THEORIE-TAG 2023 A. Lin, G. Zetzsche (Eds.): Theorietag 2023, Kaiserslautern, 4. – 6.10.2023
RPTU & MPI-SWS Kaiserslautern, Technical Report, pp. 25–28.

Infinite Nyldon words

Pamela Fleischmann[(A)] Annika Huch[(A)] Dirk Nowotka[(A)]

[(A)]Department of Computer Science, Kiel University, Germany
{fpa,dn}@informatik.uni-kiel.de,stu216885@mail.uni-kiel.de

Abstract

In this work we investigate infinite Nyldon words. Those are defined by reversing the lexicographical order in the infinite Lyndon factorisation by Siromoney et al. They showed that each infinite word can be uniquely, lexicographically non-increasingly factorised into either a finite sequence of (finite) Lyndon words followed by one infinite Lyndon word or an infinite sequence of (finite) Lyndon words. Here, we can observe several similarities to the behaviour of finite Nyldon words (in detail examined by Charlier et al.). We show that each infinite word, has a unique infinite Nyldon factorisation. Further, we state structural results on infinite Nyldon words as a characterisation of their suffixes and a standard factorisation.

1. Introduction

A *Lyndon word* is defined as a non-repetitive word which is the smallest amongst its cyclic rotations. This class of words is strongly studied in, e.g., [2, 3, 7, 5]. One major result by [2], today known as the Chen-Fox-Lyndon theorem shows that Lyndon words factorise the free monoid, i.e., for each word in the free monoid there exists a unique, lexicographically non-increasing factorisation into Lyndon words. In a Mathoverflow post from November 2014 Grinberg raised the question how the factorisation of the free monoid changes when investigating the above factorisation from Chen, Fox and Lyndon w.r.t. a reversed lexicographical order, i.e., a lexicographically non-decreasing factorisation where each factor is smaller or equal than its successor. Hence, the notion of *Nyldon words* developed as the set of those words which are either letters or cannot be factorised into at least two, lexicographically non decreasing Nyldon factors. For example, the word $\mathtt{10}$ is Nyldon since $\mathtt{1}$ and $\mathtt{0}$ are Nyldon words and thus, the only factorisation of $\mathtt{10}$ is $\mathtt{1} \cdot \mathtt{0}$ which is lexicographically decreasing. Further, $\mathtt{110}$ is not Nyldon since $\mathtt{1} \cdot \mathtt{10}$ is its non-decreasing factorisation into Nyldon words. In [1] they investigated Nyldon words and showed that they also form a unique factorisation of the free monoid. Further, they give a standard factorisation of Nyldon words (similar to the well-known standard factorisation for Lyndon words). It turns out that Nyldon words are harder to grasp because they miss the property of being the smallest word under its cyclic rotations (neither they are the largest). Thus, the properties of Lyndon words cannot be immediately transferred to Nyldon words which strengthens the relevance of their investigation. An example are the *infinite Nyldon words*, i.e., infinite words that cannot be lexicographically non-decreasingly factorised into (1) a finite sequence of (finite) Nyldon words followed by an infinite Nyldon word, or (2) an infinite sequence of (finite) Nylon words (cf. [6] for results on infinite Lyndon words). We present several properties to deeper understand and characterise these infinite words. First, we show that this infinite Nyldon factorisation is unique for each infinite Nyldon word. Further, each infinite Nyldon word is lexicographically larger than its infinite Nydon suffixes. Using these results we present a standard factorisation for infinite Nyldon words.

