Nyldon words

Joint work with Émilie Charlier (ULiège)
and Manon Philibert (Université Aix-Marseille)

Manon Stipulanti (Hofstra University) BAEF Fellow

Joint Mathematics Meetings (JMM2020)
Special Session on Sequences, Words and Automata
Denver, Colorado Convention Center (USA)
January 15, 2020

Basics

- A is a *finite alphabet* with a total order $<$.

  Example: $A = \{0, 1\}$ with $0 < 1$

- A^* is the set of all finite words over A.

  Example: $A^* = \{0, 1\}^* = \{\varepsilon, 0, 1, 00, 01, 10, 11, \ldots\}$

- A^+ is the set of all nonempty finite words over A.

  Example: $A^+ = \{0, 1\}^+ = \{0, 1, 00, 01, 10, 11, \ldots\}$

- ε is the *empty word*.
- The *length* of $w \in A^*$ is the number of letters of w, denoted by $|w|$.

  Example: $|101| = 3$

- $u \in A^*$ is a *prefix* of $w \in A^*$ if $\exists v \in A^*$ s.t. $w = uv$.
  $v \in A^*$ is a *suffix* of $w \in A^*$ if $\exists u \in A^*$ s.t. $w = uv$.

  Example: $\underline{100}01010\underline{11}$

- A prefix (resp., suffix) x of w is *proper* if $x \neq w$.

The lexicographic order

The order $<$ on the letters of A induces a total order $<_{\text{lex}}$ on A^*:

Lexicographic order

$$x <_{\text{lex}} y \quad \Leftrightarrow \quad x \in \text{Pref}(y) \setminus \{y\}$$
$$\text{or } \exists p \in A^* \text{ and } a, b \in A \text{ s.t. } \begin{cases} a < b \\ pa \in \text{Pref}(x) \\ pb \in \text{Pref}(y) \end{cases}$$

Notation: $x \leq_{\text{lex}} y$ if $x <_{\text{lex}} y$ or $x = y$.

Example: $0 <_{\text{lex}} 00 <_{\text{lex}} 011$

Lyndon words

Definition: A word x is a *power* if $x = u^k$, for a word u and $k \geq 2$.
A word is *primitive* if it is not a power.

Lyndon word

A finite word w over A is *Lyndon* if it is primitive and lexicographically minimal among its conjugates, i.e., for all $u, v \in A^+$ such that $w = uv$, we have $w <_{\text{lex}} vu$.

Over $\{0, 1\}$:

0, 1	0, 1
00, 01, 10, 11	01
000, 001, 010, 011, 100, 101, 110, 111	001, 011
0000, 0001, 0010, 0011, 0100, 0101, 0110, 0111, 1000, 1001, 1010, 1011, 1100, 1101, 1110, 1111	0001, 0011, 0111

Chen-Fox-Lyndon theorem, 1958

Theorem

Every finite word w over A can be uniquely factorized/written as

$$w = \ell_1 \ell_2 \cdots \ell_k$$

with $k \in \mathbb{N}$ and $(\ell_i)_{1 \leq i \leq k}$ a lexicographically non-increasing sequence of Lyndon words, i.e., $\ell_1 \geq_{\text{lex}} \cdots \geq_{\text{lex}} \ell_k$.

Some Lyndon factorizations:

$$0110101 = (011)(01)(01)$$
$$0100010110 = (01)(0001011)(0)$$
$$000100111001 = (000100111001)$$

Nyldon words

Recursive definition of Lyndon words

- Letters are Lyndon.
- A finite word is Lyndon if and only if it cannot be factorized into a lexicographically non-increasing sequence of shorter Lyndon words.

Question: What happens if we reverse the lexicographic order?

Nyldon words (recursive definition)

- Letters are Nyldon.
- A finite word is Nyldon if and only if it cannot be factorized into a lexicographically non-decreasing sequence of shorter Nyldon words.

Origin of Nyldon words

Nyldon words were defined by Darij Grinberg in a post on Mathoverflow in November 2014, along with three questions:

Q1. For all $n \geq 0$, how many length-n Nyldon words are there?

