Algebraic properties of substitution words over infinite alphabets

Dirk Frettlöh

Joint work with Jan Mazáč

Technische Fakultät
Universität Bielefeld

Numeration 2026

Nancy 2. June 2026

Word substitutions on finite alphabets are a well-studied topic.

Example: $\mathcal{A} = \{a, b\}$, $\varrho : a \mapsto aba, \quad \varrho : b \mapsto ababa$.

$$a \mapsto aba \mapsto abaababaaba \mapsto abaababaabaabaababaabaababaabaabaababaaba \mapsto \cdots$$

Word substitutions on finite alphabets are a well-studied topic.

Example: $\mathcal{A} = \{a, b\}$, $\varrho : a \mapsto aba, \quad \varrho : b \mapsto ababa$.

$$a \mapsto aba \mapsto abaababaaba \mapsto abaababaabaababaabaababaabaabaababaaba \mapsto \cdots$$

Hull: the dynamical system (X_ϱ, σ), where X_ϱ is the set of all bi-infinite words generated by ϱ, and σ is the left shift.

Word substitutions on finite alphabets are a well-studied topic.

Example: $\mathcal{A} = \{a, b\}$, $\varrho : a \mapsto aba, \quad \varrho : b \mapsto ababa$.

$$a \mapsto aba \mapsto abaababaaba \mapsto abaababaabaabababaabaababaabaabababaaba \mapsto \cdots$$

Hull: the dynamical system (X_ϱ, σ), where X_ϱ is the set of all bi-infinite words generated by ϱ, and σ is the left shift.

But this is not the topic of this talk.

Important property: ϱ is *primitive*. That is, there is k such that $\varrho^k(x)$ contains all letters, for any $x \in \mathcal{A}$.

- substitution matrix $\begin{pmatrix} 2 & 3 \\ 1 & 2 \end{pmatrix}$,
- inflation factor $\lambda = 2 + \sqrt{3}$,
- relative frequencies: $\mathrm{freq}(a) = \frac{3}{3+\sqrt{3}}$, $\mathrm{freq}(b) = \frac{\sqrt{3}}{3+\sqrt{3}}$.

Recently (?) substitutions over infinite alphabets gained attention.

Goal:

- Consider word substitutions over an infinite alphabet $\mathcal{A} = \mathbb{N}$.
- Study which new phenomena do occur.

See for instance Sébastien Ferenczi (2006): Substitution dynamical systems on infinite alphabets.

Recently (?) substitutions over infinite alphabets gained attention.

Goal:

- Consider word substitutions over an infinite alphabet $\mathcal{A} = \mathbb{N}$.
- Study which new phenomena do occur.

See for instance Sébastien Ferenczi (2006): Substitution dynamical systems on infinite alphabets.
Example: The drunken man substitution (adapted to $\mathcal{A} = \mathbb{N}$)

$$\begin{aligned} 0 &\mapsto 0 \quad 1 \\ 1 &\mapsto 0 \quad 2 \\ 2 &\mapsto 1 \quad 3 \\ &\vdots \\ i &\mapsto i-1 \quad i+1 \end{aligned}$$

For convenience, we denote the letters by numbers.

The substitution "matrix" for the drunken man substitution:

$$\begin{pmatrix} 1 & 1 & 0 & 0 & 0 & \cdots \\ 1 & 0 & 1 & 0 & 0 & \cdots \\ 0 & 1 & 0 & 1 & 0 & \\ 0 & 0 & 1 & 0 & 1 & \ddots \\ \vdots & & \ddots & \ddots & \ddots & \ddots \end{pmatrix}$$

(Compare random walk on $\mathbb{N}$, expected location)

"eigenvalue" 2, "eigenvector" for 2: $(1, 1, 1, 1, \ldots)^T$.

Eigenvector "normalized": $(0, 0, 0, 0, \ldots)^T$.
Hence $\text{freq}(x) = 0$ for all $x \in \mathcal{A}$.

The substitution “matrix” for the drunken man substitution:

$$\begin{pmatrix} 1 & 1 & 0 & 0 & 0 & \cdots \\ 1 & 0 & 1 & 0 & 0 & \cdots \\ 0 & 1 & 0 & 1 & 0 & \\ 0 & 0 & 1 & 0 & 1 & \ddots \\ \vdots & & \ddots & \ddots & \ddots & \ddots \end{pmatrix}$$

(Compare random walk on $\mathbb{N}$, expected location)

“eigenvalue” 2, “eigenvector” for 2: $(1,1,1,1,\ldots)^T$.

