Transfinite Lyndon words

Luc Boasson and Olivier Carton

LIAFA – CNRS and Université Paris Diderot

ALGA-2016 Marseille

Outline

History

Lyndon words: introduced by Lyndon in 1954.
Their enumeration gives the Witt's formula for the dimension of the homogeneous component $\mathcal{L}_n(A)$ of the free Lie algebra.
If $\psi_k(d)$ is the number of such words of length d over an alphabet of size k, then

$$\sum_{d|n} d\psi_k(d) = k^n$$

By the Möbius inversion formula,

$$\psi_k(n) = \frac{1}{n} \sum_{d|n} \mu(d) k^{\frac{n}{d}}$$

Conjugacy

Definition (Conjugacy)

$w = u\,v$ $\qquad$ $w' = v\,u$

Two words w and w' are

- conjugates if $w = uv$ and $w' = vu$ for $u, v \in A^*$.
- proper conjugates if $w = uv$ and $w' = vu$ for $u, v \in A^+$.

Conjugacy class: $[w] = \{vu : w = uv\}$.

Examples

$aabab$ and $abaab$ are conjugates.

$[aabab] = \{aabab, ababa, babaa, abaab, baaba\}$.

$abab$ is a proper conjugate of itself $abab$

$aabab$ is not a proper conjugate of itself

Primitivity and lexicographic ordering

Definition (Primitive word)

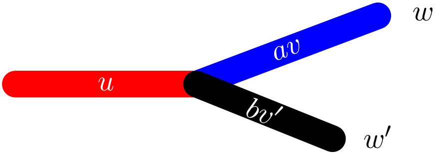
$w \neq$ 

The word w is primitive if it not equal to u^k for $k \geqslant 2$.

Examples

$aabab$ is primitive but $abab$ is not primitive.

Definition (Lexicographic ordering)



$$w < w' \quad \text{if} \quad \begin{cases} w = uav \text{ and } w' = ubv' \text{ for } a < b \\ w = u \text{ and } w' = uv \end{cases}$$

Lyndon words

Definition (Lyndon word)

A word $w \in A^*$ is Lyndon if w is strictly smaller for the lexicographic ordering than all its proper conjugates.

Examples

- a and b are Lyndon: no proper conjugate
- $aabab$ is Lyndon: $aabab < abaab < ababa < baaba < babaa$
- $abab$ is not Lyndon: $abab = abab$
- ba is not Lyndon: $ab < ba$

Alternative definitions

Proposition

The word w is Lyndon iff w is primitive and smaller that all its conjugates.

Examples

- $aabab$ is Lyndon: $aabab < abaab < ababa < baaba < babaa$
- $abab$ is not Lyndon: $abab = (ab)^2$ is not primitive
- ba is not Lyndon: $ab < ba$

Proposition

The word w is Lyndon iff w is (strictly) smaller that all its proper suffixes.

Examples

- $aabab$ is Lyndon: $aabab < ab < abab < b < bab$
- ba is not Lyndon: $a < ba$

Factorization

Theorem (Lyndon 1954)

Each word $w \in A^*$ has a unique factorization $w = u_1 \cdots u_n$ where each u_i is Lyndon and $u_1 \geqslant u_2 \geqslant \cdots \geqslant u_n$ (non-increasing).

D.E. Knuth suggested to call Lyndon words prime words.

Examples

$$aabab = aabab$$

$$ababa = ab \cdot ab \cdot a$$

$$babaa = b \cdot ab \cdot a \cdot a$$

$$abaab = ab \cdot aab$$

$$baaba = b \cdot aab \cdot a$$

Theorem (Duval 1980)

This factorization can be computed in linear time.

Ingredients

- Existence
  - Each word has a (maybe increasing) factorization in Lyndon words, namely $u = a_1 \cdot a_2 \cdots a_k$ where $a_i \in A$.
  - Fact: If u and v Lyndon and $u < v$ then uv is also Lyndon:

$$abaab = a \cdot b \cdot a \cdot a \cdot b$$
$$abaab = ab \cdot a \cdot a \cdot b$$
$$abaab = ab \cdot a \cdot ab$$
$$abaab = ab \cdot aab$$

- Unicity
  If $u = u_1 \cdots u_n$ is a factorization in Lyndon words such that $u_1 \geqslant u_2 \geqslant \cdots \geqslant u_n$, then
  - u_1 is the longest prefix which is a Lyndon word,
  - u_n is the smallest suffix for the lexicographic ordering.

Transfinite words

Definition

A transfinite word is a sequence $(a_\beta)_{\beta<\alpha}$ of symbols indexed by ordinals less than a given ordinal α called its length.

