Profinite automata

Eric Rowland[*]
University of Liege, Belgium
Hofstra University, Hempstead, NY, USA

Reem Yassawi
Trent University, Peterborough, Canada
IRIF, CNRS UMR 8243, Université Paris-Diderot, France

November 12, 2016

Abstract

Many sequences of p-adic integers project modulo p^α to p-automatic sequences for every $\alpha \geq 0$. Examples include algebraic sequences of integers, which satisfy this property for every prime p, and some cocycle sequences, which we show satisfy this property for a fixed p. For such a sequence, we construct a profinite automaton that projects modulo p^α to the automaton generating the projected sequence. In general, the profinite automaton has infinitely many states. Additionally, we consider the closure of the orbit, under the shift map, of the p-adic integer sequence, defining a shift dynamical system. We describe how this shift is a letter-to-letter coding of a shift generated by a constant-length substitution defined on an uncountable alphabet, and we establish some dynamical properties of these shifts.

1 Introduction

A *substitution* (or *morphism*) on an alphabet $\mathcal{A}$ is a map $\theta : \mathcal{A} \to \mathcal{A}^*$, extended to $\mathcal{A}^{\mathbb{N}}$ by concatenation. A substitution is *length-k* (or *k-uniform*) if, for each $a \in \mathcal{A}$, the length of $\theta(a)$ is k. The extensive literature on substitutions has traditionally focused on the case where $\mathcal{A}$ is finite. Some exceptions include recent work, for example in [Fer06] and [Mau06]. Substitutions on a countably infinite alphabet have been used to describe lexicographically least sequences on $\mathbb{N}$ avoiding certain patterns [GPS09, RS12], and they have been used in the combinatorics literature to enumerate permutations avoiding patterns [Wes96].

[*] Supported by a Marie Curie Actions COFUND fellowship.

In this article we present a new construction for constant-length substitutions on an uncountable alphabet. Our motivation comes from the following classical results. Let $(a(n))_{n\geq 0}$ be an *automatic sequence* (see Definition 2.2). Cobham's theorem (Theorem 2.6) characterizes an automatic sequence as the coding, under a letter-to-letter map, of a fixed point of a constant-length substitution. Christol's theorem (Theorem 2.8) characterizes p-automatic sequences for prime p; they are precisely the sequences whose generating function is algebraic over a finite field of characteristic p. The following can be viewed as a generalization of one direction of Christol's characterization.

Theorem 1.1 ([Chr74, Theorem 32], [DL87, Theorem 3.1]). *Let $(a(n))_{n\geq 0}$ be a sequence of p-adic integers such that $\sum_{n\geq 0} a(n)x^n$ is algebraic over $\mathbb{Z}_p(x)$, and let $\alpha \geq 0$. Then $(a(n) \bmod p^\alpha)_{n\geq 0}$ is p-automatic.*

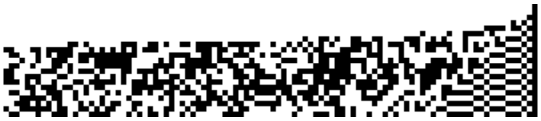
Thus certain p-adic integer sequences (and, in particular, integer sequences) have the property that they become p-automatic when reduced modulo p^α, for every $\alpha \geq 0$. More generally, the diagonal of a multivariate rational power series is p-automatic when reduced modulo p^α, and one can explicitly compute an automaton for $(a(n) \bmod p^\alpha)_{n\geq 0}$ for all but finitely many primes p [RY15, Theorem 2.1].

Fix a prime p, and let $(a(n))_{n\geq 0}$ be a sequence such that $(a(n) \bmod p^\alpha)_{n\geq 0}$ is p-automatic for every $\alpha \geq 0$. For each α, there is a finite automaton generating $(a(n) \bmod p^\alpha)_{n\geq 0}$. In Lemma 3.1 we show that these automata can be chosen in a compatible way; namely, their inverse limit exists. In this way we obtain a *profinite automaton* (Definition 3.3) generating the sequence $(a(n))_{n\geq 0}$.

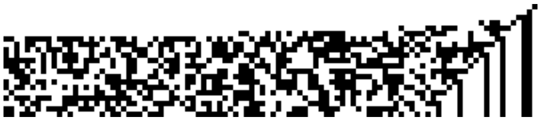
We can obtain other inverse limit objects from a p-adic integer sequence in a similar way. In particular, Cobham's theorem guarantees a length-p substitution θ_α such that $(a(n) \bmod p^\alpha)_{n\geq 0}$ is a coding of a fixed point of θ_α. Each substitution θ_α is a substitution on a finite alphabet, but their inverse limit is a profinite substitution on an alphabet that is, in general, uncountable (Theorem 4.2). This alphabet has a natural coding to the set $\mathbb{Z}_p$ of p-adic integers, and the sequence $(a(n))_{n\geq 0}$ is the coding of a fixed point of the profinite substitution.

With this profinite substitution, we obtain a shift (Theorem 4.1), as in the classical finite-alphabet case. This shift is the closure of the orbit, under the shift map, of a fixed point (or coding of a fixed point) of the profinite substitution. One feature of profinite substitutions is that their shifts live in a compact topological space, and many of the classical results on primitive substitution shifts carry through to our setting. For example, if we assume that each substitution θ_α is primitive, then the profinite shift is both minimal (Proposition 4.3) and uniquely ergodic (Corollary 4.6). With the same assumption, the maximal equicontinuous factor of the profinite shift is an odometer, and any measurable eigenvalue is continuous (Theorem 4.7). Finally, profinite substitutions are recognizable (Theorem 4.9). However it is not clear when a p-adic integer sequence generates primitive substitutions θ_α. In one example we consider, not only are the substitutions θ_1 and θ_2 not primitive, but their shifts contain shift-periodic sequences (Example 3.2).

In Section 4.3, we discuss how profinite substitutions allow us to view certain p-adic limits in terms of attractor sets for dynamical systems. For example, consider the Fibonacci sequence $F(n)_{n\geq 0} = 0, 1, 1, 2, 3, 5, 8, 13, \dots$. The following graphic shows the hundred least significant binary digits of $F(2^n)$ for each $0 \leq n \leq 20$, where 0 is represented by a white cell, 1 is represented by a black cell, and digits increase in significance to the left.



As 2-adic integers, the two limit points of the sequence $F(2^n)_{n\geq 0}$ are $\pm\sqrt{-\frac{3}{5}}$ in $\mathbb{Z}_2$ [RY16]. Similar behavior is seen for the sequence $C(n)_{n\geq 0} = 1, 1, 2, 5, 14, 42, 132, 429, \dots$ of Catalan numbers, where $C(n) = \frac{1}{n+1}\binom{2n}{n}$. The following shows the binary digits of $C(2^n)$ for $0 \leq n \leq 20$.



As this graphic suggests, the sequence $C(2^n)_{n\geq 0}$ converges in $\mathbb{Z}_2$ [MMR14]. These limits are elements in the limit sets of certain profinite shifts. In particular, such a limit is a component of an inverse limit of θ_α-periodic points.

The outline of the article is as follows. In Section 2 we recall the major facts about automatic sequences that we shall use. In Section 3 we construct profinite automata as inverse limits of finite automata. In Section 4 we define profinite substitutions and their shifts, and we establish various dynamical properties of the latter. In Section 4.4 we study a family of sequences that arise in a dynamical context (from cocycle maps), which are not algebraic, but which for a single prime p generate an inverse limit substitution dynamical system (Theorem 4.12).

2 Background on automatic sequences

In this section we establish notation and give the necessary properties of automatic sequences.

2.1 Finite automata and automatic sequences

We first give the formal definition of an automaton.

Definition 2.1. Let $k \geq 2$. A *k-deterministic automaton with output* (k-DAO) is a 6-tuple $(\mathcal{S}, \Sigma_k, \delta, \overline{s}, \mathcal{A}, \tau)$, where $\mathcal{S}$ is a set of "states", $\overline{s} \in \mathcal{S}$ is the *initial state*, $\Sigma_k = \{0, 1, \ldots, k-1\}$, $\mathcal{A}$ is an alphabet, $\tau : \mathcal{S} \to \mathcal{A}$ is the *output function*, and $\delta : \mathcal{S} \times \Sigma_k \to \mathcal{S}$ is the *transition function*.

In this section we shall be concerned with *finite* automata.

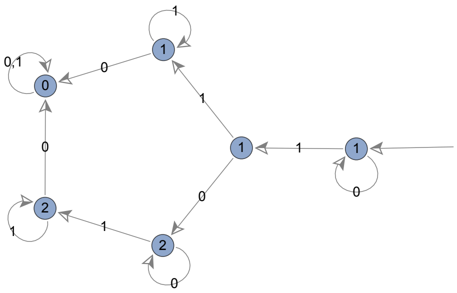
Definition 2.2. Let $k \geq 2$. A *k-deterministic finite automaton with output* (k-DFAO) is a k-DAO whose set $\mathcal{S}$ of states is finite.

The function δ extends to the domain $\mathcal{S} \times \Sigma_k^*$ by defining $\delta(s, \epsilon) := s$ for the empty word ϵ and recursively defining $\delta(s, n_\ell n_{\ell-1} \cdots n_0) := \delta(\delta(s, n_\ell), n_{\ell-1} \cdots n_0)$. Given a natural number n and an integer $k \geq 2$, we write $\operatorname{rep}_k(n) = n_\ell \cdots n_1 n_0 \in \Sigma_k^*$ for the standard base-k representation of n where $n = n_\ell k^\ell + \cdots + n_1 k + n_0$ and $n_\ell \neq 0$. We can feed $\operatorname{rep}_k(n)$, beginning with the most significant digit n_ℓ, into an automaton as follows. (Recall that the standard base-k representation of 0 is the empty word.)

Definition 2.3. A sequence $(a(n))_{n \geq 0}$ of elements in $\mathcal{A}$ is *k-automatic* if there is a k-DFAO $\mathcal{M} = (\mathcal{S}, \Sigma_k, \delta, \overline{s}, \mathcal{A}, \tau)$ such that $a(n) = \tau(\delta(\overline{s}, \operatorname{rep}_k(n)))$ for each $n \geq 0$.

For ease of notation we will write $a(n)_{n \geq 0}$ for $(a(n))_{n \geq 0}$. When we obtain an automatic sequence $a(n)_{n \geq 0}$ as in Definition 2.3 by reading the most significant digit first, we say that the automaton generates $a(n)_{n \geq 0}$ in *direct reading*. In this article our automata will always generate sequences in direct reading. (This is in contrast to the automata in [RY15], where our algorithms produced machines that generated the desired sequence by reading the least significant digit first.)

Example 2.4. Consider the following automaton for $k = 2$. Each of the six states is represented by a vertex, labelled with its output under τ. Edges between vertices illustrate δ. The unlabelled edge points to the initial state.



