AUTOMATIC SEQUENCES ARE ORTHOGONAL TO APERIODIC MULTIPLICATIVE FUNCTIONS

MARIUSZ LEMAŃCZYK
Faculty of Mathematics and Computer Science, Nicolaus Copernicus University
Chopin Street 12/18
87-100 Toruń, Poland

CLEMENT MÜLLNER*
Institut für Diskrete Mathematik und Geometrie, TU Wien
Wiedner Hauptstr. 8-10
1040 Wien, Austria

Abstract. Given a finite alphabet \( A \) and a primitive substitution \( \theta : A \to A^\lambda \) (of constant length \( \lambda \)), let \((X_\theta, S)\) denote the corresponding dynamical system, where \( X_\theta \) is the closure of the orbit via the left shift \( S \) of a fixed point of the natural extension of \( \theta \) to a self-map of \( A^\mathbb{Z} \). The main result of the paper is that all continuous observables in \( X_\theta \) are orthogonal to any bounded, aperiodic, multiplicative function \( u : \mathbb{N} \to \mathbb{C} \), i.e.
\[
\lim_{N \to \infty} \frac{1}{N} \sum_{n \leq N} f(S^n x)u(n) = 0
\]
for all \( f \in C(X_\theta) \) and \( x \in X_\theta \). In particular, each primitive automatic sequence, that is, a sequence read by a primitive finite automaton, is orthogonal to any bounded, aperiodic, multiplicative function.

Introduction. Throughout the paper, by an automatic sequence \((a_n)_{n \geq 0} \subset \mathbb{C}\), we mean a continuous observable in a substitutional system \((X_\theta, S)\), i.e. \( a_n = f(S^n x) \), \( n \geq 0 \), for some \( f \in C(X_\theta) \) and \( x \in X_\theta \). Here, we assume that \( \theta : A \to A^\lambda \) is a substitution of constant length \( \lambda \in \mathbb{N} \) (over the alphabet \( A \)), and we let \((X_\theta, S)\) denote the corresponding subshift of the full shift \((A^\mathbb{Z}, S)\) (see Remark 2 and Section 2 for details).

By \( \mu : \mathbb{N} \to \{-1, 0, 1\} \) we denote the classical M"obius function: \( \mu(p_1 \ldots p_k) = (-1)^k \) for different primes \( p_1, \ldots, p_k \), \( \mu(1) = 1 \) and \( \mu(n) = 0 \) whenever \( n \) is divisible by a square greater than 1. In connection with the celebrated Sarnak's conjecture [39] on M"obius orthogonality of zero entropy systems, i.e.
\[
\lim_{N \to \infty} \frac{1}{N} \sum_{n \leq N} f(S^n x)\mu(n) = 0
\] {#1}

2020 Mathematics Subject Classification. Primary: 37A45, 11N37, 11B85, 68R15, 11A25; Secondary: 37A05.

Key words and phrases. Automatic sequences, multiplicative functions, Sarnak conjecture, symbolic dynamical systems.

* Corresponding author: Clemens Müller.

1 This is a slight extension of the classical notion of automatic sequence meant as a sequence obtained by a finite code of a fixed point of a substitution, see e.g. [3], [14], [38] to see also a relationship with sequences read by finite automata.
for each zero entropy dynamical system \((X, S)\), all \(f \in C(X)\) and \(x \in X\) (for more information on the subject, see the survey article [12]), it has been proved in [37] that all automatic sequences \((a_n)\) satisfy (1). This triggers the question whether (1) remains true if we replace \(\mu\) by another arithmetic function. The Möbius function is an example of an arithmetic function which is multiplicative (\(\mu(mn) = \mu(m)\mu(n)\) whenever \(m, n\) are coprime), hence, it is natural to ask whether automatic sequences are orthogonal to each zero mean,\(^2\) bounded, multiplicative function \(u\). That said, one realizes immediately that the answer to such a question is negative as periodic functions are automatic sequences and there are many examples of periodic, multiplicative functions.\(^3\) Besides, even amongst non-periodic automatic sequences there are examples of (completely) multiplicative, zero mean functions \([2]\), [40], [43], see also recent [24], [26] for a significant progress on characterizing such sequences. On the other hand, it has been proved in [13] that many automatic sequences given by so-called bijective substitutions are orthogonal to all zero mean, bounded, multiplicative functions (in [9], it is proved that they are orthogonal to the Möbius function).

A stronger requirement than zero mean of \(u\) which one can consider is that of aperiodicity, that is

\[
\lim_{N \to \infty} \frac{1}{N} \sum_{n \leq N} u(an + b) = 0
\]

for each \(a, b \in \mathbb{N}\), i.e. \(u\) has a mean, equal to zero, along each arithmetic progression. Many classical multiplicative functions are aperiodic, e.g. \(\mu\) or the Liouville function \(\lambda\): \(\lambda(n) = (-1)^{\Omega(n)}\), where \(\Omega(n)\) denotes the number of prime divisors (counted with multiplicities) of \(n\).

The aim of the present paper is to prove the following (\(h(\theta)\) and \(c(\theta)\) stand, respectively, for the height and the column number of the substitution \(\theta\), see Section 2):

**Theorem 0.1.** Let \(\theta\) be a primitive substitution of constant length \(\lambda\). Then, each automatic sequence \(a_n = f(S^nx), n \geq 0, \) in \((X_\theta, S)\) is orthogonal to any bounded, aperiodic, multiplicative function \(u : \mathbb{N} \to \mathbb{C}\), i.e.

\[
\lim_{N \to \infty} \frac{1}{N} \sum_{n \leq N} a_nu(n) = 0. \tag{2}
\]

More precisely:

(i) If \(c(\theta) = h(\theta)\) then each automatic sequence \((f(S^nx))_{n \geq 0}\) is orthogonal to any bounded, aperiodic, arithmetic function \(u\).\(^4\)

(ii) If \(c(\theta) > h(\theta)\) then the automatic sequences \((f(S^nx))_{n \geq 0}\) for which the spectral measure of \(f\) is continuous are orthogonal to all bounded, multiplicative functions. If the spectral measure is discrete then (i) applies. All other automatic sequences in \((X_\theta, S)\) are orthogonal to all bounded, aperiodic, multiplicative functions.

We have already mentioned that examples of automatic sequences which are multiplicative functions are known, but they are quite special. For example, in [40],

\(^2\)Recall that a sequence \(u : \mathbb{N} \to \mathbb{C}\) has zero mean if \(M(u) := \lim_{N \to \infty} \frac{1}{N} \sum_{n \leq N} u(n)\) exists and equals zero.

\(^3\)Indeed, examples of periodic multiplicative functions are given by: Dirichlet characters, or \(n \mapsto (-1)^{n+1}\).

\(^4\)We emphasize that no multiplicativity on \(u\) is required.
it is proved that completely multiplicative, never vanishing functions “produced” by finite automata are limits in the upper density of periodic sequences, that is, they are Besicovitch rationally almost periodic.\(^5\) In fact, we can strengthen this result by showing that they are even Weyl rationally almost periodic which is a consequence of Theorem 0.1:

**Corollary 1.** All multiplicative, automatic sequences produced by primitive automata\(^6\) are Weyl rationally almost periodic. Furthermore, the automaton/subscription that produces such an automatic sequence can be chosen to have equal column number and height.

The main problem in this paper belongs to number theory and combinatorics, and the proof of Möbius orthogonality for automatic sequences in [37] relied on combinatorial properties of sequences produced by automata and an application of the (number theoretic) method of Mauduit and Rivat [33], [34]. However, the problem of orthogonality of sequences is deeply related to classical ergodic theory, namely, to Furstenberg’s disjointness of dynamical systems, see [12] for an exhaustive presentation of this approach. Our strategy in this paper is to apply ergodic theory tools to address Theorem 0.1. Specifically, we make use of some old results on the centralizer of substitutional systems [22] and, more surprisingly, we find an ergodic interpretation of the combinatorial approach from [37] in terms of joinings of substitutional systems. Finally, we relativize some of the arguments from [13] to reach the goal. For example, the reason behind the orthogonality of automatic sequences to all bounded, multiplicative functions (whenever the spectral measure of an automatic sequence is continuous) in (ii) of Theorem 0.1 is that the essential centralizer of substitutional systems is finite [22]. This “small size” of the essential centralizer puts serious restrictions on possible joinings between (sufficiently large) different prime powers \(S^p, S^q\) of \(S\) and allows one to use the numerical DKBSZ criterion on the orthogonality of numerical sequences with bounded, multiplicative functions (see Section 1.5).

The ergodic theory approach (together with number-theoretic tools) turns out to be very effective in some attempts to prove Sarnak’s conjecture. The recent results of Frantzikinakis and Host [16] show: if instead of orthogonality (1), we ask for its weaker, namely, logarithmically-averaged version

\[
\lim_{N \to \infty} \frac{1}{\log N} \sum_{n \leq N} \frac{1}{n} f(S^n x) \mu(n) = 0
\]

then this orthogonality indeed holds for all zero entropy \((X, S)\) systems whose set of ergodic (invariant) measures is countable: in particular, it applies to substitutional systems. Moreover, in (3), we can replace \(\mu\) by any so-called strongly aperiodic

---

\(^5\) A sequence \((b_n)\) taking values in a finite set \(B\) is called *Besicovitch rationally almost periodic* if it can be approximated by periodic sequences (with values in \(B\)) in the pseudo-metric

\[
d_B(x, y) = \limsup_{N \to \infty} \frac{1}{N} \left| \{1 \leq n \leq N : x(n) \neq y(n)\} \right|
\]

When, above, \(d_B\) is replaced with the pseudo-metric

\[
d_W(x, y) = \limsup_{N \to \infty} \sup_{\ell \geq N} \frac{1}{\ell} \left| \{n \leq \ell < \ell + N : x(n) \neq y(n)\} \right|
\]

we speak about *Weyl rationally almost periodic* (WRAP) sequences.

\(^6\) Sequences produced by finite automata take only finitely many values and are, therefore, bounded.
function, see [32] and Section 2.5 in [17]. However, so far, the approach through logarithmic averaging does not seem to answer the main conjecture (1) (see also [20] and [41]).

The ergodic theory approach allows for a further extension of the notion of automatic sequences, which suggests one more natural question. A uniquely ergodic topological dynamical system $(Y, T)$ (with a unique invariant measure $\nu$) is called MT- substitutional if there is a primitive substitution $\theta : A \to A^\lambda$ such that the measure-theoretic dynamical systems $(Y, \nu, T)$ and $(X_\theta, \mu_\theta, S)$ are measure-theoretically isomorphic. Then, any sequence $a_n = g(T^n y)$, $n \geq 0$ (with $g \in C(Y)$ and $y \in Y$), can be called MT- automatic. It is natural to ask whether (2) holds not only for any automatic sequence, but more generally for any MT-automatic sequences as well.

To cope with the MT-automatic case we need to be able to control the behaviour of $u$ on short intervals, i.e.,

$$\lim_{K \to \infty} \frac{1}{b_K} \sum_{k<K} \left| \sum_{b_k \leq n < b_{k+1}} u(n) \right| = 0 \quad (4)$$

(for each $(b_k)$ satisfying $b_{k+1} - b_k \to \infty$) which is much stronger than the requirement that $M(u) = 0$. It turns out however that (4) is satisfied for all non-pretentious multiplicative functions, see [31], [32]; in particular, it is satisfied for the Möbius function $\mu$.

**Corollary 2.** Any MT-automatic sequence $(a_n)$ satisfies (2) for any multiplicative, bounded, aperiodic $u : \mathbb{N} \to \mathbb{C}$ satisfying (4). Furthermore, for each MT-substitutional system $(Y, T)$ and any $g \in C(Y)$, we have

$$\frac{1}{N} \sum_{n \leq N} g(T^n y) u(n) \to 0 \text{ when } N \to \infty$$

uniformly in $y \in Y$. In particular, the uniform convergence takes place in $(X_\theta, S)$ for any primitive substitution $\theta$.

Corollary 2 answers Question 43 from [1] (see Lemma 8.1 below) on the validity of Möbius orthogonality in all uniquely ergodic models of substitutional systems.

It is clear that Möbius orthogonality of a dynamical system $(X, S)$ implies Möbius orthogonality of any topological factor of $(X, S)$. Thus, one of the key ideas to prove Theorem 0.1, given a substitutional system $(X_\theta, S)$, is to build successive continuous extensions of it, which are also given by substitutions, where for the largest extension, Theorem 0.1 will be easier to handle. For the largest extension, each continuous function $f$ on it, can be decomposed as $f = f_1 + f_2$ (with both $f_i$ continuous), where $f_1$ is highly structured implying orthogonality to all bounded aperiodic functions. While, $f_2$ is very unstructured and more difficult to handle. To show its orthogonality to all aperiodic multiplicative (bounded) functions, we will show the relative independence (over the common Kronecker-factor) of different primes powers of the shift, so that we make use of a numerical criterion of orthogonality (see Theorem 1.3) applicable for $f_2(S^n x)$ for each $x$ in the largest extension.

---

7For example, sequences given by pieces of automatic sequences in $(X_\theta, S)$: $g(S^n x_k)$ for $b_k \leq n < b_{k+1}$, $k \geq 1$, where $b_1 = 1$ and $b_{k+1} - b_k \to \infty$, for $g \in C(X_\theta)$ and $(x_k) \subset X_\theta$, are MT-automatic, see Section 7.

8As a byproduct of this procedure, we will clear up Remark 9.1 from [38], p. 229, about the form of a cocycle in a skew product representation of $(X_\theta, \mu_\theta, S)$ over the Kronecker factor.
The structure of the paper is as follows. In Section 1 we give an overview of ergodic theory necessities for, e.g. joinings, extensions and odometers. We recall basics on (dynamical systems associated with) substitutions of constant length in Section 2. This includes a classical description of the Kronecker factor for substitutions of constant length. In Section 3, for two special cases of $\theta$, we give the proofs of Theorem 0.1. They either follow the lines of proof of orthogonality of $f_j$ alone or represent a simplified version of the proof of the general case of $f$ above. In Section 4 we present some more ergodic theory prerequisites like joinings of powers of finite extensions of odometers. Moreover, we give a more detailed overview of the main ideas and the structure of the proof in Section 4.3. We finally build the aforementioned extensions of $(X_\theta, S)$ in Section 5. We finish the paper by proving Theorem 0.1 and Corollary 1 in Section 6 and Corollary 2 in Section 7.

1. **Ergodic theory necessities.** Below, we provide some basics of ergodic theory needed for the proof of Theorem 0.1. We refer the reader to [19] for more information on ergodic theory, in particular, on the theory of joinings.

1.1. **Joinings and disjointness.** By a (measure-theoretic) dynamical system we mean $(X, B, \mu, S)$, where $(X, B, \mu)$ is a standard Borel probability space and $S : X \to X$ is an a.e. bijection which is bimeasurable and measure-preserving. If no confusion arises, we will speak about $S$ itself and call it an *automorphism.*

**Remark 1.** Each homeomorphism $S$ of a compact metric space $X$ determines many (measure-theoretic) dynamical systems $(X, B(X), \mu, S)$ with $\mu \in M(X, S)$, where $M(X, S)$ stands for the set of Borel probability measures on $X$ ($B(X)$ stands for the $\sigma$-algebra of Borel sets of $X$). Recall that by the Krylov-Bogolyubov theorem, $M(X, S) \neq \emptyset$, and moreover, $M(X, S)$ endowed with the weak-$*$ topology becomes a compact metrizable space. The set $M(X, S)$ has a natural structure of a convex set (in fact, it is a Choquet simplex) and its extremal points are precisely the ergodic measures. We say that the topological system $(X, S)$ is *uniquely ergodic* if it has only one invariant measure (which must be ergodic). The system $(X, S)$ is called *minimal* if it does not contain a proper subsystem (equivalently, the orbit of each point is dense).

**Remark 2.** The basic systems considered in the paper are *subshifts* whose definition we now recall. Let $A$ be a finite, nonempty set (alphabet). By a block (or word) over $A$, we mean $B \in A^n$ (for some $n \geq 0$) and $n =: |B|$ is the length of $B$. Hence $B = (a_{i_0}, a_{i_1}, \ldots, a_{i_{n-1}})$ with $a_{i_k} \in A$ for $k = 0, \ldots, n - 1$. We will also use the following notation: $B = B_0B_1\ldots B_{n-1}$, where $B_j = a_{i_j}$ for $j = 0, \ldots, n - 1$. If $0 \leq i \leq j < n$ then we write $B[i, j]$ for $B_iB_{i+1}\ldots B_j$, in particular, notationally, $B[i] = B_i$. We say that the (sub)block $B[i, j]$ *appears* in $B$. The notation we have just presented has its natural extension to infinite sequences.

Given $\eta \in A^\mathbb{N}$, we can define

$$X_\eta := \{x \in A^\mathbb{Z} : \text{each block appearing in } x \text{ appears in } \eta\}.$$ 

It is not hard to see that $X_\eta$ is closed and $S$-invariant, where $S : A^\mathbb{Z} \to A^\mathbb{Z}$ is the left shift, i.e.

$$S((x_n)_{n \in \mathbb{Z}}) = (y_n)_{n \in \mathbb{Z}}, \text{ where } y_n = x_{n+1}, \text{ } n \in \mathbb{Z}.$$

---

9 In what follows we will also use the notation $S \in \text{Aut}(X, B, \mu)$, where $\text{Aut}(X, B, \mu)$ stands for the Polish group of all automorphisms of $(X, B, \mu)$. The topology is given by the strong operator topology of the corresponding unitary operators $U_S, U_S f := f \circ S$ on $L^2(X, B, \mu)$. 

---
Then the dynamical system \((X, S)\) is called a subshift (given by \(\eta\)). If \(\eta\) has the property that each block appearing in it reappears infinitely often (and these are the substitutional systems that are considered in this paper) then there exists \(\eta_k \in \mathbb{A}^Z\) satisfying: \(\eta_k = \eta_k\) for each \(k \geq 0\) and \(X_\eta = \{S^m \eta : m \in \mathbb{Z}\}\).

Assume that \(T \in \text{Aut}(Y, C, \nu)\) is another automorphism. Then \((X, B, \mu, S)\) and \((Y, C, \nu, T)\) are isomorphic if for some (invertible\(^{10}\)) \(W : (X, B, \mu) \rightarrow (Y, C, \nu)\), we have equivariance \(W \circ S = T \circ W\).

Given \((X, B, \mu, S)\) and \((Y, C, \nu, T)\), we may consider the set \(J(S, T)\) of joinings of automorphisms \(S\) and \(T\). (When \(S = T\), we speak about self-joinings of \(S\).) Namely, \(\kappa \in J(S, T)\) if \(\kappa\) is an \(S \times T\)-invariant probability measure on \(B \otimes C\) with the projections \(\mu\) and \(\nu\) on \(X\) and \(Y\), respectively. Note that the projections maps \(p_X : X \times Y \rightarrow X\), \(p_Y : X \times Y \rightarrow Y\) settle factor maps between the dynamical systems

\[
(X \times Y, B \otimes C, \kappa, S \times T) \text{ and } (X, B, \mu, S), (Y, C, \nu, T), \text{ respectively.}
\]

If both \(S\) and \(T\) are ergodic, the subset \(J^*(S, T)\) of ergodic joinings (i.e. of those \(\rho \in J(S, T)\) for which the system \((X \times Y, B \otimes C, \rho, S \times T)\) is ergodic) is non-empty; in fact, the ergodic decomposition of a joining consists (a.e.) of ergodic joinings.

If \((X, B, \mu, S)\) and \((Y, C, \nu, T)\) are isomorphic via \(W : (X, B, \mu) \rightarrow (Y, C, \nu)\), then \(W\) yields the corresponding graph joining \(\mu_W \in J(S, T)\) determined by

\[
\mu_W(B \times C) = \mu(B \cap W^{-1} C).
\]

Then, \((X \times Y, B \otimes C, \mu_W, S \times T)\) is isomorphic to \((X, B, \mu, S)\) (hence \(\mu_W\) is ergodic if \(\mu\) is).

The automorphisms \(S\) and \(T\) are called disjoint if the only joining of \(S\) and \(T\) is product measure \(\mu \otimes \nu\), i.e. \(J(S, T) = \{\mu \otimes \nu\}\). We will then write \(S \perp T\). Note that if \(S \perp T\) then at least one of these automorphisms must be ergodic.

We recall that an automorphism \(R \in \text{Aut}(Z, D, \kappa)\) has discrete spectrum if \(L^2(Z, D, \kappa)\) is spanned by the eigenfunctions of the unitary operator \(U_R : f \mapsto f \circ R\) on \(L^2(Z, D, \kappa)\). Assuming ergodicity, we have:

If \(R\) has discrete spectrum then each of its ergodic self-joining is graphic. \((5)\)

Each automorphism \(S \in \text{Aut}(X, B, \mu)\) has a maximal factor which has discrete spectrum. It is called the Kronecker factor of \(S\).

Joinings are also considered in topological dynamics (cf. Remark 12). If \(S_i\) is a homeomorphism of a compact metric space \(X_i, i = 1, 2\) then each \(S_1 \times S_2\)-invariant, closed subset \(M \subset X_1 \times X_2\) with the full natural projections, is called a topological joining of \(S_1\) and \(S_2\).

### 1.2. Discrete suspensions.

Given \(S \in \text{Aut}(X, B, \mu)\) and \(h \in \mathbb{N}\), consider the probability space \((\widetilde{X}, \overline{B}, \overline{\mu})\), where \(\widetilde{X} = X \times \{0, 1, \ldots, h-1\}\) and \(\overline{\mu} = \mu \otimes \rho_h\), where \(\rho_h\) is the normalized uniform measure on \(Z/hZ \cong \{0, 1, \ldots, h - 1\}\). We define now \(\widetilde{S}\) on \((\widetilde{X}, \overline{\mu})\) by setting

\[
\widetilde{S}(x,j) = (x,j + 1) \text{ whenever } x \in X, \; j = 0, 1, \ldots, h - 2
\]

and

\[
\widetilde{S}(x, h - 1) = (Sx, 0) \text{ for } x \in X.
\]

\(^{10}\)Without invertibility, we say that \(T\) is a factor of \(S\) and call \(W\) a factor map.
It is not hard to see that $\tilde{S} \in \text{Aut}(X, \tilde{B}, \tilde{\mu})$ and it is called the $h$- discrete suspension of $S$ (it is ergodic if and only if $S$ is).

