Factorizations of Cyclic Words

Volker Diekert[*] Tero Harju[†] Dirk Nowotka[*]

January 12, 2007

Abstract

Weinbaum [Proc. AMS, 109(3):615–619, 1990] showed that for each letter a in a primitive word w, there exists a conjugate $w' = uv$ of w such that both u and v are uniquely positioned in the cyclic word w such that u begins and ends with a and v neither begins nor ends with a. We give a generalization of this result using iterative methods.

1 Introduction

Let A be an alphabet and let A^* be the set of all words over A, where ε denotes the empty word. A word $w \in A^*$ is called *primitive*, if it is not a proper power of another word, i.e., $w = x^k$ implies $k = 1$. A word w' is a *conjugate* of w, if there are words u and v such that $w = uv$ and $w' = vu$, where u or v can be empty. A word f is a (*proper*) *factor* of w if $w = ufv$ (and $uv \neq \varepsilon$). A factor f is *uniquely positioned* in (the cyclic word) w if there exists a unique conjugate w' of w such that $w' = fv$.

Weinbaum showed in [2] that for each letter a occurring in a primitive word w, there exists a conjugate $w' = uv$ of w such that both u and v are uniquely positioned in w, and u begins and ends with a and v neither begins nor ends with a. In this paper, we give a short proof of Weinbaum's result and generalize it. In Theorem 4 we consider suitable pairs of factors of primitive words instead of letters.

In Section 2 we prove the existence of Weinbaum factorizations by employing Lyndon words w.r.t. specific lexicographic orders. In Section 3 we consider Weinbaum factorizations without orderings of the alphabets. The main result of the paper states that a Weinbaum factorization of a word w

[*]University of Stuttgart, Germany; {diekert,nowotka}@fmi.uni-stuttgart.de
[†]University of Turku, Finland; harju@utu.fi

can be found be iterating a given relation W at most $\log_2(n)$ many times for $n = |w|$. In Section 4 we consider the Fibonacci words as an example of the theory. Finally, in Section 5 the effect of Weinbaum factorizations on long words is discussed.

We end this section with some more notation. The length of a word $w \in A^*$ is denoted by $|w|$. Let $w = uv$ for some, possibly empty, 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 $v \leq_{\mathrm{s}} w$. We say that f *occurs* in w if f is a factor of w.

We say that a word f is a *cyclic factor* of w if $f \leq_{\mathrm{p}} w'$ for some conjugate w' of w. Note that f is a cyclic factor of w if and only if $|f| \leq |w|$ and f is a factor of w^2.

Two words u and v are said to *intersect in* (the cyclic word) w, if w^3 has a factor xyz such that $u = xy$ and $v = yz$ or $u = yz$ and $v = xy$, where x, y and z are nonempty words, or if u is a factor of v or v is a factor of u. A word f is called a *marker* if f does not intersect nontrivially with itself in w. We notice that if f is a marker in w with $f \leq_{\mathrm{p}} w$, then w has a unique factorization of the form $w = fw_1fw_2 \dots fw_k$ for some, possibly empty, words w_i with $fw_if \notin (A^* \setminus \{\varepsilon\})f(A^* \setminus \{\varepsilon\})$.

Omitted proofs from this extended abstract will appear in the full version of this paper.

2 Weinbaum Factorizations

Let w be a primitive word with a conjugate w', and let f be a word. Then $w' = uv$ is a *Weinbaum factorization of w for f* if u and v are uniquely positioned in w, and $u \in fA^* \cap A^*f$ and $v \notin fA^* \cup A^*f$. This coincides with Weinbaum's original definition in the case where f is a single letter. However, we propose a stronger, more symmetric definition of a factorization for two factors f and g such that $w' = uv$ is called a *Weinbaum factorization of w for f and g* if u and v are uniquely positioned in w, and $u \in (fA^* \cap A^*f) \setminus (gA^* \cup A^*g)$ and $v \in (gA^* \cap A^*g) \setminus (fA^* \cup A^*f)$. Note that we may have the cases where $u = f$ or $v = g$.