2. Preliminaries

Let $\mathbb{N} = \{1,2,\ldots\}$, define $[m] = \{1,\ldots,m\}$. For the standard definitions of combinatorics on words, especially for the whole background of Lyndon words, we refer to [4, Chapter 5]. Denote by $\Sigma^\omega := \{\mathtt{a}_1\mathtt{a}_2\cdots \mid \mathtt{a}_i \in \Sigma, 1 \leq i\}$ the set of all right infinite concatenations over Σ. Further, a tuple $f = (w_1,\ldots,w_k) \in (\Sigma^*)^k$ is called a *factorisation* of the (finite) word $w \in \Sigma^*$ if $w = w_1\cdots w_k$. For a *factorisation* of an infinite word $w \in \Sigma^\omega$ there are two options: (1) a factorisation $g = (w_1,w_2,\ldots)$ with $w_i \in \Sigma^*$ for $i \in \mathbb{N}$ such that $w = w_1w_2\cdots$, or (2) a factorisation $h = (w_1,\ldots,w_{n-1},w_n)$ with $w_i \in \Sigma^*$ for $i \in [n-1]$ and $w_n \in \Sigma^\omega$ such that $w = w_1\cdots w_{n-1}w_n$. Let $\triangleleft$ be a total order on Σ. We extend this order to Σ^* by $u \triangleleft v$ for $u,v \in \Sigma^*$ iff u is a prefix of v or $u = x\mathtt{a}u'$ and $v = x\mathtt{b}v'$ with $\mathtt{a} \triangleleft \mathtt{b}$ for $\mathtt{a},\mathtt{b} \in \Sigma$ and some $u',v',x \in \Sigma^*$. This extended order is called *lexicographical order* on words over Σ and forms a total order on Σ^*. This lexicographical order can be further extended to Σ^ω. Note that the first condition, u is a prefix of v, is only applicable if $u \in \Sigma^*$ is finite. For the second condition both, u and v may belong to Σ^* or Σ^ω. A word $w \in \Sigma^*$ is a Lyndon word iff it is primitive and the lexicographically smallest in its conjugacy class. We denote the set of Lyndon words by $\mathcal{L}$. It is well known that Lyndon words factorise the free monoid: each $w \in \Sigma^*$ has a unique *Lyndon factorisation* $(\ell_1,\ldots,\ell_k)$ with $\ell_j \in \mathcal{L}$ for $j \in [k]$ and $\ell_i \trianglerighteq \ell_{i+1}$ for $i \in [k-1]$ (Chen-Fox-Lyndon Theorem). Thus, Lyndon words can be defined as those words that do not have any non-decreasing factorisation into at least two Lyndon words, i.e., they cannot be further factorised into a Lyndon factorisation with at least two Lyndon factors. In [1], the authors introduce Nyldon words. A word $w \in \Sigma^+$ is called *Nyldon* ($w \in \mathcal{N}$) if $w \in \Sigma$ or there does not exist any factorisation $(n_1,\ldots n_k)$ of w with $k \geq 2$, $n_j \in \mathcal{N}$ for $j \in [k]$ and $n_i \trianglelefteq n_{i+1}$ for $i \in [k-1]$. We now continue by defining infinite Lyndon words [6]. An infinite word $w \in \Sigma^\omega$ is *ω-Lyndon* if it has an infinite number of Lyndon prefixes. Denote the set of all infinite Lyndon words by $\mathcal{L}_\omega$. Similar to finite Lyndon words, each infinite word has a unique infinite Lyndon factorisation [6], i.e., any word $w \in \Sigma^\omega$ has a unique factorisation of the form: (1) $(\ell_1\ell_2,\ldots,\ell_k)$ where $\ell_i \in \mathcal{L}$ for $i \in [k-1]$ and $\ell_k \in \mathcal{L}_\omega$ with $\ell_i \trianglerighteq \ell_{i+1}$, or (2) $(\ell_1,\ell_2,\ldots)$ where $\ell_i \in \mathcal{L}$ and $\ell_i \trianglerighteq \ell_{i+1}$ for $i \in \mathbb{N}$. We define infinite Nyldon words similar to the finite case.

Definition 2.1 *An infinite word $w \in \Sigma^\omega$ is ω-*Nyldon *if it cannot be factorised into one of the following factorisations:*
1. $(n_1,n_2,\ldots,n_k)$ where $n_i \in \mathcal{N}$ for $i \in [k-1]$, $n_k \in \mathcal{N}_\omega$ with $n_i \trianglelefteq n_{i+1}$ and $k \geq 2$, or
2. $(n_1,n_2,\ldots)$ where $n_i \in \mathcal{N}$ and $n_i \trianglelefteq n_{i+1}$ for $i \in \mathbb{N}$.

