Uncovering hidden automatic sequences

Elżbieta Krawczyk
(joint work with Clemens Müllner)

Jagiellonian University, Kraków

Liége, 22 May 2023

Automatic sequences

Example: Thue–Morse sequence

Let $\mathcal{A} = \{0, 1\}$ and let $\varphi\colon \mathcal{A} \to \mathcal{A}^*$ be a substitution given by

$$\varphi(0) = 01, \ \varphi(1) = 10.$$

The Thue–Morse sequence

$$\mathrm{x} = 0110100110010110\ldots$$

is a fixed point of φ. It is a (purely) 2-automatic sequence.

Let $\mathrm{k} \geq 2$. A sequence $\mathrm{x} \in \mathcal{A}^{\mathbb{N}}$ is k-automatic if there exist

- a substitution $\varphi\colon \mathcal{B} \to \mathcal{B}^*$ of length k,
- a fixed point y of φ, i.e. $\varphi(\mathrm{y}) = \mathrm{y}$,
- a coding $\tau\colon \mathcal{B} \to \mathcal{A}$

such that $\mathrm{x} = \tau(\mathrm{y})$.

Substitutive sequences

One can consider general substitutions (not necessarily of constant length).

A sequence $x \in \mathcal{A}^{\mathbb{N}}$ is substitutive if there exist

- a substitution $\varphi\colon \mathcal{B} \to \mathcal{B}^+$ such that $\varphi(a)$ starts with a for some $a \in \mathcal{B}$,
- an (infinite) fixed point y of φ given by

$$y = \varphi^{\infty}(a) = \lim_{n\to\infty} \varphi^{n}(a),$$

- a coding $\tau\colon \mathcal{B} \to \mathcal{A}$

such that $x = \tau(y) = \tau(\varphi^{\infty}(a))$.

Problem A [Allouche, Dekking, Queffélec (2021)]

Let x be a fixed point of a substitution. Decide whether it is automatic.

Problem A was already explicitly stated in the book of Allouche and Shallit [Problem 3, section 7.11] (2003) and investigated much earlier by Dekking (1976).

A hidden automatic sequence [Allouche, Queffélec (in an unpublished note), Grigorchuk, Lenz, Nagnibeda, (2017)]

Lysenok substitution:

$$\mathrm{L}\colon \mathrm{a} \mapsto \mathrm{aca}, \quad \mathrm{b} \mapsto \mathrm{d}, \quad \mathrm{c} \mapsto \mathrm{b}, \quad \mathrm{d} \mapsto \mathrm{c}.$$

Used by Lysenok to give a recursive presentation by generators and relations of the first Grigorchuk group G (1985):

$$\begin{aligned} \mathrm{G} = <\mathrm{a}, \mathrm{b}, \mathrm{c}, \mathrm{d} \mid 1 &= \mathrm{a}^2 = \mathrm{b}^2 = \mathrm{c}^2 = \mathrm{d}^2 = \\ &= \mathrm{bcd} = \mathrm{L}^k((\mathrm{ad})^4) = \mathrm{L}^k((\mathrm{adacac})^4),\ k \in \mathbb{N} > . \end{aligned}$$

The fixed point of L is 2-automatic, it is a fixed point of

$$\mathrm{a} \mapsto \mathrm{ac}, \quad \mathrm{b} \mapsto \mathrm{ad}, \quad \mathrm{c} \mapsto \mathrm{ab}, \quad \mathrm{d} \mapsto \mathrm{ac}.$$

A nonautomatic sequences: gaps in the Thue–Morse sequence

Let $x = 01101001\ldots$ be the Thue–Morse sequence.
The sequence encoding the differences of the consecutive occurrences of $w = 1$ in x is again 2-automatic.

What about words $w \in L(x)$ of length ≥ 2?

Theorem [Spiegelhofer (2021)]

Let w be a factor of the Thue—Morse sequence of length at least 2. The sequence of gaps between consecutive occurrences of w in x is not automatic.

