Unbordered Factors and Lyndon Words

J.-P. Duval
Univ. of Rouen
France

T. Harju
Univ. of Turku
Finland

D. Nowotka
Univ. of Stuttgart
Germany

September 2006

Abstract

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 right extension of a word w of length n is a word wu where all factors longer than n are bordered. A right 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 right extensions. Moreover, we give a conjecture which characterizes a property of every word w which has a nontrivial right extension of length $2|w| - 2$.

1 Introduction

The objects of interest in this paper are finite words. The phenomenon of periodicity of words satisfying some non-trivial relation has been a key issue in the investigations of word properties. One prominent example is the well-known theorem by Fine and Wilf [4] relating the length of a word and its periods.

In this paper we investigate the relationship between the period of a finite word and the maximum length of its unbordered factors which is a field of research that was initiated in the late 70's and beginning of the 80's [1, 2, 3]. This line of research culminated in Duval's conjecture [2] which was finally proved by the second and third author in [6] in a sharpened form. A somewhat shorter proof was later found by Holub [7]. A right extension of an unbordered word w of length n is a word wu where all factors longer than n are bordered. We call a right extension wu of w trivial if the length of w is the period of wu. The result of [6] states that for any unbordered word w of length n, any right extension longer than or equal to $2n - 1$ is trivial.

The special case of words allowing only trivial right extensions has attracted quite some interest. So far this problem has only been solved for the special case of Sturmian words [9]. We improve this result by showing that Lyndon words have only trivial right extensions. Note that it was shown in [5] that each unbordered Sturmian word is a Lyndon word.

Finally, in Section 4 we give a conjecture describing the shape of any word w which has a nontrivial right extension of length $2|w|-2$, and we show that this conjecture implies the sharpened version of Duval's conjecture.

2 Preliminaries

We consider finite words on a finite alphabet A. 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 $w = u^k$ implies $k = 1$. 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 $w = uv$ for some words u and v. Then u is called a *prefix* of w, denoted by $u \leq w$, and v is called a *suffix* of w. If $w = ufv$, where u and v are possibly empty words, then f is a *factor* of w.

Let w be an unbordered word of length $n \geq 1$. We call wu a *right extension* of w, if every factor of wu longer than n is bordered. A right 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 right extension of w, then wu' is a right extension of w, for all $u' \leq u$. Consider for example the word $w = abaabb$, then $abaabbaaba$ is a non-trivial right extension of w. It is easy to check that the word $aabab$ has only trivial right extensions.

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

Lemma 1. *If a word w has a nontrivial right extension, then it has a nontrivial right extension of length at most $\leq 2|w|$.*

Proof. Assume that wv is a nontrivial right extension of w such that $|v| > |w|$. Take the maximum $k \geq 0$ such that $v = w^k w'$ for a nonempty word w'. Let w_0 be the maximum common prefix of w and w'. So, $w' = w_0 v'$, where v' is not empty, since wv is a nontrivial right extension. In particular, we have $|w_0| < |w|$. Now, for any word u such that $u \leq w'$ and $|w_0| < |u| \leq |w|$, the word wu is a nontrivial right extension of w with $|wu| \leq 2|w|$. □

3 Right Extensions of Lyndon Words

The main result of this paper concerns Lyndon words. Recall that two words w and w' are conjugates if $w = xy$ and $w' = yx$ for some words x and y. A primitive word w is called a *Lyndon word* if it is minimal among all its conjugates with respect to some lexicographic order. In other words, see e.g. [8], w is a Lyndon word if it is minimal among its suffixes with respect to some lexicographic order. For example, consider $w = abaabb$. Then $aabbab$ and $bbabaa$ are conjugates of w and minimal with respect to the orders $a \lhd b$ and $b \lhd a$, respectively.

Let wu be a primitive 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 $|u| \leq k$.