Note that the map $(x, j) \mapsto j$ for $(x, j) \in \tilde{X}$ yields a factor map between $\tilde{S}$ and the rotation $\tau_h : x \mapsto x + 1$ on $\mathbb{Z}/h\mathbb{Z}$. Note also that $\tilde{S}^h(x, 0) = (Sx, 0)$, so in fact we can view $\tilde{S}$ as the $h$-discrete suspension of $\tilde{S}^h|_{X \times \{0\}}$. As a matter of fact, an automorphism $R \in \text{Aut}(Z, \mathcal{D}, \kappa)$ is an $h$-discrete suspension if and only if

$$R \text{ has the rotation } \tau_h \text{ as a factor.} \quad (8)$$

Indeed, if $\pi$ settles a factor map between $R$ and $\tau_h$, then set $A_0 := \pi^{-1}(\{0\})$, note that $R^h(A_0) = A_0$ and show that $R$ is isomorphic to the $h$-discrete suspension of $R^h|_{A_0}$.

Finally, consider $(m, h) = 1$. Then, it is easy to see that:

The $h$-discrete suspension $\tilde{\tau}_m$ is isomorphic to the direct product $\tau_m \times \tau_h$. \quad (9)

1.3. Group and isometric extensions. Given $(X, B, \mu, S)$ an ergodic dynamical system, consider a measurable $\varphi : X \to G$, where $G$ is a compact metric group. Then $\varphi$ defines a cocycle: $\varphi^{(0)}(x) := 1$, $\varphi^{(m)}(x) := \varphi(S^{m-1}x) \cdots \varphi(Sx) \varphi(x)$ and $\varphi^{(-m)}(x) := (\varphi^{(m)}(S^{-m}x))^{-1}$ for $m \geq 1$. Often $\varphi$ itself is called a cocycle. Let $m_G$ denote the Haar measure of $G$. The automorphism $S_\varphi : X \times G \to X \times G$, given by

$$S_\varphi(x, y) = (Sx, \varphi(x)g),$$

is called a compact group extension of $S$. We obtain the dynamical system $(X \times G, B \otimes B(G), \mu \otimes m_G, S_\varphi)$ which need not be ergodic. For example, it is not ergodic when $\varphi(x) = \xi(Sx)^{-1}\xi(x)$ for a measurable $\xi : X \to G$, i.e. when $\varphi$ is a coboundary (indeed, the map $(x, g) \mapsto (x, \xi(x)g)$ settles an isomorphism between $S_\varphi$ and $S \times \text{Id}_G$). Note that $(S_\varphi)^m(x, y) = (S^mx, \varphi^{(m)}(x)g)$.

**Proposition 1 ([18]).** Assume that $S$ and $T$ are ergodic automorphisms on $(X, B, \mu)$ and $(Y, \mathcal{C}, \nu)$, respectively. Assume moreover that $S \perp T$ and $S_\varphi, T_\varphi$ are ergodic group extensions of $S$ and $T$, respectively ($\psi : Y \to H$). If the product measure $(\mu \otimes m_G) \otimes (\nu \otimes m_H)$ is ergodic\(^{11}\) then $S_\varphi \perp T_\varphi$.

The following results are also classical.

**Lemma 1.1.** Assume that $S_\varphi$ and $T_\varphi$ are ergodic and let $\bar{\rho} \in J^r(S_\varphi, T_\varphi)$. Then (up to a natural permutation of coordinates) $(X \times G \times Y \times H, B \otimes B(G) \otimes \mathcal{C} \otimes B(H), \bar{\rho}, S_\varphi \times T_\varphi)$ is a compact group extension of $(X \times Y, B \otimes \mathcal{C}, \rho, S \times T)$, where $\rho := \bar{\rho}|_{X \times Y}$.

Moreover, if the relatively independent extension\(^{12}\) $\hat{\rho}$ of $\rho$ is ergodic then $\bar{\rho} = \hat{\rho}$.

A group extension is a special case of so-called skew products. Assume that we have a measurable map $\Sigma : X \to \text{Aut}(Z, \mathcal{D}, \rho)$. Then we can consider $S_\Sigma$:

$$S_\Sigma(x, z) := (Sx, \Sigma(x)(z))$$

which is an automorphism of $(X \times Z, B \otimes \mathcal{D}, \mu \otimes \rho)$. If additionally, $Z$ is a compact metric space and $\Sigma(x), x \in X$, are isometric then we call $S_\Sigma$ an isometric extension.

\(^{11}\)This means that the only situation in which we lose disjointness is when the cocycles $\varphi$ and $\psi$ “add” a common non-trivial eigenvalue for $U_{S_\varphi}$ and $U_{T_\varphi}$.

\(^{12}\)This measure is defined by

$$\bar{\rho}(\tilde{A} \times \tilde{B}) = \int_{X \times Y} E(\tilde{A}|X)(x)E(\tilde{B}|Y)(y) \, d\rho(x, y)$$

for Borel subsets $\tilde{A} \subset X \times G$ and $\tilde{B} \subset X \times H$. 

Since the group $\text{Iso}(Z)$ of isometries of $Z$ considered with the uniform topology is a compact metric group (by Arzela-Ascoli theorem), it is not hard to see that each isometric extension is a factor of a group extension (especially, if we assume that the isometric extension is uniquely ergodic).\textsuperscript{13} If the isometric extension is ergodic, one can choose the group extension also ergodic.

**Remark 3.** If $S_\varphi$ is a group extension, then for each closed subgroup $F \subset G$, the automorphism $S_{\varphi,F}$, given by

$$S_{\varphi,F}(x,gF) := (Sx, \varphi(x)gF),$$

is an isometric extension of $S$; it acts on the space $X \times G/F$ considered with $\mu \otimes m_{G/F}$, where the measure $m_{G/F}$ on the homogeneous space $G/F$ is the natural image of Haar measure $m_G$.

**Remark 4.** Each finite extension is isometric. It is a factor of a group extension by $G$, where $G$ is finite.

1.4. **Odometers.** Odometers are given by the inverse limits of cyclic groups: we have $n_{t-1} | n_t$ for each $t \geq 1$ and

$$X = H(n_t/n_{t-1}) := \liminf \mathbb{Z}/n_t \mathbb{Z}$$

with the rotation $R$ by 1 on each coordinate. If for each $t$, $n_{t+1}/n_t = \lambda \geq 2$, then we speak about $\lambda$-odometer and denote it by $H_\lambda(= H(\lambda))$.

Each odometer $(X, R)$ is uniquely ergodic (with the unique measure being Haar measure $m_X$ of $X$), in particular, $R \in \text{Aut}(X, m_X)$. Then $(X, m_X, R)$ has discrete spectrum with the group of eigenvalues given by all roots of unity of degree $n_t$, $t \geq 1$. Furthermore, $R^t$ is ergodic (uniquely ergodic) iff $(r, n_t) = 1$ for each $t \geq 1$. In this case $R^t$ and $R$ are isomorphic as both are ergodic and their spectra are the same, so the claim follows by the Halmos-von Neumann theorem (e.g. [19]). It now follows from (5) that whenever $p, q \in \mathcal{P}$ are different prime numbers (by $\mathcal{P}$ we denote the set of prime numbers) not dividing any $n_t$ then each $p \in J^c(R^p, R^q)$ is a graph joining (of an isomorphism between $R^p$ and $R^q$).

1.4.1. **$h$-discrete suspensions of odometers.** Assume that $(h, n_t) = 1$ for each $t \geq 1$. Let $\bar{R}$ denote the $h$-discrete suspension of $R$. Then (cf. (9)), we obtain that

$$\bar{R} \text{ is isomorphic to } R \times \tau_h. \quad (10)$$

Indeed, both automorphisms have discrete spectrum (and are ergodic). The group of eigenvalues of $U_{\bar{R}}$ is equal to $\{e^{2\pi ij/(hn_t)} : j \in \mathbb{Z}, t \geq 0\}$, while the group of eigenvalues of $U_{R \times \tau_h}$ is generated by $\{e^{2\pi i j/n_t} : j \in \mathbb{Z}, t \geq 0\}$ and the group of $h$-roots of unity. It follows that $U_{R \times \tau_h}$ and $U_{\bar{R}}$ have the same group of eigenvalues, hence again by the Halmos-von Neumann theorem, they are isomorphic.

1.4.2. **Group extensions of odometers - special assumptions.** We will assume that $R$ is an odometer with “small spectrum”, that is, the set $\{p \in \mathcal{P} : p | n_t \text{ for some } t \geq 1\}$ is finite (this will always be the case for $\lambda$-odometers). Moreover, we will assume that $R_{\varphi}$ (more precisely, $U_{R_{\varphi}}$) has continuous spectrum on the space $L^2(X \times G, m_X \otimes m_G) \subset L^2(X, m_X)$.\textsuperscript{14} It follows that if $p \in \mathcal{P}$ is sufficiently large then $(R_{\varphi})^p$

---

\textsuperscript{13}A group extension can be defined on $X \times \text{Iso}(Z)$, acting by the formula $(x, I) \mapsto (Sx, \Sigma(x) \circ I)$; to obtain a factor map, fix a point (transitive for $\text{Iso}(Z)$) $z_0 \in Z$ and consider $(x, I) \mapsto (x, I(z_0))$.

\textsuperscript{14}If $V$ is a unitary operator on a Hilbert space $\mathcal{H}$ then it is said to have continuous spectrum if for all $x \in \mathcal{H}$ the spectral measure $\sigma_x$ is continuous; the latter measure (on $\mathbb{S}^1$) is determined by its Fourier transform: $\hat{\sigma}_x(m) := \int_{\mathbb{S}^1} z^m \, \sigma_x(z) = \langle V^m x, x \rangle$ for all $m \in \mathbb{Z}$. 


is ergodic (uniquely ergodic if \( \varphi \) is continuous). Furthermore, we assume that if \( p, q \in \mathcal{P} \) are different and sufficiently large then the only ergodic joinings between \((R_{p})^p\) and \((R_{q})^q\) are relatively independent extensions of isomorphisms between \(R^p\) and \(R^q\). By all these assumptions \((X \times G \times X \times G, \mu_{X})_{W}, (R_{p})^p \times (R_{q})^q\) is ergodic and has the same eigenvalues as the odometer \(R\).

**Lemma 1.2.** Assume that \( \varphi \) is continuous. Under the above assumptions for each \((\text{real-valued}) F \in C(X \times G)\) such that \( F \perp L^2(X, \mu_{X})\), we have
\[
\frac{1}{N} \sum_{n \leq N} F((R_{p})^p(x, g))F((R_{q})^q(x, g)) \to 0
\]
for each \((x, g) \in X \times G\) and different primes \(p, q\) sufficiently large.

**Proof.** First notice that any accumulation point \( \rho \) of \( \left( \frac{1}{N} \sum_{n \leq N} \delta_{\{(R_{p})^p \times (R_{p})^q\}^n(x, x)} \right) \) is ergodic (cf. e.g. [27]), hence is graphic. It follows that any accumulation point \( \tilde{\rho} \) of \( \left( \frac{1}{N} \sum_{n \leq N} \delta_{\{(R_{p})^p \times (R_{q})^q\}^n(x, g, x, g)} \right) \) will be the relatively independent extension of an isomorphism between \(R^p\) and \(R^q\) in view of the second part of Lemma 1.1. By the definition of the relatively independent extension, \( \int F \otimes F \, d\tilde{\rho} = 0 \) and the result follows.

**Remark 5.** We can apply the above proof to \( R_{p} \) instead of \( R_{p}^p \) and \( \tau \) being any rotation on finitely many points instead of \( R_{p}^p \). Then, similarly as in the proof of Corollary 11.37 in [12], using the fact that the spectral measure of \( F \) is continuous, we obtain
\[
\frac{1}{N} \sum_{n \leq N} F((R_{p})^p(x, g))v(n) \to 0
\]
for each \((x, g) \in X \times G\) and each periodic sequence \(v\). That is, the sequence \((F(R_{p}^p(x, g)))_{n}\) itself is aperiodic. This argument shows that we cannot expect (i) in Theorem 0.1 to hold for all automatic sequences.

1.5. **Application - DKBSZ criterion.** Recall the following result (to which we refer as the DKBSZ criterion) about the orthogonality of numerical sequences with bounded, multiplicative functions:

**Theorem 1.3** (Daboussi, Kátai [23], Bourgain, Sarnak and Ziegler [6]). Let a bounded sequence \((a_{n}) \subseteq \mathbb{C}\) satisfy
\[
\lim_{N \to \infty} \frac{1}{N} \sum_{n \leq N} a_{p_{n}} \mu_{q_{n}} = 0
\]
for each pair of distinct, sufficiently large prime numbers \(p\) and \(q\). Then
\[
\lim_{N \to \infty} \frac{1}{N} \sum_{n \leq N} a_{n} u(n) = 0
\]
for each bounded, multiplicative function \(u: \mathbb{N} \to \mathbb{C}\).

---

15 We recall that if \( W : X \to X \) settles an isomorphism of \(R^p\) and \(R^q\), then it yields an ergodic joining \((m_{X})_{W}(A \times B) := m_{X}(A \cap W^{-1} B)\); its relatively independent extension \(m_{X} \tilde{W}\) is defined by
\[
(m_{X})_{W}(\tilde{A} \times \tilde{B}) = \int_{X} \mathbb{E}(\tilde{A}|X)(x)\mathbb{E}(\tilde{B}|X)(Wx) \, dm_{X}(x)\text{ for Borel subsets } \tilde{A}, \tilde{B} \subseteq X \times G.\]
In the context of topological dynamical systems, that is, given \((X,S)\), we use this result with \(a_n = f(S^n x)\) with \(f \in C(X)\) and \(x \in X\). It is not hard to see that this criterion applies for any uniquely ergodic \((X,S)\) with the property that \(S^p \perp S^q\) (disjointness is meant if we consider the unique invariant measure \(\mu\)) and we consider \(f \in C(X)\) with \(\int_X f \, d\mu = 0\) and arbitrary \(x \in X\). However, even if we do not have disjointness of sufficiently large (prime) powers, we can apply this criterion for particular continuous functions if we control the limit joinings as we have already seen in Lemma 1.2. In fact, it guides the strategy of the proof of Theorem 0.1 as explained in Section 4.3.

2. Basics on substitutions of constant length. We refer the reader to [38], Chapters 5, 6 and 9 for most of the statements about substitutions of constant length we list below.

2.1. Subshifts determined by substitutions. Let \(\mathcal{A}\) be an alphabet (a non-empty finite set). Denote \(\mathcal{A}^* := \bigcup_{m \geq 0} \mathcal{A}^m\), where \(\mathcal{A}^m\) stands for the set of words \(w = a_0 a_1 \ldots a_{m-1}\) over \(\mathcal{A}\) of length \(|w|\) equal to \(m\) (\(\mathcal{A}^0\) consists only of the empty word). Fix \(\mathbb{N} \ni \lambda \geq 2\). By a \textit{substitution} of (constant) length \(\lambda\) we mean a map

\[\theta : \mathcal{A} \rightarrow \mathcal{A}^\lambda\]

which we also write as \(\theta(a) = \theta(a_0 a_1 \ldots \theta(a)_{\lambda-1}\) for \(a \in \mathcal{A}\). Via the concatenation of words, there is a natural extension of \(\theta\) to a map from \(\mathcal{A}^m\) to \(\mathcal{A}^{m\lambda}\) (for each \(m \geq 1\)) or from \(\mathcal{A}^*\) to itself, or even from \(\mathbb{Z}\) to itself. In particular, we can iterate \(\theta\) \(k\) times:

\[\mathcal{A} \xrightarrow{\theta} \mathcal{A}^\lambda \xrightarrow{\theta} \cdots \xrightarrow{\theta} \mathcal{A}^{\lambda^k}\]

which can be viewed as the substitution \(\theta^k\) (the \(k\)th-iterate of \(\theta\)) of length \(\lambda^k\):

\[\theta^k : \mathcal{A} \rightarrow \mathcal{A}^{\lambda^k}\]

The following formula is well-known and follows directly by definition:

\[\theta^{k+\ell}(a)_{j\lambda^k+j} = \theta^k(\theta^\ell(a)_{j'})_{j}\]  \hspace{1cm} (11)

for each \(a \in \mathcal{A}\), \(j' < \lambda^\ell\), \(j < \lambda^k\). Indeed, \(|\theta^\ell(a)| = \lambda^\ell\) and consider \(\theta^\ell(a)_{j'}\), i.e. the \(j'\)-th letter in the word \(\theta^\ell(a)\). We now let act \(\theta^k\) on \(\theta^\ell(a)\) which transforms letters in the word \(\theta^\ell(a)\) into blocks of length \(\lambda^k\), in particular the \(j'\)-th letter of \(\theta^\ell(a)\) becomes the \(j'\)-th block of length \(\lambda^k\). So counting the \(j\)-th letter in this block is the same as counting \((j'\lambda^k + j)\)-th letter in \(\theta^{k+\ell}(a)\).

The \textit{subshift} \(X_\theta \subset A^\mathbb{Z}\) is determined by all words that appear in \(\theta^k(a)\) for some \(k \geq 1\) and \(a \in \mathcal{A}\):

\[X_\theta := \{ x \in A^\mathbb{Z} : \text{for each } r < s, \text{ we have } x[r,s] = \theta^k(a)[i, i+s-r-1]\} \quad \text{for some } k \geq 1, a \in \mathcal{A} \text{ and } 0 \leq i < \lambda^k - (s-r)\}. \hspace{1cm} (12)\]

That is, \(X_\theta\) is closed and invariant for the left shift \(S\) acting on \(A^\mathbb{Z}\). Then, clearly, we have

\[X_{\theta^k} \subset X_\theta \text{ for each } k \geq 1. \hspace{1cm} (13)\]

Note also that for each \(a \in \mathcal{A}\), there exist \(k, \ell \geq 1\) such that

\[\theta^k(a)_0 = \theta^{k+\ell}(a)_0\]

from which we deduce that there is a letter \(a_0 \in \mathcal{A}\) and an \(\ell \geq 1\) such that

\[\theta^\ell(a_0)_0 = a_0. \hspace{1cm} (14)\]
Hence, by iterating the substitution $\theta^\ell$, we obtain a fixed point $u \in \mathbb{A}^\mathbb{N}$ for the map $\theta^\ell : \mathbb{A}^\mathbb{N} \to \mathbb{A}^\mathbb{N}$. This, similarly to the RHS of (12) defines a subshift $X_u \subset \mathbb{A}^\mathbb{Z}$ for which we have $X_u \subset X_{\theta}$, cf. Remark 2. In general, we do not have equalities in (13), and $X_u$ is a proper subshift of $X_{\theta^\ell}$ above a fortiori. The situation changes if we assume that $\theta$ is primitive, that is, when there exists $k \geq 1$ such that

$$ \theta^k(a) \text{ contains all letters from } \mathbb{A}. \quad (15) $$

Then, we have equalities in (13), and moreover, for each $\ell \geq 1$ if $u \in \mathbb{A}^\mathbb{N}$ satisfies $\theta^\ell(u) = u$ then $X_u = X_{\theta}$. Moreover, we have

**Proposition 2.** Assume that $\theta$ is a primitive substitution of constant length $\lambda$. Then $(X_{\theta}, S)$ is minimal. Moreover, there exists exactly one invariant measure $\mu_{\theta}$ on $X_{\theta}$.

In what follows, we assume that $\theta : \mathbb{A} \to \mathbb{A}^\mathbb{A}$ is primitive. Under this assumption, it follows directly by the Perron-Frobenius theorem that, for any letter $a \in \mathbb{A}$, the density

$$ \delta_a := \lim_{n \to \infty} \frac{\# \{ j < \lambda^n : \theta^n(a')_j = a \} }{\lambda^n} $$

exists and is independent of $a' \in \mathbb{A}$. More precisely, first, denote by $M(\theta) \in \mathbb{Z}^{|A| \times |A|}$ the incidence matrix of $\theta$, i.e.,

$$ M_{a,a'}(\theta) := | \theta(a')_a | := \text{the number of occurrences of } a \text{ in } \theta(a'). \quad (16) $$

The vector $(\delta_a)_{a \in \mathbb{A}}$ is then the unique right eigenvector for the (maximal) eigenvalue $\lambda$, i.e.

$$ \lambda \cdot \delta_a = \sum_{a' \in \mathbb{A}} M(\theta)_{a,a'} \cdot \delta_{a'} \quad (17) $$

holds for all $a \in \mathbb{A}$ and uniquely defines $(\delta_a)_{a \in \mathbb{A}}$, see [38].

Then, by what is said around (14) and (15), we can also assume that for some $a_0 \in \mathbb{A}$,

$$ \theta(a_0)_{a_0} = a_0 \quad (18) $$

and by iterating $\theta$ at $a_0$ we obtain $u \in \mathbb{A}^\mathbb{N}$ such that $\theta(u) = u$ (and $X_u = X_{\theta}$).

**2.2. The height and the column number.** We first recall the definition of the height $h(\theta)$ of $\theta$ following [38]. Therefore, for $k \geq 0$, set

$$ S_k = S_k(\theta) := \{ r \geq 1 : u[k+r] = u[k] \} \quad (19) $$

and

$$ g_k := \gcd S_k. \quad (20) $$

This allows us to define

$$ h(\theta) := \max \{ m \geq 1 : (m, \lambda) = 1, m|g_0 \}. \quad (21) $$

We list now some basic properties of $h = h(\theta)$:[16]

1. $1 \leq h \leq |\mathbb{A}|$.
2. $h = \max \{ m \geq 1 : (m, \lambda) = 1, m|g_k \}$ for every $k \geq 0$.

---

[16] Even for a bijective substitution (see Section 3.2), the height can be $> 1$, see Example 3 in [35]. Also, substitution $\hat{\Theta}$ in Example 4 enjoys the same properties.
3. If, for \( j = 0, \ldots, h - 1 \), we consider the set
\[
C_j := \{ u[j + rh] : r \geq 0 \},
\]
then, by 2., the letter \( u[j] \) can be equal to some \( u[s] \) only if \( u[s] \in C_j \). Hence, the sets \( C_j \) form a partition of \( \mathcal{A} \). If we identify, in \( u \), the letters in the same set \( C_j, j = 0, \ldots, h - 1 \), we thus obtain a periodic (of period \( h \)) sequence (cf. (8)), and \( h \) is the largest integer \( \leq |\mathcal{A}| \), coprime to \( \lambda \), with this property.