The 2-automatic sequence produced by this automaton in direct reading is

$$a(n)_{n \geq 0} = 1, 1, 2, 1, 2, 2, 0, 1, 2, 2, 0, 2, 0, 0, 0, 1, \ldots.$$

2.2 Some characterizations of automaticity

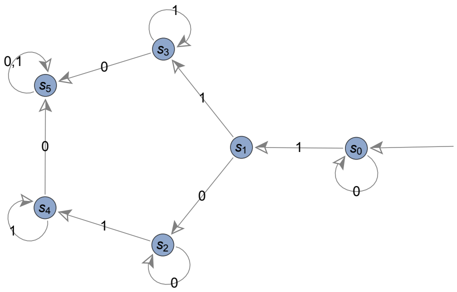
Definition 2.5. Let $\mathcal{A}$ be an alphabet, which we do not assume to be finite. A *substitution* is a map $\theta : \mathcal{A} \to \mathcal{A}^*$. The map θ extends to maps $\theta : \mathcal{A}^* \to \mathcal{A}^*$ and $\theta : \mathcal{A}^{\mathbb{N}} \to \mathcal{A}^{\mathbb{N}}$ by concatenation: $\theta(a(n)_{n\geq 0}) := \theta(a(0)) \cdots \theta(a(k)) \cdots$. If there is some k with $|\theta(a)| = k$ for each $a \in \mathcal{A}$, then we say that θ is a *length-k* substitution. A *fixed point* of θ is a sequence $a(n)_{n\geq 0} \in \mathcal{A}^{\mathbb{N}}$ such that $\theta(a(n)_{n\geq 0}) = a(n)_{n\geq 0}$. Let $\tau : \mathcal{A} \to \mathcal{B}$ be a map; it induces a *letter-to-letter projection* $\tau : \mathcal{A}^{\mathbb{N}} \to \mathcal{B}^{\mathbb{N}}$.

Given a letter s such that $\theta(s)$ starts with s, the words $\theta^n(s)$ converge, as $n \to \infty$, to a fixed point of θ. We recall Cobham's theorem [Cob72], which gives us the following characterization of automatic sequences.

Theorem 2.6. *A sequence is k-automatic if and only if it is the image, under a letter-to-letter projection, of a fixed point of a length-k substitution on a finite alphabet.*

If a sequence $a(n)_{n\geq 0}$ is generated in direct reading by $\mathcal{M} = (\mathcal{S}, \Sigma_k, \delta, \overline{s}, \mathcal{A}, \tau)$, then the transition map δ gives us the substitution described by Cobham's theorem. Namely, let $\theta(s) := \delta(s, 0)\delta(s, 1) \cdots \delta(s, k-1)$ for each state $s \in \mathcal{S}$. If $\delta(\overline{s}, 0) \neq \overline{s}$, introduce a new letter $\overline{s}'$ and let $\theta(\overline{s}') := \overline{s}'\delta(\overline{s}, 1) \cdots \delta(\overline{s}, k-1)$.

Example 2.7. Let us construct a substitution θ and projection τ that generate the sequence produced by the automaton $\mathcal{M}$ in Example 2.4 in direct reading. Removing the output function from $\mathcal{M}$ gives the following automaton, where we have named the states $s_0, s_1, \ldots, s_5$.



This automaton dictates that θ is the length-2 substitution on the alphabet $\{s_0, s_1, s_2, s_3, s_4, s_5\}$ defined by

$$\begin{aligned} \theta(s_0) &= s_0 s_1 & \theta(s_3) &= s_5 s_3 \\ \theta(s_1) &= s_2 s_3 & \theta(s_4) &= s_5 s_4 \\ \theta(s_2) &= s_2 s_4 & \theta(s_5) &= s_5 s_5. \end{aligned}$$

The sequence

$$\theta^\infty(s_0) = s_0, s_1, s_2, s_3, s_2, s_4, s_5, s_3, s_2, s_4, s_5, s_4, s_5, s_5, s_5, s_3, \ldots$$

is a fixed point of θ. The letter-to-letter projection τ is the output function for $\mathcal{M}$, namely

$$\begin{aligned} \tau(s_5) &= 0 \\ \tau(s_0) = \tau(s_1) = \tau(s_3) &= 1 \\ \tau(s_2) = \tau(s_4) &= 2. \end{aligned}$$

Therefore

$$\tau(\theta^\infty(s_0)) = 1, 1, 2, 1, 2, 2, 0, 1, 2, 2, 0, 2, 0, 0, 0, 1, \ldots$$

is the sequence produced by $\mathcal{M}$ in direct reading.

While Theorem 2.6 characterizes k-automatic sequences for all $k \geq 2$, Christol's theorem characterizes p-automatic sequences for prime p. By taking a sufficiently large finite field of characteristic p, we may assume (by choosing an arbitrary embedding) that the output alphabet is a subset of this field.

Theorem 2.8 (Christol et al. [CKMFR80]). *Let $a(n)_{n\geq 0}$ be a sequence of elements in $\mathbb{F}_{p^\alpha}$. Then $\sum_{n\geq 0} a(n)x^n$ is algebraic over $\mathbb{F}_{p^\alpha}(x)$ if and only if $a(n)_{n\geq 0}$ is p-automatic.*

3 Profinite automata generated by automatic sequences

In this section we study infinite automata that are inverse limits of finite automata.

We use the following notation for the remainder of the article. For $0 \leq \alpha \leq \beta$, let $\pi_{\alpha,\beta} : \mathbb{Z}/(p^\beta\mathbb{Z}) \to \mathbb{Z}/(p^\alpha\mathbb{Z})$ denote the natural projection map modulo p^α. The ring of p-adic integers is the inverse limit $\mathbb{Z}_p = \varprojlim \mathbb{Z}/(p^\alpha\mathbb{Z})$ in the category of rings. Let $\pi_{\alpha,\infty}$ be the projection map $\pi_{\alpha,\infty} : \mathbb{Z}_p \to \mathbb{Z}/(p^\alpha\mathbb{Z})$.

Two k-DAO's are *equivalent* if they determine the same function $\Sigma_k^* \to \mathcal{A}$ given by $w \mapsto \tau(\delta(\overline{s}, w))$. A k-DFAO with the fewest states in its equivalence class is said to be *minimal*. Given a k-DFAO $\mathcal{M}$, there is a unique minimal automaton equivalent to $\mathcal{M}$, up to renaming states [Sha08, Theorem 3.10.1]. The minimal automaton can be computed by removing unreachable states and successively identifying pairs of equivalent states, that is, states such that starting in either state and reading a word produces the same output. Note that this algorithm is typically described in the literature for k-DFA's — automata with output alphabet $\{\mathsf{True}, \mathsf{False}\}$ — but is generalized in a straightforward way to k-DFAO's.

A *morphism* from a k-DAO $\mathcal{M} = (\mathcal{S}, \Sigma_k, \delta, \overline{s}, \mathcal{A}, \tau)$ to a k-DAO $\mathcal{M}' = (\mathcal{S}', \Sigma_k, \delta', \overline{s}', \mathcal{A}', \tau')$ consists of maps $\psi : \mathcal{S} \to \mathcal{S}'$ and $\pi : \mathcal{A} \to \mathcal{A}'$ such that $\psi(\overline{s}) = \overline{s}'$, $\psi(\delta(s, i)) =$

$\delta'(\psi(s), i)$, and $\pi(\tau(s)) = \tau'(\psi(s))$ for all $s \in \mathcal{S}$ and $i \in \Sigma_k$. In this article, the map π will always be one of the projection maps $\pi_{\alpha,\beta}$.

Suppose that $(\mathcal{M}_\alpha)_{\alpha \geq 0} = (\mathcal{S}_\alpha, \Sigma_p, \delta_\alpha, \overline{s}_\alpha, \mathbb{Z}/(p^\alpha\mathbb{Z}), \tau_\alpha)_{\alpha \geq 0}$ is a sequence of finite automata such that, for each $0 \leq \alpha \leq \beta$, there is a morphism from $\mathcal{M}_\beta$ to $\mathcal{M}_\alpha$, consisting of a map $\psi_{\alpha,\beta} : \mathcal{S}_\beta \to \mathcal{S}_\alpha$ and the projection map $\pi_{\alpha,\beta}$, with the property that

$$\psi_{\alpha,\gamma} = \psi_{\alpha,\beta} \circ \psi_{\beta,\gamma} \tag{1}$$

if $0 \leq \alpha \leq \beta \leq \gamma$. Then we call the sequence $(\mathcal{M}_\alpha)_{\alpha \geq 0}$ an *inverse family of finite automata*. We remark that our automata have output in $\mathbb{Z}/(p^\alpha\mathbb{Z})$, because of our motivating examples, but a more general definition of an inverse family of finite automata is possible.

Lemma 3.1. *Let $a \in \mathbb{Z}_p^{\mathbb{N}}$ be a sequence of p-adic integers such that $a \bmod p^\alpha$ is p-automatic for each $\alpha \geq 0$. Let $\mathcal{M}_\alpha$ be the minimal automaton generating $a \bmod p^\alpha$ in direct reading. Then $(\mathcal{M}_\alpha)_{\alpha \geq 0}$ is an inverse family of finite automata.*

Proof. Let $\mathcal{M}_\alpha = (\mathcal{S}_\alpha, \Sigma_p, \delta_\alpha, \overline{s}_\alpha, \mathbb{Z}/(p^\alpha\mathbb{Z}), \tau_\alpha)$, and let $\mathcal{M}'_\alpha = (\mathcal{S}_{\alpha+1}, \Sigma_p, \delta_{\alpha+1}, \overline{s}_{\alpha+1}, \mathbb{Z}/(p^\alpha\mathbb{Z}), \pi_{\alpha,\alpha+1} \circ \tau_{\alpha+1})$ be the automaton obtained by replacing the output function $\tau_{\alpha+1}$ in $\mathcal{M}_{\alpha+1}$ with the map $s \mapsto (\tau_{\alpha+1}(s) \bmod p^\alpha)$. Then $\mathcal{M}'_\alpha$ and $\mathcal{M}_\alpha$ are equivalent. Since $\mathcal{M}_\alpha$ is minimal, minimizing $\mathcal{M}'_\alpha$ gives $\mathcal{M}_\alpha$. Since minimizing can be accomplished by successively identifying pairs of equivalent states, each element of $\mathcal{S}_{\alpha+1}$ can be assigned to a unique element of $\mathcal{S}_\alpha$; let $\psi_{\alpha,\alpha+1}$ be this map. We claim that $\psi_{\alpha,\alpha+1}$ and $\pi_{\alpha,\alpha+1}$ comprise a morphism from $\mathcal{M}_{\alpha+1}$ to $\mathcal{M}_\alpha$. Minimizing an automaton maps the initial state to the initial state, so $\psi_{\alpha,\alpha+1}(\overline{s}_{\alpha+1}) = \overline{s}_\alpha$. Minimizing preserves edge relations, so $\psi_{\alpha,\alpha+1}(\delta_{\alpha+1}(s, i)) = \delta_\alpha(\psi_{\alpha,\alpha+1}(s), i)$ for every $s \in \mathcal{S}_{\alpha+1}$ and $i \in \Sigma_k$. Finally, because of our choice of output function for $\mathcal{M}'_\alpha$, we have $\pi_{\alpha,\alpha+1}(\tau_{\alpha+1}(s)) = \tau_\alpha(\psi_{\alpha,\alpha+1}(s))$ by the equivalence of $\mathcal{M}_\alpha$ and $\mathcal{M}'_\alpha$. Thus $\psi_{\alpha,\alpha+1}$ and $\pi_{\alpha,\alpha+1}$ comprise a morphism from $\mathcal{M}_{\alpha+1}$ to $\mathcal{M}_\alpha$. Composing the maps $\psi_{\alpha,\alpha+1}$ gives maps $\psi_{\alpha,\beta}$ with the desired composition rule. □