The set of ω-Nyldon words will be denoted by $\mathcal{N}_\omega$. As an extension to the factorisations of finite words, any factorisation $(n_1,\ldots,n_k)$ or $(n'_1,n'_2,\ldots)$ of $w \in \mathcal{N}_\omega$ into Nyldon words such that $n_1 \trianglelefteq \cdots \trianglelefteq n_k$, or $n'_1 \trianglelefteq n'_2 \trianglelefteq \ldots$, respectively, is called a Nyldon factorisation *of w.*

An infinite word w is ω-Nyldon iff its infinite Nyldon factorisation is of length 1. Further, an infinite word is not ω-Nyldon, iff its infinite Nyldon factorisation is of length at least 2.

Example 2.2 *To get an intuition, we give examples for Nyldon and non-Nyldon words together with a justification regarding the Nyldon factorisation. First, consider $w = \mathtt{10}^\omega \in \mathcal{N}_\omega$. It has no factorisation of the first form of Definition 2.1 since $\mathtt{0}^\omega \notin \mathcal{N}_\omega$ ($\mathtt{0}$ is the only finite Nyldon word starting with zero, i.e., the only possible infinite Nyldon factorisation of $\mathtt{0}^\omega$ is $(\mathtt{0},\mathtt{0},\ldots)$). Further there exists no factorisation of the second form because $\mathtt{1} \ntrianglerighteq \mathtt{0}$ and $\mathtt{10}^n \ntrianglerighteq \mathtt{0}$. Thus, each possible factorisation is not increasing. As another example, one can verify that $\mathtt{101}^\omega \in \mathcal{N}_\omega$.*

Moreover, $\mathtt{010}^\omega \notin \mathcal{N}_\omega$ has a Nyldon factorisation of the first form of Definition 2.1 since $\mathtt{0} \trianglelefteq \mathtt{10}^\omega$. Further, consider $(\mathtt{10})^\omega \notin \mathcal{N}_\omega$. We know that $\mathtt{10} \in \mathcal{N}$ so $\mathtt{10}\cdots\mathtt{10}$ is an infinite Nyldon factorisation of the second form of Definition 2.1.

3. Infinite Nyldon words

In 1994, Siromoney et al. [6] have shown that several known results for Lyndon words also apply to infinite Lyndon words, e.g., the standard factorisation, and a unique infinite Lyndon factorisation. The aim of this chapter is to investigate infinite Nyldon words.

First, we want to show the uniqueness of the Nyldon factorisation for infinite words. With a similar proof this result was shown for infinite Lyndon words in [6].

Theorem 3.1 *Any infinite word $w \in \Sigma^\omega$ has a unique factorisation of the form*
1. $(n_1, n_2, \ldots, n_k)$ where $n_i \in \mathcal{N}$ for $i \in [k-1]$, $n_k \in \mathcal{N}_\omega$ and $n_i \trianglelefteq n_{i+1}$, or
2. $(n_1, n_2, \ldots)$ where $n_i \in \mathcal{N}$ and $n_i \trianglelefteq n_{i+1}$, $i \in \mathbb{N}$.

In [1] is is shown that the last factor of a words Nyldon factorisation is always the longest proper Nyldon suffix of this word. Since the notion of longest Nyldon suffixes is not applicable in the case of infinite words we adapt this result. The last factor of the Nyldon factorisation might be an infinite word. So we will show that there exists no shorter prefix such that this last factor is an infinite Nyldon word. The proof works similar to the finite case.

Proposition 3.2 *Let $w \in \Sigma^\omega$ such that its Nyldon factorisation $(n_1, \ldots, n_k)$ is finite with $n_i \in \mathcal{N}$ for $i \in [k-1]$, $n_k \in \mathcal{N}_\omega$ and $n_i \trianglelefteq n_{i+1}$ (Case 1 of Definition 2.1). Then $n_1 \cdots n_{k-1}$ is the shortest prefix of w such that $n_k \in \mathcal{N}_\omega$.*

Remark 3.3 *Note that the first factor of the infinite Nyldon factorisation is in both cases of Theorem 3.1 not necessarily the longest Nyldon prefix. For example, let $w = \mathtt{10100}^\omega \in \Sigma^\omega$. Its infinite Nyldon factorisation is $(\mathtt{10}, \mathtt{100}^\omega)$ but $\mathtt{101}$ is its longest Nyldon prefix. Second, consider $w' = (\mathtt{10})^\omega$ that decomposes into $(\mathtt{10}, \mathtt{10}, \ldots)$ although $\mathtt{101}$ is its longest Nyldon prefix.*

One useful property of infinite Nyldon words would be if, similar to finite Nyldon words, all their infinite Nyldon suffixes are smaller than the words themselves. To prove this, we need the following result. Note that the this result is not applicable to finite Nyldon words.