Moreover, if by \( \lambda(h) \) we denote the set \( \{ u[mh(\theta), mh(\theta) + h(\theta) - 1] : m \geq 0 \} \subset \mathcal{A}^h \) (which is of course finite) then we can define \( \eta : \lambda(h) \to (\lambda(h))^\lambda \) by setting
\[
\eta(w) = \theta(w)[0, h(\theta) - 1][\theta(w)[h(\theta), 2h(\theta) - 1] \ldots \theta(w)[(\lambda - 1)h(\theta), \lambda h(\theta) - 1]].
\]
Then \( \eta \) turns out to be a primitive substitution and \( h(\eta) = 1 \). Moreover, in view of 3. above, as a (measure-theoretic) dynamical system, \((X, \mu, S)\) is isomorphic to the \( h(\theta) \)-discrete suspension \( \tilde{S} \) of \((X_\mu, \mu, S)\) (in fact, \((X, S)\) is topologically isomorphic to the \( h(\theta) \)-discrete suspension of \((X_\mu, S)\), see Lemma 19 in [8]). In what follows we will denote \( \eta \) by \( \theta(h) \).

If \( \theta \) is primitive then so is \( \theta^k \) (for each \( k \geq 1 \)). Take now \( u \) a fixed point for \( \theta \). This is also a fixed point of \( \theta^k \). Moreover, \( h(\theta) = h(\theta^k) \). It follows easily that
\[
(\theta(h))^k = (\theta^k(h)). \tag{22}
\]
Note also that if \((a_0, \ldots, a_{h-1}) \in \lambda(h) \) then
\[
\theta(h)(a_0, \ldots, a_{h-1}) = \theta(a_0)\theta(a_1) \ldots \theta(a_{h-1}),
\]
where the block on the RHS is divided into blocks of length \( h \) which are elements of \( \lambda(h) \). The \( j \)-th such block \( (j < \lambda) \) begins at the position \( jh \) which is of the form
\[
jh = i\lambda + y, \quad 0 \leq y < \lambda
\]
with \( i = \lfloor jh/\lambda \rfloor \) (and \( y = jh - i\lambda \)). It follows that if
\[
\theta(h)(a_0, \ldots, a_{h-1})_j = (b_0, \ldots, b_{h-1})
\]
then \( b_0 = \theta(a_i)_{j-h\lambda} \). Replacing in this reasoning \( \theta \) by \( \theta^k \) and using (22), we obtain the following: if \((a_0, \ldots, a_{h-1}) \in \lambda(h), k \geq 1, j < \lambda^k \) and
\[
(\theta(h))^k(a_0, \ldots, a_{h-1})_j = (b_0, \ldots, b_{h-1})
\]
then
\[
b_0 = \theta^k(a_i)_{jh-i\lambda^k}, \quad \text{where} \quad i = \lfloor jh/\lambda^k \rfloor. \tag{23}
\]

Following [8], a substitution \( \theta \) is called pure if \( \theta = \theta(\mathcal{A}) \), i.e. if \( h(\theta) = 1 \).

Denote by \( c = c(\theta) \) the column number of \( \theta \). Recall its definition: we consider iterations \( \theta^k : \mathcal{A} \to \mathcal{A}^\lambda \) and each time we consider sets \( \{ \theta^k(a)_j : a \in \mathcal{A} \} \) for \( j = 0, \ldots, \lambda - 1 \), where \( \theta^k(a) = \theta^k(a)_0 \theta^k(a)_1 \ldots \theta^k(a)_{\lambda^k-1} \). Then
\[
c(\theta) := \min_{k \geq 1} \min_{0 \leq j < \lambda^k-1} |\{ \theta^k(a)_j : a \in \mathcal{A} \}|.
\]

In what follows we will make use of the following observation.

**Lemma 2.1.** Let \( \theta : \mathcal{A} \to \mathcal{A}^\lambda \) be a primitive substitution with \( c(\theta) = 1 \). Then \( \theta \) is pure, i.e. we have \( h(\theta) = 1 \).

**Proof.** By assumption, there exist \( a \in \mathcal{A}, k \in \mathbb{N} \) and \( 0 \leq j < \lambda^k \) such that \( \theta^k(x)_j = a \) for all \( x \in \mathcal{A} \). Recalling that the fixpoint of \( \theta \) is \( u \) we find that \( u[j] = u[j + \lambda^k] = a \). This shows that \( \lambda^k \in S_j(\theta) \) and thus, \( g_j \mid \lambda^k \). As \( h(\theta)|g_j \) and is coprime with \( \lambda \), we have \( h(\theta) = 1 \). \( \square \)
Lemma 2.2. Let \( \theta : \mathbb{A} \to \mathbb{A}^\lambda \) be a primitive substitution. Then
\[
c(\theta) = h(\theta) \cdot c(\theta^h).
\]
In particular, \( h(\theta) \mid c(\theta) \).

This seems to be a folklore result but we could not find a proof of it in the literature. We postpone the proof of Lemma 2.2 to Section 5.4.

The following result is well known for pure substitutions but an application of Lemma 2.2 yields the more general one below.

Proposition 3. Assume that \( \theta : \mathbb{A} \to \mathbb{A}^\lambda \) is primitive. Then, as a measure-theoretic dynamical system, \((X_\theta, \mu_\theta, S)\) is isomorphic to a \(c(\theta)\)-point extension (cf. Remark 4) of the \(\lambda\)-odometer \((H_\lambda, m_{H_\lambda}, R)\).

Proof. We have an isomorphism of \(X_\theta\) (we omit automorphisms and measures) with \(\widetilde{X}_{\theta^h}\). If \(\pi : X_{\theta^h} \to H_\lambda\) denotes the factor map which is \(c(\theta^h)\) to 1 (a.e.), then \(\pi \times Id_{\mathbb{Z}/h\mathbb{Z}}\) yields a factor map between the \(h\)-discrete suspensions \(\widetilde{X}_{\theta^h}\) and \(\widetilde{H}_\lambda\). It is again \(c(\theta^h)\) to 1 (a.e.). However, in view of (10), the suspension \(\widetilde{H}_\lambda\) is isomorphic to the direct product \(H_\lambda \times \mathbb{Z}/h\mathbb{Z}\).\(^{17}\) This makes \(H_\lambda\) a factor of \(X_\theta\) with (a.e.) fiber of cardinality \(c(\theta^h) \cdot h\). The result follows from Lemma 2.2.

Corollary 3. The dynamical system \((X_\theta, \mu_\theta, S)\) given by (primitive) \(\theta\) has purely discrete spectrum if and only if \(h(\theta) = c(\theta)\).

As a matter of fact, whenever \(\theta\) is primitive and \(h(\theta) = 1\), the system \((H_\lambda, R)\) represents the so-called maximal equicontinuous factor of \((X_\theta, S)\) (i.e., a maximal topological factor of \((X_\theta, S)\) represented by a translation on a compact Abelian group). The corresponding factor map is usually seen in the following way: Each point \(x \in X_\theta\) has a unique \(\lambda^t\)-skeleton structure. By that, one means a sequence \((j_t)_{t \geq 1}\) with \(0 \leq j_t < \lambda^t - 1\) for which, for each \(t \geq 1\), we have
\[
x[-j_t + s\lambda^t, -j_t + (s + 1)\lambda^t - 1] = \theta^t(c_s)
\]
for each \(s \in \mathbb{Z}\) and some \(c_s \in \mathbb{A}\). Now, the map \(x \mapsto (j_t)\) yields the factor map which we seek.

Notice also that if \(R \subset \mathbb{A} \times \mathbb{A}\) is an equivalence relation which is \(\theta\)-consistent, that is:
\[
(a, b) \in R \Rightarrow (\theta(a)_j, \theta(b)_j) \in R \text{ for each } j = 0, \ldots, \lambda - 1,
\]
then the quotient substitution \(\theta_R : \mathbb{A}/R \to (\mathbb{A}/R)^\lambda\) given by
\[
\theta_R([a]) := [\theta(a)_0] \ldots [\theta(a)_{\lambda-1}]
\]
is correctly defined and the dynamical system \((X_{\theta_R}, S)\) is a (topological) factor of \((X_\theta, S)\). Hence, \((X_{\theta_R}, \mu_{\theta_R}, S)\) is also a measure-theoretic factor of \((X_\theta, \mu_\theta, S)\) (clearly, \(\theta_R\) is primitive). A particular instance of a \(\theta\)-consistent equivalence relation \(R\) is given by
\[
(a, b) \in R \text{ if and only if } \theta(a) = \theta(b).
\]

\(^{17}\)The factor \(H_\lambda\) is represented in \(H_\lambda \times \mathbb{Z}/h\mathbb{Z}\) as the first coordinate \(\sigma\)-algebra and it is a factor of \(\widetilde{H}_\lambda\) represented by an invariant \(\sigma\)-algebra. However, because of ergodicity and the fact that \(H_\lambda\) has discrete spectrum, there is only one invariant \(\sigma\)-algebra in \(\widetilde{H}_\lambda\) representing \(H_\lambda\). The same argument shows that \(H_\lambda\) is a factor of \(X_\theta\) in a canonical way.
In this situation, the quotient dynamical system \((X_{\theta_0}, S)\) is in fact (topologically) isomorphic to \((X_\theta, S)\). Therefore, no harm arises if we assume that
\[
\theta \text{ is } 1-1 \text{ on letters, i.e., } \theta(a) \neq \theta(b) \text{ whenever } a \neq b.
\] (25)

Finally, we assume that
\[
\theta \text{ is aperiodic,}
\] (26)
that is, there is a non-periodic element \(x \in X_\theta\). In fact, since we assume that \(\theta\) is primitive, \(\theta\) is aperiodic if and only if \(X_\theta\) is infinite.

From now on, we consider only substitutions \(\theta\) satisfying (15), (18), (25) and (26).

Often, we will additionally assume that \(\theta\) is pure.

3. Proof of Theorem 0.1 - special cases.

3.1. Proof of Theorem 0.1 (i). The main ingredient in the proof is to show that every primitive automatic sequence (with substitution \(\theta\)) with \(c(\theta) = h(\theta)\) is WRAP. In view of Lemma 2.2, it is enough to consider the case \(c(\theta) = 1\). Indeed, \((X_\theta, S)\) is topologically isomorphic to the \(h(\theta)\)-discrete suspension of \((X_\theta, S)\) and the automatic sequence with substitution \(\eta\) being WRAP immediately implies that the \(h\)-discrete suspension is also WRAP. Hence, according to Proposition 3, we deal (from the ergodic theory point of view) with the discrete spectrum case.\(^{18}\) In fact, we will prove an even stronger property.

So we have \(\theta : A \to A^\lambda\), and we assume that for some \(a \in A\), \(\theta(a)_0 = a\). Moreover, by replacing \(\theta\) by its iterate if necessary, we can assume that
\[
|\{0 \leq j < \lambda : |\{\theta(b)_j : b \in A\}| = 1\}| =: \ell_1 \geq 1.
\]
Hence, since \(u = \theta(u)\) is a concatenation of words \(\theta(b), b \in A\), in the interval \([0, \lambda - 1]\) the number of positions \(j\), for which \(u[j + s\lambda] = u[j]\) for each \(s \geq 1\) is \(\ell_1\).

Let us pass to \(\theta^2\). We are interested in
\[
\ell_2 := |\{0 \leq j < \lambda^2 : |\{\theta^2(b)_j : b \in A\}| = 1\}|.
\]
We have \(\ell_2 \geq (\lambda - \ell_1)\ell_1 + \ell_1\lambda\lambda\). Indeed, \(\theta(\theta(b)_r)_r = c\) for all \(b \in A\) either if \(r\) contributes to the \(\ell_1\) count, for all \(s\), or if \(s\) is counted by \(\ell_1\). Inductively, if
\[
\ell_{k-1} := |\{0 \leq j < \lambda^{k-1} : |\{\theta^{k-1}(b)_j : b \in A\}| = 1\}|
\]
then
\[
\ell_k \geq (\lambda - \ell_1)\ell_{k-1} + \ell_1\lambda^{k-1}.
\] (27)
Moreover, by definition,
\[
\text{the sequence } \ell_k/\lambda^k, k \geq 1, \text{ is increasing.}
\] (28)
In view of (27), we have
\[
\frac{\ell_k}{\lambda^k} \geq \frac{\lambda - \ell_1}{\lambda - \lambda^{k-1}} + \frac{\ell_1}{\lambda},
\]
and since, by (28), the sequence \(\ell_k/\lambda^k\) is convergent, the above recurrence formula implies
\[
\lim_{k \to \infty} \frac{\ell_k}{\lambda^k} = 1.
\] (29)
Note that an interpretation of \(\ell_k\) is that it is the number of coordinates \(j\) for \(j = 0, 1, \ldots, \lambda^k - 1\) such that \(u[j + s\lambda^k] = u[j]\) for each \(s \geq 1\).

Now, (29) and Footnote 5 imply immediately that:

\(^{18}\) A prominent example in this class is the Baum-Sweet sequence given by the substitution \(a \mapsto ab, b \mapsto cb, c \mapsto bd\) and \(d \mapsto dd\). It is not hard to see \(\{\theta^j(y)_j : y \in \{a, b, c, d\}\} = \{d\}\).
Proposition 4. If $c(\theta) = 1$ then the fixed point $u = \theta(u)$ is WRAP, in particular a limit of periodic sequences of period $\lambda^k$, $k \geq 1$, with respect to $d_W$.

From the point of view of Möbius orthogonality, dynamical systems $(X_x, S)$ given by WRAP sequences $x$ have already been studied in [4] and it is proved there that all continuous observables $(f(S^ny))$ (for $f \in C(X_x)$) are orthogonal to the Möbius function $\mu$. But a rapid look at the proof in [4] shows that the only property of $\mu$ used in it was the aperiodicity of $\mu$ (what is essential in the proof is that all points $y \in X_x$ are also WRAP).

Corollary 4 ([4]). If $x \in \mathbb{A}^\mathbb{N}$ is WRAP then for each bounded $u : \mathbb{N} \to \mathbb{C}$ which is aperiodic, we have

$$\lim_{N \to \infty} \frac{1}{N} \sum_{n \leq N} f(S^ny)u(n) = 0$$

for all $f \in C(X_x)$ and $y \in X_x$. In particular, the above assertion holds for substitutional dynamical systems with $c(\theta) = 1$. 19

Remark 6. Theorem 0.1 (i) also follows from [10] but because the relations between synchronized automata and substitutions with $c(\theta) = 1$ do not seem to be explained explicitly in literature, we gave a more general and direct argument.

3.2. Proof of Theorem 0.1 in the bijective and quasi-bijective case. A substitution $\theta : \mathbb{A} \to \mathbb{A}^\lambda$ is called quasi-bijective20 if for all $n \geq n_0$ and all $j \in \{0, \ldots, \lambda^n - 1\}$, we have

$$c(\theta) = |\{\theta^n(a)_j : a \in \mathbb{A}\}|.$$

If, additionally, $c(\theta) = |\mathbb{A}|$, then we speak about a bijective substitution (sometimes, such a substitution is also called invertible). The proof of Theorem 0.1 in the bijective case is provided in [13]. Moreover, the Rudin-Shapiro substitution is also treated in [13] and the method of proof can be extended to other quasi-bijective substitutions. The proof of Theorem 0.1 also covers the general case of quasi-bijective substitutions.

4. More ergodic theory prerequisites.

4.1. Essential centralizer. Assume that $T$ is an ergodic21 automorphism of $(Y, C, \nu)$. By the centralizer $C(T)$ of $T$ we mean the group of all invertible automorphisms $V \in \text{Aut}(Y, C, \nu)$ commuting with $T$. Clearly, $\{T^n : n \in \mathbb{Z}\}$ is a normal subgroup of $C(T)$ and the group $EC(T) := C(T)/\{T^n : n \in \mathbb{Z}\}$ is called the essential centralizer of $T$.

Lemma 4.1. Assume that $EC(T)$ is finite and $C(T) = C(T^p)$ for all sufficiently large $p \in \mathcal{P}$. Then, for all sufficiently large $p, q \in \mathcal{P}$, $p \neq q$, the automorphisms $T^p$ and $T^q$ are not isomorphic.

19The automatic sequences given by observables in case $c(\theta) = 1$ represent the so-called synchronized case in [37].
20The Rudin-Shapiro substitution (see Example 3) is a prominent example of a quasi-bijective substitution which is not bijective.
21We assume also that $T$ is aperiodic: for $\nu$-a.e. $y \in Y$, the map $m \mapsto T^my$ is one-to-one on $\mathbb{Z}$. 
Proof. Since EC(T) is finite, we have

\[ C(T) = \{ T^n \circ V_i : n \in \mathbb{Z}, i = 0, 1, \ldots, K \}, \]

where \( V_0 = \text{Id} \) and \( V_i \in C(T) \setminus \{ T^j : j \in \mathbb{Z} \} \) for \( i = 1, \ldots, K \). For \( i = 1, \ldots, K \), let \( m_i \geq 1 \) be the smallest natural number such that \( V_i^{m_i} \in \{ T^j : j \in \mathbb{Z} \} \) (of course \( m_i \) is precisely the order of the coset given by \( V_i \) in \( C(T) \setminus \{ T^j : j \in \mathbb{Z} \} \)). In what follows, we consider only prime numbers \( p, q \) which are dividing no \( m_i \) for \( i = 1, \ldots, K \). Suppose that for \( p \neq q \) (sufficiently large) we have an isomorphism of \( T^p \) and \( T^q \). As obviously \( T^q \) has a \( q \)-root, it follows that there is a root of degree \( q \) of \( T^p \), i.e. there exists \( W \in \text{Aut}(Y,\mathcal{C},\nu) \) such that \( W^q = T^p \).\(^{22}\) Now, \( W \in \text{C}(T^p) \), hence (by assumption) \( W \in C(T) \). It follows that, for some \( n \in \mathbb{Z} \) and \( 0 \leq i \leq K \), \( W = T^n \circ V_i \), whence \( W^q = (T^n \circ V_i)^q \) and we obtain

\[ V_i^q = T^{p-nq} \]

which, if \( i \neq 0 \), is impossible as \( q \) is coprime to \( m_i \) and if \( i = 0 \), \( T^p = T^{nq} \) which yields \( q/p \) (or \( p = 0 \) if \( n = 0 \)), a contradiction. \(\square\)

Remark 7. Notice that in the above proof, in fact we proved that \( T^p \) cannot be isomorphic to \( U^q \) with \( U \in \text{Aut}(Y,\mathcal{C},\nu) \), in other words, we proved that \( T^p \) cannot have a \( q \)-root.

4.1.1. Centralizer of \( h \)-discrete suspensions. We assume that \( T \in \text{Aut}(Y,\mathcal{C},\nu) \) and let \( \tilde{T} \) denote its \( h \)-discrete suspension, see (6) and (7). Note that whenever \( V \in C(T) \), the formula \( \nabla(y,j) := (Vy,j) \) for \( (y,j) \in \tilde{Y} \) defines an element of the centralizer of \( \tilde{T} \). In fact, we have the following:

Proposition 5 (Cor. 1.4 in [7]). If \( T \in \text{Aut}(Y,\mathcal{C},\nu) \) is ergodic then \( C(\tilde{T}) = \{ \nabla \circ \tilde{T}^m : V \in C(T), m \in \mathbb{Z} \} \).

It follows immediately from Proposition 5 that:

Corollary 5. If \( T \in \text{Aut}(Y,\mathcal{C},\nu) \) is ergodic and \( EC(T) \) is finite, also \( EC(\tilde{T}) \) is finite.

4.1.2. Essential centralizer of substitutions of constant length. The result below has been proved by Host and Parreau in [22] for pure substitutions of constant length. However, taking into account Corollary 5 and the fact that each substitutional system is an \( h \)-discrete suspension of its pure basis (which is also given by a substitution of constant length), see Section 2, we obtain the following\(^{23}\) result.

Theorem 4.2 ([22]). Let \( \theta \) be a primitive substitution of constant length. Then (remembering that \( (X_\theta,S) \) is uniquely ergodic), \( EC(S) \) is finite if and only if \( c(\theta) > h(\theta) \).

Remark 8. If \( c(\theta) = h(\theta) \) (in particular, if \( c(\theta) = 1 \)) then we are in the synchronizing case. Thus, the spectrum of the corresponding dynamical system is discrete. Therefore, the essential centralizer is uncountable. In fact, \( S^p \) is isomorphic to \( S^q \) for all sufficiently large \( p, q \in \mathcal{P} \).

\(^{22}\)Note that if \( \tau \) is an automorphism and \( \tau = \sigma^q \) for another automorphism \( \sigma \), then \( \sigma \in C(\tau) \) as \( \sigma \circ \tau = \sigma \circ \sigma^q = \sigma^q \circ \sigma = \tau \circ \sigma \).

\(^{23}\)For the particular instance of bijective substitutions, the result was proved to hold slightly earlier in [29].
Remark 9. Assume that $V : \mathbb{A} \to \mathbb{A}$ is a bijection “commuting” with $\theta$:
\[ \theta(V(a))_j = V(\theta(a))_j \] (30)
for each $j = 0, 1, \ldots, \lambda - 1$. Then, we claim that $V$ has a natural extension to a homeomorphism $V : X_\theta \to X_\theta$ so that $S \circ V = V \circ S$. Indeed, treat $V$ as a 1-code map on $\mathbb{A}^\mathbb{Z}$ ($V(y)[m] = V(y[m])$ for all $m \in \mathbb{Z}$), so obviously, it commutes with the shift. Clearly, we only need to show that
\[ y \in X_\theta \Rightarrow Vy \in X_\theta. \]
Now, for each $i \leq j$, $y[i, j] = \theta^k(a)[m, m + (j - i)]$ for some $a \in \mathbb{A}$, $k \geq 1$ and $m \geq 0$. It follows from (30) that
\[ (Vy)[i, j] = V\theta^k(a)[m, m + (j - i)] = \theta^k(Va)[m, m + (j - i)] \]
and the claim follows. Note that if $\theta$ is primitive, then $V : X_\theta \to X_\theta$ preserves $\mu_\theta$, so $V \in C(S)$.

4.2. Joinings of powers of finite extensions of odometers.

Lemma 4.3 ([13]). Assume that $T$ acting on $(Y, C, \nu)$ is ergodic (aperiodic) and has discrete spectrum. Let $\varphi : Y \to G$ be a cocycle with $G$ finite. Assume that $T_\varphi$ is ergodic. Moreover, assume that for $p \in \mathbb{P}$ large enough the corresponding group extension $(T_\varphi)_p$ is also ergodic.\(^{24}\) Then for $p \in \mathbb{P}$ large enough, we have $C(T_\varphi) = C((T_\varphi)_p)$.

Using Lemmas 4.1 and 4.3, we obtain the following.

