TUCS

Tero Harju | Dirk Nowotka

Periods in Extensions of Words

Turku Centre *for* Computer Science

TUCS Technical Report
No 736, January 2006

Periods in Extensions of Words

Tero Harju
Turku Centre for Computer Science and Department of Mathematics,
University of Turku, FIN-20014 Turku, Finland
harju@utu.fi
Dirk Nowotka
Institute of Formal Methods in Computer Science,
University of Stuttgart, D-70569 Stuttgart, Germany
dirk.nowotka@informatik.uni-stuttgart.de

TUCS Technical Report
No 736, January 2006

Abstract

Let $\pi(w)$ denote the minimum period of the word w. Let w be a primitive word with period $\pi(w) < |w|$, and z a prefix of w. It is shown that if $\pi(wz) = \pi(w)$, then $|z| < \pi(w) - \gcd(|w|, |z|)$. Detailed improvements of this result are also proven. As a corollary we give a short proof of the fact that if u, v, w are primitive words such that u^2 is a prefix of v^2, and v^2 is a prefix of w^2, then $|w| > 2|u|$. Finally, we show that each primitive word w has a conjugate $w' = vu$, where $w = uv$, such that $\pi(w') = |w'|$ and $|u| < \pi(w)$.

Keywords: combinatorics on words, Weinbaum factorization, critical points, bordered word, primitive words

TUCS Laboratory
Discrete Mathematics for Information Technology

1 Introduction

Various aspects of periodicity play a central rôle in combinatorics on words and its applications; see Lothaire's books [8, 9, 10]. The notion of periodicity is well posed in many problems concerning algorithmic aspects of strings: in pattern matching, compression of strings, sequence analysis, and so forth.

In this paper we study extensions of words with respect to their periodicity. Let w be a word over a finite alphabet A. The length of w is denoted by $|w|$. The empty word is denoted by ε. A positive integer p is a *period* of w, if $w = (uv)^k u$ where $p = |uv|$, $k \geq 1$, and $v \neq \varepsilon$. The minimum period of w is denoted by $\pi(w)$.

For a word $w = uv$, the word u is a *prefix* of w, denoted by $u \leq_{\mathrm{p}} w$, and v is a *suffix* of w, denoted by $v \leq_{\mathrm{s}} w$. If v is nonempty, then u is a *proper prefix* of w, denoted by $u <_{\mathrm{p}} w$. A nonempty word u is a *border* of w, if u is a prefix and a suffix of w, i.e., $ux = w = yu$ for some nonempty words x and y. Each word has a unique factorization in the form $w = u^k v$, where $k \geq 1$, $v <_{\mathrm{p}} u$ and $|u| = \pi(w)$. Here u is called the *root* of w and v the *residue* of w. We denote the length $|v| \geq 0$ of the residue v by $\rho(w)$.

A word is *primitive* if it is not a power of a shorter word, i.e., if $\pi(w)$ does not divide $|w|$ properly.

Let w be a word with a nonempty residue and a prefix $z \leq_{\mathrm{p}} w$. We show that if the word wz has the same minimum period as w, that is, $\pi(wz) = \pi(w)$, then $|z| < \pi(w) - \gcd(|w|, |z|)$, where gcd denotes the greatest common divisor function. As a corollary we give a short proof of the well known result due to Crochemore and Rytter [4] stating that if u, v, w are primitive words such that $u^2 <_{\mathrm{p}} v^2 <_{\mathrm{p}} w^2$, then $u^2 <_{\mathrm{p}} w$, i.e., $|w| > 2|u|$. Finally, we strengthen the above extension result by showing that if w is a word with u as a root and w has a nonempty residue, then $\pi(wz) > \pi(w)$ for all prefixes $z \leq_{\mathrm{p}} w$ with $|z| \geq \pi(w) + \pi(u) - \rho(w) - 1$.

In the last section, we study extensions wz that force the period $\pi(wz) = |w|$. This problem is stated for unbordered conjugates. For this, let $\tau(w)$ denote the *shortest prefix* of the word w, say $w = \tau(w)u$, such that the conjugate $u\tau(w)$ is unbordered, i.e., $\pi(u\tau(w)) = |u\tau(w)|$. We show that for each primitive word w it holds that $\tau(w) < \pi(w)$.

2 Extensions of words by periods

It is clear that if u is a border of a word w, then $|w| - |u|$ is a period of w, and thus $|w| - |u| \geq \pi(w)$. A word w is said to be *bordered* (or *self-correlated* [11]), if it has a border, that is, if w has a prefix of length less than $|w|$ which is also a suffix of w. If w is not bordered, it is called *unbordered*. Clearly, a word w is unbordered if and only if $\pi(w) = |w|$.