A nonautomatic sequences: gaps in the Thue–Morse sequence

A nonautomatic sequence [Spiegelhofer (2021)]

Let B = 33423 ... be given by the substitution

$$\varphi\colon \mathrm{a} \mapsto \mathrm{a}\bar{\mathrm{a}}\mathrm{bc},\ \bar{\mathrm{a}} \mapsto \mathrm{a}\bar{\mathrm{a}}\mathrm{cb},\ \mathrm{b} \mapsto \mathrm{a}\bar{\mathrm{a}}\mathrm{bcb},\ \mathrm{c} \mapsto \mathrm{a}\bar{\mathrm{a}}\mathrm{c},$$

and the coding

$$\tau\colon \mathrm{a} \mapsto 3,\ \bar{\mathrm{a}} \mapsto 3,\ \mathrm{b} \mapsto 4,\ \mathrm{c} \mapsto 2.$$

Sequence B encodes the differences of the consecutive occurrences of the word 01 in the Thue–Morse sequence.

To show the previous Theorem, it is enough to show that B is not automatic.

How to check automaticity

What tools do we have for checking automaticity?

Incidence matrix of a substitution

Let $\varphi\colon \mathcal{A} \to \mathcal{A}^*$ be a substitution. With φ we associate the incidence matrix $\mathrm{M}_\varphi = [|\varphi(\mathrm{b})|_\mathrm{a}]_{\mathrm{a},\mathrm{b}\in\mathcal{A}}$, e.g. for

$$\varphi\colon 0 \mapsto 01, \quad 1 \mapsto 00,$$

we have

$$\mathrm{M}_\varphi = \begin{bmatrix} 1 & 2 \\ 1 & 0 \end{bmatrix}.$$

Such M_φ has a dominant (Perron–Frobenius) eigenvalue $\alpha > 0$ (i.e. $|\gamma| \leq \alpha$ for any other eigenvalue γ).

For each $\mathrm{w} \in \mathcal{A}^*$: $(|\varphi^\mathrm{n}(\mathrm{w})|)_\mathrm{n}$ = linear recurrence sequence with characteristic polynomial given by the minimal polynomial of M_φ.

Dominant eigenvalue has to be an integer

Theorem [Durand (2011)]

Let x be a substitutive sequence given by some substitution φ and coding τ. Assume x is k-automatic and not ultimately periodic. Then the dominant eigenvalue of M_φ is multiplicatively dependent with k.

It is well-known this is not sufficient for automaticity.

A nice sufficient linear condition

Theorem [Dekking (1976)]

Let $\varphi\colon \mathcal{A} \to \mathcal{A}^+$ be a substitution with a fixed point x. If the length vector ${}^t(|\varphi(a)|)_{a\in\mathcal{A}}$ is a left eigenvector of M_φ, then x is automatic.

The length vector ${}^t(|\varphi(a)|)_{a\in\mathcal{A}}$ is a left eigenvector of M_φ (corresponding to the dominant eigenvalue k) iff $|\varphi^n(a)| = c_a k^n$ for all $a \in \mathcal{A}$ for some $c_a \in \mathbb{N}$.

This condition is not necessary (there is no iff linear condition).

Other examples of hidden automatic sequences and methods:
Allouche, Dekking and Queffélec, "Hidden automatic sequences", Comb. Theory 1 (2021).

Allouche, Shallit, Yassawi, "How to prove that a sequence is not automatic", Expo. Math. 40, 1–22 (2022).

Other examples of hidden automatic sequences and methods: Allouche, Dekking and Queffélec, "Hidden automatic sequences", Comb. Theory 1 (2021).

Allouche, Shallit, Yassawi, "How to prove that a sequence is not automatic", Expo. Math. 40, 1–22 (2022).

Can we come up with a nice checkable if and only if condition for automaticity of substitutive sequences?

Main result: assumptions

We assume that our sequence x is not ultimately periodic.

We assume that our substitution φ is nondegenerate, i.e., the quotient of no two eigenvalues of the incidence matrix M_φ is a root of unity and 1 is the only root of unity which is an eigenvalue of M_φ.

