Substitutive systems and a finitary version of Cobham's theorem.

Jakub Byszewski

joint work with Jakub Konieczny and Elżbieta Krawczyk

JAGIELLONIAN UNIVERSITY IN KRAKÓW

Liège, 22 May 2023

Substitutive sequences

- $\mathcal{A}$ is a finite alphabet.
- $\varphi\colon \mathcal{A}^* \to \mathcal{A}^*$ is a substitution.
- We always assume that substitutions are growing, i.e. $\lim_{n\to\infty} |\varphi^n(a)| = \infty$ for $a \in \mathcal{A}$.
- A letter $a \in \mathcal{A}$ is prolongable if $\varphi(a) = av$ for some $v \in \mathcal{A}^*$; it gives rise to $\varphi^\omega(a) = av\varphi(v)\cdots$. Such sequences are called purely substitutive.
- A letter $a \in \mathcal{A}$ is backwards prolongable if $\varphi(a) = va$ for some $v \in \mathcal{A}^*$; it gives rise to ${}^\omega\varphi(a) = \cdots\varphi(v)va$
- A sequence is substitutive if it arises as the image of a purely substitutive sequence by a coding.
- The language $\mathcal{L}(x)$ of a sequence $x \in \mathcal{A}^\omega$ is its set of factors.

Substitutive sequences as dynamical systems

- $\mathcal{A}^\omega$ is a dynamical system with respect to the shift map (a compact space X with a continuous self-map).
- A subsystem is a nonempty closed subset invariant under the shift.
- The closed orbit $\overline{\mathrm{Orb}(x)}$ of $x \in \mathcal{A}^\omega$ consists of $y \in \mathcal{A}^\omega$ such that $\mathcal{L}(y) \subset \mathcal{L}(x)$. These subsystems are called transitive.
- A system is minimal if it has no proper subsystems.
- Call a system substitutive/automatic if it arises as the closed orbit of a substitutive/automatic sequence.
- For a substitution $\varphi\colon \mathcal{A}^* \to \mathcal{A}^*$, define the system X_φ to consist of $z \in \mathcal{A}^\omega$ such that every $w \in \mathcal{L}(z)$ is a factor of $\varphi^n(a)$ for some $n \geq 0, a \in \mathcal{A}$.
- If φ is primitive, then X_φ is minimal.
- If X_φ is minimal, then it is subsitutive, and in fact the substitution can be chosen primitive.

Theorem A

Aim: describe all (transitive) subsystems of substitutive systems.

Theorem A

Aim: describe all (transitive) subsystems of substitutive systems.

Theorem A

- Every transitive subsystem of a substitutive system is substitutive.
- Every transitive subsystem of a k-automatic system is k-automatic.

Technical assumptions on the substitution

We call a substitution $\varphi\colon \mathcal{A}^* \to \mathcal{A}^*$ idempotent if for every $a, b \in \mathcal{A}$ and $n \geq 1$ the following holds:

- b appears in $\varphi(a)$ iff b appears in $\varphi^n(a)$.
- b appears at least twice in $\varphi(a)$ iff b appears at least twice in $\varphi^n(a)$.
- the initial letter of $\varphi(a)$ is prolongable.
- if a appears in $\varphi(a)$, if b is the last letter of $\varphi(a)$ such that a appears in $\varphi(b)$, and if c is the last letter of $\varphi(b)$ such that b appears in $\varphi(c)$, then $b = c$.

Technical assumptions on the substitution

We call a substitution $\varphi\colon \mathcal{A}^* \to \mathcal{A}^*$ idempotent if for every $a, b \in \mathcal{A}$ and $n \geq 1$ the following holds:

- b appears in $\varphi(a)$ iff b appears in $\varphi^n(a)$.
- b appears at least twice in $\varphi(a)$ iff b appears at least twice in $\varphi^n(a)$.
- the initial letter of $\varphi(a)$ is prolongable.
- if a appears in $\varphi(a)$, if b is the last letter of $\varphi(a)$ such that a appears in $\varphi(b)$, and if c is the last letter of $\varphi(b)$ such that b appears in $\varphi(c)$, then $b = c$.

Reduction to idempotent substitutions

- Every substitution has a power φ^n that is idempotent.
- If X_φ is transitive, then $X_\varphi = X_{\varphi^n}$.