We begin with an application of the basic periodicity result of Fine and Wilf [6]:

Theorem 1 (Fine and Wilf). *If a word w has two periods p and q such that $|w| \geq p + q - \gcd(p, q)$, then also $\gcd(p, q)$ is a period of w.*

Note that if w has an empty residue, then $\pi(wz) = \pi(w)$ for all words $z = w^k u$ with $u \leq_{\mathrm{p}} w$ and $k \geq 0$. Therefore, in the sequel we consider words with nonempty residues. Note that each word w with a nonempty residue is primitive, and thus $\pi(w^2) = |w| > \pi(w)$.

Theorem 2. *Let w be a word with a nonempty residue and a prefix $z \leq_{\mathrm{p}} w$.*

$$\textit{If} \quad \pi(wz) = \pi(w) \quad \textit{then} \quad |z| < \pi(w) - \gcd(\pi(w), |w|)\,.$$

Proof. Clearly $\pi(wz) \geq \pi(w)$. Let $d = \gcd(\pi(w), |w|)$, and suppose that $z \leq_{\mathrm{p}} w$ satisfies $\pi(wz) = \pi(w)$. Then both $|w|$ and $\pi(w)$ are different periods of wz. If $|wz| \geq \pi(w) + |w| - d$, then Theorem 1 implies that d is a period of wz. In this case, $d = \pi(w)$, since $\pi(wz) \geq \pi(w)$, and so $\pi(w)$ divides $|w|$ contradicting primitivity of w; hence the claim follows. □ □

The following example shows that the bound given in Theorem 2 is optimal for all lengths.

Example 3. Consider the word

$$w = a^{n-1}ba$$

with the minimum period $\pi(w) = n$, and let $z = a^{n-2} \leq_{\mathrm{p}} w$. We have $\pi(wz) = n$, where $|z| = |w| - 3 = \pi(w) - \gcd(\pi(w), |w|) - 2$, since we have now $\gcd(n, n+1) = 1$.

The following example shows that the condition $|z| \geq \pi(w) - \gcd(\pi(w), |w|)$ does not imply that $\pi(wz) = |w|$.

Example 4. Consider the word

$$w = ababaabab\,.$$

Then $\pi(w) = |ababa| = 5$. Let $z = aba$. We have $|z| = \pi(w) - 2$ and

$$wz = ababa.abab.aba$$

with $\pi(w) = 5 < 7 = \pi(wz) < 9 = |w|$, since $|ababaab|$ is a period of wz.

The following result is due to Crochemore and Rytter [4]. A short proof due to Diekert is given in [9, Lemma 8.1.14]. Below we show that this result follows from Theorem 2. Note that an integer $p \leq |w|$ is a period of the word w if and only if $w \leq_{\mathrm{p}} xw$, where $x \leq_{\mathrm{p}} w$ is such that $|x| = p$.

Corollary 3. *Let u, v, w be primitive words with $u^2 <_{\mathrm{p}} v^2 <_{\mathrm{p}} w^2$. Then $|w| > 2|u|$.*

Proof. Suppose that $|w| \leq 2|u|$, and thus $w <_{\mathrm{p}} v^2 <_{\mathrm{p}} w^2$. Hence w has a nonempty residue. Let $w = vx$. Then $|x|$ is a period of v, since $vv \leq_{\mathrm{p}} ww = vxvx$ and so $v \leq_{\mathrm{p}} xv$. Now $\pi(v) \leq |x|$, and, by Theorem 2, $\pi(w) \geq |v|$, and so $\pi(w) = |v|$. However, also $|u|$ is a period of w, since $w <_{\mathrm{p}} u^2$. Therefore $|v| = \pi(w) = |u|$ gives a contradiction. □ □

For a word w with a nonempty residue, let its *maximal extension number* be defined by

$$\kappa(w) = \max\{p \mid p = |z| \text{ for a prefix } z \leq_{\mathrm{p}} w \text{ with } \pi(wz) = \pi(w)\}\,.$$

Theorem 2, $\kappa(w)$ exists and satisfies $\kappa(w) < \pi(w) - 1$. For a nonempty word w, let $w^\bullet$ denote the word from which the last letter is removed. For the proof of the following result, see Berstel and Karhumäki [1].

Lemma 4. *Let u and v be two nonempty words. If $uv^\bullet = vu^\bullet$ then there exists a word g such that $u = g^i$ and $v = g^j$ for some $i, j \geq 1$.*

We shall now have a partial improvement of Theorem 2.