Remark 1. The following holds.

$$(fA^* \cap A^*f) \setminus (gA^* \cup A^*g) \neq \varnothing \iff g \not\leq_{\mathrm{p}} f \text{ and } g \not\leq_{\mathrm{s}} f \ .$$

Indeed, if there exists a word that begins with f but does not begin with g, then $g \not\leq_{\mathrm{p}} f$. Similarly, for the suffix. On the other hand, if $(fA^* \cap A^*f) \setminus$

$(gA^* \cup A^*g) = \varnothing$ then every word that begins with f also begins with g, and hence, $g \leq_{\mathrm{p}} f$. The suffix case is symmetric.

It follows from Remark 1 that by our definition of Weinbaum factorization for two factors f and g we have that $f \not\leq_{\mathrm{p}} g$, $f \not\leq_{\mathrm{s}} g$, $g \not\leq_{\mathrm{p}} f$, and $g \not\leq_{\mathrm{s}} f$ are necessary conditions.

As shown in Theorem 4, every primitive word w admits a Weinbaum factorization uv not just for a letter a with $u \in aA^* \cap A^*a$ and $v \notin aA^* \cap A^*a$, but we can require that for suitably chosen factors f and g of w we have $u \in fA^* \cap A^*f$ and $v \in gA^* \cap A^*g$. Corollary 5 demonstrates that Theorem 4 is indeed a generalization of Weinbaum's theorem.

Example 2. Consider the word

$$w = bbaabacbaacbaaaca$$

over the alphabet $A = \{a, b, c\}$. Then w has a Weinbaum factorization for $f = ab$ and $g = ac$. Indeed, w has the conjugate, obtained by shifting the suffix a to the beginning,

$$w' = abbaabacbaacbaaac = fbafgbagbaag\,,$$

where both $fbaf$ and $gbagbaag$ are uniquely positioned in w. □

Let $\trianglelefteq$ be a total order on the alphabet A. Then $\trianglelefteq$ can be extended to a *lexicographic* order on A^* by setting $u \trianglelefteq v$ if either $u \leq_{\mathrm{p}} v$ or $xa \leq_{\mathrm{p}} u$ and $xb \leq_{\mathrm{p}} v$ where $a \neq b$ and $a \trianglelefteq b$ and $x \in A^*$. A primitive word that is the minimum among its conjugates w.r.t. $\trianglelefteq$ is called the *Lyndon word* w.r.t. $\trianglelefteq$. Note that if w is a Lyndon word then it is unbordered: $w \notin fA^* \cap A^*f$ for any nonempty word f. Indeed, assume the opposite and let f be of minimum length such that $w \in fA^* \cap A^*f$. Clearly, $w = fxf$, but then $w = fxf \trianglelefteq ffx$ implies $xf \trianglelefteq fx$, and hence, $xff \trianglelefteq fxf = w$; a contradiction.

Lemma 3. *Let $w = uv$ be a Lyndon word w.r.t. a lexicographic order $\trianglelefteq$ such that v is the maximum suffix of w w.r.t. $\trianglelefteq$. Then both u and v are uniquely positioned in w. Moreover, if v' is a cyclic factor of w such that $v \trianglelefteq v'$, then $v \leq_{\mathrm{p}} v'$.*

Proof. Assume that v' is a cyclic factor of w such that $v \trianglelefteq v'$, and let $w^2 = xv'y$ where $|x| < |w|$. Suppose first that $w = xv'y'$. Since v is the maximum suffix and $v'y' \trianglelefteq v \trianglelefteq v'$, necessarily $y' = \varepsilon$, and thus $u = x$ and $v = v'$.