Proposition 6. Assume that $T \in \text{Aut}(Y, C, \nu)$ is ergodic (aperiodic) and has discrete spectrum. Assume moreover that $T_\varphi$ is an ergodic $G$-extension ($G$ is a finite group) for which $EC(T_\varphi)$ is finite and $(T_\varphi)_p$ is ergodic for all $p \in \mathbb{P}$ sufficiently large. Then $(T_\varphi)_p$ and $(T_\varphi)_q$ are not isomorphic for all $p, q \in \mathbb{P}$ which are different and large enough.

In the case of odometers, we can prove more. Assume that $R$ is the odometer given by $\lambda$, that is, $R \in \text{Aut}(H_\lambda, m_{H_\lambda})$ is ergodic, has discrete spectrum and the eigenvalues are roots of unity of degree $\lambda^k$, $k \geq 1$. Consider $p, q \in \mathbb{P}$ relatively prime with $\lambda$. Then $R$ is isomorphic to $R^p$ (also to $R^q$), so $R^p$ is isomorphic to $R^q$. Recall also that the only ergodic joinings between $R^p$ and $R^q$ are given by graph joinings determined by isomorphisms between $R^p$ and $R^q$.

Proposition 7. Assume that $R$ is the $\lambda$-odometer. Let $\varphi : H_\lambda \to G$ be a cocycle with $G$ finite, so that $R_\varphi$ is ergodic, and $(R_\varphi)_p$ is also ergodic for $p \in \mathbb{P}$ large enough. Assume moreover that $EC(R_\varphi)_F$ is finite for each proper subgroup $F \subset G$.\(^{25}\) Then, for each pair of sufficiently large, distinct primes $p, q$, the only joinings\(^{26}\) between $(R_\varphi)_p$ and $(R_\varphi)_q$ that project onto ergodic joinings of $R^p$ and $R^q$ are relatively independent extensions of the projections.

\(^{24}\)This assumption requires small spectrum of $U_T$ in the sense of Section 1.4.2.

\(^{25}\)We recall that the factor, as an automorphism, is given by $R_{x,F}(x, gF) = (Rz, \varphi(x)gF)$. Note that, when applied to substitutions, this assumption requires $H_\lambda$ to represent the Kronecker factor, cf. Remark 8. However, if we replace $H_\lambda$ by $H_\lambda \times \mathbb{Z}/h\mathbb{Z}$, this assumption is satisfied and the proof of Proposition 7 remains the same.

\(^{26}\)Note that we drop the assumption of ergodicity, cf. the second part of Lemma 1.1.
Proof. In view of Section 1.3, \((R_\varphi)^p = R_\varphi^p\). The general theory of groups extensions (see e.g. \cite{30}, \cite{36}) tells us that if
\[
\rho \in J^c((R_\varphi)^p, (R_\varphi)^q) = J^c(R_\varphi^p, R_\varphi^q)
\]
then there exist \(F_1, F_2 \subset G\) normal subgroups of \(G\) such that an isomorphism \(W\) between \(R^p\) and \(R^q\) (we have assumed that \(\rho|_{H_\lambda \times H_\lambda}\) is the graph \((m_{H_\lambda})_W\) of \(W\)) lifts to an isomorphism \(\tilde{W}\) of the factors given by \(H_\lambda \times G/F_1\) and \(H_\lambda \times G/F_2\). But \(R_{\varphi(p)}^p \subset (R_{\varphi F_1}^p)^p\). Moreover, \(F_1, F_2\) depend on \(p, q\) but altogether we have only finitely many possibilities for \(F_1, F_2\). Assume that \(F_1\) is a proper subgroup of \(G\). We use our assumption and Lemma 4.1 together with Remark 7 to get a contradiction. It follows that \(\rho = (m_{H_\lambda})_W\) is the relatively independent extension of the graph joining \((m_{H_\lambda})_W\), see Footnote 12. We have proved that there is only one ergodic joining between \((R_\varphi)^p\) and \((R_\varphi)^q\) projecting on \((m_{H_\lambda})_W\), whence there is only one joining between \((R_\varphi)^p\) and \((R_\varphi)^q\) projecting on \((m_{H_\lambda})_W\) and the result follows. \(\square\)

Remark 10. By Mentzen’s theorem \cite{35} on (partly continuous spectrum) factors of substitutions of constant length, when \(R_\varphi\) is given by a substitution (primitive and non-synchronizing, i.e. \(h(\theta) < c(\theta)\)) then the natural factors \(R_{\varphi F}\) for \(F\) proper subgroup of \(G\) are also (up to measure-theoretic isomorphism) substitutions (cf. Footnote 25), so the assumptions of Proposition 7 will be satisfied.

4.3. Strategy of the proof of the main result. To prove Theorem 0.1, we first show that each substitutional system \((X_\Theta, S)\) is a topological factor of another substitutional system \((X_\Theta, S)\), where, from the measure-theoretic point of view, the system \((X_\Theta, S)\), which is uniquely ergodic, is isomorphic to \(R_\varphi\) satisfying the assumption of Proposition 7, see Remark 10. Moreover, we will show that the odometer of the original substitutional system is \(R\).

Let \((H_\lambda, R)\) denote the \(\lambda\)-odometer associated with \((X_\Theta, S)\). If the height of \(\hat{\Theta}\) is one, the \(\lambda\)-odometer is the maximal equicontinuous factor\(^{27}\) of \((X_\Theta, S)\). Hence, there is a continuous equivariant map \(\tau : X_\hat{\Theta} \rightarrow H_\lambda\). Moreover, we note the following observation which was already used in the proof of Lemma 1.2:

Lemma 4.4. Take different \(p, q \in \mathcal{P}\) sufficiently large. Then, for each \(y \in H_\lambda\), the point \((y, y)\) is generic for an ergodic \(R^p \times R^q\) invariant measure \(\kappa\), i.e. for each \(F \in C(H_\lambda \times H_\lambda)\), we have
\[
\frac{1}{N} \sum_{n \leq N} F(R^p y, R^q y) \rightarrow \int_{H_\lambda \times H_\lambda} F \, d\kappa.
\]

It follows that \((y, y)\) is generic for \(\kappa = (m_{H_\lambda})_W\), the graph joining given by an isomorphism \(W\) between \(R^p\) and \(R^q\).

Take any \(x \in X_\hat{\Theta}\). We study now the sequence \((p \neq q\) sufficiently large)
\[
\frac{1}{N} \sum_{n \leq N} \delta(S^p S^q)^n(x, x), \quad N \geq 1.
\]
Any of its limit points yields a joining \(\rho \in J((R_\varphi)^p, (R_\varphi)^q)\). Now, \(\rho|_{H_\lambda \times H_\lambda}\) is precisely obtained as the limit of
\[
\frac{1}{N_k} \sum_{n \leq N_k} \delta((R^p R^q)^n(\pi(x), \pi(x))\),
\]
\(^{27}\)If \(h(\hat{\Theta}) > 1\) then we must replace \(H_\lambda\) with \(H_\lambda \times \mathbb{Z}/h\mathbb{Z}\).
for a relevant subsequence \((N_k)\). But we have already noticed that any such limit must be an **ergodic** joining of \(R^p\) and \(R^q\), hence, it is a graph joining (cf. (5)). Now, we use our results to obtain that the limit of \(\frac{1}{N_k} \sum_{n \leq N_k} \delta((S^p \times S^q)^n(x,x))\) also exists and it is the relatively independent extension of the underlying graph joining between \(R^p\) and \(R^q\). We have proved the following.

**Proposition 8.** If different \(p,q \in \mathcal{P}\) are large enough, then the set
\[
\{ \rho \in J(S^p,S^q) : \rho = \lim_{k \to \infty} \frac{1}{N_k} \sum_{n \leq N_k} \delta((S^p \times S^q)^n(x,x)) \text{ for some } N_k \to \infty \}
\]
is contained in the set of relatively independent extensions of the graphs of isomorphisms between \(R^p\) and \(R^q\).

Assume now that
\[
F \in C(\hat{X}_\Theta) \text{ and } F \perp L^2(\pi^{-1}(\mathcal{B}(H_\lambda))).
\]
Again, fix \(x \in \hat{X}_\Theta\). We have
\[
\frac{1}{N_k} \sum_{n \leq N_k} F(S^p^nx)F(S^q^nx) \to \int_{\hat{X}_\Theta \times \hat{X}_\Theta} F \otimes \bar{F} \, d\rho = \int_{H_\lambda \times H_\lambda} \mathbb{E}(F \otimes \bar{F}|H_\lambda \times H_\lambda) \, d\rho|_{H_\lambda \times H_\lambda} = \int_{H_\lambda} \mathbb{E}(F|H_\lambda) \cdot \mathbb{E}(\bar{F}|H_\lambda) \circ W \, dm_{H_\lambda} = 0
\]
by the definition of the relative product.

The only thing which is missing now is to be sure that we have sufficiently many functions \(F\) satisfying (31). In fact, we aim at showing that \((\hat{X}_\Theta,S)\) has a topological factor \((X_{\Theta'},S)\), which measure-theoretically is equal to \((H_\lambda,R)\) and for each \(F \in C(\hat{X}_\Theta)\), we have
\[
F = F' + F'',
\]
where \(F' \in C(\hat{X}_{\Theta'})\) (with some abuse of notation), and \(F'' \perp L^2(\pi^{-1}(\mathcal{B}(H_\lambda)))\) (of course, \(F''\) is also continuous). This will allow us to conclude the proof of Theorem 0.1 using Theorem 1.3 for \(F''\) (as here we will deal with relatively independent joinings) and dealing separately with \(F'\).

When \(\theta\) is bijective, the existence of a “good” \(\hat{\Theta}\) is known, see Section 3.2 below: in \((\hat{X}_\Theta,S)\) we have many “good” continuous function, as the Kronecker factor of \((\hat{X}_\Theta,S)\) has a “topological” realization (see [13]). However, in the general case such an approach seems to be unknown, and in Section 5, we will show a new general construction of an extension of a substitutional system in which we will see sufficiently many continuous functions satisfying (31).

5. Substitutions joinings of substitutional systems. Assume that we have two substitutions \(\theta : \mathcal{A} \to \mathcal{A}^\lambda\) and \(\zeta : \mathcal{B} \to \mathcal{B}^\lambda\). Assume that
\[
\mathcal{A} \subset \mathcal{A} \times \mathcal{B}, \; p_\mathcal{A}(\mathcal{A}) = \mathcal{A}, \; p_\mathcal{B}(\mathcal{A}) = \mathcal{B},
\]
where \(p_\mathcal{A},\ p_\mathcal{B}\) stand for the projections on \(\mathcal{A}\) and \(\mathcal{B}\), respectively. Moreover, we assume that
\[
(a,b) \in \mathcal{A} \Rightarrow (\theta(a)_j, \zeta(b)_j) \in \mathcal{A}
\]
28In fact, by Lemma 4.4, it follows that each point \((x,x)\) is generic for one of such joinings.
for each \( j = 0, \ldots, \lambda - 1 \). Then it is easy to see that the formula
\[
\Sigma(a, b) = (\theta(a)_0, \zeta(a)_0)(\theta(a)_1, \zeta(a)_1) \ldots (\theta(a)_{\lambda-1}, \zeta(a)_{\lambda-1})
\]
defines a substitution \( \Sigma : A \rightarrow A^\lambda \) of length \( \lambda \). We can also use the notation:
\[
\Sigma = \theta \vee \zeta.
\]

**Remark 11.** We note that in general the above \( \Sigma \) need not be primitive, that is, it does not necessarily satisfy (15). Indeed, consider for example \( \zeta = \theta \) and then take for \( A \) the product set \( A \times A \). On the other hand, the “diagonal” \( \mathcal{A} := \{(a, a) : a \in A\} \) yields a primitive substitution (clearly isomorphic to \( \theta \)).

In order to see a less trivial primitive example, consider \( A = \{0, 1, 2\} \) and \( \zeta = \theta \), where (see the last example in Section 2 of [8])

\[
0 \mapsto 010, \ 1 \mapsto 102, \ 2 \mapsto 201
\]

with \( \mathcal{A} = \{(0, 1), (1, 0), (0, 2), (2, 0)\} \), which gives a primitive substitution different from the “diagonal” one.\(^{29}\) If we slightly change the definition of \( \theta \), namely:

\[
0 \mapsto 010, \ 1 \mapsto 201, \ 2 \mapsto 102
\]

(which is still primitive with \( \epsilon = h = 2 \)) then \( \Sigma \) is primitive but (18) is not satisfied.

**Remark 12.** Note that \( X_{\Sigma} \) is a topological joining of \( X_\theta \) and \( X_\zeta \). Indeed, up to a natural rearrangement of coordinates, \( X_{\theta \vee \zeta} \subset X_\theta \times X_\zeta \). Then, for every \( (x, y) \in X_{\Sigma} \), the orbits of \( x \) and \( y \) are dense in \( X_\theta \) and \( X_\zeta \), respectively.\(^{30}\) Now, the image of the natural projection \( (x, y) \mapsto x \) is contained in \( X_\theta \) and if \( \theta(u) = u \), \( \zeta(v) = v \) then \( (\theta \vee \zeta)(u, v) = (u, v) \) (after a rearrangement of coordinates). So \( u \) is in the image of \( p_{X_\theta}(x, y) = x \) and therefore

\[
p_{X_\theta} : X_{\theta \vee \zeta} \rightarrow X_\theta, \ p_{X_\zeta}(X_{\theta \vee \zeta}) = X_\zeta.
\]

Since \( p_{X_\theta} \) is continuous and equivariant, it settles a topological factor map between the relevant substitutional systems.

**Definition 5.1.** We call \( \Sigma \) a *substitution joining* of \( \theta \) and \( \zeta \) if \( \Sigma \) satisfies (15) and (18) (note that (25) and (26) are satisfied automatically as \( \theta, \zeta \) are substitutions, so they satisfy these two properties).

5.2. **Joining with the synchronizing part.** In general, when dealing with the dynamical system given by a substitution of constant length, we would like to see its Kronecker factor as a topological factor realized “in the same category of objects”, that is, realized by another substitution. This is not always possible, even in the class of bijective substitutions, see Herning’s example [21]. For the purpose of orthogonality with an arithmetic function \( u \), we need however only an extension of the original substitution which is given by another substitution (of the same length) and require that in the extended system we have a “good” realization of the Kronecker factor. This is done by a joining of \( \theta \) with its synchronizing part.

\(^{29}\)It is not hard to see that in this example, \( \theta \) is primitive and \( c(\theta) = 2 \). If by \( u \) we denote the fixed point of \( \theta \) obtained by the iterations of 0, then \( S_0 = 2 \mathbb{N} \), so \( g_0 = 2 \) and hence \( h(\theta) = 2 = c(\theta) \). This means that the dynamical system \( (X_\theta, \mu_\theta, S) \) has discrete spectrum and hence since \( \Sigma \) is also primitive, as measure-theoretic dynamical systems, \( (X_\theta, \mu_\theta, S) \) and \( (X_{\Sigma}, \mu_{\Sigma}, S) \) are isomorphic, cf. (5).

\(^{30}\)This follows by the minimality of \( (X_\theta, S) \) and \( (X_\zeta, S) \), respectively.
5.2.1. Synchronizing part of \( \theta \). Given a substitution \( \theta : \mathbb{A} \to \mathbb{A}^\lambda \), set
\[
\mathcal{X}(\theta) = \mathcal{X} := \{ M \subset \mathbb{A} : M \text{ realizes the column number of } \theta \},
\]
i.e. \( M \in \mathcal{X}(\theta) \) iff there exist \( k_M \geq 1 \), \( j_M \in \{0, \ldots, \lambda^{k_M} - 1\} \) such that \( M = \{ \theta^{k_M}(a)_{j_M} : a \in \mathbb{A} \} \) with \( |M| = c(\theta) \). We can assume without loss of generality that \( k_M = k_M' \) for all \( M, M' \in \mathcal{X} \). Note that when \( M \in \mathcal{X} \) then for each \( j = 0, \ldots, \lambda - 1 \), we have
\[
|\theta(M)_j| = c(\theta),
\]
where by \( \theta(M)_j \) we denote the set \( \{ \theta(a)_j : a \in M \} \). It follows that the formula
\[
\tilde{\theta}(M) = \theta(M)_0 \theta(M)_1 \ldots \theta(M)_{\lambda - 1}
\]
defines a substitution \( \tilde{\theta} : \mathcal{X} \to \mathcal{X}^\lambda \).

**Proposition 9.** The substitution \( \tilde{\theta} \) has the following basic properties:

(i) \( c(\tilde{\theta}) = 1 \).

(ii) \( \tilde{\theta} \) is primitive.

(iii) \( h(\tilde{\theta}) = 1 \).

**Proof.** For (i), let \( M \in \mathcal{X} \) and suppose that \( \theta^{k_M}(\mathbb{A})_{j_M} = M \). Then, for each \( M' \in \mathcal{X} \), we have \( \theta^{k_M}(M')_{j_M} = M \). The validity of (ii) follows by the same argument. Finally, (iii) follows from Lemma 2.1. \( \square \)

**Definition 5.2.** We call \( \tilde{\theta} \) the synchronizing part of \( \theta \).

Also, note that
\[
\bigcup_{M \in \mathcal{X}} M = \mathbb{A}. \tag{38}
\]
Indeed, fix \( M \in \mathcal{X} \) and let \( a \in \mathbb{A} \). Take any \( x \in M \) and (by primitivity) choose \( k \geq 1 \) so that \( \theta^k(x)_j = a \) for some \( 0 \leq j < \lambda^k \). Then \( a \in \tilde{\theta}(M)_j \) and \( \tilde{\theta}(M)_j \in \mathcal{X} \).

**Remark 13.** If the union in (38) is additionally a partition, then we obtain an equivalence relation on \( \mathbb{A} \) which is \( \theta \)-consistent (cf. (24)) and the dynamical system \( (X_{\tilde{\theta}}, S) \) is a topological factor of \( (X_{\theta}, S) \). However, in general, there is no reason for \( (X_{\tilde{\theta}}, S) \) to be a topological factor of \( (X_{\theta}, S) \).

Note that, in general, the union in (38) is not a partition as the following example shows.

**Example 1.** Consider the following substitution \( \theta \):
\[
\begin{align*}
\theta(a) &= ab \\
\theta(b) &= ca \\
\theta(c) &= ba.
\end{align*}
\]  
(39)

By looking at \( \theta^3 \), we see that \( \theta \) is primitive. Moreover, straightforward computations give that \( c(\theta) = 2 \) and \( h(\theta) = 1 \). In fact, \( \mathcal{X} = \{ \{a, b\}, \{a, c\} \} \). Therefore, the union in (38) is not a partition.

Note that the first column of (39) and all its iterates yield the set \( \{a, b, c\} \), so \( \theta \) is not quasi-bijective.

Furthermore, by replacing \( \{a, b\} \) by 0 and \( \{a, c\} \) by 1, we compute \( \tilde{\theta} \):
\[
\begin{align*}
\tilde{\theta}(0) &= 10 \\
\tilde{\theta}(1) &= 00.
\end{align*}
\]
We also have the following result:

**Proposition 10.** Let \( \theta : \mathcal{A} \to \mathcal{A}^{\lambda} \) be a substitution. Then:

\[
\theta \text{ is bijective if and only if } \tilde{\theta} \text{ is trivial, i.e. } |\mathcal{X}(\theta)| = 1. 
\] (40)

\[
\theta \text{ is quasi-bijective if and only if } X_{\tilde{\theta}} \text{ is finite, i.e. a fixed point } \tilde{u} \text{ of } \theta \text{ is periodic.} \] (41)

**Proof.** First of all, if \( \theta \) is bijective then \( \mathcal{X}(\theta) = \{ \mathcal{A} \} \), and (40) follows directly from (38). Let us pass to the proof of (41).

\( \Rightarrow \): We assume that \( \theta \) is quasi-bijective, i.e. for all \( n \geq n_0 \) and all \( j < \lambda^n \), we have

\[
|\theta^n(\mathcal{A})| = c(\theta). 
\]

We define \( M_j := \theta^{n_0}(\mathcal{A}) \) and it follows immediately that \( \tilde{\theta}^{n_0}(M) = M_j \) for all \( M \in \mathcal{X} \). Let \( \tilde{u} \) be a fixed point of \( \tilde{\theta} \). Now, for \( i \geq 0 \) and \( 0 \leq j < \lambda^{n_0} \), we find that (use (11) letting \( \ell \to \infty \)),

\[
\tilde{u}(j + i \lambda^{n_0}) = \tilde{\theta}^{n_0}(\tilde{u}(j + i \lambda^{n_0})) = \tilde{\theta}^{n_0}(\tilde{u}(j)), \quad \text{which shows that } \tilde{u} \text{ is periodic (in fact, } \lambda^{n_0} \text{ is a period of } \tilde{u}).
\]

\( \Leftarrow \): Let \( \tilde{u} \) be a periodic fixed point of \( \tilde{\theta} \) with period \( p \). First, we claim that \( \tilde{u} \) is also periodic with period \( \lambda^{n_0} \) for some \( n_0 \geq 0 \) (if so, \( \tilde{u} \) is periodic of period \( \lambda^n \) for all \( n \geq n_0 \)). Indeed, by Proposition 4 (and Proposition 9), it follows that, in the Weyl pseudo-metric \( d_W \), we can approximate \( \tilde{u} \) by periodic sequences that have period \( \lambda^n \). Thus, there exists \( n_0 \) large enough and a \( \lambda^{n_0} \)-periodic sequence \( v \) such that

\[
d_W(\tilde{u}, v) < \frac{1}{2p}.
\]

By basic properties of \( d_W \), we obtain

\[
d_W(S^{\lambda^{n_0}}(\tilde{u}), S^{\lambda^{n_0}}(v)) \leq d_W(\tilde{u}, v) < \frac{1}{2p}.
\]

Hence, remembering that \( S^{\lambda^{n_0}}(v) = v \), the triangle inequality implies

\[
d_W(\tilde{u}, S^{\lambda^{n_0}}(\tilde{u})) < \frac{1}{p}.
\]

Since \( \tilde{u} \) and \( S^{\lambda^{n_0}}(\tilde{u}) \) are both \( p \)-periodic sequences with distance less than \( 1/p \), they must coincide and the claim follows.