Theorem 5. *Let w be a word with a nonempty residue and let u be the root of w. Then*

$$\kappa(w) \leq \pi(w) + \pi(u) - \rho(w) - 2\,.$$

Proof. Let $u = vy$ where $|v| = \rho(w)$, and let x be the root of u. Assume that there exists a prefix $z \leq_{\mathrm{p}} w$ such that $\pi(wz) = \pi(w)$ and $|z| = \pi(w) + \pi(u) - \rho(w) - 1 = |wu| - |v| - 1$. By Theorem 2, we have that $\pi(u) < \rho(w)$, and thus $x <_{\mathrm{p}} u$. Now, $|vz| = |ux| - 1$ and since $vz \leq_{\mathrm{p}} ux$, we have $vz = ux^\bullet = vyx^\bullet$, and thus $z = yx^\bullet$. Also, $z = xy^\bullet$, since $z \leq_{\mathrm{p}} u$ and $y <_{\mathrm{p}} u$, for, $y <_{\mathrm{p}} z <_{\mathrm{p}} u$ and x is the root of u. By Lemma 4, $yx^\bullet = xy^\bullet$ implies that there exists a primitive word g such that $x = g^i$ and $y = g^j$ for some $i, j \geq 1$. Then $v = g^{it} g_1$ for a prefix $g_1 <_{\mathrm{p}} g$ and an integer $t \geq 0$, and so $u = vy = g^{it} g_1 g^j$. However, since x is the root of u, $u = x^r x_1$ for some $r \geq 1$ and $x_1 <_{\mathrm{p}} x$, from which it follows that $u = g^{it+j} g_1$. In order for g to be primitive, we must have $j = 0$, for otherwise g is a proper conjugate of itself. This contradicts the fact that $j \geq 1$. □ □

The bound given in Theorem 5 is optimal as shown in the following example.

Example 5. Consider the words

$$w_n = (aba)^n ab$$

where $\pi(w_n) = 3$, $\pi(u) = 2$ for the root $u = aba$ of w_n, and $\rho(w_n) = 2$. Hence, $\kappa(w) = \pi(w_n) + \pi(u) - \rho(w_n) - 2 = 1$. Indeed, the extension $w_n ab$ has a larger period than 3, namely $\pi(w_n ab) = 3n + 2$.

Also, for

$$u_n = (ab)^n aab$$

of length $2n+3$, we have $\pi(u_n) = 2n+1$, and the length $\rho(u_n)$ of the residue of u_n is 2 . Hence, $\kappa(u_n) = 2n - 1 = \pi(u_n) + \pi((ab)^n a) - \rho(u_n) - 2$.

6 Critical points and extensions

Every primitive word w has an unbordered conjugate. For instance, consider the least conjugate of w with respect to some lexicographic ordering, that is, a Lyndon conjugate of w; see e.g. Lothaire [8]. Denote by $\tau(w)$ the *shortest prefix* of w, $w = \tau(w)u$, such that the conjugate $u\tau(w)$ is unbordered. Hence $0 \leq \tau(w) < |w|$.

Lemma 6. *Each primitive word w has a factorization $w = uv$ such that the conjugate vu is unbordered and either $|u| < \pi(w)$ or $|v| < \pi(w)$.*

Proof. Let $w = u^k z$, where u is the root of w, $k \geq 1$, and $z <_{\mathrm{p}} u$. Suppose that w has no conjugate as stated in the claim. Let $w' = yu^{k-i}zu^{i-1}x$ be an unbordered conjugate of w, where $u = xy$. (Take, for instance, a Lyndon conjugate of w.) It follows that $i = k$ or $i = 1$, for otherwise yx is a border of w'. If $i = 1$, then $w' = yu^{k-1}zx$ is a required conjugate: $w' = (yu^{k-1}z)(x)$. Assume then that $i = k$, we have $w' = yzu^{k-1}x$ and thus $z <_{\mathrm{p}} x$; otherwise again yx is a border of w'. However, now $w' = (yz)(u^{k-1}x)$ is a required conjugate. □ □

In the following we say that an integer p with $1 \leq p < |w|$ is a *point* in the word 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 is a suffix of $x'x$ and u is a prefix of yy'. Let

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

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

The Critical Factorization Theorem (CFT) is a fundamental result on periodicity. It was first conjectured by Schützenberger [12] and then proved by Césari and Vincent [2]. Later it was developed into its present form by Duval [5]. We refer to [7] for a short proof of the theorem giving a technically improved version of the proof by Crochemore and Perrin [3].

Theorem 7 (CFT). *Let w be a word with at least two different letters. Then w has a critical point p such that $p < \pi(w)$.*

The following lemma rests on the CFT.