Example 3.2. We illustrate Lemma 3.1 by computing the projection map $\psi_{1,2}$ for the sequence $C(n)_{n \geq 0}$ of Catalan numbers $C(n) = \frac{1}{n+1}\binom{2n}{n}$. Catalan numbers are ubiquitous in combinatorial settings. The generating function $y = \sum_{n \geq 0} C(n)x^n$ is algebraic and satisfies $xy^2 - y + 1 = 0$. By Theorem 1.1, the sequence $(C(n) \bmod p^\alpha)_{n \geq 0}$ is p-automatic for each prime p and each $\alpha \geq 0$. Let $p = 2$. For $\alpha = 2$, the automaton $\mathcal{M}_2 = (\mathcal{S}_2, \Sigma_2, \delta_2, \overline{s}_2, \mathbb{Z}/(4\mathbb{Z}), \tau_2)$ appears in Example 2.4. It produces the sequence

$$C_2 = (C(n) \bmod 4)_{n \geq 0} = 1, 1, 2, 1, 2, 2, 0, 1, 2, 2, 0, 2, 0, 0, 0, 1, \ldots.$$

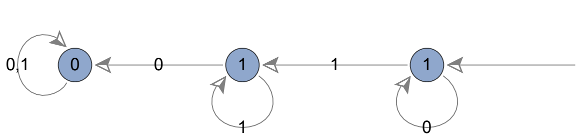
The automaton in Example 2.7 is $(\mathcal{S}_2, \Sigma_2, \delta_2, \overline{s}_2, \mathcal{S}_2, \mathrm{id})$, where the output function is the identity map and $s_0 = \overline{s}_2$. The sequence u_2 generated by this automaton is

$$u_2 = s_0, s_1, s_2, s_3, s_2, s_4, s_5, s_3, s_2, s_4, s_5, s_4, s_5, s_5, s_5, s_3, \ldots.$$

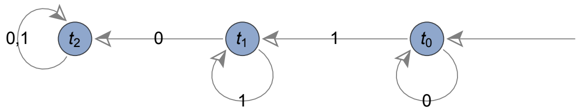
For $\alpha = 1$, the sequence

$$C_1 = (C(n) \bmod 2)_{n\geq 0} = 1, 1, 0, 1, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 1, \ldots$$

is generated by $\mathcal{M}_1 = (\mathcal{S}_1, \Sigma_2, \delta_1, \overline{s}_1, \mathbb{Z}/(2\mathbb{Z}), \tau_1)$, which is the following.



The automaton $(\mathcal{S}_1, \Sigma_2, \delta_1, \overline{s}_1, \mathcal{S}_1, \mathrm{id})$ is as follows, where we name the states t_0, t_1, t_2 and $t_0 = \overline{s}_1$.



The states s_2, s_4, s_5 in $\mathcal{M}_2$ correspond to output congruent to 0 or 2 modulo 4; since there is only one state in $\mathcal{M}_1$ whose output is 0 modulo 2, it follows that

$$\psi_{1,2}(s_2) = \psi_{1,2}(s_4) = \psi_{1,2}(s_5) = t_2.$$

One checks that $\psi_{1,2}(s_1) = \psi_{1,2}(s_3) = t_1$, and on the initial state we have $\psi_{1,2}(s_0) = t_0$.

We now give the definition of a profinite automaton. Readers familiar with inverse limits will recognize profinite automata as inverse limits in the category of p-DAO's with automaton morphisms.

Definition 3.3. Let $(\mathcal{M}_\alpha)_{\alpha\geq 0} = (\mathcal{S}_\alpha, \Sigma_p, \delta_\alpha, \overline{s}_\alpha, \mathbb{Z}/(p^\alpha\mathbb{Z}), \tau_\alpha)_{\alpha\geq 0}$ be an inverse family of finite automata with maps $\psi_{\alpha,\beta} : \mathcal{S}_\beta \to \mathcal{S}_\alpha$ and projection maps $\pi_{\alpha,\beta}$. We call $\mathcal{M} = (\mathcal{S}, \Sigma_p, \delta, \overline{s}, \mathbb{Z}_p, \tau)$ a *profinite automaton* if

1. $\mathcal{S} = \{(s_\alpha)_{\alpha\geq 0} \in \prod_{\alpha\geq 0} \mathcal{S}_\alpha : \psi_{\alpha,\beta}(s_\beta) = s_\alpha$ for each $0 \leq \alpha \leq \beta\}$,

2. $(\overline{s})_\alpha = \overline{s}_\alpha$ for each $\alpha \geq 0$,

3. $(\delta(s,i))_\alpha = \delta_\alpha(s_\alpha, i)$ for each $\alpha \geq 0$, and

4. $\pi_{\alpha,\infty}(\tau(s)) = \tau_\alpha(s_\alpha)$ for each $\alpha \geq 0$.

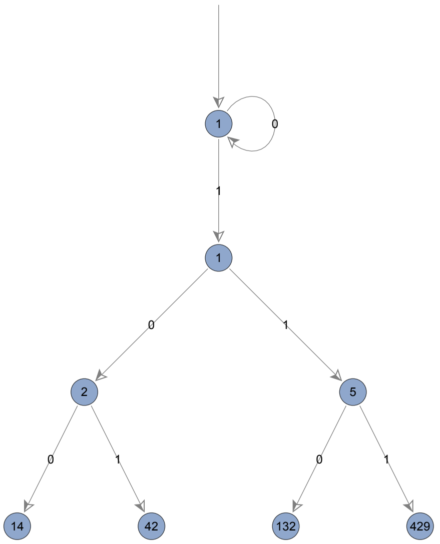
Note that the ranges of the maps δ and τ are in $\mathcal{S}$ and $\mathbb{Z}_p$ respectively because the maps $\psi_{\alpha,\beta}$ satisfy Equation (1). We define $\psi_{\alpha,\infty} : \mathcal{S} \to \mathcal{S}_\alpha$ by $\psi_{\alpha,\infty}(s) := s_\alpha$. Profinite automata satisfy the following commutative diagram.

$$\begin{tikzcd}
\mathcal{S} \arrow[dr, "\psi_{\beta,\infty}"'] \arrow[drr, "\psi_{\alpha,\infty}"] \arrow[ddd, "\tau"'] & & \\
& \mathcal{S}_\beta \arrow[r, "\psi_{\alpha,\beta}"'] \arrow[d, "\tau_\beta"'] & \mathcal{S}_\alpha \arrow[d, "\tau_\alpha"'] \\
& \mathbb{Z}/(p^\beta\mathbb{Z}) \arrow[r, "\pi_{\alpha,\beta}"] & \mathbb{Z}/(p^\alpha\mathbb{Z}) \\
\mathbb{Z}_p \arrow[ur, "\pi_{\beta,\infty}"] \arrow[urr, "\pi_{\alpha,\infty}"'] & &
\end{tikzcd}$$

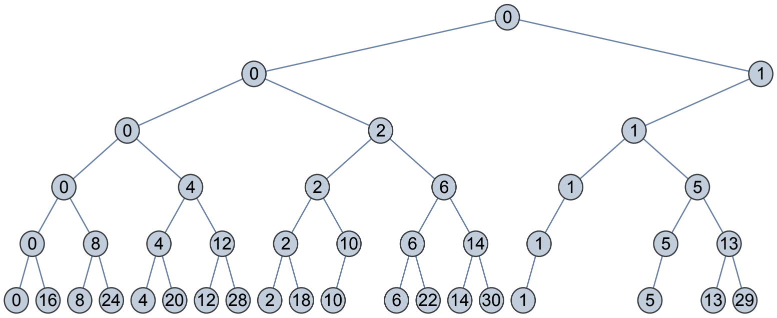
One can alternatively define profinite automata using reverse reading rather than direct reading.

By Theorem 1.1 and Lemma 3.1, algebraic sequences of p-adic integers are generated by profinite p-DAO's. But also, Theorem 1.1 has a converse in [DL87, Theorem 3.1]: a sequence generated by a profinite p-DAO is, for each α, congruent modulo p^α to an algebraic sequence of p-adic integers.

Example 3.4. The following shows the first few states in a profinite 2-DAO generating the sequence of Catalan numbers. This automaton is equivalent modulo 4 to the automaton $\mathcal{M}_2$ in Example 2.4 and equivalent modulo 2 to the automaton $\mathcal{M}_1$ in Example 3.2.



A profinite automaton can have uncountably many states, since in general each element of $\tau(\mathcal{S})$ corresponds to an infinite path in a rooted tree that encodes the residues attained by a sequence modulo each p^α. Again let $C(n) = \frac{1}{n+1}\binom{2n}{n}$ be the nth Catalan number, and let $p = 2$. Consider the rooted tree in which the vertices on level α consist of all residues j modulo 2^α such that $C(n) \equiv j \mod 2^\alpha$ for some $n \geq 0$. Two vertices on adjacent levels are connected by an edge if one residue projects to the other. Levels 0 through 5 of this tree are shown below. In particular, $C(n) \not\equiv 3 \mod 4$ for all $n \geq 0$; therefore there are only 3 vertices on level $\alpha = 2$.



The residues attained by the sequence of Catalan numbers modulo 2^α can be computed automatically [RY15, Section 3.1]. The number of states in the minimal automaton $\mathcal{M}_\alpha$ is bounded below by the number of vertices on level α.

We will henceforth use the notation $\varprojlim$ to denote the inverse limit of an inverse family. Thus for example, we shall write $\mathcal{M} = \varprojlim \mathcal{M}_\alpha$ to denote the inverse limit of an inverse family of minimal automata, and $s = \varprojlim s_\alpha$ to denote an element in the state set $\mathcal{S}$.

We say that a p-DAO is *minimal* if it contains no unreachable states and no pair of equivalent states. For finite automata, this definition of minimality coincides with the earlier definition. Up to renaming of states, profinite minimal automata have the characterization given by Theorem 3.6 below. For a p-DAO $\mathcal{M} = (\mathcal{S}, \Sigma_p, \delta, \overline{s}, \mathbb{Z}_p, \tau)$, let us use $\mathcal{M} \bmod p^\alpha$ as an abbreviation for $(\mathcal{S}, \Sigma_p, \delta, \overline{s}, \mathbb{Z}/(p^\alpha\mathbb{Z}), \pi_{\alpha,\infty} \circ \tau)$. Let $\mathcal{M}_\alpha = (\mathcal{S}_\alpha, \Sigma_p, \delta_\alpha, \overline{s}_\alpha, \mathbb{Z}/(p^\alpha\mathbb{Z}), \tau_\alpha)$ be the minimal automaton equivalent to $\mathcal{M} \bmod p^\alpha$. For each $0 \leq \alpha \leq \beta$, the automaton $\mathcal{M}_\beta$ projects modulo p^α to $\mathcal{M}_\alpha$. We say that $\mathcal{S}$ is *closed under inverse limits* if every inverse family $(s_\alpha)_{\alpha \geq 0}$, where $s_\alpha \in \mathcal{S}_\alpha$, has an inverse limit $s \in \mathcal{S}$. That is, for every sequence of states $(s_\alpha)_{\alpha \geq 0}$ where $s_\alpha \in \mathcal{S}_\alpha$ and $\pi_{\alpha,\beta}(\tau_\beta(\delta_\beta(s_\beta, w))) = \tau_\alpha(\delta_\alpha(s_\alpha, w))$ for all $0 \leq \alpha \leq \beta$ and $w \in \Sigma_p^*$, there exists a state $s \in \mathcal{S}$ such that $\pi_{\alpha,\infty}(\tau(\delta(s, w))) = \tau_\alpha(\delta_\alpha(s_\alpha, w))$ for all $\alpha \geq 0$ and for all $w \in \Sigma_p^*$.