As, by Proposition 9, \( \tilde{\theta} \) is primitive, there exists \( n \geq 0 \) such that for each \( M \in \mathcal{X} \), we can find \( j_1 < \lambda^n \) such that \( \tilde{\theta}^{n_0}(\tilde{u}[0])_{j_1} = M \). Moreover, for all \( j_0 < \lambda^{n_0} \) and all \( j < \lambda^n \) (use (11) and \( j_0 + j \lambda^{n_0} < \lambda^{n+n_0} \)), we have

\[
\tilde{u}(j_0) = \tilde{u}(j + j \lambda^{n_0}) = \tilde{\theta}^{n+n_0}(\tilde{u})(j_0 + j \lambda^{n_0}) = \tilde{\theta}^{n+n_0}(\tilde{u}[0])(j_0 + j \lambda^{n_0}) = \tilde{\theta}^{n}(\tilde{u}[0])_{j_0 + j \lambda^{n_0}}.
\]

Letting \( j = j_1 \), this shows that \( \tilde{\theta}^{n_0}(M)_{j_0} = \tilde{u}(j_0) \) for all \( M \in \mathcal{X}, j_0 < \lambda^{n_0} \). As \( \mathcal{X} \) covers \( \mathcal{A} \) in view of (38), this shows that for all \( j_0 < \lambda^{n_0} \), we have \( \tilde{\theta}^{n_0}(\mathcal{A})_{j_0} = \tilde{u}(j_0) \) and, therefore,

\[
|\tilde{\theta}^{n_0}(\mathcal{A})_{j_0}| = c(\tilde{\theta}),
\]

which concludes the proof. \( \square \)
5.2.2. Joining $\theta$ with its synchronizing part. Define

$$\mathcal{X} := \{(a, M) : a \in M, M \in \mathcal{X}\} \subset \mathcal{A} \times \mathcal{X}.$$ 

Now, in view of (38), the first compatibility condition (33) immediately follows. If $(a, M) \in \mathcal{X}$ then $a \in M$, so for each $j = 0, \ldots, \lambda - 1$, we have $\theta(a)_j \in \theta(M)_j$, and in view of (37), $\theta(a)_j \in \tilde{\theta}(M)_j$. Therefore, the second compatibility condition (34) also follows.

Set

$$\Theta := \theta \vee \tilde{\theta} : \mathcal{X} \to \mathcal{X}^\lambda,$$

$$\Theta(a, M) = (\theta(a)_0, \tilde{\theta}(M)_0)(\theta(a)_1, \tilde{\theta}(M)_1) \ldots (\theta(a)_{\lambda-1}, \tilde{\theta}(M)_{\lambda-1}).$$

Remark 15. Although, by primitivity, we can assume that there are $a \in \mathcal{A}$ and $M \in \mathcal{X}$ such that $\theta(a)_0 = a$ and $\tilde{\theta}(M)_0 = M$, it may happen that (18) is not satisfied for $\Theta$. However, as is the case for any substitution of constant length, there exists some $k \in \mathbb{N}$ such that $\Theta^k = \theta^k \vee \tilde{\theta}^k$ satisfies (18). Since the following proposition shows that $\Theta$ satisfies (15), which assures us that $X_\Theta = X_{\theta^k}$, we can assume without loss of generality that $\Theta$ satisfies (18).

Proposition 11. The substitution $\Theta$ is primitive and $h(\Theta) = h(\theta)$.

Proof. We first show that $\Theta$ is primitive. Let $(a_1, M_1), (a_2, M_2) \in \mathcal{X}$. By the definition of $\tilde{\theta}$, it follows that there are some $k \in \mathbb{N}, j < \lambda^k$ such that $\theta^k(a)_j = M_2$. Hence, $\tilde{\theta}^k(M)_j = M_2$ for every $M \in \mathcal{X}$. This ensures that there exists $a \in \mathcal{A}$ such that $\theta^k(a)_j = a_2$. Using the primitivity of $\theta$, choose $k' \in \mathbb{N}$ and $j' < \lambda^{k'}$ so that $\theta^{k'}(a_1)_j = a$. We define $M := \tilde{\theta}^{k'}(M_1)_j$ and using (11), we obtain

$$\Theta^{k+k'}((a_1, M_1))_{j' \lambda^k+j} = \Theta^k(\Theta^{k'}((a_1, M_1))_{j'}),$$

$$= \Theta^k((a, M))_j = (a_2, M_2)$$

which shows that $\Theta$ is primitive.

Let us denote by $u$ a fixed point of $\theta$, by $v$ a fixed point of $\Theta$, where $u[0] = a_0, v[0] = (a_0, M_0)$. According to (19) and the subsequent discussion, if we set

$$S_k(\theta) = \{r \geq 1 : u[k+r] = u[k]\},$$

then

$$h(\theta) = \max\{m \geq 1 : (m, \lambda) = 1, m \mid \gcd S_k(\theta)\},$$

for all $k \geq 0$.

We recall also that in the first part of the proof we showed that for $(a_1, M_1) = (a_2, M_2) = (a_0, M_0)$ (and using (11)), we obtain that

$$\Theta^{k+k'}((a_0, M_0))_{j' \lambda^k+j} = (a_0, M_0)$$

for any $k' \in \mathbb{N}, j' < \lambda^{k'}$ such that $\theta^{k'}(a_0)_j = a$. Since, $u[0, \lambda k' - 1] = \theta^{k'}(a_0)$, the above means that there exists $k \in \mathbb{N}, j < \lambda^k$ such that for all $j'$ with $u[j'] = a$, we have $u[j' \lambda^k + j] = (a_0, M_0)$. We denote by $j'_0$ the smallest such $j'$. Thus, we find

$$\lambda^k \cdot S_{j'_0}(\theta) \subset S_{j'_0 \lambda^k+j}(\Theta).$$

---

31 We always have $\Theta^k = (\theta \vee \tilde{\theta})^k = \theta^k \vee \tilde{\theta}^k = \theta^k \vee \tilde{\theta}^k$. 

Remark 14. It follows from Propositions 9 and 10 that once $\theta$ is not quasi-bijective, $(X_{\theta}, S)$ is a “realization” of $(H_{\lambda}, R)$. 

This implies that \( \gcd(S^j_0, \lambda^j + j)(\Theta) = \lambda^k \gcd(S^j_0(\theta)) \), and, therefore, \( h(\Theta) \mid h(\theta) \). Moreover, by Lemma 5.5 (below), we obtain that \( h(\theta) \mid h(\Theta) \), which completes the proof. \( \square \)

**Remark 16.** Note that from the measure-theoretic point of view, the substitutional system \( (X_{\theta \ast \theta}, S) \) (which is an extension of \( (X_\theta, \mu, S) \)) is isomorphic to \( (X_\theta, \mu, S) \). Indeed, by the unique ergodicity of \( (X_{\theta \ast \theta}, S) \), it represents an ergodic joining of \( (X_\theta, \mu, S) \) and \( (X_{\tilde{\theta}}, \mu_{\tilde{\theta}}, S) \) and the latter system has discrete spectrum contained in the discrete spectrum of \( (X_\theta, \mu, S) \). By ergodicity, the only such joining is graphic. We will elaborate upon this shortly.

Hence, using Proposition 11, the above remark and the ergodic interpretation of the column number, cf. Proposition 3, we obtain the following (which is also an immediate consequence of Lemma 5.5\( ^{33} \)):

**Corollary 6.** For a substitution \( \theta : \mathbb{A} \to \mathbb{A}^\lambda \), we have \( c(\Theta) = c(\theta) \).

**Remark 17.** Note also that \( (X_{\tilde{\theta}}, S) \) is a topological factor of \( (X_{\theta \ast \theta}, S) \) via the map:

\[
(x_n, M_n)_{n \in \mathbb{Z}} \mapsto (M_n)_{n \in \mathbb{Z}}.
\]

Assume that \( M \in \mathcal{X} \). As, for each \( j = 0, \ldots, \lambda - 1 \), the set \( \{ \theta(a)_j : a \in M \} \) has \( c(\theta) \) elements, we obtain the following:

**Lemma 5.3.** The substitution \( \theta \vee \tilde{\theta} \) has the following “relative invertibility” property: for each \( M \in \mathcal{X} \), \( (\theta \vee \tilde{\theta})(\cdot, M) \) is a bijection from \( M \) to \( \tilde{\theta}(M)_j \) for each \( j = 0, \ldots, \lambda - 1 \).

**Remark 18.** In this way, using joinings, we may explain one of the strategies in [37] which consists in representing each automatic sequence as a “combination” of the synchronized and relatively invertible parts.

### 5.2.3. Description of the subshift \( (X_{\theta \ast \theta}, S) \)

Each point of the space \( X_{\theta \ast \theta} \) is of the form \( (x_n, M_n)_{n \in \mathbb{Z}} \), where \( \overline{M} := (M_n) \in X_{\tilde{\theta}} \) with \( M_n \in \mathcal{X} \), and \( x = (x_n) \in X_\theta \), so for each \( n \in \mathbb{Z}, x_n \in M_n \subseteq \mathbb{A} \). More than that, like every substitutional system, such a point must have its (unique) \( \lambda \)-skeleton structure \( (j_i)_{i \geq 1} \) which we elaborate upon presently. First of all, by the definition of \( \theta \vee \tilde{\theta}, (j_i) \) must be the \( \lambda \)-skeleton structure of both \( \overline{M} \) and \( x \) (we make crucial use here of the fact that the skeleton structure is unique; for that, we need that the substitution \( \theta \) is non-trivial, in other words the argument makes no sense if \( \theta \) is quasi-bijective). That is, for each \( t \geq 1 \), we have

\[
\overline{M}[−j_t + s\lambda^t, −j_t + (s + 1)\lambda^t - 1] = \tilde{\theta}^t(R_s)
\]

for each \( s \in \mathbb{Z} \) and some \( R_s \in \mathcal{X} \). Fix \( t \geq 1 \) and consider \( s = 0 \). We have

\[
\overline{M}[−j_t, −j_t + \lambda^t - 1] = \tilde{\theta}^t(R_0)
\]

and \( \mathcal{X} \ni R_0 = \{r_0, \ldots, r_{c-1}\} \), where \( c = c(\theta) \) and \( r_j \in \mathbb{A} \). Hence

\[
\tilde{\theta}^t(R_0) = \begin{bmatrix}
\theta^t(r_0)_{j_0} & \ldots & \theta^t(r_0)_{j_t} & \ldots & \theta^t(r_0)_{\lambda^t-1} \\
\theta^t(r_1)_{j_0} & \ldots & \theta^t(r_1)_{j_t} & \ldots & \theta^t(r_1)_{\lambda^t-1} \\
\ldots & \ldots & \ldots & \ldots & \ldots \\
\theta^t(r_{c-1})_{j_0} & \ldots & \theta^t(r_{c-1})_{j_t} & \ldots & \theta^t(r_{c-1})_{\lambda^t-1}
\end{bmatrix}
\]

\( ^{32} \)Alternatively, \( h(\Theta) \mid h(\Theta) \) follows from the fact that \( (X_\theta, S) \) is a topological factor of \( (X_{\Theta}, S) \).

\( ^{33} \)Indeed, as \( c(\theta) = 1 \), Lemma 5.5 shows that \( c(\theta) \leq c(\Theta) \leq c(\Theta) \).
where the columns “represent” sets in $X$. We look at $\tilde{\theta}^\ell (R_0)_{ji}$, the $j_i$-th element of $\tilde{\theta}^\ell (R_0)$ which is “represented” by the $j_i$-th column of the matrix above and is equal to $M_0$ (the zero coordinate of $\tilde{M}$). Now, $x_0 \in M_0$, hence $x_0 = \theta^\ell (r_\ell)_{ji}$ for a unique $0 \leq \ell \leq c - 1$. Moreover, $x[-j_\ell,-j_\ell + \lambda^\ell - 1] = \theta^\ell (r_\ell)$. Hence, unless $j_\ell = \lambda^\ell - 1$, $x_0$ determines $x_1$.

Note that, we can reverse this reasoning in the case:

$$\min (j_t, \lambda^t - j_t - 1) \to \infty. \quad (43)$$

Indeed, in this case, given $\tilde{M}$, we can choose $x_0 \in M_0$ arbitrarily, and then successively fill in the “first” coordinate by placing there the $\ell$-th row in the matrix $(42)$, where $x_0 = \theta^\ell (r_\ell)_{ji}$. This shows that over all $\tilde{M} \in X_{\tilde{\theta}}$ we have $c$ points $(x, \tilde{M}) \in X_{\tilde{\theta}, \tilde{\theta}}$ whenever the $\lambda^t$-skeleton of $\tilde{M}$ satisfies $(43)$. What remains are points for which

either $j_{t+1} = j_t$ or $\lambda^{t+1} - j_{t+1} - 1 = \lambda^t - j_t - 1$ for $t \geq T$.

When projected down to the maximal equicontinuous factor, this condition defines a countable set, in particular, of (Haar) measure zero. For the unique measure $\mu_{\tilde{\theta}}$ we have hence a.e. a $c$-point extension, while for $\overline{M}$ which do not satisfy $(43)$, we need (at most) two coordinates for $x$ to determine all the others, so the fibers have at most $c^2$ elements.

**Example 2.** Let $A = \{a, b, c, d\}$. We consider the following substitution $\theta : A \to A^4$:

$$\theta(a) = adda, \quad \theta(b) = bccb$$

$$\theta(c) = abbc, \quad \theta(d) = badd.$$

It is not hard to see that $\theta$ is pure and $c(\theta) = 2, X = \{M_0, M_1\}$, where $M_0 = \{a, b\}$ and $M_1 = \{c, d\}$. Moreover,

$$\tilde{\theta}(M_0) = M_0 M_1 M_1 M_0$$

$$\tilde{\theta}(M_1) = M_0 M_0 M_0 M_1.$$

Let us now consider the block $\tilde{M}^{(0)} = M_0 M_0 \in X^2$, which we write as $\tilde{M}_1^{(0)} M_0^{(0)}$. Acting on it by $\tilde{\theta}$, we obtain

$$\tilde{M}_1^{(1)} = \tilde{\theta}(M_0) \tilde{\theta}(M_0) = \tilde{M}_1^{(1)} \tilde{M}_0^{(1)} \tilde{M}_0^{(1)} \tilde{M}_3^{(1)}.$$

By iterating this procedure and passing to the limit, we obtain a two-sided sequence which is a fixed point for $\tilde{\theta}$ and which we denote by $\overline{M} = \tilde{\theta}\infty (M_0), \tilde{\theta}\infty (M_0)$, the “dot” indicating the zero position of the sequence. A similar procedure can be made on the $\theta$-side, starting with $a.a, a.b, b.a$ and $b.b$. It is not hard to see that, up to a natural rearrangement of coordinates, the following four points $(\theta\infty (a), \theta\infty (a), M), (\theta\infty (a), \theta\infty (b), M), (\theta\infty (b), \theta\infty (a), M), (\theta\infty (b), \theta\infty (b), M)$ are members of $X_{\theta, \theta}$. It follows that the fiber over $\overline{M}$ has four points.

5.3. **Toward a skew-product measure-theoretic representation - making the relative invertibility clearer.** In this part we want to rename the alphabet of $\Theta$ to get a new substitution, which makes the invertible part easier to handle. Namely, our new alphabet will be the set $\{0, \ldots, c - 1\} \times X$, where $c = c(\theta)$. The only thing we need is to give a “good” identification of $\{0, \ldots, c - 1\} \times M$ with $\{(a, M) : a \in M\}$ (for $M \in X$). We start by giving a classical example.
Example 3 (Rudin-Shapiro sequence). We consider the Rudin-Shapiro substitution \( \theta \) defined by \( \mathcal{A} = \{a, b, c, d\} \) and
\[
\begin{align*}
a &\mapsto ab, \\
d &\mapsto dc \\
b &\mapsto ac, \\
c &\mapsto db.
\end{align*}
\]
We find that \( X = \{\{a, d\}, \{b, c\}\} \) and consequently for \( \Theta \), we have:
\[
\begin{align*}
(a, \{a, d\}) &\mapsto (a, \{a, d\})(b, \{b, c\}), \\
(d, \{a, d\}) &\mapsto (d, \{a, d\})(c, \{b, c\}) \\
(b, \{b, c\}) &\mapsto (a, \{a, d\})(c, \{b, c\}), \\
(c, \{b, c\}) &\mapsto (d, \{a, d\})(b, \{b, c\}).
\end{align*}
\]
Let us now fix a partial ordering on \( \mathcal{A} \) that is complete on every \( M \in X \), e.g., \( a < b, c < d \). Thus, we can identify \( (a, \{a, d\}) \cong (0, \{a, d\}) \) and \( (d, \{a, d\}) \cong (1, \{a, d\}) \) as \( a \) is the smallest element of \( \{a, d\} \) and \( d \) is the second smallest element. This gives the new substitution \( \tilde{\Theta} : \{0, \ldots, c-1\} \times X \rightarrow (\{0, \ldots, c-1\} \times X)^\lambda \); which in our example is given by:
\[
\begin{align*}
(0, \{a, d\}) &\mapsto (0, \{a, d\})(0, \{b, c\}), \\
(1, \{a, d\}) &\mapsto (1, \{a, d\})(1, \{b, c\}) \\
(0, \{b, c\}) &\mapsto (0, \{a, d\})(1, \{b, c\}), \\
(1, \{b, c\}) &\mapsto (1, \{a, d\})(0, \{b, c\}).
\end{align*}
\]
Note that \( \tilde{\theta} \) is very simple in this example, as identifying \( \{a, d\} \) with \( M_0 \) and \( \{b, c\} \) with \( M_1 \), gives \( \tilde{\theta}(M_0) = M_0 M_1 \) and \( \tilde{\theta}(M_1) = M_0 M_1 \), so we obtain a periodic sequence. We obtain the classical Rudin-Shapiro sequence by considering the fixpoint of \( \Theta \) and then applying the projection \( (i, M) \mapsto i \). (We emphasize that the classical Rudin-Shapiro sequence itself is not a fixpoint of a substitution.)

We will be dealing with \( \theta \) which is not quasi-bijective (which, via Propositions 9 and 10, guarantees that \( (X_{\tilde{\theta}}, S) \) “captures” the whole discrete spectrum of \( (H_X, m_{H_X}, R) \)). We repeat the above construction: fix a partial order on \( \mathcal{A} \) that is complete on every \( M \in X \) and identify \( (a, M) \) with \( (j, M) \) if the ordering of the elements in \( M \) according to the ordering on \( \mathcal{A} \) restricted to \( M \) yields \( a \) as the \( (j+1) \)-th smallest element and rewrite instructions using the alphabet. This yields the substitution
\[
\tilde{\Theta} : \{0, \ldots, c-1\} \times X \rightarrow ((\{0, \ldots, c-1\} \times X)^\lambda.
\]
The second coordinate is still \( \tilde{\theta} \) which corresponds to the synchronizing part and can be defined independently of the first coordinate. The first coordinate (which now depends on the second coordinate), gives the “invertible part”.

Lemma 5.4. \( \tilde{\Theta} \) is primitive, \( h(\tilde{\Theta}) = h(\theta) \) and \( c(\tilde{\Theta}) = c(\theta) \).

Proof. As \( \tilde{\Theta} \) is obtained by renaming the alphabet of \( \Theta \), this follows immediately from Proposition 11 and Corollary 6. \( \square \)

Remark 19. A point in the space \( X_{\tilde{\Theta}} \) is of the form \((y_n, M_n)_{n \in \mathbb{N}} \) with \( y_n \in \{0, \ldots, c-1\} \) and there is an equivariant map between \( X_{\tilde{\Theta}} \) and \( X_{\theta \circ \tilde{\theta}} \) given by
\[
(y_n, M_n) \mapsto (z_n, M_n), \tag{44}
\]
where \( z_n \in M_n \) is the \((y_n+1)\)-th letter in the ordering on \( M_n \). Composing this map with the projection on the first coordinate yields the topological factor \( (X_{\tilde{\theta}}, S) \).
The projection of (44) on the second coordinate yields the factor \((X_{\tilde{\theta}}, S)\). According to Subsection 5.2.3, we can now represent \((X_{\tilde{\theta}}, \mu_{\tilde{\theta}}, S)\) as a skew product over \((X_{\tilde{\theta}}, \mu_{\tilde{\theta}}, S)\) using the following: first

identifying \((i_n, M_n)_{n \in \mathbb{Z}}\) with \(((M_n)_{n \in \mathbb{Z}}, i_0) \in X_{\tilde{\theta}} \times \{0, \ldots, c - 1\}\)

(which is possible for all \(\overline{M} = (M_n)_{n \in \mathbb{Z}}\) satisfying (43), cf. Subsection 5.2.3) and then setting

\[
(M, i_0) \mapsto (SM, \sigma_{\overline{M}}(i_0)),
\]

where \(\sigma_{\overline{M}}\) (in fact, it is \(\sigma_{M_0}\)) is a permutation of \(\{0, \ldots, c - 1\}\) determined by \(M_0\); we come back to the matrix (42) in which \(M_0\) is represented as the \(j_i\)-th column. The next column is representing \(M_1\), and if we have \(z_0\) (which belongs to \(M_0\) and corresponds to \(i_0\)) in the \(i_0\)-th row then \(z_1\) is just the next element in the same row. Now, the elements in the \(j_i\)-th column have their names in \(\{0, \ldots, c - 1\}\) and this is the same for the \((j_i + 1)\)-st column. Hence \(\sigma_{\overline{M}}\) is a permutation of \(\{0, \ldots, c - 1\}\) sending an element in the \(j_i\)-th column to the neighboring one in the next column.

**Remark 20.** As from the measure-theoretic point of view, \((X_{\tilde{\theta}}, S)\) is still \((X_{\bar{\theta}}, S)\) (and \((X_{\bar{\theta}}, S)\) is the same as \((H, R)\)), the above formula (45) clears up Remark 9.1 in [38], p. 229, about the form of a cocycle representing \((X_{\bar{\theta}}, \mu_{\bar{\theta}}, S)\) as a skew product over \((H, m_{H}, R)\).

Coming back to our main problem, note that \((X_{\tilde{\theta}}, S)\) has \((X_{\bar{\theta}}, S)\) as its topological factor and it has also \((X_{\tilde{\theta}}, S)\) “representing” measure-theoretically \((H, R)\) as its topological factor, but still we do not know whether we have decomposition (32) for each \(F \in C(X_{\tilde{\theta}})\). To assure it, we will need another extension.

To conclude this section we want to discuss one particular ordering that has useful properties.