Q2. Is it true that any finite word can be uniquely factorized as a lexicographically non-decreasing sequence of shorter Nyldon words? (Chen-Fox-Lyndon theorem for Nyldon words)

Q3. Is it true that every primitive word admits exactly one Nyldon word in its conjugacy class?

We already know the answers in the Lyndon case.

First few Nyldon words on $\{0, 1\}$

0, 1	0, 1
00, 01, 10, 11	10
000, 001, 010, 011, 100, 101, 110, 111	100, 101
0000, 0001, 0010, 0011, 0100, 0101, 0110, 0111, 1000, 1001, 1010, 1011, 1100, 1101, 1110, 1111	1000, 1001, 1011
00000, 00001, 00010, 00011, 00100, 00101, 00110, 00111, 01000, 01001, 01010, 01011, 01100, 01101, 01110, 01111, 10000, 10001, 10010, 10011, 10100, 10101, 10110, 10111, 11000, 11001, 11010, 11011, 11100, 11101, 11110, 11111	10000, 10001, 10010, 10011, 10110, 10111

Counting them: $2, 1, 2, 3, 6, 9, 18, 30, 56, 99, 186, 335, 630, 1161, 2182, \ldots$

Lyndon words are lex. minimal among their conjugates.

Nyldon words are **not** lex. maximal among their conjugates.

Example: 101 is Nyldon, and $101 <_{\text{lex}} 110$.

Prefixes of Nyldon words

Apart from 0 and 1, all binary Nyldon words start with the prefix 10.

Other words are forbidden prefixes in the family of Nyldon words.

For instance,

- $00, 01$ (0 is the smallest Nyldon word)
- $11, 1010, 100100, \ldots$ (general form $10^k 10k$)
- $11011, 101011, 1001011, \ldots$ (general form $10^k 1011$)
- $10101, 1011011, 101110111, \ldots$ (general form $101^{k+1} 01^{k+1}$)
- $10011011, 1000110011, \ldots$ (general form $10^{k+2} 110^{k+1} 11$)
- $\ldots$

We found a family of forbidden prefixes:

$$F = \{p \in A^* : p = p_1 p_2 p_3,\ p_1 \text{ Nyldon },\ p_1 \leq_{\text{lex}} p_2, \\ \forall x \in A^*, p_2 p_3 x = n_1 \cdots n_k \Rightarrow |n_1| \geq |p_2|\}.$$

Question: Are there other forbidden prefixes?

Suffixes of Nyldon words

It is difficult to understand more the prefixes of Nyldon words.
What about their suffixes then?
Suffixes of Lyndon and Nyldon words share some properties.

Theorem (Lyndon)

Let $w \in A^*$. Then the following assertions are equivalent:

- w is Lyndon
- w is strictly smaller than all its proper suffixes
- w is strictly smaller than all its proper Lyndon suffixes.

Theorem (Nyldon)

If a word w is Nyldon, then w is strictly greater than all its proper Nyldon suffixes.

The converse is not true.

Example: 110 is greater than all its proper (Nyldon) suffixes (0, 10), but $110 = (1)(10)$ is not Nyldon.

The longest Nyldon suffix of a word

Theorem (Lyndon)

Let $w = \ell_1 \ell_2 \cdots \ell_k$ be the Lyndon factorization of w. Then

- ℓ_k is the longest Lyndon suffix of w
- ℓ_1 is the longest Lyndon prefix of w.

Theorem (Nyldon)

Let $w = n_1 n_2 \cdots n_k$ be **a** Nyldon factorization of w.
Then n_k is the longest Nyldon suffix of w.

There is no similar condition on prefixes in the Nyldon case.

Example: A Nyldon factorization of 10100 is $10100 = (10)(100)$ but its longest Nyldon prefix is 101.

Unicity of the Nyldon factorization

Using the results on suffixes, we proved:

Theorem (Charlier, Philibert, S., 2019)

Every finite word w over A can be uniquely factorized/written as

$$w = n_1 n_2 \cdots n_k$$

with $k \in \mathbb{N}$ and $(n_i)_{1 \leq i \leq k}$ a lexicographically non-decreasing sequence of Nyldon words, i.e., $n_1 \leq_{\text{lex}} \cdots \leq_{\text{lex}} n_k$.

