On Unique Factorizations of Primitive Words

Tero Harju
Department of Mathematics,
University of Turku, 20014 Turku, Finland
harju@utu.fi

Dirk Nowotka
Institute of Formal Methods in Computer Science,
University of Stuttgart, 70569 Stuttgart, Germany
nowotka@fmi.uni-stuttgart.de

October 2006

Abstract

We give a short proof of a result by C. M. Weinbaum [Proc. AMS, 109(3):615–619, 1990] stating that each a primitive word of length at least 2 has a conjugate $w' = uv$ such that both u and v have a unique position in the cyclic word of w.

1 Introduction

Let $w \in A^*$ be a word over the alphabet A. We say that a factor v of w has a *unique position* in w, if w has a unique conjugate having v as its prefix. For instance, if $w = abaabab \in \{a, b\}^*$, then the factor $baab$ has a unique position in w.

We shall give a short proof of the following theorem due to Weinbaum [9] using the Critical Factorization Theorem and basic properties of Lyndon words.

Theorem 1. *Let w be a primitive word of length at least 2. There is a conjugate $w' = uv$ of w such that u and v have both a unique position in w.*

Such a factorization uv of a conjugate w' is called a *Weinbaum factorization.* In the above example, where $w = abaabab$, the factor $baab$ does not induce a Weinbaum factorization, since in the conjugate $w' = baababab$ the suffix $abab$ does not have a unique position in w. Instead, a Weinbaum factorization for w is given by the conjugate $w'' = aa.babab$ with the indicated factorization.

We mention that Duncan and Howie [3] considered the extension of the present problem for free groups. Also, in his article Weinbaum [9] showed a stronger result than Theorem 1, a short proof of which is given by the present authors in [6].

Theorem 2. *Let w be a primitive word of length at least 2. Then for every letter a there is a conjugate w' such that $w' = uv$ and u and v have a unique position in w and u begins and ends in the letter a and v does not begin nor end in a.*

2 Lyndon words and the CFT

We consider finite words on a finite alphabet A. A nonempty word u is a *border* of a word w, if $w = uv = v'u$ for some words v and v'. The word w is said to be *bordered* if it has a border that is shorter than w, otherwise w is called *unbordered*. We notice that every bordered word w has a minimum border u such that u is unbordered and $|u| \leq |w|/2$, where $|w|$ denotes the length of w.

A word w is called *primitive* if $w = u^k$ implies $k = 1$. Let $w = uv$ for some words u and v. Then u is called a *prefix* of w, denoted by $u \leq_{\mathrm{p}} w$, and v is called a *suffix* of w, denoted by $v \leq_{\mathrm{s}} w$. If $w = ufv$, where u and v are possibly empty words, then f is a *factor* of w.

The Critical Factorization Theorem (CFT) is one of the main results concerning periodicity of words. A weak version of this theorem was conjectured by Schützenberger [8] and then proved by Césari and Vincent [1]. Later it was developed into its present form by Duval [4].

An integer $1 \leq p \leq n$ is a *period* of $w = a_1 a_2 \dots a_n$, where $a_i \in A$, if $a_i = a_{i+p}$ for all $1 \leq i \leq n - p$. The smallest period of w is called the *minimum period* of w, denoted by $\pi(w)$.

An integer p with $1 \leq p < |w|$ is a *point* in w. A nonempty word u is called a *repetition word* at p if $w = xy$ with $|x| = p$ and there exist words x' and y' such that $u \leq_{\mathrm{s}} x'x$ and $u \leq_{\mathrm{p}} yy'$. Let

$$\pi(w,p) = \min\bigl\{|u| \bigm| u \text{ is a repetition word at } p\bigr\}$$

denote the *local period* at point p in w. A factorization $w = uv$, with $u, v \neq \varepsilon$ and $|u| = p$, is *critical*, and p is a *critical point*, if $\pi(w,p) = \pi(w)$.

As an example, consider the word $w = abaab$ of period 3. It has two critical points, 2 and 4, indicated by dots in $ab.aa.b$. The shortest repetition words at these critical points are aab and baa, respectively.