We start by taking an arbitrary complete order \(\leq_{M_0}\) on \(M_0\) such that \(a_0\) is the minimum. By Proposition 9, \(\tilde{\theta}\) is synchronizing and primitive, thus we find \(k_0', j_0'\) such that \(\tilde{\theta}^{k_0}(M)_{j_0'} = M_0\) for all \(M \in \mathcal{X}\). As \(\theta^{k_0}_{j_0}\) permutes the elements of \(M_0\), we find by iterating \(\theta^{k_0}_{j_0}\), some \(k_0, j_0 < \lambda^{k_0}\) such that \(\theta^{k_0}(a)_{j_0} = a\) for all \(a \in M_0\) and \(\theta^{k_0}(M)_{j_0} = M_0\) for all \(M \in \mathcal{X}\).

Using (38), we extend now the ordering on \(M_0\) to a partial (but not complete in general, as below, \(a \neq b\) remain incomparable if \(\theta^{k_0}(a)_{j_0} = \theta^{k_0}(b)_{j_0}\)) ordering on \(\mathbb{A}\) by

\[
a < b \text{ if and only if } \theta^{k_0}(a)_{j_0} <_{M_0} \theta^{k_0}(b)_{j_0}.
\]

This gives a complete ordering on any \(M \in \mathcal{X}\) and, thus, allows us to define \(\bar{\theta}\).

With this ordering we obtain directly that for all \(M \in \mathcal{X}\) and \(i = 0, 1, \ldots, c - 1\), we have

\[
\bar{\theta}^{k_0}(i, M)_{j_0} = (i, M_0)
\]

(46)

which will be useful later. In particular, we see that \(a\) is the \(i\)-th largest element of some \(M \in \mathcal{X}\) if and only if \(\theta^{k_0}(a)_{j_0}\) is the \(i\)-th largest element of \(M_0\). Thus being the \(i\)-th largest element of some set \(M\) is independent of \(M\).

**Remark 21.** As the order on each \(M \in \mathcal{X}\) is fixed, we have \(\bar{\theta}^k = \bar{\Theta}^k\) (cf. footnote 31).
5.4. Column number, height and pure base – revisited. In this part we come back to connections between the column number and the height of primitive substitutions. We start with simple observations about substitution joinings.

**Lemma 5.5.** Let \( \theta : \mathcal{A} \to \mathcal{A}^\lambda \) and \( \zeta : \mathcal{B} \to \mathcal{B}^\lambda \) be substitutions of length \( \lambda \) fulfilling \((15), (18), (25)\) and \((26)\), but are not necessarily pure. For some \( \mathcal{A} \subset \mathcal{A} \times \mathcal{B} \), let \( \Sigma = \theta \vee \zeta : \mathcal{A} \to \mathcal{A}^\lambda \) be a substitution joining (in particular, we assume that \( \Sigma \) is primitive). Then,

\[
\text{lcm}(h(\theta), h(\zeta)) \mid h(\Sigma),
\]

which also yields a lower bound for \( h(\Sigma) \). Furthermore,

\[
\max(c(\theta), c(\zeta)) \leq c(\Sigma) \leq c(\theta) \cdot c(\zeta),
\]

which, in particular, yields lower and upper bounds for the column number of \( \Sigma \).

**Proof.** One sees directly that \( S_k(\Sigma) = S_k(\theta) \cap S_k(\zeta) \) and, therefore, \( g_k(\theta) \mid g_k(\Sigma) \) and \( g_k(\zeta) \mid g_k(\Sigma) \) which implies \((11)\).

As the projections of \( \mathcal{A} \) on \( \mathcal{A} \) and \( \mathcal{B} \) are full, the lower bound for \( c(\Sigma) \) is obvious. Furthermore, if \( |\theta^k(\mathcal{A})| = c(\theta) \), \( |\zeta^k(\mathcal{B})| = c(\zeta) \) then (using \((11)\)), we have

\[
\Sigma^{k+j}(\mathcal{A})_{j \cdot \lambda^k + j'} \subseteq \Sigma^{k+j}(\mathcal{A} \times \mathcal{B})_{j \cdot \lambda^k + j'} = \Sigma^{k'}(\Sigma^k(\mathcal{A} \times \mathcal{B})_{j'})_{j'} \subseteq \Sigma^{k'}(\theta^k(\mathcal{A}) \times \mathcal{B})_{j'} = \theta^{k'}(\theta^k(\mathcal{A})_{j'})_{j'} \times \zeta^k(\mathcal{B})_{j'}.
\]

This shows directly that \( |\Sigma^{k+j}(\mathcal{A})_{j \cdot \lambda^k + j'}| \leq c(\theta) \cdot c(\zeta) \) and completes the proof. \( \square \)

Let us now pass to the proof of Lemma 2.2:

**Proof of Lemma 2.2.** (We assume that the ordering on \( \mathcal{A} \) is the one described at the end of Subsection 5.3.) We have already seen that \( c := c(\theta) = c(\Theta) \) and \( h := h(\theta) = h(\Theta) \), so we consider without loss of generality the substitution \( \Theta : (\{0, \ldots, c-1\} \times X) \to (\{0, \ldots, c-1\} \times X)^\lambda \). It follows directly from the basic properties of the height that we discussed at the beginning of the paper that there exists a 1-code \( f : \mathcal{A} := (\{0, \ldots, c-1\} \times X) \to \{0, \ldots, h-1\} \) (cf. property 3. of the height in Section 2.2) such that the fixed point \( u = \Theta(u) \) is mapped to the periodic sequence

\[
01 \ldots (h-1)01 \ldots (h-1) \ldots .
\]

It follows by a simple computation (cf. \((11)\)) that

\[
u[j' \lambda^k + j] = \Theta^k(u[j'])_j
\]

whenever \( j' \geq 0 \) and \( j < \lambda^k \). As \( f(u[\ell]) \equiv \ell \mod h \) for each \( \ell \geq 0 \), it follows that (letting \( j' \) be such that \( (i, M) = u[j'] \) and \( \ell = j' \lambda^k + j \))

\[
f(\Theta^k(i, M)_j) = f(\Theta^k(u[j'])_j) = f(u[j' \lambda^k + j]) \equiv j' \lambda^k + j \mod h.
\]

By the same token, \( f(i, M) \equiv j' \mod h \), so finally

\[
f(\Theta^k(i, M)_j) \equiv f(i, M) \lambda^k + j \mod h
\]

for each \( k \geq 0 \) and \( j < \lambda^k \).

Furthermore, we know by the construction of \( \Theta \) that there exist \( k_0, j_0 \) such that \( \Theta^{k_0}(i, M)_{j_0} = (i, M_{00}) \) for every \( i \in \{0, \ldots, c-1\} \) and \( M \in X \), as noticed in \((46)\).

\[34\] One can also see this result dynamically, as both \((X_\theta, S)\) and \((X_\zeta, S)\) are topological factors of \((X_\Sigma, S)\).
Thus, in view of (48) (for \( j = j_0 \) and \( k = k_0 \)), for all \( i \in \{0, \ldots, c - 1\} \) and \( M \in \mathcal{X} \), we obtain that

\[
(f(i, M) x^{k_0} + j_0 \mod h) = f(\tilde{\Theta}^{k_0}(i, M) j_0) = f(i, M_0).
\]

This implies that \( f(i, M) \) does not depend on \( M \) (as \( \lambda \) is a multiplicatively invertible element in the ring \( \mathbb{Z}/h\mathbb{Z} \)), so \( f(i, M) = f(i, M_0) \) and \( f \) only depends on the first coordinate. Therefore, we denote \( f'(i) := f(i, M_0) \). Moreover, we recall that \( a \) is the \( i \)-th largest element of some set \( M \in \mathcal{X} \) if and only if \( \theta^{k_0}(a) j_0 \) is the \( i \)-th largest element of \( M_0 \). Thus, we can actually view \( f \) also as a map from \( \mathcal{A} \) to \( \{0, \ldots, h - 1\} \).

As the second coordinate in \( \tilde{\Theta} \) equals \( \tilde{\theta} \) is independent of the first coordinate and \( \tilde{\Theta}(\cdot, M)_{j} \) is a bijection from \( \{0, \ldots, c - 1\} \) to itself, we obtain the following formula involving the incidence matrix \( M(\tilde{\Theta}) \) (see (16)),

\[
\sum_{i' < c} M(\tilde{\Theta})_{(i, M), (i', M')} = M(\tilde{\Theta})_{M, M'}.
\]  

(49)

We now claim that

\[
\delta_{(i, M)} = \delta_{M} \cdot \frac{1}{c},
\]  

(50)
i.e. the density of \((i, M)\) for \( \tilde{\Theta} \) is equal to \( e_{(i, M)} := \frac{1}{c} \delta_{M} \). Indeed, using the definition of \( e_{(i, M)} \), (49) and (17) (for \( \tilde{\Theta} \)), we obtain

\[
\sum_{(i', M') \in \{0, \ldots, c - 1\} \times \mathcal{X}} e_{(i', M')} \cdot M(\tilde{\Theta})_{(i, M), (i', M')} = \sum_{M' \in \mathcal{X}} \frac{1}{c} \delta_{M'} \cdot \sum_{i' < c} M(\tilde{\Theta})_{(i, M), (i', M')} = \sum_{M' \in \mathcal{X}} \frac{1}{c} \delta_{M'} \cdot M(\tilde{\Theta})_{M, M'} = \lambda \cdot \frac{1}{c} \delta_{M} = \lambda \cdot e_{(i, M)}.
\]

But (by the Perron-Frobenius theorem) there is a unique solution to the above, given by \( \delta_{(i, M)} \), so \( e_{(i, M)} = \delta_{(i, M)} \) and the claim (50) follows.

Clearly, each letter \( j \in \{0, \ldots, h - 1\} \) appears in \( 01 \ldots (h - 1)01 \ldots \) with density \( \frac{1}{h} \) and, recall that, this periodic sequence is the image of \( u \) under \( f \). Using this and (50), we obtain

\[
\frac{1}{h} = \sum_{(i, M), f(i, M) = j} \delta_{(i, M)} = \sum_{i, f'(i) = j} \sum_{M} \frac{1}{c} \delta_{M} = \frac{1}{c} |\{i < c : f'(i) = j\}|.
\]  

(51)

Equation (51) shows directly that \( h | c \) and \( \#\{i < c : f'(i) = j\} \) is constant and equals \( c/h \). It remains to show that the column number of the pure base of \( \theta \) equals \( c/h \). To this end, let us first recall that \( f \) can also be defined on \( \mathcal{A} \) and that for all \( M \in \mathcal{X} \) and \( j < h \), we have

\[
\#\{a \in M : f(a) = j\} = c/h.
\]  

(52)

Recall also the construction of the pure basis \( \theta^{(h)} \) of \( \theta \). It is a substitution defined over the alphabet \( \mathcal{A}^{(h)} \) which is the set \( \{u \mid mh(\theta), mh(\theta) + h - 1 \colon m \geq 0\} \) (which of course finite). Then, we define \( \theta^{(h)} : \mathcal{A}^{(h)} \rightarrow (\mathcal{A}^{(h)})^\lambda \) by setting

\[
\theta^{(h)}(u)_{j} = \theta(u)[j h(\theta), (j + 1)h(\theta) - 1]
\]

for \( j = 0, \ldots, \lambda - 1 \).
If \( w = (w_0, \ldots, w_{h-1}) \in A^h \), then there exists some \( m \) such that \( w_i = u[\lambda h + i] \)
for \( i = 0, \ldots, h - 1 \). It follows that
\[
f(w_i) = f(u[\lambda h + i]) \equiv mh + i \equiv i \pmod{h}
\]
which gives \( f(w_i) = i \). For each \( k \geq 1 \) and \( j < \lambda^k \), we have
\[
(\theta^{(h)})^k(w_j) = \theta^k(w)[jh(\theta), (j + 1)h(\theta) - 1].
\]
(54)

Choose \( k > 1 \) such that \( \lambda^k \equiv 1 \pmod{h} \) and \( \lambda^k > (j_0 + 1)h \). Now, \( j_0 \lambda^{k-1} h_0 < \lambda^{k+k_0} \).

In view of (54) (for \( k + k_0 \) and \( j_0(\lambda^k - 1)/h + j_0 \)) and then (11) (which we may use as \( (j_0 + 1)(h - 1) < \lambda^k \)), we obtain that
\[
(\theta^{(h)})^{k+k_0}(w_0, \ldots, w_{h-1})_{j_0}(\lambda^k - 1)/h + j_0
= \theta^{k+k_0}(w_0, \ldots, w_{h-1})[j_0\lambda^{k} + j_0(h - 1), j_0\lambda^{k} + (j_0 + 1)(h - 1)]
= \theta^{k}(\theta^{k_0}(w_0))_{j_0}[j_0(h - 1), (j_0 + 1)(h - 1)].
\]

It follows from this calculation that the column number of \( \theta^{(h)} \) is not larger than the number of \( \theta^{k_0}(w_0)_{j_0} \), when we run over all possible \((w_0, \ldots, w_{h-1}) \in A^h \). But, as before (cf. (48)), \( f(\theta^{(h)}(a)) = f(a) \lambda^k + j \pmod{h} \) (for each \( a \in A \)). It follows that
\[
f(\theta^{k_0}(w_0))_{j_0} \equiv j_0 \pmod{h} \) (as \( f(w_0) = 0 \), see (53)). Moreover, by (46), \( \theta^{k_0}(a)_{j_0} \in M_0 \)
for each \( a \in A \). This shows that \( c(\theta^{(h)}) \) is not larger than the number of \( b \in M_0 \) such that \( f(b) \equiv j_0 \pmod{h} \) and hence, in view of (52), \( c(\theta^{(h)}) \leq c(\theta)/h(\theta) \).

Next, we want to give a lower bound for \( c(\theta^{(h)}) \). We fix any \( k > 0 \) and \( j < \lambda^k \).

Denote \( i = \lfloor \frac{k}{h} \rfloor \). In view of (23), it follows that we have for any \((w_0, \ldots, w_{h-1}) \in A^h \),
\[
(\theta^{(h)})^k(w_0, \ldots, w_{h-1})_j = (v_0, \ldots, v_{h-1}),
\]
where
\[
v_0 = \theta^k(w_i)_{jh-\lambda i^k}.
\]
Moreover, \( f(w_i) = i \) (see (53)). Thus, we find
\[
#(\theta^{(h)})^k(A^h)_j \geq #\theta^k(A^h)_{jh-\lambda i^k},
\]
where \( A^h := \{a \in A : f(a) = i\} \). Finally, we recall the relative bijectivity: for each \( M \in \mathcal{X} \) and \( 0 \leq j < \lambda \) we have \( |\lambda(M)_j| = c \) and, therefore, for any different \( a, a' \in M \), we have \( \theta^k(a)_j \neq \theta^k(a')_j \) for each \( j' < \lambda^k \), see also the remark before Lemma 5.3. This shows that
\[
c(\theta^{(h)}) \geq \# \{a \in M_0 : f(a) = i \} = \frac{c(\theta)}{h(\theta)},
\]
which completes the proof.

\[\Box\]

5.5. Toward a group extension. We will now aim at an extension of the substitution \( \Theta \) to a substitution \( \tilde{\Theta} \) (of length \( \lambda \)) over the alphabet \( G \times \mathcal{X} \), where \( G \) is a subgroup of the group \( \mathcal{S}_c \) of bijections on \( \{0, \ldots, c - 1\} \). Recall that for each \( j = 0, \ldots, \lambda - 1 \) and \( M \in \mathcal{X} \), \( \tilde{\Theta}(\cdot, M)_j \) is bijective in view of Lemma 5.3 and the definition of \( \Theta \). We define the corresponding permutation \( \sigma_{M,j} \in \mathcal{S}_c \) by \( \sigma_{M,j}(m) = n \) iff \( \tilde{\Theta}(n, M)_j = (m, \tilde{\Theta}(M)_j) \). This allows us to define the aforementioned group \( G \) as the group generated by the \( \sigma_{M,j} \), i.e.
\[
G := \langle \sigma_{M,j} : M \in \mathcal{X}, j \in \{0, \ldots, \lambda - 1\} \rangle.
\]
(55)
Finally, we define $\tilde{\Theta} : G \times \mathcal{X} \to (G \times \mathcal{X})^\lambda$ by setting
\[
\tilde{\Theta}(\sigma, M)_j := (\sigma \circ \sigma_{M,j}, \tilde{\theta}(M)_j),
\]
for $j = 0, \ldots, \lambda - 1$.

Repeating the same with $\theta$ replaced with $\theta^k$ (cf. Remark 21), we define the permutations $\sigma_{M,j}^{(k)}$ of $\{0, \ldots, c - 1\}$ so that
\[
\sigma_{M,j}^{(k)}(m) = n \text{ if and only if } \tilde{\theta}^k(n, M)_j = (m, \tilde{\theta}^k(M)_j).
\]
This is a very important formula, as it highlights the connection between $\tilde{\Theta}$ and $\tilde{\theta}$.

Furthermore, analogously to (56), we find that
\[
\tilde{\Theta}^k(\sigma, M)_j = (\sigma \circ \sigma_{M,j}^{(k)}, \tilde{\theta}^k(M)_j),
\]
for $j = 0, \ldots, \lambda^k - 1$.

It follows directly that $\sigma_{M,j}^{(1)} = \sigma_{M,j}$ and from (11) (applied to $\tilde{\Theta}$) that
\[
\sigma_{M,j_1 \lambda^k + j_2}^{(k_1 + k_2)} = \sigma_{M,j_1}^{(k_1)} \circ \sigma_{\theta^k(M)j_1, j_2}^{(k_2)}
\]
for $j_1 < \lambda^{k_1}$ and $j_2 < \lambda^{k_2}$.

If $\tilde{\theta}(M_0)_0 = M_0$, then clearly $\tilde{\theta}^k(M_0)_0 = M_0$ and applying (59) for $(\sigma, M_0)$, with $j_1 = j_2 = 0$, we obtain $\sigma_{M_0,0}^{(k)} = (\sigma_{M_0,0})^k$. Hence, by possibly passing to an iterate of $\tilde{\Theta}$, we can assume that
\[
\tilde{\Theta}(\sigma, M_0)_0 = (\sigma, M_0)
\]
for each $\sigma \in G$.

We recall (46), which shows the existence of some $k_0 \in \mathbb{N}$ and $j_0 < \lambda^{k_0}$ such that $\tilde{\Theta}^{k_0}(i, M)_{j_0} = (i, M_0)$ for all $i < c$ and $M \in \mathcal{X}$. This gives directly, by (57), that $\sigma_{M,j_0}^{(k_0)} = id$, or in other words
\[
\tilde{\Theta}^{k_0}(\sigma, M)_{j_0} = (\sigma, M_0)
\]
holds for all $\sigma \in G, M \in \mathcal{X}$.

The next few lemmas will be used to show that $\tilde{\Theta}$ is primitive. We start by giving some kind of a dual statement to (61).

**Lemma 5.6.** For every $M \in \mathcal{X}$ there exist $k_M \in \mathbb{N}$ and $j_M < \lambda^{k_M}$ such that $\tilde{\Theta}^{k_M}(\sigma, M)_{j_M} = (\sigma, M)$ for each $\sigma \in G$.

**Proof.** We fix $M \in \mathcal{X}$. As $\tilde{\theta}$ is primitive, we can find $k \in \mathbb{N}, j < \lambda^k$ such that $\tilde{\theta}^k(M_0)_j = M$, which gives $\tilde{\Theta}^k(\sigma, M_0)_j = (\sigma \circ \sigma_{M,j}^{(k)}, M)$ for all $\sigma \in G$. In view of (11), we obtain that
\[
\tilde{\Theta}^{k_0 + k}(\sigma, M_0)_{j \lambda^{k_0} + j_0} = \tilde{\Theta}^{k_0}(\tilde{\Theta}^k(\sigma, M_0)_j)_{j_0}
= \tilde{\Theta}^{k_0}(\sigma \circ \sigma_{M,j}^{(k)}, M)_{j_0} = (\sigma \circ \sigma_{M,j}^{(k)}, M_0).
\]

Set $\ell := ord(\sigma_{M,j}^{(k)})$. Now, by iterating (62), using (11), we obtain that
\[
\tilde{\Theta}^{k_0}(\sigma, M_0)_{j_0} = (\sigma \circ \sigma_{M,j}^{(k)})^{\ell - 1}(M_0),
\]
where \( k' = (\ell - 1)(k_0 + k) \) and \( j' = \lambda^{(\ell - 2)}(k_0 + k)(j\lambda^{k_0} + j_0) + \ldots + \lambda^{k_0+k}(j\lambda^{k_0} + j_0) \). It follows that

\[
\hat{\Theta}^{k+k'}(\sigma, M_0)_{j+k_0+j'} = \hat{\Theta}^k(\hat{\Theta}(\sigma, M_0)_{j'})_j = \hat{\Theta}^k(\sigma \circ (\sigma^{(k)}_{M,j})^{\ell-1}, M_0)_j
\]

which completes the proof. \( \square \)

Next, we find a better description for \( G \).

**Lemma 5.7.** We have

\[
G = \langle \sigma^{(k)}_{M_0,j} : k \in \mathbb{N}, j < \lambda^k, \tilde{\theta}^k(M_0)_j = M_0 \rangle.
\]

**Proof.** By (11), it is clear that any \( \sigma^{(k)}_{M_0,j} \in G \). Thus, it suffices to write any \( \sigma_{M,j} \) as \( \sigma^{(k)}_{M_0,j'} \), where \( \tilde{\theta}^k(M_0)_j = M_0 \). Using (11), Lemma 5.6, (56) and finally (61), we have

\[
\tilde{\theta}^{k_0+1+k_M}(\sigma, M_0)_{j_0+\lambda^{k_0}+\lambda^{k_0+1}j_M} = \tilde{\theta}^{k_0}(\tilde{\theta}(\tilde{\theta}(\sigma, M_0)_{j_M})_{j_0})_{j_0}
\]

which, by (58), ends the proof. \( \square \)

This allows us to give the final description for \( G \).

**Lemma 5.8.** We have

\[
G = \{ \sigma^{(k)}_{M_0,j} : k \in \mathbb{N}, j < \lambda^k, \tilde{\theta}^k(M_0)_j = M_0 \}.
\]

**Proof.** It just remains to show that

\[
\{ \sigma^{(k)}_{M_0,j} : k \in \mathbb{N}, j < \lambda^k, \tilde{\theta}^k(M_0)_j = M_0 \}
\]

is indeed a (finite) group. As each element in our group \( G \) is of finite order, all we need to show is that the multiplication of two elements \( \sigma^{(k_1)}_{M_0,j_1}, \sigma^{(k_2)}_{M_0,j_2} \), where \( \tilde{\theta}^{k_1}(M_0)_{j_1} = M_0 \), is again an element of the set. A simple computation yields

\[
\tilde{\theta}^{k_2+k_1}(\sigma, M_0)_{j_2+\lambda^{k_2}j_1} = \tilde{\theta}^{k_2}(\tilde{\theta}(\sigma, M_0)_{j_1})_{j_2} = \tilde{\theta}^{k_2}(\sigma \circ \sigma^{(k_1)}_{M_0,j_1}, M_0)_{j_2}
\]

so the result follows. \( \square \)

Thus, for any \( g \in G \), we have some \( k_g \geq 1 \) and \( j_g < \lambda^{k_g} \) such that \( g = \sigma^{(k_g)}_{M_0,j_g} \) and \( \tilde{\theta}^{k_g}(M_0)_{j_g} = M_0 \). Now, we can finally show the following result.

