Duval's Conjecture and Lyndon Words

Tero Harju
Dirk Nowotka

Turku Centre for Computer Science

TUCS Technical Reports

No 479, October 2002

Duval's Conjecture and Lyndon Words

Tero Harju
Dirk Nowotka
Turku Centre for Computer Science, TUCS,
Department of Mathematics, University of Turku

Turku Centre for Computer Science
TUCS Technical Report No 479
October 2002
ISBN 952-12-1061-3
ISSN 1239-1891

Abstract

Two words w and w' are conjugates if $w = xy$ and $w' = yx$ for some words x and y. A word $w = u^k$ is primitive if $k = 1$ for any suitable u. A primitive word w is a Lyndon word if w is minimal among all its conjugates with respect to some lexicographic order. A word w is bordered if there is a nonempty word u such that $w = uvu$ for some word v. A Duval extension of an unbordered word w of length n is a word wu where all factors longer than n are bordered. A Duval extension wu of w is called trivial if there exists a positive integer k such that $w^k = uv$ for some word v.

We prove that Lyndon words have only trivial Duval extensions. Moreover, we show that every unbordered Sturmian word is a Lyndon word which extends a result by Mignosi and Zamboni. We give a conjecture which implies a sharpened version of Duval's conjecture, namely, that for any word w of length n any Duval extension longer or equal than $2n - 1$ is trivial. Our conjecture characterizes a property of every word w which has a nontrivial Duval extension of length $2|w| - 2$.

Keywords: combinatorics on words, Duval's conjecture, Lyndon words, Sturmian words

TUCS Laboratory
Discrete Mathematics for Information Technology

1 Introduction

The relationship between the period of a finite word and the maximum length of its unbordered factors is a field of research that was initiated in the late 70's and beginning of the 80's [1, 3, 2]. This line of research culminated in Duval's conjecture [2]. A Duval extension of an unbordered word w of length n is a word wu where all factors longer than n are bordered. We call a Duval extension wu of w trivial if the length of w is the period of wu. Duval's conjecture states that for any unbordered word w of length n any Duval extension longer or equal than $2n$ is trivial. That conjecture has remained unsolved until today. Recently however, Duval's conjecture was proved for the special case of Sturmian words [4].

We show in Section 3 that Lyndon words have only trivial Duval extensions and that every unbordered Sturmian word is a Lyndon word which extends Mignosi and Zamboni's result in [4]. In Section 4 we give a conjecture describing the shape of any word w which has a nontrivial Duval extension of length $2|w|-2$, and we show that this conjecture implies a widely believed sharpened version of Duval's conjecture, namely, that any Duval extension of length $2|w|-1$ is trivial.

2 Preliminaries

Let A be a finite nonempty alphabet

Let w be an infinite word such that it contains exactly $n+1$ factors of length n for all $n \geq 0$. Then w is called a *Sturmian word*. Note, that Sturmian words are allways over a binary alphabet. A finite factor of a Sturmian word is also called Sturmian word.

We only consider finite words in the following. Let A^* denote the monoid of all finite words in A. Let $\lhd_A$ be an ordering of $A = \{a_1, a_2, \ldots, a_n\}$, say $a_1 \lhd_A a_2 \lhd_A \cdots \lhd_A a_n$. Then $\lhd_A$ induces a *lexicographic order* on A^* such that

$$u \lhd_A v \iff u \leq v \quad \text{or} \quad u = xau' \text{ and } v = xbu' \text{ with } a \lhd_A b$$

where $a, b \in A$. We write $\lhd$ for $\lhd_A$, for some alphabet A, if the context is clear.