Lyndon factorizations VS Nyldon factorizations:

$$(011)(01)(01) = 0110101 = (0)(1)(10)(101)$$
$$(01)(0001011)(0) = 0100010110 = (0)(1000)(10110)$$
$$(000100111001) = 000100111001 = (0)(0)(0)(100111001)$$

Remark: Answer to Q2 (Darij Grinberg's questions).

The sequence A001037 in OEIS

Corollary

There are equally many Lyndon and Nyldon words of each length.

In Sloane's On-Line Encyclopedia of Integer Sequences (OEIS), the sequence A001037

$$2, 1, 2, 3, 6, 9, 18, 30, 56, 99, 186, 335, 630, 1161, 2182, 4080, \ldots$$

gives the number of (binary) Lyndon words of each length.

Remark: Answer to Q1 (Darij Grinberg's questions).

Complete factorizations of the free monoid

A totally-ordered family $F \subseteq A^+$ is a *complete factorization* of the free monoid A^* if each $w \in A^*$ can be uniquely factorized/written as

$$w = x_1 x_2 \cdots x_k \tag{1}$$

with $k \in \mathbb{N}$ and $(x_i)_{1 \leq i \leq k}$ a non-increasing sequence of F.

Example:

- Lyndon words are a complete factorization for the order $<_{\text{lex}}$. (Chen-Fox-Lyndon, 1958)
- Nyldon words are a complete factorization for the order $>_{\text{lex}}$. (Charlier, Philibert, S., 2019)

A powerful tool for complete factorizations

Theorem (Schützenberger, 1965)

Let $F \subseteq A^+$ with a total order $\prec$.
Then any two of the following three conditions imply the third:

- Each $w \in A^*$ admits at least one factorization (1).
- Each $w \in A^*$ admits at most one factorization (1).
- All elements of F are primitive and each primitive conjugacy class of A^+ contains exactly one element of F.

Representatives of primitive conjugacy classes

As a consequence of the unicity of the Nyldon factorization and Schützenberger's theorem, we have:

Corollary

All Nyldon words are primitive and every primitive word admits exactly one Nyldon word in its conjugacy class.

Remark: Answer to Q3 (Darij Grinberg's questions).

Length-preserving bijection:

Nyldon → Lyndon: $w \mapsto w'$
where
w' = unique Lyndon conjugate

Easy: w' lex. min. conjugate

Lyndon → Nyldon: $w' \mapsto w$
where
w = unique Nyldon conjugate

Less easy: Mélançon algorithm

Lyndon and Nyldon words: a short comparison

Many results for Lyndon words do not have analogues for Nyldon words.

- Suffixes of Lyndon/Nyldon:

w Lyndon $\Leftrightarrow$ w is strictly smaller than all its proper (Lyndon) suffixes.

w Nyldon $\Rightarrow$ w is strictly greater than all its proper Nyldon suffixes.

The converse is not true.

Example: 110 is greater than all its proper (Nyldon) suffixes (0, 10), but $110 = (1)(10)$ is not Nyldon

- Concatenation of Lyndon/Nyldon:

u, v Lyndon with $u <_{\text{lex}} v \Rightarrow uv$ Lyndon

u, v Nyldon with $u >_{\text{lex}} v \not\Rightarrow uv$ Nyldon

Example: $10010, 1$ Nyldon with $10010 >_{\text{lex}} 1$
but $100101 = (100)(101)$ is not Nyldon

- Standard factorization:

If $w = uv$ where v = longest Lyndon proper suffix of w (standard fact.), then w Lyndon $\Leftrightarrow$ u Lyndon and $u <_{\text{lex}} v$.
Moreover v is the lex. smallest nonempty proper suffix of w.

If $w = uv$ where v = longest Nyldon proper suffix of w (standard fact.), then w Nyldon $\Leftrightarrow$ u Nyldon and $u >_{\text{lex}} v$.
But v is not necessarily the lex. greatest nonempty proper suffix of w.

Example: $w = 100$ Nyldon with $(u, v) = (10, 0)$
but the lex. greatest proper suffix of w is 00

- Širšov factorization:

If $w = uv$ where u = longest Lyndon proper prefix of w (Širšov fact.), then w Lyndon $\Leftrightarrow$ v Lyndon and $u <_{\text{lex}} v$.

