Combinatorics on words: Introduction

Anna FRID

Aix-Marseille Université, September 2020

Alphabets and words

Alphabets and words

abracadabra is a finite word over the alphabet $\{a, b, c, d, r\}$

Alphabets and words

abracadabra is a finite word over the alphabet $\{a, b, c, d, r\}$

$31415926535897932384626433832795 0\cdots$ is an infinite word over the alphabet $\{0, 1, \ldots, 9\}$

Alphabets and words

abracadabra is a finite word over the alphabet $\{a, b, c, d, r\}$

$31415926535897932384626433832795 0\cdots$ is an infinite word over the alphabet $\{0, 1, \ldots, 9\}$

Notation:

$u \in \Sigma^*$ for finite words;

Alphabets and words

abracadabra is a finite word over the alphabet $\{a, b, c, d, r\}$

$31415926535897932384626433832795 0\cdots$ is an infinite word over the alphabet $\{0, 1, \ldots, 9\}$

Notation:

$u \in \Sigma^*$ for finite words;

$\mathbf{u} \in \Sigma^\omega$ for (right) infinite words;

Alphabets and words

abracadabra is a finite word over the alphabet $\{a, b, c, d, r\}$

$31415926535897932384626433832795 0\cdots$ is an infinite word over the alphabet $\{0, 1, \ldots, 9\}$

Notation:

$u \in \Sigma^*$ for finite words;

$\mathbf{u} \in \Sigma^\omega$ for (right) infinite words;

ε is the empty word of length 0.

Classic theorem: Fine and Wilf

Theorem

Let w be a finite word with periods p and q, of length

$$|w| \geq p + q - (p, q).$$

Then w is also periodic with period (p, q).

Classic theorem: Fine and Wilf

Theorem

Let w be a finite word with periods p and q, of length

$$|w| \geq p + q - (p, q).$$

Then w is also periodic with period (p, q).

Example: $p = 5,\ q = 8,\ w = abaab \cdots$

Classic theorem: Fine and Wilf

Theorem

Let w be a finite word with periods p and q, of length

$$|w| \geq p + q - (p, q).$$

Then w is also periodic with period (p, q).

Example: $p = 5,\ q = 8,\ w = abaab \cdots$

$$abaababaaba$$

Classic theorem: square-free word

Theorem

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

$$abcacbabcbacabcacbcabcababc\cdots$$

Classic theorem: square-free word

Theorem

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.

Connections

I. Algebra

Connections

I. Algebra

The set Σ^* of finite words over the alphabet Σ is a free monoid with concatenation; the identity element is the empty word ε.

Connections

I. Algebra

The set Σ^* of finite words over the alphabet Σ is a free monoid with concatenation; the identity element is the empty word ε.

$ab \cdot bca = abbca, \qquad ab \cdot \varepsilon = \varepsilon \cdot ab = ab$

Connections

I. Algebra

The set Σ^* of finite words over the alphabet Σ is a free monoid with concatenation; the identity element is the empty word ε.

$$ab \cdot bca = abbca, \qquad ab \cdot \varepsilon = \varepsilon \cdot ab = ab$$

We can also consider other semigroups and groups.

Connections

I. Algebra

Connections

I. Algebra

Theorem (Novikov, Adian, 1968)

For every odd number n with $n > 4381$, there exist infinite, finitely generated groups of exponent n.

Exponent of a group G is the least n such that $g^n = 1$ for all $g \in G$.

Connections

I. Algebra

Theorem (Novikov, Adian, 1968)

For every odd number n with $n > 4381$, there exist infinite, finitely generated groups of exponent n.

Exponent of a group G is the least n such that $g^n = 1$ for all $g \in G$.

This is the solution of the *bounded Burnside problem for groups*, and the proof uses the existence of a square-free word over a finite alphabet.

Connections

II.a Symbolic dynamics

Connections

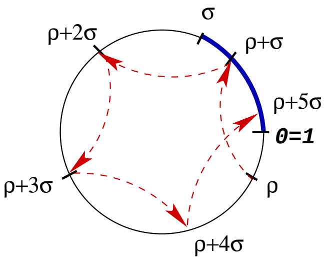
II.a Symbolic dynamics

Code of a trajectory of a point.

Connections

II.a Symbolic dynamics

Code of a trajectory of a point.

