\bigcap_{l=1}^{\infty} Y_{n,l}\right) = \lim_{l\to\infty} P(Y_{n,l}).$$

On the other hand, by noting that

$$\log\left(\prod_{k=n}^{l} (1 - P(X_k))\right) = \sum_{k=n}^{l} \log(1 - P(X_k)) \leq -\sum_{k=n}^{l} P(X_k),$$

and since $(X_n)_{n\in\mathbb{N}}$ is pairwise stochastically independent, we obtain

$$P(Y_{n,l}) = P\left(\bigcap_{k=n}^{l} (\Omega \setminus X_k)\right) = \prod_{k=n}^{l} (1 - P(X_k)) \leq \exp\left(-\sum_{k=n}^{l} P(X_k)\right) \xrightarrow[l\to\infty]{} 0.$$

We conclude

$$1 \geq P\left(\limsup_{n\to\infty} X_n\right) = 1 - P(\Omega \setminus X) \geq 1 - \sum_{n=1}^{\infty} P\left(\bigcap_{k=n}^{\infty} (\Omega \setminus X_k)\right) = 1 - 0 = 1. \quad \square$$

Corollary A.11. *Let $(X_n)_{n\in\mathbb{N}}$ be a sequence of random variables in a probability space $(\Omega, \mathcal{A}, P)$ and X be a random variable in $(\Omega, \mathcal{A}, P)$. If $\sum_{n=1}^{\infty} P(|X_n - X| > \varepsilon) < \infty$ for each $\varepsilon > 0$, then $X_n \xrightarrow[n\to\infty]{} X$ a.s.*

It is shown e.g. in [6] how Corollary A.11 follows from Theorem A.10.

A.6. The Chinese Remainder Theorem

Theorem A.12. *Let $p, q \in \mathbb{N}$ be coprime, i.e., $(p, q) = 1$. Then the system of equations*

$$x \equiv a \mod p$$
$$x \equiv b \mod q$$

has a unique solution for x modulo pq.

Remark A.13. The reverse direction is trivial: given $x \in \mathbb{Z}_{pq}$, we can reduce x modulo p and x modulo q to obtain two equations of the above form.

Proof of Theorem A.12. [2] Choose p_1, q_1 such that

$$p_1 \equiv p - 1 \mod q$$

and

$$q_1 \equiv q - 1 \mod p.$$

These must exist since p and q are coprime. We claim that if x is an integer such that

$$x \equiv aqq_1 + bpp_1 \mod pq$$

then x satisfies both equations. Indeed, modulo p we have

$$x = aqq_1 \equiv a \mod p,$$

since $qq_1 \equiv 1 \mod p$. Similarly, $x \equiv b \mod q$. Thus x is a solution for the above equations.

It remains to show no other solutions modulo pq exist . If $z \equiv a \mod p$ then $z - x$ is a multiple of p. If also $z \equiv b \mod q$, then $z - x$ is a multiple of q as well. Since $(p, q) = 1$, this implies that $z - x$ is a multiple of pq. Hence

$$z \equiv x \mod pq.$$ □

A.7. The Stone–Weierstraß Theorem

Theorem A.14 (Stone–Weierstrass)**.** *Let X be a compact* Hausdorff *space and let $\mathcal{A}$ be the $\mathbb{C}$-algebra of continuous functions $f : X \to \mathbb{C}$. Let $\mathcal{P}$ be a sub-$\mathbb{C}$-algebra of $\mathcal{A}$ such that*

- *$\mathcal{P}$ separates the points of X, i.e., $\forall x, y \in X \, \exists f \in \mathcal{P} : f(x) \neq f(y)$,*
- *$\mathcal{P}$ vanishes nowhere on X, i.e., $\forall x \in X \, \exists f \in \mathcal{P} : f(x) \neq 0$,*
- *$\mathcal{P}$ is invariant under conjugation, i.e., $\forall f \in \mathcal{P} : \overline{f} \in \mathcal{P}$.*

Then $\mathcal{P}$ is dense in $\mathcal{A}$ given the topology of uniform convergence.

For a proof of Theorem A.14 see e.g. [51].

[2] cf. [40].

A.8. Weyl's Theorem

Lemma A.15 (Bergelson–van der Corput)**.** *Let H be a* Hilbert *space and $(u_n)_{n\in\mathbb{N}}$ be a bounded sequence in H such that*

$$\lim_{M\to\infty}\frac{1}{M}\sum_{h=1}^{M}\limsup_{N\to\infty}\left|\frac{1}{N}\sum_{n=1}^{N}\langle u_{n+h},u_n\rangle_H\right|=0.$$

Then

$$\lim_{N\to\infty}\frac{1}{N}\sum_{n=1}^{N}u_n=0.$$

A proof Lemma A.15 can be found in [7].