Example 3.5. Let $p = 2$, and for each $\alpha \geq 0$ let $\iota : \mathbb{Z}/(2^\alpha\mathbb{Z}) \to \mathbb{Z}_2$ be the lifting defined by $\iota(n + 2^\alpha\mathbb{Z}) = n + 0 \cdot 2^\alpha + 0 \cdot 2^{\alpha+1} + \cdots$ for each $0 \leq n \leq 2^\alpha - 1$. Again let

$C(n)$ denote the nth Catalan number. For $n_\ell \cdots n_1 n_0 \in \Sigma_2^*$, let $\operatorname{val}_2(n_\ell \cdots n_1 n_0) = n_\ell 2^\ell + \cdots + n_1 2 + n_0$. Consider the minimal 2-DAO $\mathcal{M} = (\mathcal{S}, \Sigma_2, \delta, \overline{s}, \mathbb{Z}_2, \tau)$ defined for $w \in \Sigma_2^*$ by

$$\tau(\delta(\overline{s}, w)) = \begin{cases} \iota(C(\operatorname{val}_2(v)) \bmod 2^\alpha) & \text{if } w = 1^\alpha 0 v \text{ for some } \alpha \geq 1 \text{ and } v \in \Sigma_2^* \\ 0 & \text{otherwise.} \end{cases}$$

Since $1^\alpha 0$ is a prefix of w for at most one value of α, the automaton $\mathcal{M}$ is not over-determined. The set $\mathcal{S}$ is not closed under inverse limits, since, for each $\alpha \geq 0$, the sequence $(C(n) \bmod 2^\alpha)_{n \geq 0}$ occurs in $\mathcal{M}_\alpha$ (i.e. is obtained by reading words v starting from the state $\delta_\alpha(\overline{s}_\alpha, 1^\alpha 0)$), but $C(n)_{n \geq 0}$ does not occur in $\mathcal{M}$. This example shows that the condition that $\mathcal{S}$ is closed under inverse limits is necessary in the following theorem.

Theorem 3.6. *Let $\mathcal{M} = (\mathcal{S}, \Sigma_p, \delta, \overline{s}, \mathbb{Z}_p, \tau)$ be a minimal p-DAO. Then $\mathcal{M}$ is profinite if and only if*

1. *the automaton $(\mathcal{S}, \Sigma_p, \delta, \overline{s}, \mathbb{Z}/(p^\alpha \mathbb{Z}), \pi_{\alpha,\infty} \circ \tau)$ is equivalent to a finite automaton for each $\alpha \geq 0$, and*
2. *$\mathcal{S}$ is closed under inverse limits.*

Proof. The forward direction is clear: Since $\mathcal{M}$ is profinite, it has an inverse family $(\mathcal{M}_\alpha)_{\alpha \geq 0}$ of finite automata, and $\mathcal{M} \bmod p^\alpha$ is equivalent to $\mathcal{M}_\alpha$. Moreover, $\mathcal{S}$ is closed under inverse limits by Part 1 of Definition 3.3.

For the other direction, let $\mathcal{M}_\alpha = (\mathcal{S}_\alpha, \Sigma_p, \delta_\alpha, \overline{s}_\alpha, \mathbb{Z}/(p^\alpha \mathbb{Z}), \tau_\alpha)$ be the minimal finite automaton equivalent to $\mathcal{M} \bmod p^\alpha$. For each $0 \leq \alpha \leq \beta$, the automaton $\mathcal{M}_\beta$ projects modulo p^α to $\mathcal{M}_\alpha$, so there is a unique map $\psi_{\alpha,\beta} : \mathcal{S}_\beta \to \mathcal{S}_\alpha$ such that $\psi_{\alpha,\beta}$ and $\pi_{\alpha,\beta}$ comprise a morphism from $\mathcal{M}_\beta$ to $\mathcal{M}_\alpha$. We have $\psi_{\alpha,\gamma} = \psi_{\alpha,\beta} \circ \psi_{\beta,\gamma}$ if $0 \leq \alpha \leq \beta \leq \gamma$, so $(\mathcal{M}_\alpha)_{\alpha \geq 0}$ is an inverse family. Let $\mathcal{M}' = (\mathcal{S}', \Sigma_p, \delta', \overline{s}', \mathbb{Z}_p, \tau')$ be the profinite automaton corresponding to the inverse family $(\mathcal{M}_\alpha)_{\alpha \geq 0}$. We establish a map $\psi : \mathcal{S} \to \mathcal{S}'$ such that ψ and the identity map $\mathrm{id} : \mathbb{Z}_p \to \mathbb{Z}_p$ comprise a morphism from $\mathcal{M}$ to $\mathcal{M}'$. Define the image $\psi(s)$ of a state $s \in \mathcal{S}$ as follows. For each $\alpha \geq 0$, consider the function $\Sigma_p^* \to \mathbb{Z}/(p^\alpha \mathbb{Z})$ defined on words by $w \mapsto \pi_{\alpha,\infty}(\tau(\delta(s, w)))$, i.e. obtained by modifying $\mathcal{M} \bmod p^\alpha$ to have initial state s. Since $\mathcal{M} \bmod p^\alpha$ is equivalent to $\mathcal{M}_\alpha$, this function is also given by $w \mapsto \tau_\alpha(\delta_\alpha(s_\alpha, w))$ for some state $s_\alpha \in \mathcal{S}_\alpha$; namely, since $\mathcal{M}$ has no unreachable states, there exists a word $v \in \Sigma_p^*$ such that $\delta(\overline{s}, v) = s$, and we can take $s_\alpha = \delta_\alpha(\overline{s}_\alpha, v)$. Moreover, s_α is unique, since two distinct states in $\mathcal{M}_\alpha$ yielding the same function would be equivalent and would contradict the minimality of $\mathcal{M}_\alpha$. The sequence $(s_\alpha)_{\alpha \geq 0}$ of states forms an inverse family, and we define $\psi(s) = \varprojlim \delta_\alpha(\overline{s}_\alpha, v) \in \mathcal{S}'$. The initial state $\overline{s}$ projects to the initial state $\overline{s}_\alpha$ for each $\alpha \geq 0$; by Part 2 of Definition 3.3, $\overline{s}'$ also projects to $\overline{s}_\alpha$ for each $\alpha \geq 0$, so $\psi(\overline{s}) = \varprojlim \overline{s}_\alpha = \overline{s}'$. Moreover, for each $s \in \mathcal{S}$ and $i \in \Sigma_p$ we have

$$\psi(\delta(s,i)) = \varprojlim \delta_\alpha(\overline{s}_\alpha, vi) = \varprojlim \delta_\alpha(\delta_\alpha(\overline{s}_\alpha, v), i) = \delta'(\psi(s), i)$$

for the appropriate word v, where vi is the word v with i appended, and where the last equality follows from Part 3 of Definition 3.3. Finally, for each $s \in \mathcal{S}$ we have

$$\tau(s) = \varprojlim \tau_\alpha(\delta_\alpha(\overline{s}_\alpha, v)) = \tau'(\psi(s)),$$

for the appropriate word v, where the last equality follows from Part 4 of Definition 3.3. Therefore ψ and id comprise a morphism from $\mathcal{M}$ to $\mathcal{M}'$.

It remains to show that ψ is bijective. Suppose $s, t \in \mathcal{S}$ such that $\psi(s) = \psi(t)$. Let v, z be words such that $\delta(\overline{s}, v) = s$ and $\delta(\overline{s}, z) = t$. Since $\psi(s) = \psi(t)$, we have $\delta_\alpha(\overline{s}_\alpha, v) = \delta_\alpha(\overline{s}_\alpha, z)$ for each $\alpha \geq 0$. For all $w \in \Sigma_p^*$, we have

$$\tau(\delta(s,w)) = \varprojlim \tau_\alpha(\delta_\alpha(\overline{s}_\alpha, vw)) = \varprojlim \tau_\alpha(\delta_\alpha(\overline{s}_\alpha, zw)) = \tau(\delta(t,w)).$$

Therefore s and t are equivalent states in $\mathcal{M}$. Since $\mathcal{M}$ is minimal, it follows that $s = t$, and hence ψ is injective. To show surjectivity, let $s' \in \mathcal{S}'$. The state s' projects to a state $s_\alpha \in \mathcal{S}_\alpha$ for each $\alpha \geq 0$. Since $\mathcal{S}$ is closed under inverse limits, there exists $s = \varprojlim s_\alpha \in \mathcal{S}$ such that $\psi(s) = s'$. Therefore $\mathcal{M}$ and $\mathcal{M}'$ are isomorphic, so $\mathcal{M}$ is profinite. □

4 Substitution shifts defined by profinite automata

In this section we construct a substitution shift, usually on an uncountable alphabet, from a profinite automaton. The construction is completely analogous to the construction of substitution shifts on a finite alphabet. We generalize known dynamical properties of constant-length substitution shifts on a finite alphabet to these substitution shifts on an uncountable alphabet.

4.1 Profinite shifts and substitutions

Let $\mathcal{A}$ be a compact topological space. We endow $\mathcal{A}^\mathbb{N}$ with the product topology, so that it is a compact space, and let $\sigma : \mathcal{A}^\mathbb{N} \to \mathcal{A}^\mathbb{N}$ denote the *shift map* $\sigma(a(n)_{n\geq 0}) := a(n+1)_{n \geq 0}$; σ is a continuous mapping on $\mathcal{A}^\mathbb{N}$. If $a = a(n)_{n\geq 0} \in \mathcal{A}^\mathbb{N}$, define $X_a := \overline{\{\sigma^n(a) : n \in \mathbb{N}\}}$, the closure of the σ-orbit of a in $\mathcal{A}^\mathbb{N}$; we call (X_a, σ) the *one-sided shift generated by a*. We will be primarily interested in *substitution shifts*, namely, those where a is a fixed point of a substitution defined on an alphabet $\mathcal{A}$, and their codings.