**Proposition 12.** \( \hat{\Theta} \) is primitive.

**Proof.** As we have \( \hat{\Theta}(\sigma, M_0)_0 = (\sigma, M_0) \) for all \( \sigma \in G \), it is sufficient to show that for all \( g, h \in G \) and \( M, M' \in \mathcal{X} \) there exist \( k_1 \geq 1, j_1 < \lambda^{k_1} \) such that \( \hat{\Theta}^{k_1}(id, M_0)_{j_1} = (g, M) \) and \( k_2 \geq 1, j_2 < \lambda^{k_2} \) such that \( \hat{\Theta}^{k_2}(h, M')_{j_2} = (id, M_0) \). Indeed, for all \( k \geq k_1 + k_2 \), we then obtain that

\[
\tilde{\theta}^{k}(h, M')_{j_2\lambda^{k_1}j_1} = \tilde{\theta}^{k_1}(\tilde{\theta}^{k-k_1-k_2}(\tilde{\theta}^{k_2}(h, M')_{j_2})_0)_{j_1}
\]

and

\[
\tilde{\theta}^{k_1}(\tilde{\theta}^{k-k_1-k_2}(id, M_0)_0)_{j_1} = \tilde{\theta}^{k_1}(id, M_0)_{j_1} = (g, M).
\]
As there are only finitely many \( g, h \in G \) and \( M, M' \in \mathcal{X} \), we find that there exists some (large) \( k \) such that for all \( g, h \in G \) and \( M, M' \in \mathcal{X} \) there exists \( j \) such that \( \hat{\Theta}^k(h, M')_j = (g, M) \).

Fix now \( g \in G \) and \( M \in \mathcal{X} \). Since \( g = \sigma_{M_0,j_0}^{(k_s)} \), where \( \bar{\Theta}^{k_s}(M_0)_{j_0} = M_0 \), we have (using Lemma 5.6)

\[
\hat{\Theta}^{k M + k_s}(id, M_0)_{j M + \lambda^k M J_g} = \hat{\Theta}^{k M} (\hat{\Theta}^{k_s}(id, M_0)_{j M}) = \hat{\Theta}^{k M} (g, M)_{j M} = (g, M).
\]

Furthermore, we have (using (60))

\[
\hat{\Theta}^{k_{h-1} + k_0}(h, M')_{j_{h-1} + \lambda^{k_{h-1}} j_0} = \hat{\Theta}^{k_{h-1}}(\hat{\Theta}^{k_0}(h, M')_{j_0})_{j_{h-1}} = \hat{\Theta}^{k_{h-1}}(h, M_0)_{j_{h-1}} = (h \circ h^{-1}, M_0) = (id, M_0),
\]

which completes the proof. \( \square \)

**Proposition 13.** \((X_{\hat{\Theta}}, S)\) is a topological factor of \((X_{\hat{\Theta}}, S)\).

**Proof.** We note that a point in the space \( X_{\hat{\Theta}} \) is of the form \((\sigma_n, M_n)_{n \in \mathbb{Z}}\) and define the equivariant map

\[
(\sigma_n, M_n) \mapsto (i_n, M_n),
\]

(63)

where \( i_n \) is defined by \( \sigma_n(i_n) = 0 \). We claim that the fixed point of \( \hat{\Theta} \) that starts with \((id, M_0)\) is mapped to the fixed point of \( \hat{\Theta} \) that starts with \((0, M_0)\). Indeed, in view of (58), we have for all \( k \geq 0, n < \lambda^k \) that \( \hat{\Theta}^k(id, M_0)_n = (\sigma^{(k)}_{M_0,n}, \bar{\Theta}^k(M_0)_n) \) and, by (57), we have that \( \sigma^{(k)}_{M_0,n}(i) = 0 \) if and only if \( \bar{\Theta}^k(0, M_0)_n = (i, \bar{\Theta}^k(M_0)_n) \).

Thus, the map (63), sends the fixpoint of \( \hat{\Theta} \) starting with \((id, M_0)\) to the fixpoint of \( \hat{\Theta} \) starting with \((0, M_0)\). Now, the proof follows by the minimality of the dynamical systems under consideration. \( \square \)

As \((X_{\hat{\Theta}}, S)\) is topologically isomorphic to \((X_{\hat{\Theta}}, S)\), we obtain the following (cf. Remark 20).

**Corollary 7.** \((X_{\theta}, S)\) is a topological factor of \((X_{\hat{\Theta}}, S)\).

We have seen that \( h(\theta) = h(\Theta) = h(\hat{\Theta}) \), and \( c(\theta) = c(\Theta) = c(\hat{\Theta}) \). These equalities do not carry over to \( \hat{\Theta} \).

**Lemma 5.9.** We have \( c(\hat{\Theta}) = |G| \).

**Proof.** We recall that, by (46), there are \( k_0, j_0 \) such that

\[
\hat{\Theta}^{k_0}(\sigma, M)_{j_0} = (\sigma, M_0),
\]

which shows that \( c(\hat{\Theta}) \leq |\{(\sigma, M_0) : \sigma \in G\}| = |G| \).

Furthermore, it follows from the definition of \( \hat{\Theta} \) that for all \( M \in \mathcal{X} \), we have

\[
\hat{\Theta}^k((g, M) : g \in G)_j = \{(g \cdot \sigma^{(k)}_{M,j}, \bar{\Theta}^k(M)_j) : g \in G\} = \{(g', \bar{\Theta}^k(M)_j) : g' \in G\}.
\]

This shows directly that

\[
\hat{\Theta}^k(G \times \mathcal{X})_j = G \times \bar{\Theta}^k(\mathcal{X})_j,
\]

which implies that \( c(\hat{\Theta}) \geq |G| \). \( \square \)
By Corollary 7, it follows that the maximal equicontinuous factor of \((X_\theta,S)\) is a factor of the maximal equicontinuous factor of \((X_{\tilde{\Theta}},S)\). As \(\theta\) and \(\tilde{\Theta}\) have the same length, it follows that \(h(\theta)\mid h(\tilde{\Theta})\). Even though one could expect the equality of these two numbers, it is not the case in general. The following example shows that \(h(\Theta)\) and \(h(\theta)\) can be different (and, in general, \(h(\tilde{\Theta})\) does not divide \(c(\theta)\)).

**Example 4.** We consider the following bijective (whence \(c(\theta) = 3\)) substitution \(\theta\):
\[
\begin{align*}
\theta(a) &= aab \\
\theta(b) &= bcc \\
\theta(c) &= cba.
\end{align*}
\]
Since \(a \mapsto aab\), obviously, \(h(\theta) = 1\). Of course \(X = \{a,b,c\}\), so in fact for \(\tilde{\Theta}\) we are interested only in the first coordinate \((\tilde{\Theta}(\sigma) = \tilde{\Theta}(\sigma,M_0))\). The bijections given by the columns yield three permutations of \(\{0,1,2\}\): \(\sigma_0 = Id, \sigma_1 = (12)\) and \(\sigma_2 = (012)\). As they generate \(S_3\), we have \(G = S_3\). We also have \(\tilde{\Theta}(\sigma) = (\sigma)(\sigma \circ (12))(\sigma \circ (012))\),
for all \(\sigma \in S_3\).

We consider the fixpoint \(u\) that starts with \(id\) and find that \(u[4] = (12)(12) = id\) and \(u[26] = (012)(012)(012) = id\). Thus, see (20), we have \(g_0 = \text{gcd}(S_0) \leq 2\). Furthermore, we find that for \(n = \sum_{i=0}^{r} n_i 3^i\),
\[u[n] = \tilde{\Theta}^{r+1}(id)_n = \sigma_{n_r} \circ \ldots \circ \sigma_{n_0}.
\]
We consider the sign of a permutation and find that
\[\text{sgn}(\sigma_0) = \text{sgn}(\sigma_2) = 1, \text{sgn}(\sigma_1) = -1\]
which gives
\[\text{sgn}(u[n]) = \prod_{i=0}^{r} \text{sgn}(\sigma_{n_i}) = \prod_{i=0}^{r} (-1)^{n_i} = \prod_{i=0}^{r} (-1)^{n_i 3^i} = (-1)\sum_{i=0}^{r} n_i 3^i = (-1)^n.
\]
This shows that \(h(\tilde{\Theta}) = 2\).

This example shows that \(\tilde{\Theta}\) is much closer to \(\theta\) than \(\tilde{\Theta}\), but \(\tilde{\Theta}\) has a structural advantage, as it relies on a group \(G\). Indeed, we are able to find a representation of the height within the group \(G\).

**Lemma 5.10.** Let \(\hat{\Theta} : G \times X \rightarrow (G \times X)^\lambda\) be as described above. Then, there exists \(G_0 \leq G\) such that \(G/G_0 \cong \mathbb{Z}/h(\hat{\Theta})\mathbb{Z}\). Furthermore, if we define \(\zeta : G/G_0 \times X \rightarrow (G/G_0 \times X)^\lambda\) by
\[
\zeta(G_0 g, M) = (G_0 g \sigma_{M,j}, \tilde{\Theta}(M)_j),
\]
then \(\zeta\) is the joining of a periodic substitution \(p\) (of period \(h\)) and \(\tilde{\Theta}\), i.e. \(\zeta = p \lor \tilde{\Theta}\).

**Proof.** Similarly as at the beginning of Lemma 2.2, by setting \(h = h(\tilde{\Theta})\), there exists \(f : G \times X \rightarrow \{0, \ldots, h-1\}\) (which we also treat as a 1-code) such that the fixed point \(u\) of \(\tilde{\Theta}\) obtained by iterations of \(\hat{\Theta}\) at \((id,M_0)\) is mapped via \(f\) to \(01 \ldots (h-1)01 \ldots\) i.e. \(f(u[n]) = n \mod h\). Furthermore, we can assume without loss of generality that \(\lambda \equiv 1 \mod h\), as we can always replace \(\tilde{\Theta}\) by \(\tilde{\Theta}^t\), if necessary.

We find similarly to (48) that
\[
f(\hat{\Theta}^k(\sigma, M)_j) \equiv f((\sigma, M))\lambda^k + j \mod h
\]
for each $k \geq 0$ and $j < \lambda^k$. Using additionally (61), this gives

$$f((\sigma, M_0)) = f(\hat{\Theta}^k(\sigma, M)_0) \equiv f((\sigma, M))\lambda^k + j_0 \equiv f((\sigma, M)) + j_0 \mod h,$$

which shows that $f((\sigma, M))$ only depends on $\sigma$ ($f(\sigma) = f(\sigma, M_0)$). Since (for each $k$) $u$ begins with $\hat{\Theta}^k(id, M_0)$, by the definition of $f$, we have $f(\sigma_{M_0,j}^{(k)}) = j \mod h$.

Due to Lemma 5.8, we can write each $g_1, g_2 \in G$ as $g_i = \sigma_{M_0,j_i}^{(k_i)}$. Thus, we find

$$f(g_1g_2) = f(\hat{\Theta}^k(id, M_0)_{j_1}, j_2) \equiv f(\hat{\Theta}^k(id, M_0)_{j_1}) + j_2 \equiv j_1 + j_2 \mod h.$$

This shows, that for all $g_1, g_2 \in G$ we have

$$f(g_1g_2) = f(g_1) + f(g_2) \mod h.$$

It follows that $f$ is a group homomorphism and therefore $G_0 := f^{-1}(0)$ is a normal subgroup. We can identify $G_0g$ with $f(g)$ and the last statement follows immediately for $p : \{0, \ldots, h-1\} \rightarrow \{0, \ldots, h-1\}$ defined by $p(i) = \lambda i + j \mod h$. □

**Remark 22.** Lemma 5.10 allows us to find a representation of the maximal equicontinuous factor of $\hat{\Theta}$ in the case, where $\theta$ is not periodic. For a general substitution of constant length, we need to rely on a more complicated construction.

First, we would like to note that we can find a representation of $G$ in the centralizer of $(X_{\hat{\Theta}}, S)$. Indeed, given $\tau \in G$, consider $V_\tau : G \times \mathcal{X} \rightarrow G \times \mathcal{X}$ defined by

$$V_\tau(\sigma, M) = (\tau \circ \sigma, M).$$

Note that in view of (56), $\hat{\Theta}$ “commutes” with $V_\tau$. It follows from Remark 9 that $V_\tau$ (uniquely) extends to a homeomorphism $V_\tau$ of $X_{\hat{\Theta}}$ and commutes with $S$:

$$V_\tau((\sigma_n, M_n)_{n \in \mathbb{Z}}) = (\tau \circ \sigma_n, M_n)_{n \in \mathbb{Z}}$$

for each $(\sigma_n, M_n)_{n \in \mathbb{Z}} \in X_{\hat{\Theta}}$. It follows that we have a (finite) group $V := \{V_\tau : \tau \in G\}$ of homeomorphisms of $X_{\hat{\Theta}}$ commuting with the shift.

Next, we want to determine the dynamical system obtained by factoring $(X_{\hat{\Theta}}, S)$ by $V$. To this end, consider the map $g : (G \times \mathcal{X})^\mathbb{Z} \rightarrow (G \times \mathcal{X} \times \mathcal{X})^\mathbb{Z}$ defined by

$$g((\sigma_n, M_n)_{n \in \mathbb{Z}}) = (\sigma_{n+1}^{-1} \circ \sigma_n, M_n, M_{n+1})_{n \in \mathbb{Z}},$$

which means that $g$ is a sliding block code with (right-) radius 1 (that is, there exists some $\Phi : (G \times \mathcal{X})^\mathbb{Z} \rightarrow G \times \mathcal{X} \times \mathcal{X}$ such that $g(x)[n] = \Phi(x[n, n+1])$). We see directly that $g((\sigma_n, M_n)_{n \in \mathbb{Z}}) = g((\sigma_n', M_n')_{n \in \mathbb{Z}})$ if and only if $(\sigma_n', M_n')_{n \in \mathbb{Z}} = V_\tau((\sigma_n, M_n)_{n \in \mathbb{Z}})$ for some $\tau \in G$. We want to describe $g(X_{\hat{\Theta}})$. As $g$ is a sliding block code, it follows that $g(X_{\hat{\Theta}})$ is a (minimal) subshift. We want to show that it is actually a substitutional dynamical system, where the substitution has column number 1, i.e. it is synchronizing.

To this end, let us define a substitution $\eta : G \times \mathcal{X} \times \mathcal{X} \rightarrow (G \times \mathcal{X} \times \mathcal{X})^\lambda$ as follows:

$$\eta(\sigma, M, M')_{j} = \begin{cases} (\sigma_{M,j+1}^{-1} \circ \sigma_{M,j}, \tilde{\theta}(M)_{j+1}), & \text{for } j < \lambda - 1 \\ (\sigma_{M',0}^{-1} \circ \sigma \circ \sigma_{M,\lambda-1}, \tilde{\theta}(M)_{\lambda-1}, \tilde{\theta}(M')_{0}), & \text{for } j = \lambda - 1. \end{cases}$$

A simple computation involving (59) gives

$$\eta^k(\sigma, M, M')_{j} = \begin{cases} ((\sigma_{M,j+1}^{(k)})^{-1} \circ \sigma_{M,j}^{(k)}, \tilde{\theta}^k(M)_{j+1}), & \text{for } j < \lambda^k - 1 \\ ((\sigma_{M',0}^{(k)})^{-1} \circ \sigma \circ \sigma_{M,\lambda-1}^{(k)}, \tilde{\theta}(M)_{\lambda-1}, \tilde{\theta}(M')_{0}), & \text{for } j = \lambda^k - 1. \end{cases}$$
We denote the fixpoint of \( \tilde{\Theta} \) that starts with \((id, M_0)\) by \( u \) and denote \( u[n] = (\sigma_n, M_n) \). This allows us to reduce the alphabet of the substitution \( \eta \) from \( G \times \mathcal{X} \times \mathcal{X} \) to \( \mathbb{B} \), where

\[
\mathbb{B} := \{ (\sigma_{n+1}^{-1} \circ \sigma_n, M_n, M_{n+1}) : n \geq 0 \}.
\]

The formula above shows that \( \eta \) maps the fixed point of \( \tilde{\Theta} \) starting with \((\sigma_0, M_0)\) \((\sigma_1, M_1)\) \(\ldots\) to the fixed point of \( \eta \) starting with \((\sigma_1^{-1} \circ \sigma_0, M_0, M_1)\). Furthermore, we see directly that the restriction of \( f : X_{\tilde{\Theta}} \to \mathbb{B}^\mathbb{Z} \) is still well-defined on the fixed point of \( \tilde{\Theta} \) and therefore on \( X_{\tilde{\Theta}} \). This restriction is necessary to ensure that \( \eta \) is primitive.

Thus we showed that \( g(X_{\tilde{\Theta}}) = X_\eta \).

It remains to show that \( \eta \) has column number 1. As \( \tilde{\theta} \) has column number 1, we find by (29) some integers \( k \geq 1, j < \lambda^k - 1 \) and \( M' \in \mathcal{X} \) such that \( \tilde{\theta}^k(M)_{j} = M' \) holds for all \( M \in \mathcal{X} \). It follows that for any \( \sigma \in G \) and \( M_1, M_2 \in \mathcal{X} \), we have

\[
\eta^{2k}(\sigma, M_1, M_2)_{\lambda^k + j} = \eta^{k}(\eta^{k}(\sigma, M_1, M_2))_{j}
\]

which shows that \( c(\eta) = 1 \).

**Example 5.** We give the explicit construction of \( \eta \) for the Rudin-Shapiro sequence, which was already introduced in Example 3. We recall that,

\[
\tilde{\Theta}(0, M_0) = (0, M_0)(0, M_1), \quad \tilde{\Theta}(1, M_0) = (1, M_0)(1, M_1),
\]

\[
\tilde{\Theta}(0, M_1) = (0, M_0)(1, M_1), \quad \tilde{\Theta}(1, M_1) = (1, M_0)(0, M_1).
\]

This gives directly

\[
\tilde{\Theta}(\sigma, M_0) = (\sigma, M_0)(\sigma, M_1),
\]

\[
\tilde{\Theta}(\sigma, M_1) = (\sigma, M_0)((\sigma \circ (01), M_1).
\]

We see that \( \mathbb{B} = \{(\sigma, M, M') \in G \times \mathcal{X} \times \mathcal{X} : M \neq M'\} \) as the only fixpoint of \( \tilde{\theta} \) is given by \( M_0M_1M_0M_1 \ldots \). Hence, \( \eta \) is defined as follows:

\[
\eta(\sigma, M_0, M_1) = (id, M_0, M_1)(\sigma, M_1, M_0),
\]

\[
\eta(\sigma, M_1, M_0) = (\sigma \circ (01), M_0, M_1)(\sigma \circ (01), M_1, M_0).
\]

A simple computation yields \( \eta(\sigma, M, M')_0 = (id, M_0, M_1) \) for all \( (\sigma, M, M') \in \mathbb{B} \) which shows that \( \eta \) has column number 1.

This procedure works well to find a representation of the maximal equicontinuous factor when \( h(\tilde{\Theta}) = 1 \). However, in general we are factoring out too much and need to take a subgroup of \( V \), namely \( V_1 := \{ V_\tau : \tau \in G_0 \} \) \((G_0 \text{ comes from Lemma 5.10})\). Therefore we define \( g_\tau : (G \times \mathcal{X})^\mathbb{Z} \to (G \times \mathcal{X} \times \mathcal{X} \times \{0, \ldots, h-1\})^\mathbb{Z} \) as an “extension” of the previously mentioned function \( g : (G \times \mathcal{X})^\mathbb{Z} \to (G \times \mathcal{X} \times \mathcal{X})^\mathbb{Z} \) in the following way (see the proof of Lemma 5.10 for \( f \) and its properties):

\[
g_h((\sigma_n, M_n)_{n \in \mathbb{Z}}) = (\sigma_{n+1}^{-1} \circ \sigma_n, M_n, M_{n+1}, f(\sigma_n))_{n \in \mathbb{Z}}.
\]

We see directly that \( g((\sigma_n, M_n)_{n \in \mathbb{Z}}) = g((\sigma'_n, M'_{n})_{n \in \mathbb{Z}}) \) if and only if \((\sigma_n', M'_{n})_{n \in \mathbb{Z}} = V_\tau((\sigma_n, M_n)_{n \in \mathbb{Z}}) \) for some \( \tau \in G_0 \), as \( f(\tau \sigma) \equiv f(\tau) + f(\sigma) \mod h \).
Thus, we consider now \( \eta_h : G \times \mathcal{X} \times \mathcal{X} \times \{0, \ldots, h-1\} \to (G \times \mathcal{X} \times \mathcal{X} \times \{0, \ldots, h-1\})^3 \), where the first three coordinates coincide with \( \eta \) and the last coordinate can be seen as another substitution \( p : \{0, \ldots, h-1\} \to \{0, \ldots, h-1\}^3 \) where \( p(i)_j = \lambda i + j \mod h \).

This gives that \( \eta_h = \eta \vee p \). As we have seen that \((X_\eta, S)\) is measure-theoretically isomorphic to \((H_\lambda, m_{H_\lambda}, R)\), and \((X_\eta, S) \cong (\mathbb{Z}/h\mathbb{Z}, m_{\mathbb{Z}/h\mathbb{Z}}, \tau_h)\), we have found a factor of \( X_\Theta \) that is itself given by \( X_\eta \) and is measure-theoretically isomorphic to the maximal equicontinuous factor of \( X_\Theta \) (cf. Proposition 3).

6. Proof of Theorem 0.1, Corollary 1 and some questions.

Proof of Theorem 0.1. We have already shown (i) of Theorem 0.1. We now need to handle the case \( c(\theta) > h(\theta) \).