A nonempty word u is called a *border* of a word w, if $w = uv = v'u$ for some suitable words v and v'. We call w *bordered* if it has a border that is shorter than w, otherwise w is called *unbordered*. Note, that every bordered word w has a minimum border u such that $w = uvu$ and u is unbordered. A word w is called *primitive* if it cannot be factored such that $w = u^k$

for some $k \geq 2$. Let $w = w_{(1)}w_{(2)} \cdots w_{(n)}$ where $w_{(i)}$ is a letter, for every $1 \leq i \leq n$. Then we denote the length n of w by $|w|$. An integer $1 \leq p \leq n$ is a *period* of w, if $w_{(i)} = w_{(i+p)}$ for all $1 \leq i \leq n - p$. The smallest period of w is called the *minimum period* of w. Let $w = uv$. Then u is called a *prefix* of w, denoted by $u \leq w$, and v is called a *suffix* of w, denoted by $v \preccurlyeq w$.

Let w be a nonempty, unbordered word of length n. We call wu a *Duval extension* of w, if every factor of wu longer than n is bordered. Since a Duval extension is only defined for an unbordered word, we assume all words, we take Duval extensions of in the following, to be unbordered. A Duval extension wu of w is called *trivial*, if there exists a positive integer k such that $u \leq w^k$, that is, the minimum period of wu is n. Certainly, if wu is a Duval extension of w, then wu' is a Duval extension of w, for all $u' \leq u$.

We are concerned with nontrivial Duval extensions. The following lemma reduces our focus to Duval extensions of length less than or equal to $2n$.

Lemma 1. *If an unbordered word w of length n has a nontrivial Duval extension wv such that $|v| > |w|$, then it has a nontrivial Duval extension wu such that $|u| \leq |w|$.*

Proof. Take the maximum $k \geq 0$ such that $v = w^k w'$. Let w_0 be the maximum common prefix of w and w'. So, $w' = w_0 v'$. Clearly, v' is not empty, since wv is a nontrivial Duval extension. Now, any word u such that $u \leq w_0 v'$ and $|w_0| < |u| \leq |w|$ is a nontrivial Duval extension of w. □

Consider $w = abaabb$ and $u = aaba$ as an example for a nontrivial Duval extension of w

$$wu = abaabbaaba\ .$$

Now, every factor of wu of length 7 or more is bordered.

3 Duval Extensions of Lyndon Words

The main result of this paper concerns Lyndon words. A word w is called a *Lyndon word* if it is primitive and minimal among all its conjugates with respect to some lexicographic order. For example, consider $w = abaabb$. Then $aabbab$ and $bbabaa$ conjugates of w and minimal with respect to the order $a \lhd b$ and $b \lhd a$, respectively.

Let wu be a word with k many different letters. Surely, there are at least k many Lyndon words among all conjugates of wu since there is a Lyndon word beginning with a for each letter a. Note, that wuw contains all conjugates of wu except at most $|u| - 1$ many of them. We have that wuw contains at least one Lyndon word which is a conjugate of wu, if $|u| \leq k$.

It is clear that any prefix of a Lyndon word w is lexicographically smaller or equal to any other factor of w of the same length, and that Lyndon words are unbordered.

Theorem 2. *Lyndon words only have trivial Duval extensions.*

Proof. Let $w \in A^*$ be a Lyndon word with respect to an order $\lhd$. Certainly, w is unbordered since it is a Lyndon word. Assume contrary to the claim that there exists a nonempty word u such that wu is a nontrivial Duval extension of w. Let u be of minimum length such that $u \not\leq w$. So, either $u = va$ and $vb \leq w$ or $u = vb$ and $va \leq w$ for some $a, b \in A$ with $a \neq b$ and $a \lhd b$. Then $|u| \leq |w|$ by Lemma 1.

If $v = \varepsilon$ then $u = b$ since the first letter of w is minimal with respect to $\lhd$. Let the minimum border of wb be ayb, we have then that w is bordered with ay; a contradiction. Therefore, $v \neq \varepsilon$ in the following.