No such factorization for Nyldon.

Example: $w = 10010100100$ Nyldon with $(u, v) = (100101001, 00)$
but $v = 00 = (0)(0)$ is not Nyldon

- Comma-free codes:

$F \subseteq A^*$ *code*: $x_1 \cdots x_m = y_1 \cdots y_n$ with $x_1, \ldots, x_m, y_1, \ldots y_n \in F$
$\Leftrightarrow m = n$ and $x_i = y_i$

$F \subseteq A^*$ *comma-free* code: $w \in F^+$, $u, v \in A^*$, $uwv \in F^* \Rightarrow u, v \in F^*$

The length-n Lyndon words over A form a comma-free code
$\Leftrightarrow$ $n = 1$, or $n = 2$ and $|A| \in \{2, 3\}$, or $n \in \{3, 4\}$ and $|A| = 2$.

The length-n Nyldon words over A form a comma-free code
$\Leftrightarrow$ $n = 1$, or $n = 2$ and $|A| \in \{2, 3\}$, or $n \in \{3, 4, 5, 6\}$ and $|A| = 2$.

- Lazard factorization:

Theorem (Viennot, 1978)

A complete factorization $(F, \prec)$ of A^* is left (resp., right) Lazard
$\Leftrightarrow \forall x, y \in F$, $xy \in F \Rightarrow x \prec xy$ (resp., $xy \prec y$).

Lyndon words form a left and right Lazard factorization for $<_{\text{lex}}$.

Nyldon words form a right Lazard factorization for $>_{\text{lex}}$, but not a left Lazard factorization for $>_{\text{lex}}$.

Conclusion

Answers to Darij Grinberg's questions:

Q1. For all $n \geq 0$, there are equally many length-n Lyndon and Nyldon words.

Q2. Every finite word can be uniquely factorized as a lexicographically non-decreasing sequence of shorter Nyldon words. (Chen-Fox-Lyndon theorem for Nyldon words)

Q3. All Nyldon words are primitive and every primitive word admits exactly one Nyldon word in its conjugacy class.

Lyndon words have many (strong) properties.
Some of them have analogues in terms of Nyldon words, while some of them (many in fact) do not.

References

- L. A. Bokut, Y. Chen, and Y. Li. Lyndon-Shirshov basis and anti-commutative algebras. J. Algebra, 378:173–183, 2013.
- J. Berstel, D. Perrin, and C. Reutenauer. Codes and automata, volume 129 of Encyclopedia of Mathematics and its Applications. Cambridge University Press, Cambridge, 2010.
- É. Charlier, M. Philibert, and M. Stipulanti. Nyldon words. J. Combin. Theory. Ser. A., 167:60–90, 2019.
- K.-T. Chen, R. H. Fox, and R. C. Lyndon. Free differential calculus. IV. The quotient groups of the lower central series. Ann. of Math. (2), 68:81–95, 1958.
- D. Grinberg. "Nyldon words": understanding a class of words factorizing the free monoid increasingly. https://mathoverflow.net/questions/187451/, 2014.
- G. Mélançon. Combinatorics of Hall trees and Hall words. J. Combin. Theory Ser. A, 59(2):285–308, 1992.
- D. Perrin and C. Reutenauer. Hall sets, Lazard sets and comma-free codes. Discrete Math., 341(1):232–243, 2018.
- C. Reutenauer. Free Lie algebras, volume 7 of London Mathematical Society Monographs. New Series. The Clarendon Press, Oxford University Press, New York, 1993. Oxford Science Publications.
- M.-P. Schützenberger. On a factorisation of free monoids. Proc. Amer. Math. Soc., 16:21–24, 1965.
- A. I. Širšov. On free Lie rings. Mat. Sb. N.S., 45(87):113–122, 1958.
- A. I. Širšov. On the bases of a free Lie algebra. Algebra i Logika Sem., 1(1):14–19, 1962.
- G. Viennot. Algèbres de Lie libres et monoïdes libres, volume 691 of Lecture Notes in Mathematics. Springer, Berlin, 1978.