Minimal subsystems

Minimal subsystems

- A substitutive system has only finitely many minimal subsystems.
- Assume $\varphi\colon \mathcal{A}^* \to \mathcal{A}^*$ is idempotent. Then every minimal subsystem of X is of the form

$$X_b = \{x \in \mathcal{A}^\omega \mid \text{ every } w \in \mathcal{L}(x) \text{ is a factor of some } \varphi^n(b)\}$$

for some $b \in \mathcal{A}$.

Example 1

- $\mathcal{A} = \{0,1,2,3\}$, $\varphi\colon \mathcal{A}^* \to \mathcal{A}^*$,

$$\varphi(0) = 12, \quad \varphi(1) = 11, \quad \varphi(2) = 23, \quad \varphi(3) = 32.$$

- $X = X_\varphi$. Aim: describe all (transitive) subsystems of X.
- $X_0 = X$, $X_1 = \{1^\omega\}$, $X_2 = X_3 = \mathrm{TM}$ (the Thue–Morse system on letters $2,3$).
- $y = {}^\omega\varphi(1)\varphi^\omega(2) = \cdots 1111.23323223\ldots.$
- For $n \in \mathbb{Z}$, let $y_{[n,\infty)} = y_n y_{n+1} \cdots$ and $Y_n = \overline{\mathrm{Orb}(y_{[n,\infty)})}$.
- $Y_0 = Y_1 = Y_2 = \cdots = \mathrm{TM}$.
- For $n < 0$, Y_n is the union of TM and n extra points: $y_{-n} \mapsto y_{-(n-1)} \mapsto \cdots \mapsto y_{-1} \mapsto y_0 \in \mathrm{TM}$.
- The subsystems of X are the following ones:

$$X, X_1, \mathrm{TM} \text{ and } Y_n \text{ for } n < 0$$

Example 2

- $\mathcal{A} = \{0, 1, 2, 3\}$, $\tau\colon \mathcal{A}^* \to \mathcal{A}^*$,

$$\tau(0) = 01023, \quad \tau(1) = 12, \quad \tau(2) = 22, \quad \tau(3) = 33.$$

- $X = X_\tau$. Aim: describe all (transitive) subsystems of X.
- $X_0 = X$, $X_1 = \{12^\omega, 2^\omega\}$, $X_2 = \{2^\omega\}$, $X_3 = \{3^\omega\}$.
- Let $U_k = \{2^n 3^\omega \mid n \le k\}$, $V_k = \{3^n 2^\omega \mid n \le k\}$, $k \ge 0$.
- Put $v = 01$, $w = 23$,

$$\begin{aligned} z &= \cdots \tau^2(v)\tau(v)v.0w\tau(w)\tau^2(w)\ldots \\ &= \cdots 010231201.0232^2 3^2 2^4 3^4 2^8 3^8 \cdots \end{aligned}$$

- $Z_n = \overline{\mathrm{Orb}(z_{[n,\infty)})} = \mathrm{Orb}(z_{[n,\infty)}) \cup \{3^k 2^\omega, 2^k 3^\omega \mid k \ge 0\}$.
- The transitive subsystems of X are the following ones:

$$X, X_1, U_k, V_k \text{ for } k \ge 0, Z_n \text{ for } n \in \mathbb{Z}$$

Theorem B

Theorem B (simplified)

Let $\varphi\colon \mathcal{A}^* \to \mathcal{A}^*$ be an idempotent substitution. Let $\mathrm{y} \in \mathrm{X}_\varphi$ and let Y be the orbit closure of y. Then one of the following conditions holds:

1. <u>either</u> $\mathrm{Y} = \mathrm{X}_\mathrm{b}$ for some $\mathrm{b} \in \mathcal{A}$;
2. <u>or</u> there exist a backwards prolongable letter a and a prolongable letter b such that y is a suffix of ${}^\omega\varphi(\mathrm{a})\varphi^\omega(\mathrm{b})$.
3. <u>or</u> there exists a letter a such that $\varphi(\mathrm{a}) = \mathrm{v_a a w_a}$ for some words $\mathrm{v_a}$ and $\mathrm{w_a}$ such that $\mathrm{w_a} \neq \epsilon$ and y is a suffix of

$$\cdots \varphi^2(\mathrm{v_a})\varphi(\mathrm{v_a})\mathrm{v_a a w_a}\varphi(\mathrm{w_a})\varphi^2(\mathrm{w_a})\cdots .$$

Theorem B

Theorem B (simplified)

Let $\varphi\colon \mathcal{A}^* \to \mathcal{A}^*$ be an idempotent substitution. Let $y \in X_\varphi$ and let Y be the orbit closure of y. Then one of the following conditions holds:

1. <u>either</u> $Y = X_b$ for some $b \in \mathcal{A}$;
2. <u>or</u> there exist a backwards prolongable letter a and a prolongable letter b such that y is a suffix of ${}^\omega\varphi(a)\varphi^\omega(b)$.
3. <u>or</u> there exists a letter a such that $\varphi(a) = v_a a w_a$ for some words v_a and w_a such that $w_a \neq \epsilon$ and y is a suffix of

$$\cdots \varphi^2(v_a)\varphi(v_a)v_a a w_a \varphi(w_a)\varphi^2(w_a)\cdots .$$

Corollary

Either $Y = X_b$ or y is itself substitutive.

Cobham's theorem

Theorem (Cobham, 1969)

If $k, l \geq 2$ are mutiplicatively independent, then a sequence is simultaneously k-automatic and l-automatic if and only if it is ultimately periodic.

Cobham's theorem

Theorem (Cobham, 1969)

If $k, l \geq 2$ are mutiplicatively independent, then a sequence is simultaneously k-automatic and l-automatic if and only if it is ultimately periodic.

Let $k, l \geq 2$ be multiplicatively independent, let x be a k-automatic sequence and let y be an l-automatic. Assume x, y are not ultimately periodic.

Philosophy (Shallit)

Not only is $x \neq y$, but the common factors of x and y cannot be too complicated.

Cobham's theorem

Theorem (Cobham, 1969)

If $k, l \geq 2$ are mutiplicatively independent, then a sequence is simultaneously k-automatic and l-automatic if and only if it is ultimately periodic.

Let $k, l \geq 2$ be multiplicatively independent, let x be a k-automatic sequence and let y be an l-automatic. Assume x, y are not ultimately periodic.

Philosophy (Shallit)

Not only is $x \neq y$, but the common factors of x and y cannot be too complicated.

In fact Mol–Rampersad–Shallit–Stipulanti (2018) and Krawczyk (2023) got explicit bounds on the length of a common prefix.

Example: what kind of common factors can one get?

- $\mathcal{A} = \{0, 1, 2\}$.
- $\varphi\colon \mathcal{A}^* \to \mathcal{A}^*$, $\varphi(0) = 012$, $\varphi(1) = 111$, $\varphi(2) = 222$.
- $\mathrm{x} = \varphi^\omega(0) = 0121^3 2^3 1^9 2^9 1^{27} 2^{27} \cdots$ is 3-automatic.

Example: what kind of common factors can one get?

- $\mathcal{A} = \{0, 1, 2\}$.
- $\varphi\colon \mathcal{A}^* \to \mathcal{A}^*$, $\varphi(0) = 012$, $\varphi(1) = 111$, $\varphi(2) = 222$.
- $\mathrm{x} = \varphi^\omega(0) = 0121^3 2^3 1^9 2^9 1^{27} 2^{27} \cdots$ is 3-automatic.
- $\tau\colon \mathcal{A}^* \to \mathcal{A}^*$, $\tau(0) = 0121$, $\tau(1) = 1111$, $\tau(2) = 2222$.
- $\mathrm{y} = \tau^\omega(0) = 0121^5 2^4 1^{20} 2^{16} 1^{80} 2^{64} \cdots$ is 4-automatic.

Example: what kind of common factors can one get?

- $\mathcal{A} = \{0, 1, 2\}$.
- $\varphi\colon \mathcal{A}^* \to \mathcal{A}^*$, $\varphi(0) = 012$, $\varphi(1) = 111$, $\varphi(2) = 222$.
- $\mathrm{x} = \varphi^\omega(0) = 0121^3 2^3 1^9 2^9 1^{27} 2^{27} \cdots$ is 3-automatic.
- $\tau\colon \mathcal{A}^* \to \mathcal{A}^*$, $\tau(0) = 0121$, $\tau(1) = 1111$, $\tau(2) = 2222$.
- $\mathrm{y} = \tau^\omega(0) = 0121^5 2^4 1^{20} 2^{16} 1^{80} 2^{64} \cdots$ is 4-automatic.
- $\mathrm{X}_\varphi = \mathrm{Orb}(\mathrm{x}) \cup \{2^{\mathrm{n}} 1^\omega \mid \mathrm{n} \geq 0\} \cup \{1^{\mathrm{n}} 2^\omega \mid \mathrm{n} \geq 0\}$.
- $\mathrm{X}_\tau = \mathrm{Orb}(\mathrm{y}) \cup \{2^{\mathrm{n}} 1^\omega \mid \mathrm{n} \geq 0\} \cup \{1^{\mathrm{n}} 2^\omega \mid \mathrm{n} \geq 0\}$.