Case 1: Suppose $u = va$. Then $w = vbz$. We have that va is not a factor of w since va is lexicographically smaller than vb. Therefore, the minimum border of any factor $w'va$ of wu, where $w' \preccurlyeq w$ and $|w'va| > |w|$, is smaller than $|va|$. Moreover, we have that b occurs in v otherwise $v = a^k$ for some $k \geq 1$, and we have that $bzva$ is longer than $|w|$ and has a border that ends in bxa^{k+1} and $va = a^{k+1}$ occurs in w; a contradiction. Let $v = v'ca^k$, with $c \in A$ and $c \neq a$ and $k \geq 0$, and let $U = ca^k$ for the sake of a simplified notation.

The suffix $s_1 = Ubzu$ of $wu = vbzu = v'Ubzv'Ua$ is of length greater than $|w|$ and therefore has a minimum border $x_1a = Ubv_1Ua$. We have now that $v = w_1x_1 = w_1Ubv_1U$ and $s_1 = Ubzu = x_1az' = Ubv_1Uaz'$, and hence

$$wu = vbzu = w_1Ubv_1Ubzu = w_1Ubv_1s_1 = w_1Ubv_1Ubv_1Uaz' \, .$$

Note, that $|v| > |x_1|$ since va does not occur in w. Let $wu = w_1s_2$. Then $s_2 = x_1bzu = Ubv_1Ubzu$ is a suffix of wu with $|s_2| > |s_1| > |w|$, and hence, it has a minimum border x_2a. We have $|v| > |x_2|$, since va does not occur in w, and also $|x_2| > |x_1|$, otherwise x_1a is bordered and therefore not the minimum border of s_1. Inductively, we obtain an infinite sequence $x_1, x_2, \ldots$ of border words for suffixes $s_1, s_2, \ldots$ of wu such that $|x_1| < |x_2| < \cdots$ and $|s_1| < |s_2| < \cdots$ and we have a contradiction since w is finite.

Case 2: Suppose $u = vb$. Then $w = vaz$. By assumption, wu has a border word xb. Clearly, $x \neq \varepsilon$ and $|xb| \leq |u|$, otherwise w is bordered. So, $xb \leq w$ and $x \preccurlyeq v$ and xa is a factor but not a prefix of w. But, xa is lexicographically smaller than the prefix xb, and hence, w is not a Lyndon word; a contradiction. □

Mignosi and Zamboni proved in [4] that unbordered Sturmian words, that is unbordered, finite factors of Sturmian words, only have trivial Duval extensions. Proposition 4 below shows that Theorem 2 extends that result since every unbordered Sturmian word is a Lyndon word.

Let $\tau\colon A^* \to B^*$ be a morphism, and $\lhd_A$ and $\lhd_B$ be orders on A and B, respectively, such that

$$a_1 \lhd_A a_2 \implies \tau(a_1) \lhd_B \tau(a_2) \tag{1}$$

for every $a_1, a_2 \in A$, and $\tau(a)$ is a Lyndon word w.r.t. $\lhd_B$ for every $a \in A$.

Lemma 3. *If $w \in A^*$ is a Lyndon word, then $\tau(w)$ is a Lyndon word.*

Proof. Let $|w| = n$. Assume $\tau(w)$ is not a Lyndon word. So, $\tau(w) = xy$ such that yx is minimal w.r.t. $\lhd_B$, and x and y are not empty.

If $x = \tau\big(w_{(1)}w_{(2)}\cdots w_{(i)}\big)$ and $y = \tau\big(w_{(i+1)}w_{(i+2)}\cdots w_{(n)}\big)$ with $1 \leq i < n$, then we have an immediate contradiction by (1).