Eigenvector ”normalized”: $(0,0,0,0,\ldots)^T$.
Hence $\mathrm{freq}(x) = 0$ for all $x \in \mathcal{A}$.

Problems:

- Not each infinite ”matrix” gives rise to a proper substitution (for instance, all tile frequencies 0, or negative frequencies.)
- What does “primitive” mean here?

Mañibo-Rust-Walton (2025): conditions for substitutions for infinitely many letters ("compact alphabets") with a proper inflation factor, and distinct tile frequencies.

- If the alphabet $\{0, 1, 2, \ldots, \} \cup \{\infty\}$ is compact,
- the substitution "matrix" is quasicompact,
- the substitution ϱ is continuous,
- and primitive,

then the hull (the infinite words) is uniquely ergodic.
In particular, all letters have well defined positive frequencies.

Mañibo-Rust-Walton (2025): conditions for substitutions for infinitely many letters ("compact alphabets") with a proper inflation factor, and distinct tile frequencies.

- If the alphabet $\{0, 1, 2, \dots, \} \cup \{\infty\}$ is compact,
- the substitution "matrix" is quasicompact,
- the substitution ϱ is continuous,
- and primitive,

then the hull (the infinite words) is uniquely ergodic.
In particular, all letters have well defined positive frequencies.

All terms and proofs are deep and tricky.

But this is not the topic of this talk.

However: the drunken man substitution is neither primitive nor quasicompact.

A first good example: Prototiles $0, 1, 2, 3, \ldots$ and ∞.

$$\begin{array}{rlll} 0 \mapsto & 0 & 1 & \\ i \mapsto & 0 & i-1 & i+1 \\ \vdots & \vdots & & \\ \infty \mapsto & 0 & \infty & \infty \end{array}$$

The substitution has a proper inflation factor: $\lambda = \frac{5}{2}$, and well-defined (distinct) frequencies $\mathrm{freq}(i)$:

$$\mathrm{freq}(i) = \frac{1}{2^{i+1}}$$

A first good example: Prototiles $0, 1, 2, 3, \ldots$ and ∞.

$$\begin{array}{rcccc} 0 & \mapsto & 0 & 1 & \\ i & \mapsto & 0 & i-1 & i+1 \\ \vdots & & \vdots & & \\ \infty & \mapsto & 0 & \infty & \infty \end{array}$$

The substitution has a proper inflation factor: $\lambda = \frac{5}{2}$, and well-defined (distinct) frequencies $\mathrm{freq}(i)$:

$$\mathrm{freq}(i) = \frac{1}{2^{i+1}}$$

This example kindled our interest.

Generalize the example above:

Let $(a_i)_i = (a_0, a_1, a_2, \ldots)$ with $1 \leqslant a_i \leqslant N$ for some $N \in \mathbb{N}$.

$$\text{Let } A = \begin{pmatrix} a_0 & 1 + a_1 & a_2 & a_3 & a_4 & \cdots \\ 1 & 0 & 1 & 0 & 0 & \cdots \\ 0 & 1 & 0 & 1 & 0 & \\ 0 & 0 & 1 & 0 & 1 & \ddots \\ \vdots & & \ddots & \ddots & \ddots & \ddots \end{pmatrix}$$

Generalize the example above:

Let $(a_i)_i = (a_0, a_1, a_2, \ldots)$ with $1 \leqslant a_i \leqslant N$ for some $N \in \mathbb{N}$.

$$\text{Let } A = \begin{pmatrix} a_0 & 1+a_1 & a_2 & a_3 & a_4 & \cdots \\ 1 & 0 & 1 & 0 & 0 & \cdots \\ 0 & 1 & 0 & 1 & 0 & \\ 0 & 0 & 1 & 0 & 1 & \ddots \\ \vdots & & \ddots & \ddots & \ddots & \ddots \end{pmatrix}$$

For instance, $a_i = 1$ for all $i \geqslant 0$ is the example above.

This yields a family of tile substitutions:

$$\begin{array}{rlll} 0 \mapsto & 0^{a_0} & 1 & \\ i \mapsto & 0^{a_i} & i-1 & i+1 \\ \vdots & \vdots & & \end{array}$$

This allowed us to prove:

Theorem (F-Garber-Mañibo 2024)

For any $\lambda > 2$ there is a primitive substitution with infinitely many prototiles having λ as inflation factor.

(The proof uses the family above)

This allowed us to prove:

Theorem (F-Garber-Mañibo 2024)

For any $\lambda > 2$ there is a primitive substitution with infinitely many prototiles having λ as inflation factor.