Traditionally, $\mathcal{A}$ is a finite alphabet, with the discrete topology. We will also consider shifts defined on non-finite alphabets. Some of these alphabets will be subsets of $\mathbb{Z}_p$, with the topology defined by the p-adic absolute value, which coincides with the inverse limit topology if we consider $\mathbb{Z}_p$ as the inverse limit of the topological rings $\mathbb{Z}/(p^\alpha\mathbb{Z})$. If $\mathcal{A}$ is a closed subset of $\mathbb{Z}_p$, then $\mathcal{A}$ inherits the subspace topology from $\mathbb{Z}_p$. We extend the projection maps $\pi_{\alpha,\beta}$ and $\pi_{\alpha,\infty}$ termwise to sequences in $(\mathbb{Z}/(p^\beta\mathbb{Z}))^{\mathbb{N}}$ and $\mathbb{Z}_p^{\mathbb{N}}$. If $a \in \mathbb{Z}_p^{\mathbb{N}}$ we let $a_\alpha := \pi_{\alpha,\infty}(a)$ be the sequence obtained by reducing each term modulo p^α. We also extend output functions τ and morphism maps ψ to sequences termwise.

For the rest of the article we continue to use the notation of Definition 3.3. Namely, we suppose that we have a sequence $a \in \mathbb{Z}_p^{\mathbb{N}}$ of p-adic integers such that a_α is p-automatic for each $\alpha \geq 0$. We let $\mathcal{M}_\alpha = (\mathcal{S}_\alpha, \Sigma_p, \delta_\alpha, \overline{s}_\alpha, \mathbb{Z}/(p^\alpha\mathbb{Z}), \tau_\alpha)$ be the minimal automaton generating a_α. Then Lemma 3.1 tells us that $(\mathcal{M}_\alpha)_{\alpha\geq 0}$ is an inverse family of finite automata. For $0 \leq \alpha \leq \beta$, we let the morphism from $\mathcal{M}_\beta$ to $\mathcal{M}_\alpha$ be given by the map $\psi_{\alpha,\beta} : \mathcal{S}_\beta \to \mathcal{S}_\alpha$ and the projection $\pi_{\alpha,\beta}$. Finally, we let $u_\alpha \in \mathcal{S}_\alpha^{\mathbb{N}}$ be the sequence generated by $(\mathcal{S}_\alpha, \Sigma_p, \delta_\alpha, \overline{s}_\alpha, \mathcal{S}_\alpha, \mathrm{id})$, so that $a_\alpha = \tau_\alpha(u_\alpha)$.

The following theorem extends the results of Lemma 3.1 to shifts.

Theorem 4.1. *Suppose that $a \in \mathbb{Z}_p^{\mathbb{N}}$ is a sequence of p-adic integers such that a_α is p-automatic for each $\alpha \geq 0$. Then there is a sequence u, generated by a profinite automaton, and a letter-to-letter projection $\tau : X_u \to X_a$ with $\tau(u) = a$, such that for each $0 \leq \alpha \leq \beta$ the following diagram commutes.*

$$\begin{tikzcd}
(X_u,\sigma) \arrow[dr, "\psi_{\beta,\infty}"'] \arrow[drr, "\psi_{\alpha,\infty}"] \arrow[ddd, "\tau"'] & & \\
& (X_{u_\beta},\sigma) \arrow[r, "\psi_{\alpha,\beta}"'] \arrow[d, "\tau_\beta"'] & (X_{u_\alpha},\sigma) \arrow[d, "\tau_\alpha"'] \\
& (X_{a_\beta},\sigma) \arrow[r, "\pi_{\alpha,\beta}"] & (X_{a_\alpha},\sigma) \\
(X_a,\sigma) \arrow[ur, "\pi_{\beta,\infty}"] \arrow[urr, "\pi_{\alpha,\infty}"'] & &
\end{tikzcd}$$

Proof. Let $0 \leq \alpha \leq \beta$. As the maps $\psi_{\alpha,\beta}$, $\pi_{\alpha,\beta}$, and τ_β are extended termwise to sequences, they commute with the shift. Since

$$\psi_{\alpha,\beta}(u_\beta) = \psi_{\alpha,\beta}(\delta_\beta(\overline{s}_\beta, \mathrm{rep}_p(k))_{k\geq 0}) = \delta_\alpha(\overline{s}_\alpha, \mathrm{rep}_p(k))_{k\geq 0} = u_\alpha,$$

it follows that $\psi_{\alpha,\beta}$ maps X_{u_β} to X_{u_α}. That the map $\pi_{\alpha,\beta}$ sends X_{a_β} to X_{a_α} is proved analogously. Finally, as $\pi_{\alpha,\beta} \circ \tau_\beta = \tau_\alpha \circ \psi_{\alpha,\beta}$ is true on the orbit of u_β, it extends to an equality on X_{u_β}. Thus the inner displayed diagram commutes.

Define

$$X := \varprojlim X_{u_\alpha} = \left\{ x = \varprojlim x_\alpha : x_\alpha \in X_{u_\alpha} \text{ and } \psi_{\alpha,\beta}(x_\beta) = x_\alpha \text{ for each } 0 \leq \alpha \leq \beta \right\}.$$

The space X lives in $\prod_{\alpha\geq 0} X_{u_\alpha}$ and inherits the subspace topology from the product topology on $\prod_{\alpha\geq 0} X_{u_\alpha}$. Since for each $\alpha \geq 0$, X_{u_α} is Hausdorff and σ-invariant, then is X closed and σ-invariant. Define $u := \varprojlim u_\alpha$; then $u \in X$, and so $X_u \subset X$.

To show that $X \subset X_u$, let $x = \varprojlim x_\alpha \in X$. We will show that for any open set U containing x, there is some $n \in \mathbb{N}$ with $\sigma^n(u) \in U$. Since each X_{u_α} is a shift defined on a finite alphabet, it is metrizable; denote this metric by d, where points are close if and only if they agree on a sufficiently large initial block. Any open set U containing x contains a set of the form $B = \Big(B_0(x_0, \epsilon) \times \cdots \times B_\beta(x_\beta, \epsilon) \times \prod_{\gamma>\beta} X_{u_\gamma}\Big) \cap X$ where $B(x, \epsilon)$ is a ball of radius ϵ centered at x. Since $x_\beta \in X_{u_\beta}$, there is some n such that $\sigma^n(u_\beta) \in B_\beta(x_\beta, \epsilon) \cap X_{u_\beta}$. If $\alpha \leq \beta$, then $d(\sigma^n(u_\alpha), x_\alpha) = d(\sigma^n \psi_{\alpha,\beta}(u_\beta), \psi_{\alpha,\beta}(x_\beta)) = d(\psi_{\alpha,\beta}\sigma^n(u_\beta), \psi_{\alpha,\beta}(x_\beta)) < \epsilon$. In other words, $\sigma^n(u) \in B$, and it follows that $X \subset X_u$.

For $x \in X_u$, define $\tau(x) := \varprojlim \tau_\alpha(x_\alpha)$. Let $a := \tau(u)$; then similar to above, the set $Y := \varprojlim \tau_\alpha(X_{u_\alpha})$ is closed and σ-invariant, and $Y = X_a$. Now the verification that the diagram commutes follows from the definitions of $X = X_u$ and $Y = X_a$. □

We remark that if u_α is the sequence generating by $\mathcal{M}_\alpha$, and $\mathcal{M}$ is the profinite automaton defined by $(\mathcal{M}_\alpha)_{\alpha\geq 0}$ with state set $\mathcal{S}$, then the sequence $u = \varprojlim u_\alpha$ can be identified with an element of $\mathcal{S}^{\mathbb{N}}$, where the nth component is the inverse limit of the nth component of u_α. Since the interpretation will be clear from context, we pass between these two objects without further remarks, just as we identify elements of $\mathbb{Z}_p^{\mathbb{N}}$ with elements of $\prod_{\alpha\geq 0}(\mathbb{Z}/(p^\alpha\mathbb{Z}))^{\mathbb{N}}$.

Recall that a finite automaton defines a substitution as described after Theorem 2.6.

Theorem 4.2. *Let $(\mathcal{M}_\alpha)_{\alpha\geq 0} = (\mathcal{S}_\alpha, \Sigma_p, \delta_\alpha, \overline{s}_\alpha, \mathbb{Z}/(p^\alpha\mathbb{Z}), \tau_\alpha)_{\alpha\geq 0}$ be an inverse family of finite automata whose inverse limit is the profinite automaton $\mathcal{M} = (\mathcal{S}, \Sigma_p, \delta, \overline{s}, \mathbb{Z}_p, \tau)$. For each $\alpha \geq 0$, let θ_α be the substitution on $\mathcal{S}_\alpha$ corresponding to $\mathcal{M}_\alpha$, with fixed point u_α. Let $u = \varprojlim u_\alpha \in \mathcal{S}^{\mathbb{N}}$. Then there is a length-p substitution $\theta : \mathcal{S} \to \mathcal{S}^p$, with $\theta(u) = u$, where for each $0 \leq \alpha \leq \beta$ the following diagram commutes.*

$$\begin{tikzcd}
X_u \arrow[dr, "\psi_{\beta,\infty}"'] \arrow[drr, "\psi_{\alpha,\infty}"] \arrow[ddd, "\theta"'] & & \\
& X_{u_\beta} \arrow[r, "\psi_{\alpha,\beta}"'] \arrow[d, "\theta_\beta"'] & X_{u_\alpha} \arrow[d, "\theta_\alpha"'] \\
& X_{u_\beta} \arrow[r, "\psi_{\alpha,\beta}"] & X_{u_\alpha} \\
X_u \arrow[ur, "\psi_{\beta,\infty}"] \arrow[urr, "\psi_{\alpha,\infty}"'] & &
\end{tikzcd}$$

Proof. For each $s = \varprojlim s_\alpha \in \mathcal{S}$, we will define $\theta(s)$, a word of length p. If $0 \le i \le p-1$, we use $\theta(s)_i$ to refer to the i-th letter of $\theta(s)$. Define

$$\theta(s)_i := \varprojlim \theta_\alpha(s_\alpha)_i.$$

Then for $k \ge 0$ and $0 \le i \le p-1$,

$$u(pk+i) = \varprojlim u_\alpha(pk+i) = \varprojlim \theta_\alpha u_\alpha(k)_i = \theta(u(k))_i,$$

which implies that $\theta(u) = u$. As in Definition 2.5, the substitution $\theta : \mathcal{S} \to \mathcal{S}^p$ extends to the substitution $\theta : X_u \to X_u$. To verify that the inner diagram commutes, we note that

$$\psi_{\alpha,\beta}\theta_\beta\sigma^j(u_\beta) = \psi_{\alpha,\beta}\sigma^{jp}(u_\beta) = \sigma^{jp}(u_\alpha) = \theta_\alpha\sigma^j(u_\alpha) = \theta_\alpha\psi_{\alpha,\beta}\sigma^j(u_\beta),$$

and the continuity of $\psi_{\alpha,\beta}$ and θ_β gives us the result. The commutativity of the outer diagram follows in a similar way. □

We call the substitution θ of Theorem 4.2 a *profinite substitution*, and the shift (X_u, σ) of Theorem 4.1 a *profinite substitution shift.*

4.2 Dynamical properties of profinite substitution shifts

Recall that the substitution θ on the finite alphabet $\mathcal{A}$ is *primitive* if there is some $k \in \mathbb{N}$ such that for all $a, a' \in \mathcal{A}$, the word $\theta^k(a)$ contains a'. For primitive substitutions on a finite alphabet, any word that appears in any $x \in X_u$ appears in x with bounded gaps [Que87, Chapter 5.2], and X_u is *minimal*: X_u is the σ-orbit closure of any point in X_u. An application of the argument in the second part of the proof of Theorem 4.1 yields the following.