Theorem 3 (Critical Factorization Theorem)**.** *Each word w with $|w| \geq 2$ has at least one critical factorization $w = uv$, with $u, v \neq \varepsilon$ and $|u| < \pi(w)$, i.e., $\pi(w, |u|) = \pi(w)$.*

Let $\lhd$ be an ordering of $A = \{a_1, a_2, \dots, a_n\}$, say $a_1 \lhd a_2 \lhd \cdots \lhd a_n$. Then $\lhd$ induces a *lexicographic order* on A^* such that

$$u \lhd v \iff u \leq_{\mathrm{p}} v \quad \text{or} \quad u = xau' \text{ and } v = xbu' \text{ with } a \lhd b$$

where $a, b \in A$. A suffix v of w is called *maximum* w.r.t. $\lhd$ if $v' \lhd v$ for any suffix v' of w. Let $\lhd^{-1}$ denote the *inverse order*, $a_n \lhd^{-1} \cdots \lhd^{-1} a_2 \lhd^{-1} a_1$, of $\lhd$.

We refer to [5] for a short proof of the CFT giving a technically improved version of the proof by Crochmore and Perrin [2]. This proof gives the following

Theorem 4. *Let $w \in A^*$ be a word with $|w| \geq 2$, and let $\lhd$ be an order on A. Let $\{\alpha, \beta\}$ be the set of the maximum suffixes of w w.r.t. $\lhd$ and $\lhd^{-1}$ such that $|\beta| < |\alpha|$, say $\alpha = u\beta$. Then $|u|$ is a critical point of w.*

Recall that two words w and w' are *conjugates* if $w = uv$ and $w' = vu$ for some words u and v. A primitive word w is called a *Lyndon word* if it is minimal among all its conjugates with respect to the lexicographic order $\lhd$ (for some $\lhd$). In other words, see e.g. [7], 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.

The following result is well known.

Lemma 5. *Each Lyndon word is unbordered. In particular, every primitive word has an unbordered conjugate.*

Proof. Assume that uvu is a Lyndon word w.r.t. $\lhd$, where u is nonempty. Then $uvu \lhd uuv$ and so $vu \lhd uv$, when the common prefix is removed. This gives a contradiction, $vuu \lhd uvu$. □ □

We notice that Lemma 5 follows also from the Critical Factorization Theorem. Indeed, consider a critical point of the word w^2 for which $\pi(w) = |w|$ whenever w is primitive. Thus w^2 does have an unbordered factor of length $\pi(w^2)$ and hence this factor is a conjugate of w.

Moreover, each word having k many different letters has at least k Lyndon words among all conjugates, since there is a Lyndon word beginning with a for each letter a.

Lemma 6. *Let $\lhd$ be an order on the alphabet A. If $u \lhd v$ and $u \lhd^{-1} v$ then $u \leq_{\mathrm{p}} v$.*

Proof. Assume that u is not a prefix of v and let w be the longest common prefix of u and v, i.e., $u = wau'$ and $v = wbv'$, where $a \lhd b$ for some $a, b \in A$. It follows that $bv' \lhd^{-1} au'$ and thus also $v \lhd^{-1} u$, as required. □ □

3 Weinbaum points

If $w = uv$ is a Weinbaum factorization then $|u|$ is called a *Weinbaum point* of w. For instance, the word $w = abaababba$ has three Weinbaum points 4, 5 and 6, since $abaa.babba$, $abaab.abba$, and $abaaba.bba$ are all Weinbaum factorizations of w.

We prove first an result that does not refer to conjugates of a word. We say that u and v *intersect*, if u and v overlap or one is a factor of the other.

