Combinatorics on words: Pattern avoidance

Anna FRID

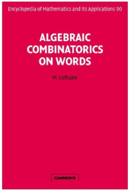
Aix-Marseille Université, September 2020

The main source

M. Lothaire, *Algebraic Combinatorics on Words*. Cambridge Univ. Press, 2002.

Chapter 3, Pattern avoidance (J. Cassaigne)

available online



Classic theorem: square-free word

Theorem (Thue, 1906)

There exists a word over the ternary alphabet which does not contain two consecutive equal factors.

$$abcacbabcbacabcacbcabcababc\cdots$$

Classic theorem: square-free word

Theorem (Thue, 1906)

There exists a word over the ternary alphabet which does not contain two consecutive equal factors.

$$abcacbabcbacabcacbcabcababc\cdots$$

No *abcabc* and in general, no XX for any finite X.

A bit of terminology

A *factor* of a finite or infinite word w is a finite word u such that $w = puw'$.
The length of u is denoted by $|u|$.

$$011010011001\cdots$$

A bit of terminology

A *factor* of a finite or infinite word w is a finite word u such that $w = puw'$.
The length of u is denoted by $|u|$.

$$01\textcolor{red}{10100}11001\cdots$$

A *square* is a word of the form XX, $X \in \Sigma^+$.

сносно, *bonbon*

To *avoid* a pattern = never contain such factors

Avoiding squares

No binary infinite word avoids squares

Avoiding squares

No binary infinite word avoids squares

What about ternary alphabet ?

Avoiding squares

No binary infinite word avoids squares

What about ternary alphabet ?

abacaba is a dead end

Avoiding squares

No binary infinite word avoids squares

What about ternary alphabet ?

abacaba is a dead end

$abcacbabcbacabcacb \cdots$ looks fine

Classic theorem: square-free word

Theorem (Thue, 1906)

There exists a word over the ternary alphabet which does not contain two consecutive equal factors.

Classic theorem: square-free word

Theorem (Thue, 1906)

There exists a word over the ternary alphabet which does not contain two consecutive equal factors.

Thue's original construction was bulky; we use another one.

Morphisms

Let Γ, Δ be alphabets. A *morphism* is a map $\varphi : \Gamma^* \to \Delta^*$ such that

$$\varphi(xy) = \varphi(x)\varphi(y).$$

Morphisms

Let Γ, Δ be alphabets. A *morphism* is a map $\varphi : \Gamma^* \to \Delta^*$ such that

$$\varphi(xy) = \varphi(x)\varphi(y).$$

A morphism is uniquely defined by words $\varphi(g)$, $g \in \Gamma$.

$$\varphi(a_1 a_2 \cdots a_n) = \varphi(a_1)\varphi(a_2) \cdots \varphi(a_n)$$

Thue-Morse morphism

The Thue-Morse morphism $\mu : \{0,1\}^* \to \{0,1\}^*$

$$\mu(0) = 01, \mu(1) = 10$$

Thue-Morse morphism

The Thue-Morse morphism $\mu : \{0,1\}^* \to \{0,1\}^*$

$$\mu(0) = 01, \mu(1) = 10$$

Тогда, например, $\mu(1101) = 10\ 10\ 01\ 10$.

Fixed points of morphisms

If $\varphi(a) = ax$ and the lengths $|\varphi^k(a)| \to \infty$, then there exists a limit of the sequence

$$a \to \varphi(a) \to \varphi^2(a) \to \varphi^3(a) \to \cdots .$$

Fixed points of morphisms

If $\varphi(a) = ax$ and the lengths $|\varphi^k(a)| \to \infty$, then there exists a limit of the sequence

$$a \to \varphi(a) \to \varphi^2(a) \to \varphi^3(a) \to \cdots .$$

The limit:

$$\mathbf{u} = \varphi(\mathbf{u}).$$

Fixed points of morphisms

If $\varphi(a) = ax$ and the lengths $|\varphi^k(a)| \to \infty$, then there exists a limit of the sequence

$$a \to \varphi(a) \to \varphi^2(a) \to \varphi^3(a) \to \cdots .$$

The limit:

$$\mathbf{u} = \varphi(\mathbf{u}).$$

$$\mathbf{u} = \varphi^\omega(a).$$

Thue-Morse word

$$\mu(0) = 01, \mu(1) = 10$$

$$0 \to 01 \to 01\ 10 \to 0110\ 1001 \to 01101001\ 10010110 \to \cdots$$

The Thue-Morse word

$$\mathbf{t} = \varphi^{\infty}(a) = 0110\ 1001\ 1001\ 0110\ 1001\ 0110\ 0110\ 1001 \cdots$$