Each substitution has some power which is nondegenerate.

Main result

Let x be a fixed point of a nondegenerate substitution $\varphi\colon \mathcal{A} \to \mathcal{A}^*$. Assume that x is not ultimately periodic. Let $\mathcal{L}(x)$ denote the set of all words appearing in x.

Theorem I [K., Müllner]

The following conditions are equivalent:

1. The sequence x is automatic.
2. For all $v \in \mathcal{L}(x)$ we have

$$|\varphi^n(v)| = c_v k^n + d_v \text{ for } n \geq |\mathcal{A}|,$$

for some $c_v, d_v \in \mathbb{Q}$ and $k \in \mathbb{N}$ with d_v, $v \in \mathcal{L}(x)$ taking only finitely many values.

Condition (2), in particular, implies that $|\varphi^n(a)| = c_a k^n + d_a$ for $n \geq 0$ for all $a \in \mathcal{A}$.

Main result

For each $a \in \mathcal{A}$ and $m \geq 1$, let $P_a(m)$ be the set of nonempty prefixes v of $\varphi^m(a)$ such that va is a prefix of $\varphi^m(a)$.

For example, let

$$\varphi(a) = bab \quad \varphi(b) = ab.$$

Then $P_a(1) = \{b\}$ and $P_a(2) = \{abb, abbab\}$ since

$$\varphi^2(a) = abbabab.$$

Main result

For each $a \in \mathcal{A}$ and $m \geq 1$, let $P_a(m)$ be the set of prefixes v of $\varphi^m(a)$ such that va is a prefix of $\varphi^m(a)$.

Theorem I (continued) [K., Müllner]

The following conditions are equivalent:

1. The sequence x is automatic.
2. For each $a \in \mathcal{A}$, $|\varphi^n(a)| = c_a k^n + d_a$ for $n \geq |\mathcal{A}|$ with $c_a, d_a \in \mathbb{Q}$ and
$$|\varphi^n(v)| = c_v k^n \quad \text{for} \quad n \geq |\mathcal{A}|$$
for all $v \in P_m(a)$, $a \in \mathcal{A}$, and $1 \leq m \leq |\mathcal{A}|$.

Here, $n \geq |\mathcal{A}|$ can be replaced by $n \geq s$, where s denotes the size of the largest Jordan block of the incidence matrix M_φ corresponding to the eigenvalue 0.

Main result

Now assume that x is any substitutive sequence given as the image by a coding τ of a fixed point of a nondegenerate substitution $\varphi\colon \mathcal{A} \to \mathcal{A}^*$.

If x is aperiodic (i.e., the orbit closure of x has no periodic points) Theorem I is true. In particular, automaticity of x depends only on the substitution φ (and not on the coding τ).

It is not true in the general case (example later).

A version for uniformly recurrent sequences

Theorem [Dekking (1976)]

Let $\varphi\colon \mathcal{A} \to \mathcal{A}^+$ be a substitution with a fixed point x. If the length vector ${}^t(|\varphi(\mathrm{a})|)_{\mathrm{a}\in\mathcal{A}}$ is a left eigenvector of M_φ, then x is automatic.

For uniformly recurrent x, Dekking's criterion applied to the return substitution $\varrho\colon \mathrm{R}_\mathrm{a} \to \mathrm{R}_\mathrm{a}^*$ essentially gives an iff condition for automaticity of x.

A version for uniformly recurrent sequences

Let $x = \tau(\varphi^\infty(a))$ be a substitutive sequence. Assume x is uniformly recurrent and nonperiodic.

Let $\varrho\colon R_a \to R_a^*$ be the return substitution to the letter a and let M_ϱ denote the incidence matrix of ϱ.

Corollary

The following conditions are equivalent:

1. x is automatic;
2. ${}^t(|\varphi^s(w)|)_{w\in R_a}$ is a left eigenvector of M_ϱ, where s is the size of the largest Jordan block of M_φ corresponding to the eigenvalue 0.