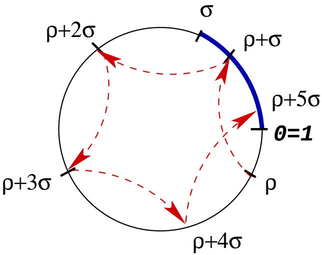


$w = 0{\color{blue}1}000{\color{blue}1}\cdots$

Connections

II.a Symbolic dynamics

Code of a trajectory of a point.



$w = 010001\cdots$

Rotation words are periodic or *Sturmian*.

Connections

II.b Discrete dynamics: shifts spaces

The shift operator σ:

$$\sigma(a[0]a[1]a[2]\cdots a[n]\cdots) = a[1]a[2]a[3]\cdots a[n+1]\cdots$$

Shift space = a closed set of infinite words invariant under σ

Connections

II.b Discrete dynamics: shifts spaces

The shift operator σ:

$$\sigma(a[0]a[1]a[2]\cdots a[n]\cdots) = a[1]a[2]a[3]\cdots a[n+1]\cdots$$

Shift space = a closed set of infinite words invariant under σ

Many properties of a *uniformly recurrent* infinite word depend only on its subshift.

Connections

III. Discrete geometry

Connections

III. Discrete geometry

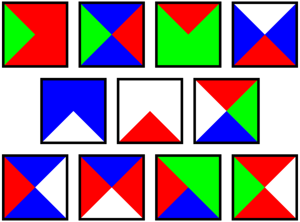
Tilings are like multidimensional infinite words

Connections

III. Discrete geometry

Tilings are like multidimensional infinite words

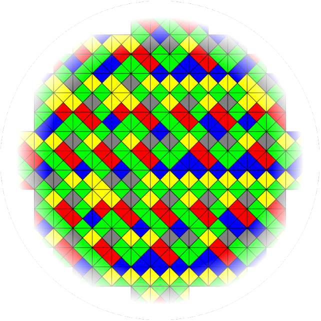
Example: Wang tiles



Connections

III. Discrete geometry

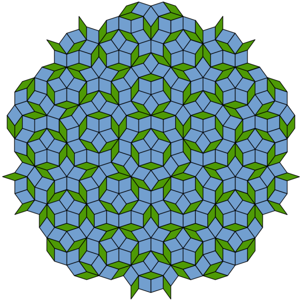
Tilings are like multidimensional infinite words
Example: Wang tiles



Connections

III. Discrete geometry

Tilings are like multidimensional infinite words
Another example: Penrose tilings



Connections

IV. Number theory
Decimal (or k-ary) expansions of irrational numbers are infinite words.
What are their properties?

$$\sqrt{2} = 1.41421356237309504880168872420969807856967187 5 \cdots$$

Does 7 occur in this word an infinite number of times?

Connections

IV. Number theory
Decimal (or k-ary) expansions of irrational numbers are infinite words.
What are their properties?

$$\sqrt{2} = 1.4142135623730950488016887242096980785696718750\cdots$$

Does 7 occur in this word an infinite number of times?

NOBODY KNOWS

Connections

IV. Number theory

Conjecture

All k-ary expansions of irrational algebraic numbers are normal, meaning that every finite pattern of length n occurs in this expansion with the expected limiting frequency k^n.

Connections

IV. Number theory

Conjecture

All k-ary expansions of irrational algebraic numbers are normal, meaning that every finite pattern of length n occurs in this expansion with the expected limiting frequency k^n.

This conjecture is not proven for *any* irrational algebraic number.

Connections

IV. Number theory

Theorem (Adamczewski and Bugeaud, 2007)

The complexity function $p(n)$, defined as the number of different patterns of length n, of the k-ary expansion of every irrational algebraic number satisfies

$$\liminf_{n\to\infty} \frac{p(n)}{n} = \infty.$$

Connections

IV. Number theory

Theorem (Adamczewski and Bugeaud, 2007)

The complexity function $p(n)$, defined as the number of different patterns of length n, of the k-ary expansion of every irrational algebraic number satisfies

$$\liminf_{n \to \infty} \frac{p(n)}{n} = \infty.$$

The proof uses combinatorics on words.

Other connections

- Formal languages

Other connections

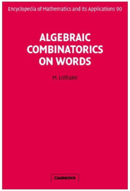
- Formal languages
- Algorithms on strings

Other connections

- Formal languages
- Algorithms on strings
- Applied algorithms: bioinformatics etc.

Main sources

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



available online

Main sources

Jean-Paul Allouche, Jeffrey Shallit, *Automatic Sequences — Theory, Applications, Generalizations*. Cambridge Univ. Press, 2003.