Example: what kind of common factors can one get?

- $\mathcal{A} = \{0, 1, 2\}$.
- $\varphi\colon \mathcal{A}^* \to \mathcal{A}^*$, $\varphi(0) = 012$, $\varphi(1) = 111$, $\varphi(2) = 222$.
- $\mathrm{x} = \varphi^\omega(0) = 0121^3 2^3 1^9 2^9 1^{27} 2^{27} \cdots$ is 3-automatic.
- $\tau\colon \mathcal{A}^* \to \mathcal{A}^*$, $\tau(0) = 0121$, $\tau(1) = 1111$, $\tau(2) = 2222$.
- $\mathrm{y} = \tau^\omega(0) = 0121^5 2^4 1^{20} 2^{16} 1^{80} 2^{64} \cdots$ is 4-automatic.
- $\mathrm{X}_\varphi = \mathrm{Orb}(\mathrm{x}) \cup \{2^{\mathrm{n}} 1^\omega \mid \mathrm{n} \geq 0\} \cup \{1^{\mathrm{n}} 2^\omega \mid \mathrm{n} \geq 0\}$.
- $\mathrm{X}_\tau = \mathrm{Orb}(\mathrm{y}) \cup \{2^{\mathrm{n}} 1^\omega \mid \mathrm{n} \geq 0\} \cup \{1^{\mathrm{n}} 2^\omega \mid \mathrm{n} \geq 0\}$.
- The common factors of x and y are exactly the words in

$$\mathcal{L}({}^\omega 12^\omega) \cup \mathcal{L}({}^\omega 21^\omega) \cup \mathcal{L}(0121^3).$$

Finitary version of Cobham's theorem

Theorem C

Let $k, l \geq 2$ be multiplicatively independent integers, let $\mathcal{A}$ be an alphabet, and let $U \subset \mathcal{A}^*$. The following conditions are equivalent:

Finitary version of Cobham's theorem

Theorem C

Let $k, l \geq 2$ be multiplicatively independent integers, let $\mathcal{A}$ be an alphabet, and let $U \subset \mathcal{A}^*$. The following conditions are equivalent:

(a) there exist a k-automatic sequence x and an l-automatic sequence y such that U is the set of common factors of x and y;

Finitary version of Cobham's theorem

Theorem C

Let $k, l \geq 2$ be multiplicatively independent integers, let $\mathcal{A}$ be an alphabet, and let $U \subset \mathcal{A}^*$. The following conditions are equivalent:

(a) there exist a k-automatic sequence x and an l-automatic sequence y such that U is the set of common factors of x and y;

(b) the set U is a finite union of sets of the form $\mathcal{L}({}^{\omega}vuw^{\omega})$, where u, v, w are (possibly empty) words over $\mathcal{A}$.

Finitary version of Cobham's theorem

Theorem C

Let $k, l \geq 2$ be multiplicatively independent integers, let $\mathcal{A}$ be an alphabet, and let $U \subset \mathcal{A}^*$. The following conditions are equivalent:

(a) there exist a k-automatic sequence x and an l-automatic sequence y such that U is the set of common factors of x and y;

(b) the set U is a finite union of sets of the form $\mathcal{L}({}^{\omega}vuw^{\omega})$, where u, v, w are (possibly empty) words over $\mathcal{A}$.

It is easy to show that the second property implies the first one.

Proof of the finitary version of Cobham's theorem

- $k, l \geq 2$ are multiplicatively independent
- $x \in \mathcal{A}^\omega$ is k-automatic, $X = \overline{\mathrm{Orb}(x)}$
- $y \in \mathcal{A}^\omega$ is l-automatic, $Y = \overline{\mathrm{Orb}(y)}$.

Proof of the finitary version of Cobham's theorem

- $k, l \geq 2$ are multiplicatively independent
- $x \in \mathcal{A}^\omega$ is k-automatic, $X = \overline{\mathrm{Orb}(x)}$
- $y \in \mathcal{A}^\omega$ is l-automatic, $Y = \overline{\mathrm{Orb}(y)}$.

Step I

Any $z \in X \cap Y$ is ultimately periodic.

Proof of the finitary version of Cobham's theorem

- $k, l \geq 2$ are multiplicatively independent
- $x \in \mathcal{A}^\omega$ is k-automatic, $X = \overline{\mathrm{Orb}(x)}$
- $y \in \mathcal{A}^\omega$ is l-automatic, $Y = \overline{\mathrm{Orb}(y)}$.