Every Lyndon word w is unbordered. Indeed, let w be a Lyndon word according to the order $\lhd$. Then it is easy to see that if $w = uvu$ for some words u and v, then $uuv \lhd uvu$, and hence u must be the empty word.

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

Proof. Let w be a Lyndon word with respect to an order $\lhd$. In particular, w is unbordered. Assume contrary to the claim that there exists a nonempty word u such that wu is a nontrivial right extension of w. Assume that u is chosen so that it is of minimum length. By Lemma 1, we have $|u| \leq |w|$, and since wu is a nontrivial right extension, we have $u \nleq w$. Hence 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 $a \lhd b$.

If $v = \varepsilon$ then $u = b$ since the first letter of w is minimal with respect to $\lhd$. Let the shortest 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. Let $v = v'v''$ be such that $v''a$ is the longest unbordered suffix of va. Consider the suffix $v''bzva$ of wu. We have that $|v''bzva| > |w|$. Let s be the shortest border of $v''bzva$. Now $|s| > |v''|$ otherwise $v''a$ is bordered. Moreover, $|s| \neq |v''a|$ since $a \neq b$. If $|v''a| < |s| \leq |va|$ then $v''a$ is not the longest unbordered suffix of va; a contradiction. But, if $|s| \geq |va|$ then va occurs in w and w is not a Lyndon word; a contradiction.

Case 2: Suppose $u = vb$. Then $w = vaz$. Since va is a prefix of a Lyndon word and $a \lhd b$, we have that wvb is also a Lyndon word, and hence, unbordered. This contradicts the assumption that wu is a right extension. □

The situation is different for words that are not Lyndon words. Consider for example $w = abaabb$ and $u = aaba$ and the nontrivial right extension

$$wu = abaabbaaba\,.$$

Now, every factor of wu of length 7 is bordered. Note also, that $|w| > |u|$.

However, Theorem 2 gives the following corollary.

Corollary 3. *Let $wvwu$ be a nontrivial right extension of wv. Then vw is not a Lyndon word.*

corollary

Proof. Assume vw is a Lyndon word. Then vwu is a trivial right extension of vw by Theorem 2, and hence, $u \leq (vw)^k$ for some $k \geq 1$. But now, we have $wu \leq (wv)^{k+1}$ and $wvwu$ is a trivial right extension; a contradiction. □

Corollary 3 implies the following lemma which is an improvement of Lemma 1 and will be used in the following section.

Lemma 4. *If a primitive word w of length n has a nontrivial right extension wv such that $|v| > |w|$, then it has a nontrivial right extension wu such that $n-2 \leq |u| \leq n$ and $u \leq v$.*

Proof. Note first that every right extension a of a single letter is necessarily trivial. So, we can assume that $n > 1$. Let w_0 be the longest common prefix of w and v, and let $w = w_0 w'$ and $v = w_0 v'$. If $|w_0| < n-2$, each prefix u of v with $n-2 \leq |u| \leq n$ gives that wu is a nontrivial right extension of w. Assume thus that $|w_0| \geq n-2$, and so that $|w'| \leq 2$. Since at least two different letters occur in w (for it is primitive and $n > 1$), there are two occurrences of Lyndon words which are conjugates of w in $w_0 w' w_0$. It follows from Corollary 3 that every right extension of w that has $w_0 w' w_0$ as a prefix is trivial because $w_0 w' w_0 v'$ can be factored into $\bar{w}\bar{v}\bar{w}\bar{u}$ where $\bar{v}\bar{w}$ is a Lyndon conjugate of w in $w_0 w' w_0$ and $\bar{w}$ and $\bar{v}$ correspond to their unbarred versions in Corollary 3. In particular, wv is a trivial right extension of w; a contradiction. □

4 A Conjecture

It was a longstanding conjecture by Duval [2] that every right extension wu of an unbordered word w with $|u| \geq |w|$ is trivial. Actually, a stronger result was proved in [6]. Namely,

Theorem 5. *Every nontrivial right extension wu of an unbordered word w satisfies $|wu| \leq 2|w| - 2$.*

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

Nontrivial right extensions of w of (optimal) length $2|w| - 2$ seem to be all of special shape. We propose that following conjecture.