Proposition 4.3. *Let u be a fixed point of a profinite substitution $\theta = \varprojlim \theta_\alpha$. If θ_α is length-p and primitive for each $\alpha \ge 0$ then (X_u, σ) is minimal.*

Definition 4.4 ([Cho58, Definitions 5 and 6])**.** Let $\{(X_\alpha, \mathcal{B}_\alpha, \mu_\alpha) : \alpha \ge 0\}$ be a family of compact Borel measure spaces, and for each $0 \le \alpha \le \beta$ let $\psi_{\alpha,\beta} : X_\beta \to X_\alpha$ be a continuous map. If $\mu_\alpha(A) = \mu_\beta(\psi_{\alpha,\beta}^{-1}(A))$ for each $A \in \mathcal{B}_\alpha$ and each $0 \le \alpha \le \beta$, then $\{(X_\alpha, \mathcal{B}_\alpha, \mu_\alpha) : \alpha \ge 0\}$ is called an *inverse family of (compact) topological measure spaces.*

Given an inverse family of compact topological measure spaces, let

$$X := \{x = \varprojlim x_\alpha : x_\alpha \in X_\alpha \text{ and } \psi_{\alpha,\beta}(x_\beta) = x_\alpha \text{ for each } 0 \le \alpha \le \beta\}.$$

Let $\psi_{\alpha,\infty} : X \to X_\alpha$ be the natural projection map, and let $\mathcal{B}_\alpha^* := \psi_{\alpha,\infty}^{-1}(\mathcal{B}_\alpha)$. Let $\mathcal{B}$ be the σ-algebra generated by $\bigcup_\alpha \mathcal{B}_\alpha^*$. Then $\mu(B) := \mu_\alpha(\psi_{\alpha,\infty}B)$ for $B \in \mathcal{B}_\alpha^*$ defines a finitely additive set function on $\bigcup_\alpha \mathcal{B}_\alpha^*$. If μ has an extension to $\mathcal{B}$, then we say that $(X, \mathcal{B}, \mu)$ is the *inverse limit* of $\{(X_\alpha, \mathcal{B}_\alpha, \mu_\alpha) : \alpha \ge 0\}$.

Theorem 4.5 ([Cho58, Theorems 2.2 and 2.3])**.** *Let $\{(X_\alpha, \mathcal{B}_\alpha, \mu_\alpha) : \alpha \geq 0\}$ be an inverse family of compact topological measure spaces. Then the inverse limit $(X, \mathcal{B}, \mu)$ exists. If μ_α is Baire for each α, then μ is Baire.*

Let θ be a primitive substitution on a finite alphabet. Then any θ-fixed point u, (or any θ-periodic point u in the case where there are no θ-fixed points) generates the substitution shift (X_u, σ), which is independent of the fixed point chosen. With the Borel σ-algebra $\mathcal{B}$, this shift is uniquely ergodic [Mic74].

Corollary 4.6. *Let p be prime. Let u be a fixed point of a profinite substitution $\theta = \varprojlim \theta_\alpha$. Suppose that for each $\alpha \geq 0$, the substitution θ_α is length-p and primitive. Then (X_u, σ) is uniquely ergodic.*

Proof. Let u_α be the fixed point of θ_α such that $\{(X_{u_\alpha}, \sigma) : \alpha \geq 0\}$ is the inverse family of substitution shifts generated by $(\theta_\alpha)_{\alpha \geq 0}$. As θ_α is primitive, we let μ_α denote the unique σ-invariant measure for $(X_{u_\alpha}, \mathcal{B}_\alpha, \sigma)$ with $\mathcal{B}_\alpha$ the Borel σ-algebra. Then $\mu_\alpha = \mu_\beta \circ \pi_{\alpha,\beta}^{-1}$ on cylinder sets, so that $\mu_\alpha = \mu_\beta \circ \pi_{\alpha,\beta}^{-1}$, i.e. $\{(X_{u_\alpha}, \mathcal{B}_\alpha, \mu_\alpha) : \alpha \geq 0\}$ is an inverse family of topological measure spaces. We apply Theorem 4.5 to the inverse family $\{(X_{u_\alpha}, \mathcal{B}_\alpha, \mu_\alpha, \sigma) : \alpha \geq 0\}$ to conclude that the inverse limit $\mu = \varprojlim \mu_\alpha$ exists on the σ-algebra $\mathcal{B}$ generated by $\bigcup_\alpha \mathcal{B}_\alpha^*$, with $\mathcal{B}_\alpha^* := \psi_{\alpha,\infty}^{-1}(\mathcal{B}_\alpha)$. Furthermore since each μ_α is σ-invariant, then μ is σ-invariant on each $\mathcal{B}_\alpha^*$ and this implies that μ is σ-invariant on $\mathcal{B}$.

Let ν be any other σ-invariant measure on $(X_u, \mathcal{B})$. Then for any α, $\nu_\alpha(A) := \nu(\psi_{\alpha,\infty}^{-1}(A))$, $A \in \mathcal{B}_\alpha$ defines a measure on X_{u_α}. This measure ν_α is also σ-invariant, and the unique ergodicity of (X_{u_α}, σ) for each $\alpha \geq 0$ implies that $\nu_\alpha = \mu_\alpha$ for each $\alpha \geq 0$, which gives $\nu = \mu$. □

Next we extend results of Dekking [Dek78] concerning the discrete spectrum of constant-length substitution shifts. We refer to his article for all relevant definitions and to [Dow05] for the definition of an odometer. The *maximal equicontinuous factor* of a topological dynamical system is a rotation on a compact abelian group that is determined by the collection of *continuous* eigenvalues of the system. An eigenvalue is continuous if there exists a continuous eigenfunction for that eigenvalue. The set of continuous eigenvalues is generally a proper subset of the set of measurable eigenvalues. For primitive constant-length substitution shifts, the continuous and measurable eigenvalues coincide [Dek78].

Theorem 4.7. *Let p be prime. Let u be a fixed point of a profinite substitution $\theta = \varprojlim \theta_\alpha$, with each θ_α length-p and primitive. Let u_α be a fixed point of θ_α, so that $(X_u, \mathcal{B}, \mu, \sigma) = \varprojlim (X_{u_\alpha}, \mathcal{B}_\alpha, \mu_\alpha, \sigma)$. Then the maximal equicontinuous factor of (X_u, σ) is an odometer, and every measurable eigenvalue is a continuous eigenvalue.*

Proof. First suppose that there exists α such that the sequence u_α is not eventually periodic; then for each $\beta \geq \alpha$, u_β is not eventually periodic. In this case Dekking's results [Dek78] tell us that there exists some α_0 such that for each $\alpha \geq \alpha_0$, $(\mathbb{Z}_p \times \mathbb{Z}/(h_\alpha \mathbb{Z}), +(1,1))$ is the maximal equicontinuous factor of (X_{u_α}, σ), for some h_α coprime to p, and also that every measurable eigenvalue is a continuous eigenvalue. Note that $h_\alpha \mid h_{\alpha+1}$ for each $\alpha \geq 0$. Without loss of generality, we suppose that $\alpha_0 = 1$. Let $(\mathcal{Z}, +1)$ be the odometer formed by the sequence $(h_{\alpha+1}/h_\alpha)_{\alpha \geq 0}$. Then $(\mathbb{Z}_p \times \mathbb{Z}/(h_\alpha \mathbb{Z}), +(1,1))_{\alpha \geq 0}$ forms an inverse limit system in the category of group rotations, and $(\mathbb{Z}_p \times \mathcal{Z}, +(1,1))$ is its inverse limit. Hence $(\mathbb{Z}_p \times \mathcal{Z}, +(1,1))$ is an equicontinuous factor of the inverse limit system (X_u, σ). If for each $\alpha \geq 0$, u_α is eventually periodic, with period h_α, then the above argument follows through, except that in this case $(\mathcal{Z}, +1)$ is an equicontinuous factor of (X_u, σ).

Given a nontrivial measurable eigenfunction f of $(X_u, \mathcal{B}, \mu, \sigma)$, with eigenvalue λ, let $f_\alpha := \mathbb{E}(f \mid \mathcal{B}^*_\alpha)$ denote the conditional expectation of f given $\mathcal{B}^*_\alpha$. Then f_α is isomorphic to an eigenfunction of $(X_{u_\alpha}, \mathcal{B}_\alpha, \sigma, \mu_\alpha)$ with eigenvalue λ, and f_α is nontrivial for large α. For primitive substitutions, every measurable eigenvalue is a continuous eigenvalue. Also, $f_\alpha \to f$ μ-almost everywhere, so that any measurable eigenvalue of $(X_u, \mathcal{B}, \mu, \sigma)$, and hence any continuous eigenvalue, has already contributed to the maximal equicontinuous factor $(\mathbb{Z}_p \times \mathbb{Z}/(h_\alpha \mathbb{Z}), +(1,1))$ (or $(\mathbb{Z}/(h_\alpha \mathbb{Z}), +1)$) of (X_{u_α}, σ) for large enough α. This completes the proof that $(\mathbb{Z}_p \times \mathcal{Z}, +(1,1))$ (or $(\mathcal{Z}, +1)$) is the maximal equicontinuous factor of (X_u, σ). □

Example 4.8. The Fibonacci sequence $F = F(n)_{n \geq 0} = 0, 1, 1, 2, 3, 5, 8, 13, \ldots$ is periodic when reduced modulo m for any $m \geq 1$; hence for any prime p and any $\alpha \geq 0$, $(F(n) \bmod p^\alpha)_{n \geq 0}$ is p-automatic. Note that because of the repeated $F(1) = F(2) = 1$, the letter-to-letter coding τ for which $F = \tau(u)$ is not the identity map. For example, let $p = 2$, consider the alphabet $\mathcal{S} = \{0, s, 1, 2, 3, 5, 8, 13, \ldots\}$, where we are using integers as state names for convenience. Let $\tau(s) = 1$ and $\tau(m) = m$ for all $m \in \mathbb{N}$. Then the sequence $u = 0, s, 1, 2, 3, 5, 8, 13, \ldots$ satisfies $F = \tau(u)$. Moreover, the profinite substitution θ satisfies

$$\begin{aligned} \theta(0) &= 0\, s \\ \theta(s) &= 1\, 2 \\ \theta(F(m)) &= F(2m)\, F(2m+1) \quad \text{for } m \geq 2, \end{aligned}$$

and u is a fixed point of θ. For each α, the sequence $F \bmod 2^\alpha$ is the coding of a primitive substitution θ_α. For example, the fixed point of the substitution $\theta_2(0) = \theta_2(2) = 0a$, $\theta_2(a) = \theta_2(3) = b2$, $\theta_2(b) = \theta_2(c) = 3c$ and projects to $F \bmod 4 = 0, 1, 1, 2, 3, 1 \ldots$ via the coding $\tau_2(a) = \tau_2(b) = \tau_2(c) = 1$.