Definition (alternative for countable ones)

The class of transfinite words is the smallest class of "words"

- containing the symbols a, b, ...
- closed under finite concatenation and ω-concatenation: $(x, y \mapsto xy$ and $x_0, x_1, x_2, \ldots \mapsto x_0 x_1 x_2 \cdots)$

Examples

- finite words like a, $aabab$, ... of length 1 and 5
- infinite words like ab^ω, $(ab)^\omega$ and $aba^2ba^3b\cdots$ of length ω
- $ab^\omega b$ and $ab^\omega(ab)^\omega aba$ of length $\omega+1$ and $\omega \cdot 2 + 3$
- $(ab^\omega)^\omega a$ and $ab^\omega a^2 b^\omega a^3 b^\omega \cdots$ of length ω^2+1 and ω^2.
- and many more like $(ab)^{\omega^\omega}$ of length ω^ω, ...

Transfinite Lyndon words

Definition

A transfinite word is Lyndon iff it is primitive (not equal to u^α for $\alpha \geqslant 2$) and smaller than all its proper suffixes.

Examples

- $aabab$ is Lyndon
- ab^ω is Lyndon: $ab^\omega < b^\omega = b^\omega = \cdots$
- $abab^2ab^3\cdots$ is Lyndon: $abab^2\cdots < ab^2ab^3\cdots <$
- $(ab)^\omega b$ is Lyndon: $(ab)^\omega b < b < (ba)^\omega b$
  This word is equal to its proper suffix $(ab)^{-1}[(ab)^\omega b]$.
- $(ab)^\omega$ is not Lyndon: not primitive
- $aba^2ba^3b\cdots$ is not Lyndon: $a^2ba^3ba^4b\cdots < aba^2ba^3b\cdots$

Factorizations

$$(ab)^\omega = ab \cdot ab \cdot ab \cdots$$
$$aba^2a^3b \cdots = ab \cdot a^2b \cdot a^3b \cdots$$

Problem

$$aba^2a^3b \cdots b = ab \cdot a^2b \cdot a^3b \cdots b \text{ is wrong because } ab < b$$

Factorization theorem

Theorem

Any transfinite word x has a unique locally non-increasing factorization $x = \prod_{\beta<\alpha} u_\beta$ in Lyndon words.

- *Locally non-increasing* a is a relaxation of *non-increasing*.
- It only allows increase at limits where cofinally strict decreases occur before.

Examples

- $aba^2ba^3b\cdots = ab \cdot a^2b \cdot a^3b \cdots$: $ab > a^2b > a^3b \cdots$
- $aba^2ba^3b\cdots b = ab \cdot a^2b \cdot a^3b \cdots b$: $ab > a^2b > \cdots < b$ is locally non-increasing
- $(ab)^\omega = ab \cdot ab \cdots = (ab)^\omega$
- $(ab)^\omega b$ is Lyndon: $ab = ab = \cdots < b$ is not locally non-increasing
- $(ba)^\omega = b \cdot ab \cdot ab \cdots = b \cdot (ab)^\omega$: $ab = ab \cdots < b$
- $(ba)^\omega b = b \cdot (ab)^\omega b$: $b > (ab)^\omega b$

Rational words

Definition (Rational words)

The class of rational words is the smallest class of "words"

- containing the symbols a, b, ...
- closed under finite concatenation and ω-iteration: $(x, y \mapsto xy$ and $x \mapsto xxx\cdots = x^\omega)$

Examples

- finite words like a, $aabab$
- ultimately periodic infinite words like $(bba)^\omega$
- $aba^2ba^3b\cdots$ is not rational
- $(a^\omega b)^\omega a^\omega$

Factorization of rational words

Theorem

For any rational word x, there exists a finite decreasing sequence of rational prime words $u_1 > \cdots > u_n$ and ordinals $\alpha_1, \ldots, \alpha_n$ less than ω^ω such that $x = u_1^{\alpha_1} \cdots u_n^{\alpha_n}$.

Examples

- $$(a^\omega b)^\omega a^\omega = a^\omega b \cdot a^\omega b \cdots a \cdot a \cdot a \cdots = (a^\omega b)^\omega \cdot a^\omega = \|(a^\omega b|)^\omega\|(a|)^\omega\|$$
- $$(bba)^\omega = b \cdot b \cdot abb \cdot abb \cdots = b^2(abb)^\omega = \|b|b\|a(bb|a)^\omega\|$$
- $$(b^\omega a^\omega)^\omega = b \cdot b \cdot b \cdots a^\omega b^\omega \cdot a^\omega b^\omega \cdot a^\omega b^\omega \cdots = b^\omega \cdot (a^\omega b^\omega)^\omega = \|(b|)^\omega\|(a^\omega b^\omega|)^\omega\|$$

Computation of this factorization

Definition (Transformation τ)

$$\tau(a) = a$$
$$\tau(ee') = \tau(e)\tau(e')$$
$$\tau(e^\omega) = \tau(e)\tau(e)^\omega$$

Examples

$$\tau((bba)^\omega) = bba(bba)^\omega$$
$$\tau((a^\omega b)^\omega a^\omega) = aa^\omega b(aa^\omega b)^\omega aa^\omega$$

Theorem

The factorization of a rational word x given by an expression e can described by inserting $\|$ and $|$ in $\tau(e)$ and these insertions can be computed in cubic time in the size of $\tau(e)$.