Let then $w = xv'_1 = v'_2y$ where $v' = v'_1v'_2$ with $v'_2 \neq \varepsilon$ so that $w^2 = xv'y$. Assume that $v'_1 \neq v$. If $|v'_1| > |v|$ then $v \trianglelefteq v'$ implies $v \trianglelefteq v'_1$ contradicting the maximality of v. If $|v'_1| < |v|$ then $v'_1 \trianglelefteq v \trianglelefteq v'_1 v'_2$ $(= v')$ implies that $v = v'_1 v_2$ for some $v_2 \neq \varepsilon$ with $v_2 \trianglelefteq v'_2$. Thus $v_2 \not\leq_{\mathrm{p}} v'_2$, for otherwise $w \in v_2 A^* \cap A^* v_2$ would imply that w is not a Lyndon word. But now, $v_2 u v'_1$ is a conjugate of w and $v_2 u v'_1 \trianglelefteq v'_2 y = w$ contradicts the assumption that w is a Lyndon word. Consequently, $v \trianglelefteq v'$ implies that $v \leq_{\mathrm{p}} v'$ and v is unique positioned in w.

Finally, consider the occurrences of the prefix u. Let $w^2 = xuy$ where $0 < |x| \leq |w|$. Let $w^2 = xuv'y'$ where $|v'| = |v|$. We have $v \trianglelefteq v'$ because uv' is a conjugate of w and w is a Lyndon word. Now v' is a cyclic factor of w, and hence, by the above, $v = v'$ and v' is uniquely positioned in w. This means that $w = x$ and therefore also u is uniquely positioned in w. □

For the statement of Theorem 4 we introduce the notion of complementary marker. Let w be a primitive word and f be a factor in w such that there is a conjugate w' of w with $w' = u_1 v_1 u_2 v_2 \cdots u_k v_k$ where $u_i \in fA^* \cap A^* f$ for all i and f is not a factor of v_i for any $1 \leq i \leq k$. A factor g of w is a *complementary marker* for f in w, if

(1) g does not intersect in w with any u_i,

(2) f and g do not intersect in w, and

(3) g is not a proper factor of any v_i.

Note that if g is a complementary marker of f, then g is also a marker, and $g = v_i$ for at least one index i. For every primitive word w with $|w| > 1$, there is a factor f for which a complementary marker exists in w. Indeed, choose any $f \in A$ and $u_i \in f^*$. Then g can be chosen to be a word of maximum length between two occurrences of the letter f.

In the following proof, we let $\bar{z}$ denote a letter corresponding to a word z.

Theorem 4. *Let w be a primitive word and let f be a factor of w such that a complementary marker g exists for f in w. Then w has a Weinbaum factorization for f and g.*

Proof. Since we consider conjugates of words, and g is a marker, we may assume that $w = gz_1gz_2 \cdots gz_k$ where $k \geq 1$ such that $z_i \in fA^* \cap A^* f$ and g is not a factor of any z_i. Let $B = \{\bar{g}, \bar{z}_i \mid i = 1, 2, \ldots, k\}$ be a new alphabet corresponding to the words g and z_i. We may assume that $x = \bar{g}\bar{z}_1\bar{g}\bar{z}_2 \cdots \bar{g}\bar{z}_k$ is a Lyndon word w.r.t. a lexicographic order $\trianglelefteq$ on B^* such that $\bar{g}$ is the minimum in B and if z_i occurs in z_j, then $\bar{z}_i \trianglelefteq \bar{z}_j$ for all $1 \leq i, j \leq k$.

Let t is the maximum suffix of x w.r.t $\trianglelefteq$, say $x = st$. Then $s = \bar{g}\bar{z}_1 \cdots \bar{g}\bar{z}_{m-1}\bar{g}$ and $t = \bar{z}_m\bar{g} \cdots \bar{z}_{k-1}\bar{g}\bar{z}_k$, where $\bar{z}_m$ is the maximum element w.r.t. $\trianglelefteq$. By Lemma 3, the prefix s is uniquely positioned in x, and hence also the corresponding prefix $v = gz_1 \cdots gz_{m-1}g$ of w is uniquely positioned in w, since the factor g serves as a marker. Also, $v \in gA^* \cap A^*g$.