Let $\ell(m)$ denote the (minimal) period length of $(F(n) \bmod m)_{n \geq 0}$. For prime p, Wall [Wal60, Theorem 5] showed that if e is the smallest positive integer such that $\ell(p^e) \neq \ell(p)$, then $\ell(p^\alpha) = p^{\alpha+1-e}\ell(p)$ for $\alpha \geq e$. For $p = 2$ we have $\ell(2) = 3$ and $e = 2$. For each θ_α then, (X_{u_α}, σ) is conjugate to the finite group $\big(\mathbb{Z}/(2^{\alpha-1}\mathbb{Z}) \times \mathbb{Z}/(3\mathbb{Z}), +(1,1)\big)$, and so $(\mathbb{Z}_2 \times \mathbb{Z}/(3\mathbb{Z}), +(1,1))$ is the maximal equicontinuous factor of (X_u, σ). Furthermore, the maximal equicontinuous factor of (X_a, σ), which must be contained in that of (X_u, σ), is also $(\mathbb{Z}_2 \times \mathbb{Z}/(3\mathbb{Z}), +(1,1))$.

We end by showing that profinite substitutions are *recognizable*. Let (X_u, σ) be the shift generated by a fixed point of the profinite substitution θ. We say that θ is *recognizable* if for any $y \in X_u$ there is a unique way to write $y = \sigma^k(\theta(x))$ with $x \in X_u$ and $0 \leq k < |\theta(x(0))|$. Mossé [Mos92] showed that if θ is primitive and generates an aperiodic fixed point u, and X_u is the *two-sided* shift generated by θ, then θ is recognizable. The two-sided shift can be thought of in two equivalent ways. It can be defined as the set of bi-infinite sequences, all of whose subwords are a subword of $\theta^n(a)$ for some $n \geq 1$ and some letter a. It can also defined as the *natural extension* of the one-sided shift. In other words, if $\tilde{X}_u$ is the one-sided shift, then the two-sided shift X_u is defined as

$$X_u := \{x = \varprojlim x_\alpha : x_\alpha \in \tilde{X}_u \text{ and } \sigma(x_{\alpha+1}) = x_\alpha \text{ for each } \alpha \geq 0\}.$$

If $(\tilde{X}_{u_\alpha}, \sigma)_{\alpha \geq 0}$ is an inverse family of one-sided shifts, then their natural extensions $(X_{u_\alpha}, \sigma)_{\alpha \geq 0}$ form an inverse family of (two-sided) shifts.

Theorem 4.9. *Let p be prime, and let θ be a profinite substitution $\theta = \varprojlim \theta_\alpha$ that generates an inverse family of two-sided shifts $\{(X_{u_\alpha}, \sigma) : \alpha \geq 0\}$. Suppose that each θ_α is length-p and primitive. Then θ is recognizable.*

Proof. Let $x = \varprojlim x_\alpha \in \varprojlim X_{u_\alpha}$. Mossé's theorem tells us that each x_α can be written in a unique way as $x_\alpha = \sigma^{k_\alpha}\theta_\alpha(y_\alpha)$ with $0 \leq k_\alpha < p$. Then

$$\sigma^{k_\alpha}\theta_\alpha(y_\alpha) = x_\alpha = \psi_{\alpha,\beta}(x_\beta) = \psi_{\alpha,\beta}\sigma^{k_\beta}\theta_\beta(y_\beta) = \sigma^{k_\beta}\psi_{\alpha,\beta}\theta_\beta(y_\beta),$$

where the final equality follows since $\psi_{\alpha,\beta}$ is defined termwise. Hence, since $\psi_{\alpha,\beta}$ commutes with θ_β for each $0 \leq \alpha \leq \beta$, we have $k_\beta = k_\alpha$ and $\psi_{\alpha,\beta}(y_\beta) = y_\alpha$. Let $k := k_\alpha$ and $y := \varprojlim y_\alpha$; then $x = \sigma^k\theta(y)$. If $x = \sigma^{k'}\theta(y')$, then for each $\alpha \geq 0$ we have $x_\alpha = \psi_{\alpha,\infty}(x) = \psi_{\alpha,\infty}(\sigma^{k'}\theta(y')) = \sigma^{k'}\theta_\alpha(\psi_{\alpha,\infty}(y'))$, so that uniqueness implies that $k = k'$ and $y_\alpha = \psi_{\alpha,\infty}(y')$. □

4.3 Limit sets of profinite substitutions

Given a profinite substitution θ, which defines a shift (X_u, σ), we define the *limit set* $\mathcal{L}(\theta) := \bigcap_{n \geq 0} \theta^n(X_u)$. The limit set is a nonempty compact set. If θ is a substitution on a finite alphabet, then $\mathcal{L}(\theta)$ contains only the periodic points of θ, of which there are finitely many. However, when we consider substitutions on an infinite alphabet, $\mathcal{L}(\theta)$ consists of inverse limits of θ_α-periodic points, and the period lengths can increase. For example, let a be the coding of a fixed point u of a profinite substitution θ. Define $\varphi : \mathbb{Z}_p^{\mathbb{N}} \to \mathbb{Z}_p$ by $\varphi(x(n)_{n\geq 0}) = x(0)$. We have, for natural numbers n, k, and $r < p^n$,

$$a(kp^n + r) = \varphi\sigma^{kp^n+r}(a) = \varphi\sigma^{kp^n+r}(\tau(u)) = \varphi\sigma^r\tau\sigma^{kp^n}(u) = \varphi\sigma^r\tau\Big(\theta^n\sigma^k(u)\Big),$$

so that

$$a(kp^n + r) \in \varphi\sigma^r\tau\left(\bigcap_{j\leq n}\theta^j(X_u)\right).$$

Hence if $\lim_{n\to\infty} a(kp^n + r)$ exists, then $\mathcal{L}(\theta)$ contains points other than u. For example, as mentioned in Section 1, the sequence $C(2^n)_{n\geq 0}$ converges in $\mathbb{Z}_2$, where $C(n)$ is the nth Catalan number. More generally, we have the following.

Proposition 4.10 ([MMR14, Corollary 3.1])**.** *Let p be prime, and let $C(n)$ be the nth Catalan number. For each $k, r \in \mathbb{Z}$ with $k \geq 1$, the limit* $\lim_{n\to\infty} C(kp^n + r)$ *exists in $\mathbb{Z}_p$.*

We have a similar result for the Fibonacci sequence. More generally, we have such limits for any sequence satisfying a linear recurrence

$$a(n+\ell) + c_{\ell-1}a(n+\ell-1) + \cdots + c_1a(n+1) + c_0a(n) = 0$$

with constant coefficients $c_i \in \mathbb{Z}_p$. The *characteristic polynomial* of this sequence is $x^\ell + \cdots + c_1x + c_0$.

Proposition 4.11 ([RY16, Corollary 11])**.** *Let p be prime, and let $a(n)_{n\geq 0}$ be a constant-recursive sequence of p-adic integers with monic characteristic polynomial $g(x) \in \mathbb{Z}_p[x]$. There exists an integer $f \geq 1$ such that, for each $k, r \in \mathbb{Z}$ with $k \geq 1$, the limit* $\lim_{n\to\infty} a(kp^{fn} + r)$ *exists and is algebraic over $\mathbb{Q}_p$.*

A suitable integer f can be given explicitly as follows. Let K be a degree-d splitting field of $g(x)$ over $\mathbb{Q}_p$ with ramification index e; then we can take $f = d/e$. For example, for the Fibonacci sequence and $p = 2$ as pictured in Section 1, we obtain the value $f = 2$, and the two limit points are $\pm\sqrt{-\frac{3}{5}}$ in $\mathbb{Z}_2$.

Unlike Proposition 4.10, the limit in Proposition 4.11 comes from an *approximate twisted interpolation* of the sequence $a(n)_{n\geq 0}$ to the relevant extension of $\mathbb{Q}_p$. Amice and Fresnel [AF72] give an alternate characterization of sequences which have twisted interpolations.

4.4 Cocycle sequences

The examples we have worked with so far consist of sequences of p-adic integers whose generating function is algebraic over $\mathbb{Z}_p(x)$. In this section we show that certain *cocycle sequences* are codings of the fixed point of a length-p profinite substitution. Let M_p be the $p \times p$ matrix all of whose entries are 1. This matrix is the incidence matrix for the substitution θ^* on the alphabet $\mathbb{Z}/(p\mathbb{Z})$ defined as $\theta^*(j) = 01\cdots(p-1)$ for each $j \in \mathbb{Z}/(p\mathbb{Z})$, whose fixed point is periodic.

We say that θ is *aperiodic* if it has a fixed point which is not periodic. Let θ be any aperiodic substitution on $\mathbb{Z}/(p\mathbb{Z})$ whose incidence matrix is M_p. Let us assume also that $\theta(0)$ starts with 0, and let $u = 0\cdots$ be the fixed point starting with 0. Let μ be the unique measure that is preserved by σ. Then the shift (X_u, σ) has a *Bratteli–Vershik* representation (X_B, φ_θ) [VL92], where B is a *Bratteli diagram* and φ_θ is a *Vershik map:* We briefly describe these objects. The Bratteli diagram B is an infinite directed graph, and for our example, we illustrate B in Figure 1 for the case $p = 3$.

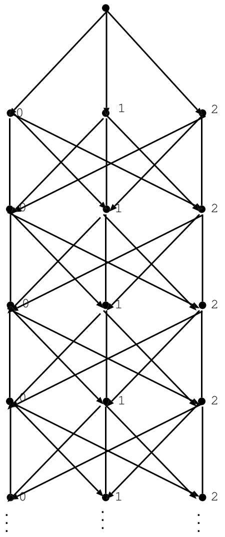


Figure 1: The Bratteli diagram associated with M_3.

Apart from the root vertex at the top of the diagram, there are p vertices at each level n, which we label $0, \ldots, p-1$, moving from left to right. The levels are indexed by increasing indices n as we move down in the diagram, $n = 0, 1, \ldots$; we do not think of the root vertex as occupying a level. The edge structure for B is determined by the matrix M_p. Namely, the number of edges from vertex i on level $n-1$ to vertex j on level n is the (i, j) entry of M_p. The substitution θ defines a linear order on the incoming edges to any vertex: if the vertex is labelled j and $\theta(j) := i_0 \cdots i_{p-1}$, then we give the edge with source i_k the label k. Let X_B be the set of infinite paths in B starting at the root vertex. Such a path is labelled $x = x_0, x_1, \ldots$ where x_n is the label of the edge from level n to level $n+1$. The linear order on the incoming edges to a vertex defines a partial order on X_B. Namely, we can compare two infinite paths x and x' in B if and only if they eventually agree:

if n is the smallest integer such that x and x' agree from level n onwards, then we write $x < x'$ if $x_{n-1} < x'_{n-1}$. If x is a non-maximal path (i.e. one of its edges is not maximally labelled) then it has a *successor* in this order. Namely, if n is the smallest integer such that x_n is not a maximal edge, then the successor of x agrees with x from level $n+1$, is minimal up to level n, and the edge between level n and level $n+1$ is the successor of x_n.

Thus any substitution θ with incidence matrix M_p defines a partial ordering of X_B and this determines a *Vershik map* $\varphi_\theta : X_B \to X_B$ where $\varphi_\theta(x)$ is defined to be the successor of x in the ordering determined by θ. Note that φ is not defined on the set of maximal paths, but this is a finite set, and here we define it arbitrarily.