So, there exists an i, where $1 \leq i \leq n$, and $\tau(w_{(i)}) = v_1v_2$ such that $x = \tau\big(w_{(1)}w_{(2)}\cdots w_{(i-1)}\big)v_1$ and $y = v_2\tau\big(w_{(i+1)}w_{(i+2)}\cdots w_{(n)}\big)$ and $v_1, v_2 \neq \varnothing$. That implies $v_2 \lhd_B v_1v_2$, and we have $v_1 = u^j$ and $v_2 = u^k$, for some primitive u and $j, k \geq 1$, since v_1v_2 is a Lyndon word by assumption. But now, either

$$v_1yxv_1^{-1} \lhd_B yx \qquad \text{or} \qquad v_2^{-1}yxv_2 \lhd_B yx$$

a contradiction. □

Proposition 4. *Every unbordered Sturmian word is a Lyndon word.*

Proof. Let $u \in \{a, b\}$ be an unbordered Sturmian word. Assume u begins with a and ends with b without restriction of generality. The case is clear if $u = ab^k$ for some $k \geq 1$. Assume a occurs at least twice in u. Then $u = ab^kvab^{k+1}$ and u can be factored into ab^k and ab^{k+1} for some $k \geq 1$. Let $\tau\colon \{a, b\}^* \to \{a, b\}^*$ such that $\tau(a) = ab^k$ and $\tau(b) = ab^{k+1}$. Now, let $w = \tau(u)$ and we have that w is an unbordered Sturmian word that begins with a and ends in b. By induction w is a Lyndon word w.r.t. $a \lhd b$ and u is a Lyndon word w.r.t. $\lhd$ by Lemma 3. □

However, Lyndon words are not the only words that have only a trivial Duval extension. Consider

$$ababbaabb \qquad \text{and its reverse} \qquad bbaabbaba$$

which both have no nontrivial Duval extension and are not Lyndon words. Note, that these examples are the only words up to isomorphism that are of minimal length in a binary alphabet.

Finally in this section, let us consider the following corollary of Theorem 2 which will be used in section 4.

Corollary 5. *Let $wvwu$ be a nontrivial Duval extension of wv. Then vw is not a Lyndon word.*

Proof. Assume vw is a Lyndon word. Then vwu is a trivial Duval extension of vw, and hence, $u \leq (vw)^k$ for some $k \geq 1$. But now, we have $\lambda(wvwu) = |wv| = \mu(wvwu)$ and $wvwu$ is a trivial Duval extension; a contradiction. □

4 On Duval's Conjecture

It is a longstanding conjecture by Duval [2] that it is always the case that $|w| \geq |u|$ for a nontrivial Duval extension wu of w.

Conjecture 6 (Duval). *Every Duval extension wu where $|u| \geq |w|$ is trivial.*

Actually, it is believed that a stronger version of that conjecture is true, see also [4]. Namely, every Duval extension wu where $|u| \geq |w| - 1$ is trivial.

The sharpened Duval's conjecture cannot be strengthened further, as the following example shows. Let $w = a^iba^{i+j}bb$, then $u = a^{i+j}ba^i$ gives a nontrivial Duval extension $wu = a^iba^{i+j}bba^{i+j}ba^i$ of w of length $2|w| - 2$.

Nontrivial Duval extensions of w of length $2|w|-2$ seem to be of a special shape. We propose tha following conjecture.

Conjecture 7. *Let $w = w'ab^k$ for some $k \geq 1$. If wu is a nontrivial Duval extension of w of length $2|w| - 2$, then b^k does not occur in w'.*

The following theorem shows that Conjecture 7 implies the sharpened Duval's conjecture.

Theorem 8. *If for every nontrivial Duval extension wv of w of length $2|w| - 2$, with $w = w'ab^k$ for some $k \geq 1$, we have that b^k does not occur in w', then every Duval extension wu of w where $|u| \geq |w| - 1$ is trivial.*

Proof. Let w be an unbordered word of length $n \geq 2$ such that $w = w'ab^k$ for some $k \geq 1$. Assume wu is a nontrivial Duval extension of w such that $|u| \geq n - 1$. Let p be the leftmost position where w is different from u, that is, $u_{(1)}u_{(2)} \cdots u_{(p-1)} \leq w$ and $w_{(p)} \neq u_{(p)}$. If $u > n$, we can assume that there exists a nontrivial Duval extension wu' with $|u'| \leq n$ and $u' \leq u$ by Lemma 1. So, let's assume that $n - 1 \leq |u| \leq n$.