Thue-Morse word

$$\mu(0) = 01, \mu(1) = 10$$

$$0 \to 01 \to 01\ 10 \to 0110\ 1001 \to 01101001\ 10010110 \to \cdots$$

The Thue-Morse word

$$\mathbf{t} = \varphi^{\infty}(a) = 0110\ 1001\ 1001\ 0110\ 1001\ 0110\ 0110\ 1001 \cdots$$

$$\mathbf{t} = \mu(\mathbf{t}) = \mu^{\omega}(0).$$

Thue-Morse word

$$\mu(0) = 01, \mu(1) = 10$$

$$0 \to 01 \to 01\ 10 \to 0110\ 1001 \to 01101001\ 10010110 \to \cdots$$

The Thue-Morse word

$$\mathbf{t} = \varphi^{\infty}(a) = 0110\ 1001\ 1001\ 0110\ 1001\ 0110\ 0110\ 1001 \cdots$$

$$\mathbf{t} = \mu(\mathbf{t}) = \mu^{\omega}(0).$$

Thue 1912; Morse 1921; Euwe 1929; Prouhet 1851

Overlap-free word

An *overlap* is a word of the form $aXaXa$, where a is a letter and X is a word (perhaps empty).

Overlap-free word

An *overlap* is a word of the form $aXaXa$, where a is a letter and X is a word (perhaps empty).

$$entente = e\ nt\ e\ nt\ e$$

Overlap-free word

An *overlap* is a word of the form $aXaXa$, where a is a letter and X is a word (perhaps empty).

$$entente = e\ nt\ e\ nt\ e$$

Theorem (Thue, 1912)

The Thue-Morse word is overlap-free.

Proof

$$\mathbf{t} = \mu^\omega(0) = 0110\ 1001\ 1001\ 0110\ 1001\ 0110\ 0110\ 1001 \cdots$$

Theorem (Thue, 1912)

The Thue-Morse word is overlap-free.

PROOF.

- 000, 111, 01010, 10101 are not factors of $\mathbf{t}$.
- Let *axaxa* be the *shortest* overlap in $\mathbf{t}$; then $|axaxa| \geq 5$, so $|axa| \geq 3$, so $|axa|$ contains 00 or 11, so the length of $|ax|$ is even.
- Taking every second symbol of *axaxa*, we observe a twice shorter overlap appearing in $\mathbf{t}$, a contradiction. □

A square-free word

Consider the morphism $\varphi : \{a, b, c\}^* \to \{a, b, c\}^*$:

$$\varphi : \begin{cases} a \mapsto abc, \\ b \mapsto ac, \\ c \mapsto b. \end{cases}$$

The fixed point $\mathbf{u} = abc\ ac\ b\ abcb\ ac\ abcacbac\ abcb \cdots$ of φ is square-free.

A square-free word

Consider the morphism $\varphi : \{a, b, c\}^* \to \{a, b, c\}^*$:

$$\varphi : \begin{cases} a \mapsto abc, \\ b \mapsto ac, \\ c \mapsto b. \end{cases}$$

The fixed point $\mathbf{u} = abc\ ac\ b\ abcb\ ac\ abcacbac\ abcb \cdots$ of φ is square-free.

Proof. $\mathbf{u} = \varphi(\mathbf{u})$; $\mathbf{t} = \mu(\mathbf{t})$.
Consider π: $\pi(a) = 011$, $\pi(b) = 01$, $\pi(c) = 0$. Then

- $\pi \circ \varphi = \mu \circ \pi$.
- $\mu(\pi(\mathbf{u})) = \pi(\varphi(\mathbf{u})) = \pi(\mathbf{u})$.
- $\pi(\mathbf{u}) = \mathbf{t}$.

If $\mathbf{u}$ contains a square XX, then $\mathbf{t}$ contains an overlap $\pi(X)\pi(X)a$. A contradiction. □

Generalizations

- Pattern avoidance;

Generalizations

- Pattern avoidance;
- Avoiding fractional powers (Dejean conjecture recently proven)

Generalizations

- Pattern avoidance;
- Avoiding fractional powers (Dejean conjecture recently proven)
- The number of words of length n avoiding something;

Generalizations

- Pattern avoidance;
- Avoiding fractional powers (Dejean conjecture recently proven)
- The number of words of length n avoiding something;
- Abelian powers, additive powers etc.

Pattern avoidance

A *pattern* is a finite word over the alphabet of variables.
A square = the pattern XX.
An occurrence of a pattern $P \in \Delta^* = \{X, Y, \cdots\}^*$ is a word $h(P)$ for some morphism $h : \Delta \to \Sigma^+$.

Example

$aba\ bb\ bb\ aba$ is an occurrence of $XYYX$.