(The proof uses the family above)

Corollary

There are a lot of substitution tilings with transcendental inflation factor λ.

This allowed us to prove:

Theorem (F-Garber-Mañibo 2024)

For any $\lambda > 2$ there is a primitive substitution with infinitely many prototiles having λ as inflation factor.

(The proof uses the family above)

Corollary

There are a lot of substitution tilings with transcendental inflation factor λ.

But this is not the topic of this talk.

Topic of this talk are the following questions:

1. When are all letter frequencies contained in $\mathbb{Q}(\lambda)$?
2. Which properties of $(a_i)_i$ yield...
   2.1 inflation factors λ being algebraic numbers?
   2.2 inflation factors λ being algebraic integers?
   2.3 inflation factors λ being algebraic units?

In the finite alphabet case: 1, 2.1, and 2.2 are always true.

Regarding Question 1:

In the finite case we always have $\mathrm{freq}(x) \in \mathbb{Q}(\lambda)$.

Not here.

[1] For instance in the example above, $a_i = 1$ for all i: geometric series, $\mu = \frac{1}{2}$.

Regarding Question 1:

In the finite case we always have $\text{freq}(x) \in \mathbb{Q}(\lambda)$.

Not here.
There is a magic number $0 < \mu < 1$ such that [1]

$$\frac{1}{\mu} = \sum_{i=0}^{\infty} a_i \mu^i$$

Then $\lambda = \mu + \frac{1}{\mu}$.

[1] For instance in the example above, $a_i = 1$ for all i: geometric series, $\mu = \frac{1}{2}$.

Regarding Question 1:

In the finite case we always have $\text{freq}(x) \in \mathbb{Q}(\lambda)$.

Not here.
There is a magic number $0 < \mu < 1$ such that [1]

$$\frac{1}{\mu} = \sum_{i=0}^{\infty} a_i \mu^i$$

Then $\lambda = \mu + \frac{1}{\mu}$.

And the right eigenvector is always

$$(1 - \mu)(1, \mu, \mu^2, \mu^3, \ldots)^T.$$

Hence all frequencies are always contained in $\mathbb{Q}(\mu)$.

[1] For instance in the example above, $a_i = 1$ for all i: geometric series, $\mu = \frac{1}{2}$.

But there are examples such that the tile frequencies are not contained in $\mathbb{Q}(\lambda)$.

Let $(a_i)_i = (3, 17, 0, 17, 0, 17, 0, 17, \ldots)$. Then

$$\mu = 3 - 2\sqrt{2}, \quad \frac{1}{\mu} = \frac{1}{3 - 2\sqrt{2}} = 3 + 2\sqrt{2}.$$

But there are examples such that the tile frequencies are not contained in $\mathbb{Q}(\lambda)$.

Let $(a_i)_i = (3, 17, 0, 17, 0, 17, 0, 17, \ldots)$. Then

$$\mu = 3 - 2\sqrt{2}, \quad \frac{1}{\mu} = \frac{1}{3 - 2\sqrt{2}} = 3 + 2\sqrt{2}.$$

Therefore $\lambda = 3 - 2\sqrt{2} + 3 + 2\sqrt{2} = 6$. Hence $\mathbb{Q}(\lambda) = \mathbb{Q}$.

But for instance $\mathrm{freq}(0) = 2\sqrt{2} - 2 \notin \mathbb{Q}$.

Regarding Question(s) 2:

Theorem

Let the sequence $(a_i)_i$ be eventually periodic, with prefix length q and period p. Then μ and λ are algebraic numbers.

Regarding Question(s) 2:

Theorem

Let the sequence $(a_i)_i$ be eventually periodic, with prefix length q and period p. Then μ and λ are algebraic numbers.

Easy to see: consider $(a_i)_i = (3, 0, 2, 0, 0, 2, 0, 0, 2, 0, 0, \ldots)$. Then

$$
\begin{aligned}
& \frac{1}{\mu} = \sum_{i=0}^{\infty} a_i \mu^i = 3 + \sum_{i=0}^{\infty} 2\mu^{3i+2} \\
\iff \quad & \frac{1}{\mu} = 3 + 2\mu^2 \sum_{i=0}^{\infty} (\mu^3)^i \\
\iff \quad & \frac{1}{\mu} = 3 + \frac{2\mu^2}{1-\mu^3} \\
\iff & 1 - \mu^3 = 3\mu(1-\mu^3) + 2\mu^3
\end{aligned}
$$

The other direction is not true. Also easy to see:

Consider $(a_i)_i = (1, 1, 1, 1, 1, \ldots)$, yielding $\mu = \frac{1}{2}$.