Lemma 7. *Let $w = uv$ be an unbordered word with a critical point $|u|$. Then u and v do not intersect.*

Proof. Let $w = uv$ be any factorization of the unbordered word w to nonempty intersecting words u and v. We can assume that $|u| \leq |v|$ without loss of generality. Note that $\pi(w) = |w|$, since w is unbordered. If $u = u's$ and $v = sv'$ for a nonempty word s, then $\pi(w, |u|) \leq |s| < |w|$, and so $|u|$ is not a critical point. Similarly, if $u = su'$ and $v = v's$, then s would be a border of w; a contradiction. Finally, if $v = sut$, then $\pi(w, |u|) \leq |su| < |w|$, and again $|u|$ is not a critical point. These cases prove the claim. □ □

Theorem 8. *Let w be an unbordered word with $|w| \geq 2$. Then every critical point of w is a Weinbaum point.*

Proof. Since w is unbordered, we have $\pi(w) = |w|$. It follows that if $x \leq_{\mathrm{p}} w$ is any prefix of w, say $w = xy$, then $w \neq x_2 y x_1$ for all factorizations $x = x_1 x_2$ of x where x_1 is nonempty. Therefore, if x occurs only as a prefix in w, it has a unique position in (the cyclic word) w. The same conclusion holds for the suffixes of w.

Let then $w = uv$ be such that $p = |u|$ is a critical point in w. By Lemma 7, u and v do not intersect, and by the above, it follows that u and v have a unique position in w. Therefore $|u|$ is a Weinbaum point of w. □ □

Corollary 9. *Let w be a primitive word with $|w| \geq 2$. There is a conjugate uv of w such that u and v have a unique position in w.*

Proof. Let w' be an unbordered conjugate of w which exists by Lemma 5 since w is primitive. By Theorem 8, w' does have a Weinbaum point. □ □

In the next corollary we consider the strong version of Weinbaum's theorem for binary words. Note that there we can have that $u = a$ and/or $v = b$.

Corollary 10. *Let $A = \{a, b\}$ be a binary alphabet, and let $w \in A^*$ be a primitive word with $|w| \geq 2$. Then there exists a conjugate w' of w such that $w' = uv$, where u and v have a unique position in w and $u \in aA^* \cap A^*a$ and $v \in bA^* \cap A^*b$.*

Proof. Let w' be a Lyndon word of w with respect to the order $a \lhd b$. Then w' begins with the letter a and ends with b. Let then $w' = uv$, where v is the maximum suffix of w' with respect to the order $\lhd^{-1}$. By Lemma 6, w' is the maximum suffix w.r.t. $\lhd$, and hence, by Theorem 4, $p = |u|$ is a critical point of w', and also a Weinbaum point by Theorem 8. By the choice of v, $a \leq_{\mathrm{s}} u$ and $b \leq_{\mathrm{p}} v$. This proves the claim. □ □

As an example, for *non-binary* words consider the Lyndon word $w = abccac$. The order of the letters is given by $a \lhd b \lhd c$. The Weinbaum points in w are 2 ($w = ab.ccac$; a critical point) and 3 ($w = abc.cac$; a noncritical point). These are not of the strong form required for binary words by the proof of Corollary 10. However, w does have a conjugate $w' = acabcc$ in which we have a Weinbaum factorization $w = aca.bcc$ as required by Theorem 2.

References

[1] Y. Césari and M. Vincent. Une caractérisation des mots périodiques. *C. R. Acad. Sci. Paris Sér. A*, 286:1175–1177, 1978.

[2] M. Crochemore and D. Perrin. Two-way string-matching. *J. ACM*, 38(3):651–675, 1991.

[3] A. J. Duncan and J. Howie. Weinbaum's conjecture on unique subwords of nonperiodic words. *Proc. Amer. Math. Soc.*, 115(4):947–954, 1992.

[4] J.-P. Duval. Périodes et répétitions des mots de monoïde libre. *Theoret. Comput. Sci.*, 9(1):17–26, 1979.

[5] T. Harju and D. Nowotka. Density of critical factorizations. *Theor. Inform. Appl.*, 36(3):315–327, 2002.

[6] T. Harju and D. Nowotka. Uniquely occurring factors in primitive words. *in preparation*, 2005.

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

[8] M.-P. Schützenberger. A property of finitely generated submonoids of free monoids. In *Algebraic theory of semigroups (Proc. Sixth Algebraic Conf., Szeged, 1976)*, volume 20 of *Colloq. Math. Soc. János Bolyai*, pages 545–576. North-Holland, Amsterdam, 1979.

[9] C. M. Weinbaum. Unique sunwords in nonperiodic words. *Proc. Amer. Math. Soc.*, 109(3):615–619, 1990.