Theorem A.16 (Weyl)**.** *Let P be a non-constant polynomial with coefficients in $\mathbb{Z}$. Then for each $k\in\mathbb{Z}\setminus\{0\}$ and every $\alpha\in\mathbb{R}\setminus\mathbb{Q}$ we have*

$$\lim_{N\to\infty}\frac{1}{N}\sum_{n=1}^{N}e^{2\pi i\alpha P(n)k}=0.$$

Proof. First, let $\deg P=1$, where $\deg P$ denotes the degree of P. Then there are $a,b\in\mathbb{Z}$, $a\neq 0$, so that $P(n)=an+b$. Hence

$$\frac{1}{N}\sum_{n=1}^{N}e^{2\pi i\alpha P(n)k}=\frac{1}{N}e^{2\pi i\alpha bk}\sum_{n=1}^{N}e^{2\pi i\alpha ank}\xrightarrow[N\to\infty]{}0.$$

Now assume that for $\deg P\leq d\in\mathbb{N}$ the assertion is already shown and let $\deg P=d+1$. Define a sequence $(u_n)_{n\in\mathbb{N}}\subset\mathbb{C}$ by $u_n\coloneqq e^{2\pi i\alpha P(n)k}$. Then

$$\langle u_{n+h},u_n\rangle=e^{2\pi i\alpha Q_h(n)k},$$

where $Q_h(n)\coloneqq P(n+h)-P(n)$ is a polynomial with $\deg Q_h=d$. Hence, by the induction hypothesis we have

$$\frac{1}{N}\sum_{n=1}^{N}\langle u_{n+h},u_n\rangle=\frac{1}{N}\sum_{n=1}^{N}e^{2\pi i\alpha Q_h(n)k}\xrightarrow[N\to\infty]{}0.$$

This implies

$$\limsup_{N\to\infty}\left|\frac{1}{N}\sum_{n=1}^{N}\langle u_{n+h},u_n\rangle\right|=0$$

for all $h\in\mathbb{N}$, and therefore

$$\lim_{M\to\infty}\frac{1}{M}\sum_{h=1}^{M}\limsup_{N\to\infty}\left|\frac{1}{N}\sum_{n=1}^{N}\langle u_{n+h},u_n\rangle_H\right|=0.$$

Hence, by Lemma A.15 and the choice of $(u_n)_{n\in\mathbb{N}}$ the assertion follows. □

Bibliography

[1] Abramov, Leonid M.: On the Entropy of a Flow. In: *Ten Papers on Functional Analysis and Measure Theory* Bd. 49, No. 2. Providence, Rhode Island : American Mathematical Society, 1966, S. 167–170

[2] Adler, R. L. ; Konheim, A. C. ; McAndrew, M. H.: Topological entropy. In: *Transactions of the American Mathematical Society* Bd. 114. Providence, Rhode Island : American Mathematical Society, 1965, S. 309–319

[3] Anzai, Hirotada: Ergodic Skew Product Transformations on the Torus. In: *Osaka Mathematical Journal* Bd. 3, No. 1. Osaka : Departments of Mathematics of Osaka University and Osaka City University, 1951, S. 83–99

[4] Apostol, Tom M.: *Introduction to Analytic Number Theory.* New York - Heidelberg - Berlin : Springer, 1976 (Undergraduate Texts in Mathematics)

[5] Bär, Christian ; Becker, Christian: C*-algebras. In: *Quantum Field Theory on Curved Spacetime; Lecture Notes in Physics* Bd. 786. Berlin - Heidelberg : Springer, 2009, S. 1–37

[6] Bauer, Heinz: *Wahrscheinlichkeitstheorie.* 5. durchges. und verb. Berlin : Walter de Gruyter, 2002

[7] Bergelson, Vitaly ; March, Peter ; Rosenblatt, Joseph: *Convergence in Ergodic Theory and Probability.* Reprint 2011. Berlin : Walter de Gruyter, 1996

[8] Bourgain, Jean: *On the correlation of the Moebius function with random rank-one systems.* 2011. – arXiv:1112.1032v1

[9] Bourgain, Jean ; Sarnak, Peter ; Ziegler, Tamar: *Disjointness of Mobius from horocycle flows.* 2011. – arXiv:1110.0992v1

[10] Chowla, Sarvadaman: The Riemann hypothesis and Hilbert's tenth problem. In: *Mathematics and Its Applications* Bd. 4. New York : Gordon and Breach Science Publishers, 1965

[11] Conway, John B.: *A Course in Functional Analysis.* Second Edition. New York - Heidelberg - Berlin : Springer, 1997 (Graduate Texts in Mathematics)

[12] Cornfeld, Isaak P. ; Fomin, Sergei V. ; Sinai, Yakov G.: *Ergodic Theory.* Berlin - Heidelberg : Springer, 1982