Conjecture 6. *Let $w = w' a b^k$ for letters $a \neq b$ and integer $k \geq 1$. If w has a nontrivial right extension of length $2|w| - 2$, then b^k does not occur in w'.*

The following theorem shows that Conjecture 6 implies Theorem 5.

Theorem 7. *Conjecture 6 implies that every right extension wu of an unbordered w with $|u| \geq |w| - 1$ is trivial.*

Proof. Assume that Conjecture 6 holds. Let w be an unbordered word of length $n \geq 2$. Since w is unbordered, we may write $w = w' a b^k$ for some $a \neq b$ and $k \geq 1$. For the sake of getting a contradiction, assume that wu is a nontrivial right extension of w such that $|u| \geq n - 1$. By Lemma 4, we may assume that $|u| \leq n$. Now, let p be the leftmost position where w is different from u.

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 right extension wu' of w.

Case 1: If $p < n-1$. Let wu' be a nontrivial right extension of w with $u' = u_{(1)}u_{(2)}\cdots u_{(n-2)}$. We apply conjecture 6. Then wu' is a nontrivial right 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, $u_{(q+1)}u_{(q+2)}\cdots u_{(n-1)}$ is a suffix of $w'ab^{k-1}$ and $w_{(q)} \neq u_{(q)}$.

Let $w_{(q)}r$ be the largest unbordered prefix of $w_{(q)}w_{(q+1)}\cdots w_{(n-1)}$. Consider the factor $w_{(q)}w_{(q+1)}\cdots w_{(n)}u_{(1)}u_{(2)}\cdots u_{(q)}r$ which is longer than n, and let s be its shortest border. We have that $|s| > |w_{(q)}r|$ otherwise $w_{(q)}r$ is bordered. Moreover, $|s| > |w_{(q)}w_{(q+1)}\cdots w_{(n-1)}|$ otherwise $w_{(q)}r$ is not the largest unbordered prefix. Hence, $w_{(q)}w_{(q+1)}\cdots w_{(n-1)}b \leq s$ and b^k occurs in u; a contradiction.

Case 2: If $p \geq n-1$. Then $w = \bar{w}w_{(n-1)}w_{(n)}$ and $u = \bar{w}u'$, where $u' \neq \varepsilon$. Since there are at least two different letters in wu, we have that $\bar{w}w_{(n-1)}w_{(n)}\bar{w}$ contains at least one Lyndon word which is a conjugate of w. By Corollary 3 wu is a trivial right extension; a contradiction. $\square$

A short proof of Conjecture 6 would give a shorter proof to the sharpened Duval's Conjecture, Theorem 5.

Acknowledgments

The authors would like to thank the anonymous referees for their detailed comments and suggestions which greatly helped to improve this paper.

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] N. J. Fine and H. S. Wilf. Uniqueness theorem for periodic functions. *Proc. Amer. Math. Soc.*, 16:109–114, 1965.

[5] T. Harju and D. Nowotka. Minimal duval extensions. *Internat. J. Foundations of Comput. Sci.*, 15:349–354, 2004.

[6] T. Harju and D. Nowotka. Periodicity and unbordered words. In V. Diekert and M. Habib, editors, *21st Annual Symp. Theoret. Aspects of Comput. Sci.*, volume 2996 of *Lecture Notes in Comput. Sci.*, pages 294–304, 2004.

[7] S. Holub. A proof of the extended Duvals conjecture. *Theoret. Comput. Sci.*, 339(1):61–67, 2005.

[8] M. Lothaire. *Combinatorics on Words*, volume 17 of *Encyclopedia of Mathematics*. Addison-Wesley, Reading, Massachusetts, 1983.

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