If θ is an aperiodic primitive substitution, then $(X_u, \mathcal{B}, \mu, \sigma)$ is measurably conjugate to $(X_B, \overline{\mathcal{B}}, \overline{\mu}, \varphi_\theta)$ with $\overline{\mathcal{B}}$ the σ-algebra generated by cylinder sets on X_B and $\overline{\mu}$ the image of μ via the conjugacy. If θ is a periodic substitution (as θ^* is), then (X_B, φ_θ) is topologically conjugate to the p-adic odometer $(\mathbb{Z}_p, +1)$. In this latter case, if the finite path x has edges labelled by the base-p expansion of m, then $\varphi^n_{\theta^*}(x)$ is the finite path whose edges are labelled by the base-p expansion of $m+n$. We refer the reader to [VL92] for details.

Note that the ordering induced by θ^* on B has the special property that an edge labelled i has as source a vertex labelled i.

Suppose that θ is aperiodic and $\theta(0)$ starts with 0. Let 0^∞ denote the minimal path in B that runs through the vertices labelled 0, and let the sequence $s(n)_{n\geq 0}$ of natural numbers be defined by

$$\varphi^n_\theta(0^\infty) = \varphi^{s(n)}_{\theta^*}(0^\infty).$$

The sequence $s(n)_{n\geq 0}$ is called a *cocycle.*

Theorem 4.12. *Let p be prime. Let M_p be the $p \times p$ matrix all of whose entries are 1. Let θ be an aperiodic substitution with incidence matrix M_p, and let $\theta^*(j) = 01\cdots(p-1)$ for each $j \in \mathbb{Z}/(p\mathbb{Z})$. Suppose $\theta(0)$ starts with 0. Let $s(n)_{n\geq 0}$ be the cocycle defined by θ and θ^*. Then $(s(n) \bmod p^\alpha)_{n\geq 0}$ is the fixed point of a length-p substitution θ_α for every $\alpha \geq 0$.*

Proof. Note that if the finite path $x_0x_1\cdots x_k$ passes through the vertices $v_0v_1\cdots v_{k+1}$, and if $x_0x_1\cdots x_k = \varphi^n_\theta(00\cdots 0)$, then $s(n) = \sum_{j=0}^{k+1} p^j v_j$, and $s(n) \bmod p^\alpha = \sum_{j=0}^{\alpha-1} p^j v_j$.

Given $\alpha \geq 0$, we define a substitution θ_α on $\mathbb{Z}/(p^\alpha\mathbb{Z})$ of length-p as follows. Given $j = j_0p^0 + j_1p^1 + \cdots + j_{\alpha-1}p^{\alpha-1} \in \mathbb{Z}/(p^\alpha\mathbb{Z})$, define $\theta_\alpha(j) = \theta(j_0) + p(j_0p^0 + j_1p^1 + \cdots + j_{\alpha-2}p^{\alpha-2})$, where here we are adding $p(j_0p^0 + j_1p^1 + \cdots + j_{\alpha-2}p^{\alpha-2})$ to each entry in the word $\theta(j_0)$. We claim that $(s(n) \bmod p^\alpha)_{n\geq 0}$ is a fixed point of θ_α.

To see this, we need to show that for each n, $\theta_\alpha(s(n) \bmod p^\alpha) = (s(pn), s(pn+1), \ldots, s(pn+p-1)) \bmod p^\alpha$. To get $s(pn+\ell) \bmod p^\alpha$, we need the first α vertices through which the path $\varphi_\theta^{pn+\ell}(00\cdots 0)$ runs. Suppose that the path $\varphi_\theta^{n}(00\cdots 0)$ passes through the vertices $v_0, v_1, \ldots, v_{\alpha-1}$, so that $s(n) \bmod p^\alpha = \sum_{j=0}^{\alpha-1} v_j p^j$. Recall that we use the notation $\theta(a)_j$ to denote the j-th letter of $\theta(a)$. Then the path $\varphi_\theta^{pn+\ell}(00\cdots 0)$ starts at the vertex labelled $\theta(v_0)_\ell$, followed by $v_0, \ldots, v_{\alpha-2}$ at levels $1, \ldots, \alpha-1$ of the diagram respectively. In other words, $s(pn+\ell) \bmod p^\alpha = \theta(v_0)_\ell + p\sum_{j=0}^{\alpha-2} v_j p^j = \theta_\alpha(s(n) \bmod p^\alpha)_\ell$, as desired. □

Remark 4.13. Since cocycle sequences are bijections of $\mathbb{N}$, it is very easy to define the cocycle sequence as the fixed point of a length-p substitution on $\mathbb{N}$. For example, if $p = 2$ and $\phi(0) = 01$, $\phi(1) = 10$ is the Thue–Morse substitution, then it has as transition matrix M_2 and its cocycle sequence

$$s(n)_{n\geq 0} = 0, 1, 3, 2, 7, 6, 4, 5, 15, 14, 12, 13, 8, 9, 11, 10, \ldots$$

is the fixed point of the length-2 substitution θ on $\mathbb{N}$ defined by

$$\theta(m) = \begin{cases} (2m)\,(2m+1) & \text{if } m \text{ is even} \\ (2m+1)\,(2m) & \text{if } m \text{ is odd.} \end{cases}$$

In particular, $s(n)_{n\geq 0}$ projects modulo 2 to the Thue–Morse sequence. However, it can be shown that $s(n)_{n\geq 0}$ is 2-regular in the sense of Allouche and Shallit [AS92]; namely, we have the recurrence

$$\begin{aligned} s(4n) &= -2s(n) + 3s(2n) \\ s(4n+1) &= -2s(n) + 2s(2n) + s(2n+1) \\ s(4n+2) &= -2s(n) + 3s(2n+1) \\ s(4n+3) &= -2s(n) + s(2n) + 2s(2n+1). \end{aligned}$$

It follows that $(s(n) \bmod k)_{n\geq 0}$ is 2-automatic for every $k \geq 2$ [AS92, Corollary 2.4]. Therefore, by a theorem of Cobham, for a prime $p \neq 2$ the sequence $(s(n) \bmod p^\alpha)_{n\geq 0}$ is not p-automatic unless it is eventually periodic. Moreover, the generating function $\sum_{n\geq 0} s(n)x^n$ is not rational, so it follows from a result of Bézivin [Béz94, BCR13] that $s(n)_{n\geq 0}$ is not algebraic, nor is it the diagonal of a rational function.

Acknowledgments

The authors acknowledge the hospitality and support of LaCIM, Montréal and LIAFA, Université Paris 7. We also thank the referee for excellent comments and suggestions.

References

[AF72] Yvette Amice and Jean Fresnel. Fonctions zêta p-adiques des corps de nombres abeliens réels. *Acta Arith.*, 20:353–384, 1972.

[AS92] Jean-Paul Allouche and Jeffrey Shallit. The ring of k-regular sequences. *Theoretical Computer Science*, 98(2):163–197, 1992.

[BCR13] Jason P. Bell, Michael Coons, and Eric Rowland. The rational–transcendental dichotomy of Mahler functions. *Journal of Integer Sequences*, 16:Article 13.2.10, 2013.

[Béz94] Jean-Paul Bézivin. Sur une classe d'equations fonctionnelles non linéaires. *Funkcialaj Ekvacioj*, 37:263–271, 1994.

[Cho58] J. R. Choksi. Inverse limits of measure spaces. *Proc. London Math. Soc. (3)*, 8:321–342, 1958.

[Chr74] Gilles Christol. Éléments analytiques uniformes et multiformes. *Séminaire Delange-Pisot-Poitou. Théorie des nombres*, 1(15):1–18, 1973-1974.

[CKMFR80] Gilles Christol, Teturo Kamae, Michel Mendès France, and Gérard Rauzy. Suites algébriques, automates et substitutions. *Bull. Soc. Math. France*, 108(4):401–419, 1980.

[Cob72] Alan Cobham. Uniform tag sequences. *Math. Systems Theory*, 6:164–192, 1972.

[Dek78] F. M. Dekking. The spectrum of dynamical systems arising from substitutions of constant length. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete*, 41(3):221–239, 1977/78.

[DL87] Jan Denef and Leonard Lipshitz. Algebraic power series and diagonals. *J. Number Theory*, 26(1):46–67, 1987.

[Dow05] Tomasz Downarowicz. Survey of odometers and Toeplitz flows. In *Algebraic and topological dynamics*, volume 385 of *Contemp. Math.*, pages 7–37. Amer. Math. Soc., Providence, RI, 2005.

[Fer06] Sébastien Ferenczi. Substitution dynamical systems on infinite alphabets. *Ann. Inst. Fourier (Grenoble)*, 56(7):2315–2342, 2006.

[GPS09] Mathieu Guay-Paquet and Jeffrey Shallit. Avoiding squares and overlaps over the natural numbers. *Discrete Mathematics*, 309:6245–6254, 2009.

[Mau06] Christian Mauduit. Propriétés arithmétiques des substitutions et automates infinis. *Ann. Inst. Fourier (Grenoble)*, 56(7):2525–2549, 2006.

[Mic74] Pierre Michel. Stricte ergodicité d'ensembles minimaux de substitution. *C. R. Acad. Sci. Paris Sér. A*, 278:811–813, 1974.

[MMR14] Alexandra Michel, Andrew Miller, and Joseph Rennie. On p-adic limits of subsequences of the Catalan numbers. *2014 MSRI-UP Research Reports*, 2014. `http://www.msri.org/system/cms/files/81/files/original/Research_Reports_2014_MSRI-UP_(Single_File).pdf`.

[Mos92] Brigitte Mossé. Puissances de mots et reconnaissabilité des points fixes d'une substitution. *Theoret. Comput. Sci.*, 99(2):327–334, 1992.

[Que87] Martine Queffélec. *Substitution dynamical systems—spectral analysis*, volume 1294 of *Lecture Notes in Mathematics*. Springer-Verlag, Berlin, second edition, 1987.

[RS12] Eric Rowland and Jeffrey Shallit. Avoiding 3/2-powers over the natural numbers. *Discrete Mathematics*, 312:1282–1288, 2012.

[RY15] Eric Rowland and Reem Yassawi. Automatic congruences for diagonals of rational functions. *Journal de Théorie des Nombres de Bordeaux*, 27:245–288, 2015.

[RY16] Eric Rowland and Reem Yassawi. p-adic asymptotic properties of constant-recursive sequences. *Indagationes Mathematicae*, to appear, 2016. `http://arxiv.org/abs/1602.00176`.

[Sha08] Jeffrey Shallit. *A Second Course in Formal Languages and Automata Theory*. Cambridge University Press, New York, NY, USA, 2008.

[VL92] A. M. Vershik and A. N. Livshits. Adic models of ergodic transformations, spectral theory, substitutions, and related topics. In *Representation theory and dynamical systems*, volume 9 of *Representation theory and dynamical systems*, pages 185–204. Representation theory and dynamical systems, 1992.

[Wal60] D. D. Wall. Fibonacci series modulo m. *The American Mathematical Monthly*, 67:525–532, 1960.

[Wes96] Julian West. Generating trees and forbidden subsequences. *Discrete Math.*, 157(1-3):363–374, 1996.