[13] Dartyge, Cecile ; Tenenbaum, Gerald: Sommes des chiffres de multiples d'entiers (Sums of digits of multiples of integers). In: *Annales de l'institut Fourier* Bd. 55, No. 7. Association des Annales de l'Institut Fourier, 2005

[14] Davenport, Harold: On some infinite series involving arithmetical functions (II). In: *The Quarterly Journal of Mathematics* Bd. os-8 Issue 1. Oxford : Oxford University Press, 1937, S. 313–320

[15] Davenport, Harold: *Multiplicative Number Theory.* Second Edition. New York - Heidelberg - Berlin : Springer, 1980 (Graduate Texts in Mathematics)

[16] Denker, Manfred: *Einführung in die Analysis dynamischer Systeme.* Berlin - Heidelberg - New York : Springer, 2006

[17] Downarowicz, Tomasz: *Entropy in Dynamical Systems.* Cambridge : Cambridge University Press, 2011

[18] Durrett, Rick: *Probability - Theory and Examples.* Cambridge : Cambridge University Press, 2010

[19] Einsiedler, Manfred ; Schmidt, Klaus: *Dynamische Systeme - Ergodentheorie und topologische Dynamik.* Basel : Birkhäuser, 2014

[20] Einsiedler, Manfred ; Ward, Thomas: *Ergodic Theory - with a view towards Number Theory.* Berlin - Heidelberg : Springer, 2010 (Graduate Texts in Mathematics)

[21] El Abdalaoui, El H. ; Kasjan, Stanislaw ; Lemanczyk, Mariusz: *0-1 sequences of the Thue-Morse type and Sarnak's conjecture.* 2013. – arXiv:1304.3587v2

[22] El Abdalaoui, El H. ; Kulaga-Przymus, Joanna ; Lemanczyk, Mariusz ; de la Rue, Thierry: *The Chowla and the Sarnak conjectures from ergodic theory point of view.* 2014. – arXiv:1410.1673v2

[23] El Abdalaoui, El H. ; Lemanczyk, Mariusz ; de la Rue, Thierry: On spectral disjointness of powers for rank-one transformations and Möbius orthogonality. In: *Journal of Functional Analysis* Bd. 266, Issue 1. Amsterdam : Elsevier, 2014, S. 284–317

[24] Ferenczi, Sebastien ; Kulaga-Przymus, Joanna ; Lemanczyk, Mariusz ; Mauduit, Christian: *Substitutions and Möbius disjointness.* 2015. – arXiv:1507.01123v1

[25] Forys, Magdalena: On Sequence Entropy of Thue-Morse Shift. In: *Schedae Informaticae* Bd. 22. Krakow : Wydawnictwo Uniwersytetu Jagiellonskiego, 2013, S. 12–25

[26] Furstenberg, Harry: Ergodic behavior of diagonal measures and a theorem of Szemeredi on arithmetic progressions. In: *Journal d'Analyse Mathematique* Bd. 31, Issue 1. Jerusalem : Hebrew University of Jerusalem, 1977, S. 204–256

[27] Gauss, Carl F.: *Disquisitiones Arithmeticae... - Primary Source Edition.* Berlin : BiblioLife, 2014

[28] Green, Ben ; Tao, Terence: The Möbius function is strongly orthogonal to nilsequences. In: *Annals of Mathematics* Bd. 175, No. 2. Princeton : Princeton University and the Institute for Advanced Study, 2012

[29] Harting, Donald G.: The Riesz Representation Theorem Revisited. In: *The American Mathematical Monthly* Bd. 90, No. 4. Washington, D.C. : Mathematical Association of America, 1983, S. 227–280

[30] Heuser, Harro: *Funktionalanalysis - Theorie und Anwendung*. 4. durchges. Aufl. 2006. Wiesbaden : Vieweg+Teubner Verlag, 2006

[31] Hochman, Michael: Notes on ergodic theory. (2012). `math.huji.ac.il/~mhochman/courses/ergodic-theory-2012/notes.final.pdf`

[32] Iwaniec, Henryk ; Kowalski, Emmanuel: Analytic Number Theory. In: *Colloquium Publications* Bd. 53. Providence, Rhode Island : American Mathematical Society, 2004

[33] Jacobs, Konrad ; Jungnickel, Dieter: *Einführung in die Kombinatorik*. 2. öllig neu bearb. und erw. A. Berlin : Walter de Gruyter, 2004

[34] Jänich, Klaus: *Topologie*. New York - Heidelberg - Berlin : Springer, 2013

[35] Jia, Chao H.: The distribution of square-free numbers. In: *Science in China Series A: Mathematics* Bd. 36, No. 2. Chinese Academy of Sciences and National Natural Science Foundation of China, 1993