The other direction is not true. Also easy to see:

Consider $(a_i)_i = (1,1,1,1,1,\ldots)$, yielding $\mu = \frac{1}{2}$.

Because $2 \cdot \frac{1}{4} = \frac{1}{2}$, we can use "carries":

$$\begin{aligned} 2 =& 1 + 1 \cdot \frac{1}{2} + 1 \cdot \frac{1}{4} + 1 \cdot \frac{1}{8} + 1 \cdot \frac{1}{16} + \cdots \\ =& 1 + 0 \cdot \frac{1}{2} + 3 \cdot \frac{1}{4} + 0 \cdot \frac{1}{8} + 3 \cdot \frac{1}{16} + \cdots \end{aligned}$$

So $(1,0,3,0,3,0,3,\ldots)$ also yields $\mu = \frac{1}{2}$.

The other direction is not true. Also easy to see:

Consider $(a_i)_i = (1, 1, 1, 1, 1, \dots)$, yielding $\mu = \frac{1}{2}$.

Because $2 \cdot \frac{1}{4} = \frac{1}{2}$, we can use "carries":

$$\begin{aligned} 2 =& 1 + 1 \cdot \frac{1}{2} + 1 \cdot \frac{1}{4} + 1 \cdot \frac{1}{8} + 1 \cdot \frac{1}{16} + \cdots \\ =& 1 + 0 \cdot \frac{1}{2} + 3 \cdot \frac{1}{4} + 0 \cdot \frac{1}{8} + 3 \cdot \frac{1}{16} + \cdots \end{aligned}$$

So $(1, 0, 3, 0, 3, 0, 3, \dots)$ also yields $\mu = \frac{1}{2}$.

In fact, we may change any pair 1,1 into 0,3 without changing μ. In particular we can obtain sequences $(a_i)_i$ which are not eventually periodic.

OK. Regarding Question 2.2 (λ being an algebraic *integer* or not)

Theorem

Let the sequence $(a_i)_i$ be eventually periodic, with prefix length q and period p. Then μ and λ are algebraic integers if and only if $a_{p+q-1} - a_{q-1} = \pm 1$. In this case μ is a unit as well.

OK. Regarding Question 2.2 (λ being an algebraic *integer* or not)

Theorem

Let the sequence $(a_i)_i$ be eventually periodic, with prefix length q and period p. Then μ and λ are algebraic integers if and only if $a_{p+q-1} - a_{q-1} = \pm 1$. In this case μ is a unit as well.

The last results were about the sequences. The following results regard the nature of μ and λ only.

So let μ (hence λ) be algebraic numbers.

OK. Regarding Question 2.2 (λ being an algebraic *integer* or not)

Theorem

Let the sequence $(a_i)_i$ be eventually periodic, with prefix length q and period p. Then μ and λ are algebraic integers if and only if $a_{p+q-1} - a_{q-1} = \pm 1$. In this case μ is a unit as well.

The last results were about the sequences. The following results regard the nature of μ and λ only.

So let μ (hence λ) be algebraic numbers.

Since $\lambda = \mu + \frac{1}{\mu}$, we have $\mu^2 - \lambda\mu + 1 = 0$.

Hence μ is (at most) quadratic over $\mathbb{Q}(\lambda)$.

OK. Regarding Question 2.2 (λ being an algebraic *integer* or not)

Theorem

Let the sequence $(a_i)_i$ be eventually periodic, with prefix length q and period p. Then μ and λ are algebraic integers if and only if $a_{p+q-1} - a_{q-1} = \pm 1$. In this case μ is a unit as well.

The last results were about the sequences. The following results regard the nature of μ and λ only.

So let μ (hence λ) be algebraic numbers.

Since $\lambda = \mu + \frac{1}{\mu}$, we have $\mu^2 - \lambda\mu + 1 = 0$.

Hence μ is (at most) quadratic over $\mathbb{Q}(\lambda)$.

So either $\mathbb{Q}(\mu) = \mathbb{Q}(\lambda)$, or $[\mathbb{Q}(\mu) : \mathbb{Q}(\lambda)] = 2$.

Lemma

Either $\mathbb{Q}(\mu) = \mathbb{Q}(\lambda)$, *or* $[\mathbb{Q}(\mu) : \mathbb{Q}(\lambda)] = 2$.
In other words, either $\deg(\mu) = 2\deg(\lambda)$, *or* $\deg(\mu) = \deg(\lambda)$.