Theorem I does not hold in general

Consider a substitution

$$\varphi(\alpha) = \alpha\mathrm{ab}\alpha, \quad \varphi(\mathrm{a}) = \mathrm{abbbba}, \quad \varphi(\mathrm{b}) = \mathrm{aa}$$

and a coding

$$\tau(\alpha) = \alpha, \quad \tau(\mathrm{a}) = \tau(\mathrm{b}) = 0.$$

and the corresponding automatic sequence

$$\mathrm{x} = \tau(\varphi^{\omega}(\alpha)) = \alpha 00 \alpha 0^{8} \alpha 00 \alpha 0^{32} \alpha \ldots$$

Theorem I does not hold in general

Consider a substitution

$$\varphi(\alpha) = \alpha\mathrm{ab}\alpha, \quad \varphi(\mathrm{a}) = \mathrm{abbbba}, \quad \varphi(\mathrm{b}) = \mathrm{aa}$$

and a coding

$$\tau(\alpha) = \alpha, \quad \tau(\mathrm{a}) = \tau(\mathrm{b}) = 0.$$

and the corresponding automatic sequence

$$\mathrm{x} = \tau(\varphi^{\omega}(\alpha)) = \alpha 00 \alpha 0^{8} \alpha 00 \alpha 0^{32} \alpha \ldots$$

We have

$$|\varphi^{\mathrm{n}}(\mathrm{a})| = \frac{4}{3} \cdot 4^{\mathrm{n}} - \frac{1}{3}(-2)^{\mathrm{n}}, \quad |\varphi^{\mathrm{n}}(\mathrm{b})| = \frac{2}{3} \cdot 4^{\mathrm{n}} + \frac{1}{3}(-2)^{\mathrm{n}}, \quad |\varphi^{\mathrm{n}}(\alpha)| = 4^{\mathrm{n}}.$$

The fixed point $\mathrm{x}' = \varphi^{\omega}(\alpha)$ is not automatic. However, the sequence x is 2-automatic: it is a fixed point of a substitution

$$\varphi'(\alpha) = \alpha 00 \alpha, \quad \varphi'(0) = 0000.$$

General case

Call a representation $(\varphi, \mathcal{A}, \alpha, \tau)$ of a substitutive sequence x minimal if x cannot be given by some $(\varphi', \mathcal{A}', \alpha', \tau')$ with $|\mathcal{A}'| < |\mathcal{A}|$.

Is it true that if a substitutive sequence x given by a minimal representation $(\varphi, \mathcal{A}, \alpha, \tau)$ is automatic, then all letters have growth of the type $ck^n + d$?

Recognizability in purely automatic systems

Let $\varphi\colon \mathcal{A} \to \mathcal{A}^*$ be a substitution of constant length k with a fixed point x. Let X be the system generated by x.

Theorem [Berthé, Steiner, Thuswaldner, Yassawi (2020)]

For each nonperiodic $y \in X$, there exists a unique $0 \le c < k$ and $y' \in X$ such that

$$y = T^c(\varphi(y')).$$

Hence, for each $n \ge 1$ and nonperiodic $y \in X$, there exists a unique $0 \le c_n < k^n$ and $y^{(n)} \in X$ such that

$$y = T^{c_n}(\varphi^n(y^{(n)})).$$

Recognizability in purely automatic systems

For a sequence $x = (x_n)_n$ and a word $w \in \mathcal{L}(x)$ of length $t \geq 1$ we let

$$N(x, w) = \{n \in \mathbb{N} \mid x_{[n,n+t)} = w\}$$

denote the set of all occurrences of w in x.

Let $\mathcal{L}_{per}(x)$ be the union of languages of all periodic subsystems of the system generated by x. A "compactness argument" gives:

Corollary

For each $w \in \mathcal{L}(x) \setminus \mathcal{L}_{per}(x)$ we have

$$N(x, w) - N(x, w) \equiv 0 \bmod k^{n(|w|)},$$

for some nondecreasing $n(|w|) \in \mathbb{N}$, $n(|w|) \to \infty$ as $|w| \to \infty$.