Again by Lemma 3, the word $t = \bar{z}_m\bar{g} \cdots \bar{z}_{k-1}\bar{g}\bar{z}_k$ is uniquely positioned in x, but now it is not so immediate that the position of $u = z_mg \cdots z_{k-1}gz_k$ is unique in w. The factor z_m corresponding to the maximum $\bar{z}_m$ serves as a marker, and thus there is a cyclic factor u' in w with $u' = z_mg \cdots z_{k-1}gz_\ell$ where $z_k \leq_{\mathrm{p}} z_\ell$. But then $t \trianglelefteq \bar{z}_m\bar{g} \cdots \bar{z}_{k-1}\bar{g}\bar{z}_\ell$. By Lemma 3, this implies $\bar{z}_k = \bar{z}_\ell$, and so $u = u'$ and $x = s\bar{z}_m\bar{g} \cdots \bar{z}_{k-1}\bar{g}\bar{z}_\ell$. This means $w = vu$ and u is uniquely positioned in w. Finally, we obtain from $z_m, z_k \in fA^* \cap A^*f$ that also $u \in fA^* \cap A^*f$. □

The following corollary is a slightly generalized version of Weinbaum's original theorem.

Corollary 5. *Let w be a primitive word and a^m be a cyclic factor of w for some $m \geq 1$. Then w has a Weinbaum factorization for a^m.*

Proof. Since w is primitive, there is a conjugate $w' = a^{n_1}v_1a^{n_2}v_2 \cdots a^{n_k}v_k$ of w such that $n_i \geq m$, a^m does not occur in v_i and $v_i \notin aA^* \cup A^*a$. Hence, there exists a complementary marker g of a^m in w. The claim follows from Theorem 4. □

3 An Iterative Construction

It was shown in the previous section that Weinbaum factorizations can be constructed directly by a mapping from A to a new alphabet B and taking the maximum suffix of a Lyndon word w.r.t. some lexicographic order on B^*. The choice of the lexicographic order is crucial there. In this section we consider Weinbaum factorizations from a different point of view. We do not require orderings of the alphabets here. The main result of this section is Theorem 11 which shows that a Weinbaum factorization of a word w can be found be iterating a special relation W (as defined below) at most $\log_2(n)$ many times where $n = |w|$.

In the following, let w be a fixed primitive word. Let f be a proper factor of w. We define the set $G(f)$ of factors of w such that $g \in G(f)$ if g is a cyclic factor of w that is preceded and followed by f and g does not intersect with f. More precisely

$$G(f) = \{g \mid g \neq \varepsilon,\ fgf \text{ is a factor of } w^2 \text{ and } f \text{ occurs only as a prefix and suffix in } fgf\}\,.$$

Clearly $|fg| \leq |w|$ for all $g \in G(f)$ by the property required from the factor f.

Example 6. The set $G(f)$ can be empty even for short factors f. For instance, consider $w = (aab)^k aaaba$ with $k \geq 2$, and $f = aabaa$. Here each factor fgf has a third occurrence of f. □

We define the subset $W(f)$ of the set of factors of w as follows

$$W(f) = \{g \in G(f) \mid g \text{ does not occur in any other element of } G(f), \\ \text{and } g \text{ does not intersect with } f \text{ in } w\}.$$

A word f is a *Weinbaum factor* of w if $W(f) \neq \varnothing$. Note that a Weinbaum factor is not necessarily a marker (and vice versa).

Example 7. Let $w = abababb$. Then $f = aba$ is not a marker in w but $W(f) = \{bb\}$. Now, ab is a marker in w but $G(ab) = \{b\}$ and $W(ab) = \varnothing$.

Let us consider some properties of $W(f)$. We say that a word g is *clipped* by f in w if for all conjugates w' of w where $g \leq_{\mathrm{p}} w'$ it follows that $w' \in gfA^* \cap A^*f$.