Unavoidable patterns

The pattern X is unavoidable;

Unavoidable patterns

The pattern X is unavoidable;

the pattern XYX is unavoidable;

Unavoidable patterns

The pattern X is unavoidable;

the pattern XYX is unavoidable;

If P is unavoidable and X is a new variable which does not occur in P, then PXP is unavoidable.

Zimin patterns

Let us define Z_n by induction:

$$Z_0 = X_1,$$
$$Z_n = Z_{n-1} X_n Z_{n-1}.$$

Theorem

All patterns Z_n are unavoidable.

Zimin (1979,1982), Bean, Ehrenfeucht, McNulty (1979).

Unavoidable patterns

Theorem (Zimin (1979,1982), Bean, Ehrenfeucht, McNulty (1979))

A pattern is avoidable if and only if it is irreducible.

Unavoidable patterns

Theorem (Zimin (1979,1982), Bean, Ehrenfeucht, McNulty (1979))

A pattern is avoidable if and only if it is irreducible.

+a whole theory of pattern avoidance

Abelian powers

Let $|u|_a$ be the number of *a*s in *u*.
The Parikh vector of *u*: $\psi(u) = (|u|_a, |u|_b, \ldots)$.

Abelian powers

Let $|u|_a$ be the number of *a*s in u.
The Parikh vector of u: $\psi(u) = (|u|_a, |u|_b, \ldots)$.
X and X' are *abelian equivalent*, if $\psi(X) = \psi(X')$.
Then XX' is an *abelian square*.

Abelian powers

Let $|u|_a$ be the number of *a*s in u.
The Parikh vector of u: $\psi(u) = (|u|_a, |u|_b, \ldots)$.
X and X' are *abelian equivalent*, if $\psi(X) = \psi(X')$.
Then XX' is an *abelian square*.

Example

abbc cbab

Abelian squares?

Erdős, 1961

Are abelian squares avoidable over a finite alphabet?

Abelian squares?

Erdős, 1961

Are abelian squares avoidable over a finite alphabet?

Evdokimov, 1962

Yes, over the alphabet of 25 letters.

Abelian squares?

Erdős, 1961

Are abelian squares avoidable over a finite alphabet?

Evdokimov, 1962

Yes, over the alphabet of 25 letters.

Three letters are not enough; five are (Pleasants, 1965). What about 4 letters?

Keränen word

Theorem (Keränen, 1992)

The fixed point $\mathbf{w}$ *of the uniform symmetric morphism* φ *defined by*

$$\varphi(a) = abcacdcbcdcadcdbdabacabadbabcbdbcbacbcdcacbab$$
$$dabacadcbcdcacdbcbacbcdcacdcbdcdadbdcbca$$

avoids abelian squares.

Additive powers

Consider an alphabet subset of $\{0, 1, \cdots, n\}$.

Additive powers

Consider an alphabet subset of $\{0, 1, \cdots, n\}$.

Question

Can we avoid k consecutive words of the same sum?

Additive powers

Consider an alphabet subset of $\{0, 1, \cdots, n\}$.

Question

Can we avoid k consecutive words of the same sum?

0103 4

Additive powers

Consider an alphabet subset of $\{0, 1, \cdots, n\}$.

Question

Can we avoid k consecutive words of the same sum?

0103 4

Answer

NO (a corollary of the Szemerédi theorem)

Additive powers

Consider an alphabet subset of $\{0, 1, \cdots, n\}$.

Question

Can we avoid k consecutive words of the same sum and the same length?

Additive powers

Consider an alphabet subset of $\{0, 1, \cdots, n\}$.

Question

Can we avoid k consecutive words of the same sum and the same length?

Additive powers

Consider an alphabet subset of $\{0, 1, \cdots, n\}$.

Question

Can we avoid k consecutive words of the same sum and the same length?

Answer

YES for $k = 4$: There exists a binary word avoiding abelian 4-powers (Dekking, 1979).

Additive powers

Consider an alphabet subset of $\{0, 1, \cdots, n\}$.

Question

Can we avoid k consecutive words of the same sum and the same length?

Answer

YES for $k = 4$: There exists a binary word avoiding abelian 4-powers (Dekking, 1979).

What about $k = 3$? $k = 2$?

Additive cubes

Theorem (J. Cassaigne, J. Currie, L. Schaeffler, J. Shallit, 2011)

The fixed point of the morphism

$$\varphi : \begin{cases} 0 \mapsto 03, \\ 1 \mapsto 43, \\ 3 \mapsto 1, \\ 4 \mapsto 01. \end{cases}$$

over the alphabet $\{0, 1, 3, 4\}$ *avoids additive cubes.*

$$031430110343430310110 11 \cdots$$

Open problem

Are additive *squares* avoidable over a finite alphabet of integers?