Lemma

Either $\mathbb{Q}(\mu) = \mathbb{Q}(\lambda)$, *or* $[\mathbb{Q}(\mu) : \mathbb{Q}(\lambda)] = 2$.
In other words, either $\deg(\mu) = 2\deg(\lambda)$, *or* $\deg(\mu) = \deg(\lambda)$.

Theorem

If $\deg(\mu) = 2\deg(\lambda)$ *then* μ *and* λ *are algebraic integers.*

Lemma

Either $\mathbb{Q}(\mu) = \mathbb{Q}(\lambda)$, *or* $[\mathbb{Q}(\mu) : \mathbb{Q}(\lambda)] = 2$.
In other words, either $\deg(\mu) = 2\deg(\lambda)$, *or* $\deg(\mu) = \deg(\lambda)$.

Theorem

If $\deg(\mu) = 2\deg(\lambda)$ *then* μ *and* λ *are algebraic integers.*

Theorem

The minimal polynomial of μ *is self-reciprocal if and only if* $\deg(\mu) = 2\deg(\lambda)$.

Recall: a polynomial is self-reciprocal, if the coefficient vector is a palindrome. That is, if $p(x) = a_n x^n + a_{n-1}x^{n-1} + \cdots + a_1 x + a_0$, then $p(x) = a_0 x^n + a_1 x^{n-1} + \cdots + a_{n-1}x + a_n$.

E.g., $2x^3 - 3x^2 - 3x + 2$ *is* self-reciprocal, but $2x^3 - 3x^2 - 3x + 1$ is not.

So, regarding the frequencies being contained in $\mathbb{Q}(\lambda)$ or not:

Remember: we always have $\mathrm{freq}(x) \in \mathbb{Q}(\mu)$.
So, if $\deg(\lambda) = \deg(\mu)$, hence $\mathbb{Q}(\lambda) = \mathbb{Q}(\mu)$, then $\mathrm{freq}(x) \in \mathbb{Q}(\lambda)$.

So, regarding the frequencies being contained in $\mathbb{Q}(\lambda)$ or not:

Remember: we always have $\mathrm{freq}(x) \in \mathbb{Q}(\mu)$.
So, if $\deg(\lambda) = \deg(\mu)$, hence $\mathbb{Q}(\lambda) = \mathbb{Q}(\mu)$, then $\mathrm{freq}(x) \in \mathbb{Q}(\lambda)$.

This shows one direction. With some more effort one can show:

Theorem

$\deg(\lambda) = \deg(\mu)$ *if and only if for all* $x \in \mathcal{A}$ *holds:* $\mathrm{freq}(x) \in \mathbb{Q}(\lambda)$.

So, regarding the frequencies being contained in $\mathbb{Q}(\lambda)$ or not:

Remember: we always have $\mathrm{freq}(x) \in \mathbb{Q}(\mu)$.
So, if $\deg(\lambda) = \deg(\mu)$, hence $\mathbb{Q}(\lambda) = \mathbb{Q}(\mu)$, then $\mathrm{freq}(x) \in \mathbb{Q}(\lambda)$.

This shows one direction. With some more effort one can show:

Theorem

$\deg(\lambda) = \deg(\mu)$ *if and only if for all* $x \in \mathcal{A}$ *holds:* $\mathrm{freq}(x) \in \mathbb{Q}(\lambda)$*.*

So we obtain a full characterization:

Corollary

The minimal polynomial of μ *is self-reciprocal if and only if*
$2\deg(\lambda) = \deg(\mu)$ *if and only if*
not for all $x \in \mathcal{A}$ *holds:* $\mathrm{freq}(x) \in \mathbb{Q}(\lambda)$*.*

So, regarding the frequencies being contained in $\mathbb{Q}(\lambda)$ or not:

Remember: we always have $\mathrm{freq}(x) \in \mathbb{Q}(\mu)$.
So, if $\deg(\lambda) = \deg(\mu)$, hence $\mathbb{Q}(\lambda) = \mathbb{Q}(\mu)$, then $\mathrm{freq}(x) \in \mathbb{Q}(\lambda)$.

This shows one direction. With some more effort one can show:

Theorem

$\deg(\lambda) = \deg(\mu)$ *if and only if for all* $x \in \mathcal{A}$ *holds:* $\mathrm{freq}(x) \in \mathbb{Q}(\lambda)$.

So we obtain a full characterization:

Corollary

The minimal polynomial of μ *is self-reciprocal if and only if*
$2\deg(\lambda) = \deg(\mu)$ *if and only if*
not for all $x \in \mathcal{A}$ *holds:* $\mathrm{freq}(x) \in \mathbb{Q}(\lambda)$.

* *
*

Thank you!