Lemma 8. *Let f be a Weinbaum factor of w. Then we have:*

1. *Each $g \in W(f)$ is clipped by f in w.*
2. *If $g \in W(f)$ then $G(g) \neq \varnothing$.*
3. *If $g \in W(f)$ then $W(g) \subseteq fA^* \cap A^*f$.*
4. *$W^{2n}(f) \subseteq fA^* \cap A^*f$ for all $n \geq 1$.*
5. *$W^{2n-1}(f) \cap (fA^* \cup A^*f) = \varnothing$ for all $n \geq 1$.*

Lemma 9. *Let f be a Weinbaum factor of w. Then each $g \in W(f)$ is a marker and a Weinbaum factor of w.*

Lemma 10. *Let f be a marker in w. Then for each $g \in W(f)$, either $W(g) = \{f\}$ or $W(g) \subseteq fA^*f$.*

For the construction of a Weinbaum factorization we iterate the operation W on a given factor f of w. We are not interested in the set of the solutions but only in a single factorization. Let therefore $\overline{W}$ be an arbitrary choice function on the set of Weinbaum factors of w that selects some element from $W(f)$, that is, $\overline{W}(f) = g$ for some $g \in W(f)$.

It follows from Lemma 9, and from the fact that there are only finitely many factors of w, that for every Weinbaum factor f of w exists an n such that $\overline{W}^n(f) = \overline{W}^k(f)$ for some $k < n$. The following theorem elaborates on this observation and gives our main result about Weinbaum factorizations.

Theorem 11. *Let w be a primitive word of length m, and let f be a marker. There exists an integer $n \leq \log_2(m)/2$ such that $\overline{W}^{2i}(f) = \overline{W}^{2i-2}(f)$ for all $i \geq n$. Moreover, $\overline{W}^{2n}(f)\,\overline{W}^{2n-1}(f)$ is a Weinbaum factorization of w for f and $\overline{W}(f)$.*

4 An Example Related to Fibonacci Words

The following example provides a sequence of words with a large number of iterations of $\overline{W}$ in order to find a Weinbaum factorization. This example also gives an interesting connection to the Fibonacci words.

Consider the following sequence $\{f_i\}_{i\geq 0}$ of words over the binary alphabet $\{a, b\}$:

$$f_0 = \varepsilon, \quad f_1 = a, \quad f_2 = b, \quad \text{and} \quad f_{i+1} = f_{i-1}f_{i-2}f_{i-1} \quad (i \geq 2)\,.$$

We have for example $f_3 = aa$, $f_4 = bab$, $f_5 = aabaa$, and so forth. Let

$$w_n = f_n f_{n-1}\,.$$

For example $w_1 = a$, $w_2 = ba$, $w_3 = aab$, $w_4 = babaa$, and so on. We will show in the following that we need $\mathcal{O}(\log |w_n|)$ many iterations of $\overline{W}$ to obtain a Weinbaum factorization of w_n for a.

Let $\{F_i\}_{i\geq 1}$ where

$$F_0 = 1, \quad F_1 = 1, \quad \text{and} \quad F_{i+1} = F_i + F_{i-1} \quad (i \geq 1)$$

denote the Fibonacci numbers. We have $|w_n| = F_n$. We also observe that $\{f_i\}_{i\geq 0}$ is similar to the set $\{h_i\}_{i\geq 1}$ of Fibonacci words defined by

$$h_1 = b, \quad h_2 = a, \quad \text{and} \quad h_{i+1} = h_i h_{i-1} \quad (i \geq 2)$$

where we have $f_{2i} = bh^{\bullet}_{2i}$ and $f_{2i-1} = ah^{\bullet}_{2i-1}$, for all $i \geq 1$, where $x^{\bullet}$ denotes x without its last letter.

Let us start with some observations about $\{f_i\}_{i\geq 0}$.