Step I

Any $z \in X \cap Y$ is ultimately periodic.

Proof: Let $Z = \overline{\mathrm{Orb}(z)} \subset X \cap Y$. By Theorem A there exists a k-automatic sequence x' and an l-automatic sequence y' such that $Z = \overline{\mathrm{Orb}(x')} = \overline{\mathrm{Orb}(y')}$. By Fagnot's generalisation of Cobham's theorem x', y' and z are ultimately periodic.

Proof of the finitary version of Cobham's theorem

We call a nonempty factor u of x cyclic if $u^\omega \in X$. Since X has only finitely many minimal subsystems, it has finitely many primitive cyclic factors.

Proof of the finitary version of Cobham's theorem

We call a nonempty factor u of x cyclic if $u^\omega \in X$. Since X has only finitely many minimal subsystems, it has finitely many primitive cyclic factors.

Step II

Let $u, v, w \in \mathcal{A}^*$ and let $S = \{n \geq 0 \mid vu^n w \in \mathcal{L}(X)\}$. Then

1. either $S = \mathbb{N}$, and moreover in this case either v is a suffix or w is a prefix of a power of u;
2. or S is a finite union of sets of the form $\{ak^{mn} + b \mid n \geq 0\}$ for some $a, b \in \mathbb{Q}$, $m \geq 1$.

Proof of the finitary version of Cobham's theorem

Corollary

If v is not a suffix and w is not a prefix of a power of u, then $vu^n w$ is a common factor of x and y only for finitely many n.

Proof: Use the description of the set of n such that $vu^n w$ is a factor of x (resp. y) given in Step II. Such sets have finite intersections since $ak^n + bl^m = c$ has only finitely many solutions in n, m.

Proof of the finitary version of Cobham's theorem

Corollary

If v is not a suffix and w is not a prefix of a power of u, then $vu^n w$ is a common factor of x and y only for finitely many n.

Proof: Use the description of the set of n such that $vu^n w$ is a factor of x (resp. y) given in Step II. Such sets have finite intersections since $ak^n + bl^m = c$ has only finitely many solutions in n, m.
These steps are effective.

Step III: Conclude by compactness

Let ℓ be the maximal length of a primitive cyclic common factor. Write any common factor t of x and y in the form

$$\mathrm{t} = \mathrm{v}_0 \mathrm{u}_1^{\mathrm{n}_1} \mathrm{v}_1 \mathrm{u}_2^{\mathrm{n}_2} \cdots \mathrm{v}_{\mathrm{s}-1} \mathrm{u}_\mathrm{s}^{\mathrm{n}_\mathrm{s}} \mathrm{v}_\mathrm{s} \tag{1}$$

for some integer $\mathrm{s} \geq 0$, integers $\mathrm{n}_\mathrm{i} \geq 0$, and words u_i, v_i such that:

1. u_i are primitive cyclic common factors,
2. v_i have length $\leq \ell$,
3. some minimality conditions to make the representation unique.

One then proves that the values of $\mathrm{n}_2, \ldots, \mathrm{n}_{\mathrm{s}-1}$ and s need to be bounded by a constant not depending on t, since otherwise we could construct an infinite non-ultimately periodic sequence in $\mathrm{X} \cap \mathrm{Y}$, contradicting Steps I and II. It is easy to conclude.

Effectivity of the result

Question

Is the above theorem effective?

Effectivity of the result

Question

Is the above theorem effective?

Theorem (Krawczyk, 2023)

There exists an algorithm that, given a k-automatic sequence x and an l-automatic sequence y, produces a finite set of words u_i, v_i, w_i over $\mathcal{A}$ such that the set of common factors of x and y is

$$\bigcup_i \mathcal{L}({}^{\omega}v_i u_i w_i^{\omega}).$$

Effectivity of the result

Question

Is the above theorem effective?

Theorem (Krawczyk, 2023)

There exists an algorithm that, given a k-automatic sequence x and an l-automatic sequence y, produces a finite set of words u_i, v_i, w_i over $\mathcal{A}$ such that the set of common factors of x and y is

$$\bigcup_i \mathcal{L}({}^\omega v_i u_i w_i^\omega).$$

More precisely, there exists a computable constant C (depending only on k, l and the numbers of states of the automata generating x and y) such that the lengths of u_i, v_i, w_i are bounded by C.