To this end, we have defined \( \theta \vee \tilde{\theta} \) extending \( \theta \) (and being primitive), where \((X_\tilde{\theta}, S)\) "represents" (measure-theoretically) \((H_\lambda, m_{H_\lambda}, R)\) factor of \((X_\theta, S)\) if \( \theta \) is not quasi-bijective or it is finite. Via an isomorphic copy \( \tilde{\Theta} \) of \( \theta \vee \tilde{\theta} \), we finally have got the substitution \( \tilde{\Theta} \) which is primitive by Proposition 12. Moreover, by Corollary 7, \((X_\tilde{\Theta}, S)\) has \((X_\theta, S)\) as its topological factor, and therefore \( c(\tilde{\Theta}) > h(\tilde{\Theta}) \) (it cannot be synchronized).\(^{35}\) We only need to prove (ii) of Theorem 0.1 for \((X_\tilde{\Theta}, S)\).

Via (65), we have shown that there is a compact group, a copy of the group \( G \), namely \( V = \{V_\tau : \tau \in G\} \), in the centralizer of \( S \) which consists of homeomorphisms. Because of Lemma 5.9 and Proposition 3, the factor \( \sigma \)-algebra \( B(V) \) of subsets which are fixed by all elements of \( V \) “represents” (measure-theoretically) \((H_\lambda, m_{H_\lambda}, R)\). Of course, the \( L^2 \)-space for \((X_\tilde{\Theta}, \mu_{\tilde{\Theta}}, S)\) is spanned by continuous functions. Now, the conditional expectation of \( F \in C(X_\tilde{\Theta}) \) with respect to \( B(V) \) is given by \( \frac{1}{|G|} \sum_{\tau \in G} F \circ V_\tau \), so it is still a continuous function. It follows that each continuous function \( F \) is represented as \( F_1 + F_2 \), where \( F_1 \) is continuous and orthogonal to the \( L^2(B(V)) \), and \( F_2 \) is represented by a continuous function that belongs to \( L^2(B(V)) \). If we knew that \( H_\lambda \) represents the Kronecker factor of \((X_\tilde{\Theta}, \mu_{\tilde{\Theta}}, S)\), via Proposition 7 we could apply Section 4.3 to conclude the proof. But, as we have already noticed, we cannot control \( h = h(\tilde{\Theta}) \) and the Kronecker factor \( K \) of \((X_\tilde{\Theta}, \mu_{\tilde{\Theta}}, S)\) is given by \((H_{\lambda} \times \mathbb{Z}/h\mathbb{Z}, m_{H_{\lambda}} \otimes m_{\mathbb{Z}/h\mathbb{Z}}, R \times \tau_h)\). But \( B(V) \subset K \), so by a result of Veech [42], there is a compact subgroup \( \mathcal{V}_1 \subset V \) such that \( K \) is the \( \sigma \)-algebra of subsets fixed by all elements of \( \mathcal{V}_1 \). (We are actually able to give a concrete description of \( \mathcal{V}_1 \) as \( \{V_\tau : \tau \in G_0\} \), where \( G_0 \) was defined in Lemma 5.10.) So, we can now repeat all the above arguments (cf. Footnote 25) with \( V \) replaced by \( \mathcal{V}_1 \) to complete the proof.

Proof of Corollary 1. We begin with a 1-code \( F : X_\theta \to \mathbb{C} \) (that is, \( F(y) \) depends only on the 0-coordinate \( y[0] \) of \( y \in X_\theta \)). We assume that for some fixed point \( y \in X_\theta \), the sequence \( (F(S^ny))_{n \in \mathbb{N}} \) is multiplicative. Such an \( F \) is a continuous function and \((X_\theta, S)\) is a topological factor of \((X_\tilde{\Theta}, S)\), say \( \pi : X_\tilde{\Theta} \to X_\theta \) settle a factor map. Now, \( F \circ \pi \in C(X_\tilde{\Theta}) \) and (as in the proof of Theorem 0.1) we have \( F \circ \pi = F_1 + F_2 \), where both \( F_1, F_2 \) are continuous, \( F_1 \in L^2(B(V)) \) and \( F_2 \perp L^2(B(V)) \). Take \( x \in \pi^{-1}(y) \) that can be chosen as a fixed point of \( \tilde{\Theta} \). By

\(^{35}\) It has also \((X_\tilde{\Theta}, S)\) as its topological factor.
Theorem 0.1, \[
\frac{1}{N} \sum_{n \leq N} F_2(S^n x) F(S^n(\pi(x))) \to 0
\] since \((F(S^n x))\) is multiplicative. On the other hand, by unique ergodicity, \[
\frac{1}{N} \sum_{n \leq N} F_2(S^n x) F(S^n(\pi(x))) \to \int_{X_\Theta} F \circ \pi \, d\mu_\Theta.
\] It follows that \(F \circ \pi \perp F_2\), which implies \(F_2 = 0\). It follows that the spectral measure of \(F \circ \pi = F_1\) is discrete.

Let us first assume that \(h(\hat{\Theta}) = 1\). First we note that \(\pi\) is actually a 1-code, which shows that \(F \circ \pi\) is also a 1-code. Indeed, \(\pi\) is given by the composition \((\sigma_n, \tilde{M}_n)_{n \in \mathbb{Z}} \mapsto (i_n, \tilde{M}_n)_{n \in \mathbb{Z}} \mapsto (z_n)_{n \in \mathbb{Z}}\) where \(i_n\) is defined by \(\sigma(i_n) = 0\) and \(z_n\) is the \(i_n + 1\)-st element of \(\tilde{M}_n\) (see Remark 19 and Proposition 13). Furthermore, we have an explicit formula for \(F_1\):

\[
F_1(x) = \frac{1}{|G|} \sum_{\tau \in G} (F \circ \pi) \circ V_\tau(x).
\] (66)

As \(F \circ \pi\) is a 1-code and \(V_\tau\) is also a 1-code, it follows that \(F_1\) is a 1-code as well. Moreover, by (66), we obtain that

\[
F_1(\sigma, M) = \frac{1}{|G|} \sum_{\tau \in G} (F \circ \pi)(\tau \sigma, M) = \frac{1}{|G|} \sum_{\sigma' \in G} (F \circ \pi)(\sigma', M).
\]

Thus, \(F_1(\sigma, M)\) only depends on \(M\) and we see that (with \(y = (\sigma, M)\))

\[
F(S^n y) = F_1(S^n M) = F_1(M_n).
\]

As we have shown that \(F_1 \in C(X_\hat{\Theta})\) is a 1-code and \(\hat{\Theta}\) is synchronizing by Proposition 9 (i), the result follows.

Consider now the case \(h(\hat{\Theta}) \geq 1\). We recall that there exists some \(f : G \times \mathcal{X} \to \{0, \ldots, h - 1\}\) such that \(f(y|n]) \equiv n \operatorname{mod} h\). As we have seen at the end of Section 5.5, we can identify \(V_1\) as \(\{V_\tau : \tau \in G_0\}\). Now, we find that \(F_1(\sigma, M)\) depends on \(M\) and \(\sigma G_0\), or equivalently on \(M\) and \(f(\sigma, M) = f(\sigma G_0, M)\). This gives that

\[
F(S^n y) = F_1((n \operatorname{mod} h), M_n).
\]

As \((M_n)_{n \in \mathbb{N}}\) and \((n \operatorname{mod} h)_{n \in \mathbb{N}}\) are both WRAP (see Remark 17 and Proposition 4), we see directly that \((F_1((n \operatorname{mod} h), M_n))_{n \in \mathbb{N}}\) is also WRAP. Furthermore, we define a substitution \(p : \{0, \ldots, h - 1\} \to \{0, \ldots, h - 1\}\), by setting:

\[
p(i) = i\lambda + j \operatorname{mod} h,
\]

for all \(j < \lambda\) and \(i \in \{0, \ldots, h - 1\}\). We immediately compute that \(c(p) = h(p) = h\). Moreover, we find that \(((n \operatorname{mod} h), M_n))_{n \in \mathbb{N}}\) is the fixpoint of \(p \vee \tilde{\Theta}\). We find by Lemma 5.5 that

\[
\max(c(p), c(\tilde{\Theta})) \leq c(p \vee \tilde{\Theta}) \leq c(p)c(\tilde{\Theta})
\]

and since \(c(\tilde{\Theta}) = 1\), this gives \(c(p \vee \tilde{\Theta}) = h\). Furthermore, we have

\[
\text{lcm}(h(p), h(\tilde{\Theta}))(h(p \vee \tilde{\Theta})
\]

and as \(h(\tilde{\Theta}) = 1\) this gives that \(h \leq h(p \vee \tilde{\Theta})\). However, Lemma 2.2 implies that \(h(p \vee \tilde{\Theta}) \leq c(p \vee \tilde{\Theta})\) and therefore \(h(p \vee \tilde{\Theta}) = h\). It follows that \((F(S^n y))_{n \in \mathbb{N}}\) is given
as a 1-code of a fixpoint of a substitution of constant length for which the column number equals the height, or in other words (see Corollary 3, it has purely discrete spectrum.

Remark 23. If we want to obtain in the assertion a weaker conclusion, namely that such functions are Besicovitch rationally almost periodic, another proof of Corollary 1 can be obtained in the following way.

We first recall that if

\[ M := \{ v : \mathbb{N} \to \mathbb{C} : |v| \leq 1 \text{ and } v \text{ is multiplicative} \} \]

then to check that \( v \) has a mean along an arbitrary arithmetic progression, it is enough to show that \( v \cdot \chi \) has a mean for each Dirichlet character \( \chi \), see e.g. Proposition 3.1 in [5]. Now, the argument used in the proof of Lemma 2 in [40] works well and it shows that a function \( v \in M \) taking finitely many values has a mean along any arithmetic progression. Furthermore, Theorem 1.3 in [5] says that \( v \) is either Besicovitch rationally almost periodic or it is uniform. Moreover, uniformity is equivalent to aperiodicity by [15]. So if our \( v \) is not Besicovitch rationally almost periodic, then first \( \lim sup_{N \to \infty} \frac{1}{N} \sum_{n \leq N} |v_n| > 0 \) but also \( v \) is an aperiodic multiplicative automatic sequence, which is in conflict with Theorem 0.1.

7. Proof of Corollary 2. Let \( u : \mathbb{N} \to \mathbb{C} \) be a bounded multiplicative function. Following [1], we say that a topological dynamical system \((Y,T)\) satisfies the strong \( u \)-OMO property\(^{37}\) if for each increasing sequence \((b_k)_{k \geq 1} \subset \mathbb{N}, b_1 = 1, b_{k+1} - b_k \to \infty\), each sequence \((y_k) \subset Y\) and each \( f \in C(Y)\), we have

\[
\lim_{K \to \infty} \frac{1}{b_K} \sum_{k \leq K} \left| \sum_{b_k \leq n < b_{k+1}} f(T^n y_k) u(n) \right| = 0 \tag{67}
\]

Substituting \( f = 1 \) above, we see that \( u \) has to satisfy (4).

Clearly, given \((Y,T)\), (67) implies (2). In fact (for \( u = \mu \)), in the class of zero entropy systems, they are equivalent: it is proved in [1] that Sarnak’s conjecture holds if and only if all zero entropy systems enjoy the strong MOMO property.\(^{38}\)

Lemma 7.1. For each primitive substitution \( \theta \), the system \((X_\theta, S)\) has the strong \( u \)-OMO property for each bounded, aperiodic, multiplicative \( u : \mathbb{N} \to \mathbb{C} \) satisfying (4).

Proof. We only need to prove the result for \((X_\Theta, S)\). Then, the result follows immediately from Theorem 25 of [1] or, more precisely, from its proof. Indeed, all ergodic rotations satisfy the strong \( u \)-OMO property (Corollary 28 in [1]) and the only ergodic joinings between \( S^p \) and \( S^q \) whenever \( p \neq q \) are large enough, are relative products over isomorphisms of Kronecker factors. Moreover, the structure of the space of continuous functions allows us to represent each \( F \in C(X_\Theta) \) as \( F = F_1 + F_2 \), with both \( F_i \) continuous, \( F_1 \) measurable with respect to the Kronecker factor, and \( F_2 \) orthogonal to \( L^2 \) of that factor. Finally, we apply the reasoning from the proof of Theorem 25 of [1] to \( F_2 \). \( \square \)

\(^{36}\)Note that \( \mu^2 \) is Besicovitch rationally almost periodic but is not Weyl rationally almost periodic, e.g. [4].

\(^{37}\)Instead of the strong \( \mu \)-OMO property, we speak about the strong MOMO property. The acronym MOMO stands for Mӧbius Orthogonality of Moving Orbits.

\(^{38}\)Moreover, no positive entropy system has the strong MOMO property. We recall (see [11]) that there are positive entropy systems \((Y,T)\) satisfying (1) for all \( f \in C(Y)\).
Proof of Corollary 2. It follows from Lemma 7.1 and the equivalence of Properties 1 and 3 in Main Theorem in [1] that all MT-substitutional systems satisfy the strong \(u\)-OMO property for each bounded, aperiodic, multiplicative \(u\) satisfying (4). The uniform convergence follows now from Theorem 7 in [1].

Note that among MT-substitutional systems there are uniquely ergodic models which are topologically mixing [28]. In such models the maximal equicontinuous factor must be trivial, that is, there is no topological “realization” of the Kronecker factor.

Theorem 0.1 tells us that for each primitive substitution \(\theta\) and each bounded, multiplicative, aperiodic \(u\), (2) holds for each \(f \in C(X_\theta)\) and arbitrary \(x \in X_\theta\). We do not know however whether this convergence is uniform in \(x\). Note that Corollary 2 gives the positive answer if \(u\) satisfies additionally (4).

Remark 24. In the recent paper [25], it is proved that all \(q\)-multiplicative sequences are either almost periodic, which roughly would correspond to (i) of Theorem 0.1, or are orthogonal to all bounded, multiplicative functions. This makes some overlap (some automatic sequences are \(q\)-multiplicative) with our main result but the classes considered in [25] and in the present paper are essentially different.

Remark 25. Both of the main results of the paper, Theorem 0.1 and Corollary 1, remain valid if we drop the assumption of primitivity of substitutions. This will be treated in detail in a forthcoming paper.

Acknowledgments. The first author is supported by Narodowe Centrum Nauki grant UMO-2019/33/B/ST1/00364. The second author is supported by European Research Council (ERC) under the European Union’s Horizon 2020 research and innovation programme under the Grant Agreement No 648132, by the project F55-02 of the Austrian Science Fund FWF which is part of the Special Research Program “Quasi-Monte Carlo Methods: Theory and Applications” and by Project I1751 (FWF), called MUDERA (Multiplicativity, Determinism, and Randomness). The authors are very grateful to the anonymous referee for the careful reading of this paper and the constructive input.

REFERENCES

[1] H. El Abdalaoui, J. Kulaga-Przymus, M. Lemańczyk and T. de la Rue, Möbius disjointness for models of an ergodic system and beyond, Israel J. Math., 228 (2018), 707–751.
[2] J.-P. Allouche and L. Goldmakher, Mock characters and the Kronecker symbol, J. Number Theory, 192 (2018), 356–372.
[3] J.-P. Allouche and J. Shallit, Automatic Sequences. Theory, Applications, Generalizations, Cambridge University Press, Cambridge, 2003.
[4] V. Bergelson, J. Kulaga-Przymus, M. Lemańczyk and F. K. Richter, Rationally almost periodic sequences, polynomial multiple recurrence and symbolic dynamics, Ergodic Theory Dynam. Systems, 39 (2019), 2332–2383.
[5] V. Bergelson, J. Kulaga-Przymus, M. Lemańczyk and F. Richter, A structure theorem for level sets of multiplicative functions and applications International Math. Research Notices, 5 (2020), 1300–1345.
[6] J. Bourgain, P. Sarnak and T. Ziegler, Disjointness of Möbius from horocycle flows, From Fourier analysis and number theory to Radon transforms and geometry, Springer, New York, 2013, 67–83.
[7] A. Danienko and M. Lemańczyk, Spectral multiplicities for ergodic flows, Discrete Contin. Dyn. Syst. 33 (2013), 4271–4289.
[8] F. M. Dekking, The spectrum of dynamical systems arising from substitutions of constant length, Z. Wahrscheinlichkeitstheorie und Verw. Gebiete, 41 (1977/78), 221–239.
[9] M. Drmota, Subsequences of automatic sequences and uniform distribution, in *Uniform Distribution and Quasi-Monte Carlo Methods*, Radon Series in Computational and Applied Mathematics, Vol. 15, De Gruyter, Berlin, 2014, 87–104.

[10] J.-M. Deshouillers, M. Drmota and C. Mülhner, Automatic sequences generated by synchronizing automata fulfill the Sarnak conjecture, *Studia Math.*, 231 (2015), 83–95.

[11] T. Downarowicz and J. Serafin, Almost full entropy subshifts uncorrelated to the Möbius function, *Int. Math. Res. Not. IMRN*, (2019), no. 11, 3459–3472.

[12] S. Ferenczi, J. Kułaga-Przymus, M. Lemańczyk, Sarnak’s conjecture: what’s new, in *Ergodic Theory and Dynamical Systems in their Interactions with Arithmetic and Combinatorics*, Springer, Cham, 2018, 163–235.

[13] S. Ferenczi, J. Kułaga-Przymus, M. Lemańczyk and C. Mauduit, Substitutions and Möbius disjointness, in *Ergodic Theory, Dynamical Systems, and the Continuing Influence of John C. Oxtoby*, Amer. Math. Soc., Providence, RI, 2016, 151–174.

[14] N. P. Fogg, *Substitutions in Dynamics, Arithmetics and Combinatorics*, Lecture Notes in Mathematics, Vol. 1794, Springer-Verlag, Berlin, 2002.

[15] N. Frantzikinakis and B. Host, Higher order Fourier analysis of multiplicative functions and applications, *J. Amer. Math. Soc.* 30 (2017), 67–157.

[16] N. Frantzikinakis and B. Host, The logarithmic Sarnak conjecture for ergodic weights, *Ann. of Math.*, 187 (2018), 869–931.

[17] N. Frantzikinakis and B. Host, Furstenberg systems of bounded multiplicative functions and applications, *International Mathematics Research Notices*, 2 (2020), 1073–7928.

[18] H. Furstenberg, Disjointness in ergodic theory, minimal sets, and a problem in Diophantine approximation, *Math. Systems Theory* 1 (1967), 1–49.

[19] E. Glasner, *Ergodic theory via joinings*, in *Mathematical Surveys and Monographs*, Vol. 101, American Mathematical Society, Providence, RI, 2003.

[20] A. Gomilko, D. Kwietniak and M. Lemańczyk, Sarnak’s conjecture implies the Chowla conjecture along a subsequence, in *Ergodic Theory and Dynamical Systems in their Interactions with Arithmetic and Combinatorics*, Lecture Notes in Mathematics, Vol. 2213, Springer, Cham, 2018.

[21] J. L. Herning, *Spectrum and Factors of Substitution Dynamical Systems*, Ph.D dissertation, George Washington University, 2013, 93 pp.

[22] B. Host and F. Parreau, Homomorphismes entre systèmes dynamiques définis par substitutions *Ergodic Theory Dynam. Systems*, 9 (1989), 469–477.

[23] I. Kátai, A remark on a theorem of H. Daboussi, *Acta Math. Hungar.*, 47 (1986), 223–225.

[24] O. Klurman and P. Kurlberg, A note on multiplicative automatic sequences, II, *Bulletin of the London Mathematical Society*, 52 (2020), 185–188.

[25] J. Konieczny, Möbius orthogonality for $q$-multiplicative sequences, preprint, *arXiv:1808.06196*.

[26] J. Konieczny, On multiplicative automatic sequences, *Bulletin of the London Mathematical Society*, 52 (2020), 175–184.

[27] J. Kułaga-Przymus and M. Lemańczyk, The Möbius function and continuous extensions of rotations, *Monath. Math.*, 178 (2015), 553–582.

[28] E. Lehrer, Topological mixing and uniquely ergodic models, *Israel J. Math.*, 57 (1987), 239–255.

[29] M. Lemańczyk and M. K. Mentzen, On metric properties of substitutions, *Compositio Math.*, 65 (1988), 241–263.

[30] M. Lemańczyk and M. K. Mentzen, Compact subgroups in the centralizers of natural factors of an ergodic group extension of a rotation determine all factors, *Ergodic Theory Dynam. Systems*, 10 (1990), 763–776.

[31] K. Matomäki and M. Radziwill, Multiplicative functions in short intervals, *Ann. of Math.*, 183 (2016), 1015–1056.

[32] K. Matomäki, M. Radziwill and T. Tao, An averaged form of Chowla’s conjecture, *Algebra Number Theory*, 9 (2015), 2167–2196.

[33] C. Mauduit and J. Rivat, Sur un problème de Gelfond : La somme des chiffres des nombres premiers, *Ann. of Math.*, 171 (2010), 1591–1646.

[34] C. Mauduit and J. Rivat, Prime numbers along Rudin-Shapiro sequences, *J. Eur. Math. Soc.*, 17 (2015), 2595–2642.

[35] M. K. Mentzen, Invariant sub-$σ$-algebras for substitutions of constant length, *Studia Math.*, 92 (1989), 257–273.
[36] M. K. Mentzen, Ergodic properties of group extensions of dynamical systems with discrete spectra, *Studia Math.*, 101 (1991), 19–31.

[37] C. Müllner, Automatic sequences fulfill the Sarnak conjecture, *Duke Math. J.*, 166 (2017), 3219–3290.

[38] M. Queffélec, *Substitution Dynamical Systems - Spectral Analysis*, second edition, Lecture Notes in Mathematics, Vol. 1294, Springer-Verlag, Berlin, 2010.

[39] P. Sarnak, Three lectures on the Möbius function, randomness and dynamics, http://publications.ias.edu/sites/default/files/Mobius%20lectures%20Summer%202010.pdf.

[40] J.-C. Schlage-Puchta, Completely multiplicative automatic functions, *Integers*, 11 (2011), A31, 8 pp.

[41] T. Tao, The logarithmically averaged and non logarithmically averaged Chowla conjectures https://terrytao.wordpress.com/2017/10/20/the-logarithmically-averaged-and-non-logarithmically-averaged-chowla-conjectures/

[42] W. A. Veech, A criterion for a process to be prime, *Monatsh. Math.*, 94 (1982), 335–341.

[43] S. Yazdani, Multiplicative functions and k-automatic sequences, *J. Théor. Nombres Bordeaux*, 13 (2001), 651–658.

Received August 2019; revised April 2020.

E-mail address: mlem@mat.umk.pl
E-mail address: clemens.muellner@tuwien.ac.at