[36] Kawan, Christoph: Vorlesungsskript Dynamische Systeme. (2014). `http://www.fim.uni-passau.de/fileadmin/files/lehrstuhl/wirth/Publikationen_Dr._Christoph_Kawan/Skript_DS.pdf`

[37] Knopp, Marvin ; Robins, Sinai: Easy Proofs of Riemann's Functional Equation and of Lipschitz Summation. In: *Proceedings of the American Mathematical Society* Bd. 129, No. 7. Providence, Rhode Island : American Mathematical Society, 2001, S. 1915–1922

[38] Kulaga-Przymus, Joanna ; Lemanczyk, Mariusz: *The Möbius function and continuous extensions of rotations*. 2014. – arXiv:1310.2546v2

[39] Liu, Jianya ; Sarnak, Peter: *The Möbius function and distal flows*. 2013. – arXiv:1303.4957v3

[40] Lynn, Benn: The Chinese Remainder Theorem. `https://crypto.stanford.edu/pbc/notes/numbertheory/crt.html`

[41] MacCluer, Barbara: *Elementary Functional Analysis*. 1st Edition. 2nd Printing. 2008. Berlin - Heidelberg : Springer, 2008 (Graduate Texts in Mathematics)

[42] Mauduit, Christian ; Rivat, Joel: Prime numbers along Rudin-Shapiro sequences. (2013). `http://iml.univ-mrs.fr/~rivat/preprints/PNT-RS.pdf`

[43] Montgomery, Hugh L. ; Vaughan, Robert C.: Multiplicative Number Theory: I. Classical Theory. In: *Cambridge studies in advanced mathematics* Bd. 97. Cambridge - New York : Cambridge University Press, 2006

[44] Moreira, Joel: Ergodic Decomposition. (2013). `https://joelmoreira.wordpress.com/2013/09/20/ergodic-decomposition`

[45] Newman, Donald J.: Simple analytic proof of the prime number theorem. In: *The American Mathematical Monthly* Bd. 87, No. 9. Washington, D.C. : Mathematical Association of America, 1980, S. 693–696

[46] Pedersen, Gert K.: *Analysis Now.* Berlin - Heidelberg : Springer, 1989 (Graduate Texts in Mathematics)

[47] Petersen, Karl E.: *Ergodic Theory.* London : Cambridge University Press, 1983

[48] Queffelec, Martine: Substitution Dynamical Systems-Spectral Analysis. In: *Lecture Notes in Mathematics* Bd. 1294. New York - Heidelberg - Berlin : Springer, 1987

[49] Richardson, Leonard: Examples of Dual Spaces from Measure Theory. `https://www.math.lsu.edu/~rich/L_p.pdf`

[50] Rokhlin, Vladimir A.: Entropy of metric automorphism. In: *Doklady Akademii Nauk SSSR (Proceedings of the USSR Academy of Sciences)* Bd. 124. Moscow : Academy of Sciences of the USSR, 1959, S. 980–983

[51] Rudin, Walter: *Principles of Mathematical Analysis.* 3rd International edition. New York : McGraw-Hill, 1976

[52] Sarig, Omri: Lecture Notes on Ergodic Theory. (2008). `http://www.math.psu.edu/sarig/506/ErgodicNotes.pdf`

[53] Sarnak, Peter: Three Lectures on the Mobius Function Randomness and Dynamics. (2010). `http://publications.ias.edu/sites/default/files/MobiusFunctionsLectures(2).pdf`

[54] Shannon, Claude E. ; Wyner, A. D. ; Sloane, N. J. a.: *Claude Elwood Shannon - collected papers.* New York : IEEE Press, 1993

[55] Tao, Terence: 245B, Notes 11: The strong and weak topologies. In: *What's new (Blog)* (2009). `https://terrytao.wordpress.com/2009/02/21/245b-notes-11-the-strong-and-weak-topologies`

[56] Tao, Terence: 254A, Notes 1: Concentration of measure. In: *What's new (Blog)* (2010). `https://terrytao.wordpress.com/2010/01/03/254a-notes-1-concentration-of-measure`

[57] Tao, Terence: The Katai-Bourgain-Sarnak-Ziegler orthogonality criterion. In: *What's new (Blog)* (2011). `https://terrytao.wordpress.com/2011/11/21/the-bourgain-sarnak-ziegler-orthogonality-criterion`

[58] Tao, Terence: The Chowla conjecture and the Sarnak conjecture. In: *What's new (Blog)* (2012). `https://terrytao.wordpress.com/2012/10/14/the-chowla-conjecture-and-the-sarnak-conjecture`

[59] Titchmarsh, E. C.: *The Theory of the Riemann Zeta-function.* Second Edition. Oxford : Clarendon Press, 1986

[60] Walters, Peter: *An Introduction to Ergodic Theory.* Berlin - Heidelberg : Springer, 1982 (Graduate Texts in Mathematics)