We want: a version for general (non purely) automatic sequences and control over $n(|w|)$.

Let x be a k-automatic sequence. Let $\mathcal{L}_{per}(x)$ be the union of languages of all periodic subsystems of the system generated by x.

Theorem [K., Müllner]

Let x be a k-automatic sequence. There exist integers $t \geq 1$, $M \geq 0$ and a finite set $F \subset \mathbb{Z}(k^t - 1)^{-1}$ such that for each $w \in \mathcal{L}(x) \setminus \mathcal{L}_{per}(x)$ long enough

$$N(x, w) - N(x, w) \subset F \bmod k^{l(w)-M},$$

where

$$l(w) = \lfloor \log_k |w| \rfloor \ \text{(i.e., } k^{l(w)} \leq |w| < k^{l(w)+1}\text{)}.$$

The statement clearly does not hold for words $w \in \mathcal{L}_{per}(x)$.

How we use it

Let x be a fixed point of a substitution $\varphi\colon \mathcal{A} \to \mathcal{A}^*$. Assume x is k-automatic.

Let $u \in \mathcal{L}(x)$. A word w is called a return word to u if wu lies in $\mathcal{L}(x)$ and u is a prefix of wu. For example, if u = aa, then

$$x = \mathrm{aacb|aaccc|a|aabb|aa}\ldots$$

Claim C

There exists a finite set $F \subset \mathbb{Q}$ such that for each nonperiodic $z \in \mathcal{L}(x)$ (i.e., $\varphi^n(z) \notin \mathcal{L}_{per}(x)$ for large n) and each return word w to z with the same growth type as z we have

$$|\varphi^n(w)| = c_w k_z^n + d_w, \quad n \geq |\mathcal{A}| \tag{1}$$

for some integer k_z multiplicatively dependent with k, $c_w \in \mathbb{Q}$, and $d_w \in F$.

How we use it

1. For each $n \geq 1$, $\varphi^n(w)$ is a return word to $\varphi^n(z)$.
2. $|\varphi^n(w)| \in N(x, \varphi^n(z)) - N(x, \varphi^n(z))$ $n \geq 1$.
3. $|\varphi^n(w)| \in F \bmod k^{l(\varphi^n(z))-M}$ for n big enough, where $l(v) = \lfloor \log_k |v| \rfloor$ for a word v.
4. $|\varphi^n(w)| = c_w(n)k^{l(\varphi^n(z))-M} + d_w(n) = c_w(n)k^{m(n)} + d_w(n)$ with $c_w(n) \in \mathbb{Z}(k^t - 1)^{-1}$, $d_w(n) \in F$, and $m(n) = l(\varphi^n(z)) - M$ an increasing sequence of integers.
5. The sequences $c_w(n)$ and $d_w(n)$ take only finitely many values.
6. $u(n) = |\varphi^n(w)|$ is a nondegenerate linear recurrence sequence and so $u(n) = c_w k_z^n + d_w$ for some $k_z \geq 2$ multiplicatively dependent with k and $c_w, d_w \in \mathbb{Q}$, $d_w \in F$.

How we use it

Claim C

There exists a finite set $F \subset \mathbb{Q}$ such that for each nonperiodic $z \in \mathcal{L}(x)$ (i.e., $\varphi^n(z) \notin \mathcal{L}_{per}(x)$ for large n) and each return word w to z with the same growth type as z we have

$$|\varphi^n(w)| = c_w k_z^n + d_w, \quad n \geq |\mathcal{A}| \tag{2}$$

for some integer k_z multiplicatively dependent with k, $c_w \in \mathbb{Q}$, and $d_w \in F$.

Claim C + some combinatorics on words implies that

$$|\varphi^n(v)| = c_v k^n + d_v, \quad n \geq |\mathcal{A}| \tag{3}$$

for all $v \in \mathcal{L}(x)$ with d_v taking finitely many values (i.e., condition (2) in main theorem). □