Lemma 3.4 *Let $w = ps \in \mathcal{N}_\omega$. If $s \in \mathcal{N}_\omega$ then $p \in \mathcal{N}$.*

Example 3.5 *Consider $w = \mathtt{101100}^\omega \in \mathcal{N}_\omega$. Now, $\mathtt{100}^\omega \in \mathcal{N}_\omega$ and $\mathtt{101} \in \mathcal{N}$ (Proposition 3.4).*

Theorem 3.6 *Let $w \in \mathcal{N}_\omega$. Then for all infinite Nyldon suffixes $s \in \mathcal{N}_\omega$ of w, $s \triangleleft w$ holds.*

Remark 3.7 *Note, that a standard factorisation for infinite Nyldon words cannot work similar to the standard factorisation of finite Nyldon words. The problem is that the standard factorisation of finite Nyldon words relies on Nyldon suffixes. For example, the infinite Nyldon words $\mathtt{10}^\omega$ and $\mathtt{101}^\omega$ both do not have any infinite Nyldon suffix (neither $\mathtt{0}^\omega$, $\mathtt{1}^\omega$, nor $\mathtt{01}^\omega$ are Nyldon).*

Lemma 3.8 *Let $w \in \Sigma^*$ with an infinite Nyldon factorisation $(n_1, n_2, \ldots)$. Then there exists no proper infinite Nyldon suffix $s <_s w$.*

This allows us to introduce an adapted version of the standard factorisation to the finite case.

Theorem 3.9 *Let $p \in \mathcal{N}$ and $s \in \mathcal{N}_\omega$. We have $w \in \mathcal{N}_\omega$ iff one of the following holds:*
1. There exists no proper infinite Nyldon suffix s of w, or
2. if p is the shortest proper Nyldon prefix of ps such that $s \in \mathcal{N}_\omega$ then $p \triangleright s$.

4. Conclusion

In this work we introduced infinite Nyldon words and started their investigation. We can observe that infinite Nyldon words resemble the finite Nyldon words in many aspects, i.e., the uniqueness of the infinite Nyldon factorisation and the fact that each Nyldon word is smaller than all its Nyldon suffixes. When investigating a standard factorisation, we need to add the adaption that an infinite Nyldon word might not have any infinite Nyldon suffix. For further research it would be interesting to determine the infinite Nyldon factorisations of several famous infinite words like the Thue-Morse word, the Fibonacci word and Sturmian words in general.

References

[1] É. Charlier, M. Philibert, M. Stipulanti, Nyldon words. *J. Comb. Theory, Ser. A* **167** (2019), 60–90.

[2] K. T. Chen, R. H. Fox, R. C. Lyndon, Free differential calculus, IV. The quotient groups of the lower central series. *Annals of Mathematics* (1958), 81–95.

[3] J. Duval, Mots de Lyndon et Périodicité. *RAIRO Theor. Informatics Appl.* **14** (1980) 2, 181–191.

[4] M. Lothaire, *Combinatorics on Words*. Cambridge Mathematical Library, Cambridge University Press, 1997.

[5] R. Siromoney, L. Mathew, A Public Key Cryptosystem Based on Lyndon Words. *Inf. Process. Lett.* **35** (1990) 1, 33–36.

[6] R. Siromoney, L. Mathew, V. R. Dare, K. G. Subramanian, Infinite Lyndon Words. *Inf. Process. Lett.* **50** (1994) 2, 101–104.

[7] K. G. Subramanian, R. Siromoney, L. Mathew, Lyndon Trees. *Theor. Comput. Sci.* **106** (1992) 2, 373–383.