Lemma 12. *The following holds for all $n \geq 2$ and $i \leq n$.*

1. $f_{n-2}f_{n-3}\cdots f_0 \leq_{\mathrm{p}} f_n$,

2. $f_i \leq_{\mathrm{p}} f_n \iff i \equiv n \pmod 2$,

3. $|f_n| = |f_{n-2}f_{n-1}|$ *and* $f_n \neq f_{n-2}f_{n-1}$.

The next lemma shows that every f_i in w_n with $i+1 < n$ is a marker.

Lemma 13. *f_i does not intersect itself in f_n for all $n > i$.*

The next two lemmas show that $G(f_i) = \{f_{i-1}, f_{i+1}\}$ for all $i+1 < n$.

Lemma 14. *f_i does not occur in f_{i+1}.*

Lemma 15. *If $f_i g f_i$ occurs in f_n such that f_i is not a factor of g, then we have $g \in \{\varepsilon, f_{i-1}, f_{i+1}\}$.*

Consider now w_n for some fixed n. Proposition 16 follows straightforwardly from Lemma 15.

Proposition 16. *$W(f_i) = \{f_{i+1}\}$, for all $1 \leq i < n$, and $W(f_n) = \{f_{n-1}\}$.*

Accordingly $\overline{W}$ is defined by $\overline{W}(f_i) = f_{i+1}$ and $\overline{W}(f_n) = f_{n-1}$, and hence, $w_n = \overline{W}^{n-1}(a)\overline{W}^{n-2}(a)$. We have

$$|w_n| = \left[\frac{\phi^n}{\sqrt{5}}\right]$$

where $\phi = (1+\sqrt{5})/2$ is the golden ratio and $[x]$ denotes the nearest integer of x. Hence, we need $\mathcal{O}(\log |w_n|)$ many steps to reach a Weinbaum factorization of w_n for a.

5 Weinbaum Factorizations for Long Factors

In this section we investigate some cases for which a factor f a word w does or does not admit a Weinbaum factorization.

Proposition 17. *Let $f \in A^*$, $a \in A$ and $|A| \geq 2$. Let m be the maximum exponent such that a^m occurs in f. Then $w = fa^n$ admits a Weinbaum factorization for f if and only if $f \notin aA^* \cup A^*a$ and $n > m$*

The following proposition is well-known to follow from the Fine-Wilf Theorem; see for example [1].

Proposition 18. *Let $f \in A^*$ be a nonempty word. Then there is at most one letter a in A such that fa is not primitive.*

Weinbaum's theorem states that every primitive word w admits a Weinbaum factorization for all letters a that occur in w. Moreover, we have seen in Corollary 5 that this is true even for all a^m, with $m \geq 0$, that occur in w. However, the next observation follows straightforwardly from Propositions 17 and 18 and shows that this is the best we can expect.

Observation 19. *Let $f \in A^*$ be a nonempty word where the letters $a_1, \ldots, a_k$ occur for $k \geq 2$. Then at least $k-1$ of the words fa_i are primitive, but none of them admits a Weinbaum factorization for f.*

By definition, Weinbaum factorizations for given words f and g can exist only if $(fA^* \cap A^*f) \setminus (gA^* \cup A^*g) \neq \emptyset$ and $(gA^* \cap A^*g) \setminus (fA^* \cup A^*f) \neq \emptyset$. In the following we call a pair (f, g) satisfying these conditions a *Weinbaum candidate* for short. For all Weinbaum candidates there exist Weinbaum factorizations. In fact this is not a rare event at all, they exist in all long enough random words. This is the gist of the next proposition.

Proposition 20. *Let $k, m \geq 1$ be constants. Denote by $\Pr(n, k, m)$ the probability that a word w of length n (under the uniform distribution) is primitive and that it admits for all Weinbaum candidates (f, g) with $|fg| \leq m$ at least k different Weinbaum factorizations for f and g. Then $\Pr(n, k, m)$ converges exponentially fast to 1 if n tends to infinity.*

References

[1] T. Harju, V. Halava, and L. Ilie. Periods and binary words. *J. Combin. Theory Ser. A*, 89:298–303, 2000.

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