We can assume that $|u| = n - 1$ if $p \leq n - 1$ since any prefix u' of u such that $|u'| \geq p$ gives a nontrivial Duval extension wu' of w.

Case 1: If $p < n - 1$. Let wu', with $u' = u_{(1)}u_{(2)} \cdots u_{(n-2)}$. We apply conjecture 7. Then wu' is a nontrivial Duval extension of length $2n - 2$, and hence, b^k does not occur in w'. Neither does b^k occur in u, since if $u''b^k \leq u$ then $wu''b^k$ is unbordered; a contradiction. Let $u = u_0ab^\ell$ for some $0 \leq \ell < k$. If $\ell < k-1$ then b^ku_0a is longer than n and unbordered; a contradiction. Assume $\ell = k-1$. Let q be the rightmost position where $w'ab^{k-1}$ is different from u, that is, $u_{(q+1)}u_{(q+2)}\cdots u_{(n-1)} \preccurlyeq w'ab^{k-1}$ and $w_{(q)} \neq u_{(q)}$.

We have that

$$w_0 = w_{(q)}u_{(q+1)}u_{(q+2)}\cdots u_{(n-1)}bu_{(1)}u_{(2)}\cdots u_{(q)}$$

is of length $n+1$, and hence, bordered by some word v of minimum length m such that $1 < m < n-q$, since $w_{(q)} \neq u_{(q)}$ and b^k does not occur in u. Note, that v is unbordered since it is of minimum length. We have $u_{(q+i)} = v_{(i+1)}$ for all $1 \leq i \leq m-1$. Consider,

$$w_1 = v_{(1)}v_{(2)}\cdots v_{(m)}u_{(n-m-1)}\cdots u_{(n-1)}bu_{(1)}u_{(2)}\cdots u_{(q)}v_{(2)}v_{(3)}\cdots v_{(m)}$$

which is a factor of wu and $|w_0| < |w_1|$. Let v' be the shortest border of w_1 of length m'. Then $m < m' < n-q$ since v is unbordered and b^k does not occur in u. Again, we have that $u_{(q+i)} = v'_{(i+1)}$ for all $1 \leq i \leq m'-1$. By induction, we get an infinite sequence w_0, w_1, w_2, $\ldots$ such that

$$|w_0| < |w_1| < |w_2| < \cdots$$

which contradicts the finiteness of wu.

Case 2: If $p \geq n-1$. Then $w = w'w_{(n-1)}w_{(n)}$ and $u = w'u'$, where $u' \neq \varepsilon$. Since there are at least two different letters in wu, we have that $w'w_{(n-1)}w_{(n)}w'$ contains at least one Lyndon word which is a conjugate of w. By Corollary 5 wu is a trivial Duval extension; a contradiction. □

Acknowledgement

The authors would like to thank Julien Cassaigne for his comments and a fruitful discussion.

References

[1] R. Assous and M. Pouzet. Une caractérisation des mots périodiques. *Discrete Math.*, 25(1):1–5, 1979.

[2] J.-P. Duval. Relationship between the period of a finite word and the length of its unbordered segments. *Discrete Math.*, 40(1):31–44, 1982.

[3] A. Ehrenfeucht and D. M. Silberger. Periodicity and unbordered segments of words. *Discrete Math.*, 26(2):101–109, 1979.

[4] F. Mignosi and L. Q. Zamboni. A note on a conjecture of Duval and Sturmian words. *Theor. Inform. Appl.*, 36(1):1–3, 2002.

Turku Centre for Computer Science
Lemminkäisenkatu 14
FIN-20520 Turku
Finland

http://www.tucs.fi

University of Turku
- **Department of Information Technology**
- **Department of Mathematics**

Åbo Akademi University
- **Department of Computer Science**
- **Institute for Advanced Management Systems Research**

Turku School of Economics and Business Administration
- **Institute of Information Systems Science**