Lemma 8. *Let w be an unbordered word with $|w| \geq 2$, and let $w = uv$ be such that $p = |u|$ is any critical point of w. Then also the conjugate vu is unbordered.*

Proof. Without loss of generality we can assume that $|u| \leq |v|$. Now $\pi(w) = |w|$, since w is unbordered. Assume, contrary to the claim, that the word vu is bordered. We have two cases to consider. (1) Assume that $v = sv'$ and $u = u's$ for a nonempty word s. Then $\pi(w, |u|) \leq |s| < |w|$ contradicting the assumption that $|u|$ is a critical point. (2) Assume that $v = sut$. Then $\pi(w, |u|) \leq |su| < |w|$, and again $|u|$ is not a critical point; a contradiction. These cases prove the claim. □ □

The following theorem states the main result of this section.

Theorem 9. *Let w be a primitive word. Then $\tau(w) < \pi(w)$.*

Proof. Suppose first that $\pi(w) > |w|/2$. Assume that $w = xyz$, where $|xy| = \pi(w)$, $z <_{\mathrm{p}} xy$, and $|x|$ is a critical point of w such that $|x| < \pi(w)$ provided by Theorem 7. Suppose that the conjugate $w' = yzx$ is bordered, and let u be its shortest border. Since $|x|$ is a critical point in w and u is a local repetition at $|x|$ in w, we have $|u| \geq \pi(w)$, and hence $|u| \geq |yx|$. Since u is unbordered, it does not overlap with itself, and therefore $|yzx| \geq 2|u|$, which implies that $|yzx| \geq 2|yx|$ and hence $|z| \geq |yx|$; a contradiction. Hence the conjugate $w' = yzx$ is unbordered, and so $\tau(w) < \pi(w)$.

Assume then that $\pi(w) < |w|/2$, and et u be the root of w. Then $w = u^k z$ where $\pi(w) = |u|$ and $z <_{\mathrm{p}} u$ and $k \geq 2$.

Assume that $\tau(w) \geq \pi(w)$, and thus that $\tau(w) > \pi(w)$. By Lemma 6, there exists an unbordered conjugate $w' = vu^{k-1}t$ of w, where $v \leq_{\mathrm{s}} w$ such that $|v| < \pi(w)$. Consider a critical point p of w', say $w' = gh$, where $|g| = p$.

First, v is a suffix of uz, and thus the critical point p is not in v, i.e., $p > |v|$, since $\pi(w') = |w'|$ and v occurs in $u^{k-1}t$. Similarly, $p < |vu|$, since all suffixes of w' starting from a position $q \geq |vu|$ occur in w' starting from the point $q - |u|$ and thus there is a local repetition at point q of length at most $|u|$. Now we have $|v| < |g| < |vu|$ and the conjugate hg is unbordered by Lemma 8. Let $u = rs$ such that $g = vr$. Then $hg = su^{k-1}zr$ and $1 \leq |r| < |u|$ as required. □ □

The following example illustrates that it is not enough to just consider critical points for proving Theorem 9.

Example 7. It is not true that a conjugate vu with respect to a critical point $|u|$ of $w = uv$ is unbordered. Consider for instance the word $w = abcbababcbabab$, where $\pi(w) = 6$, and $p = 3$ is a critical point, but the corresponding conjugate $w' = bababcbabababc$ has a border $bababc$.

Note that we always have $\pi(w^k z) \leq |w|$ for prefixes $z \leq_{\mathrm{p}} w$ and non-negative integers k. Theorem 9 gives a complementary result to Theorem 2 and 5.

Corollary 10. *Let w be a word with a nonempty residue and a prefix $z \leq_{\mathrm{p}} w$.*

$$\textit{If} \quad |z| \geq \pi(w) \quad \textit{then} \quad \pi(wz) = |w| \, .$$

Proof. Let $|z| \geq \pi(w)$. By Theorem 9, w has an unbordered conjugate $w' = vu$ where $w = uv$ and $|u| < \pi(w)$. Then we have $\pi(wu) = |w|$ for the extension wu, since $\pi(wu)$ is at least the length of the longest unbordered factor of wu. The claim follows now from $wu \leq_{\mathrm{p}} wz$. □

The following example elaborates on the differences between Theorem 2 and Corollary 10.

Example 8. Consider the word

$$w = aaabaa$$

for which $|w| = 6$ and $\pi(w) = 4$ and $\gcd(\pi(w), |w|) = 2$ so that we get $\pi(w) - \gcd(\pi(w), |w|) = 2$. We have $\pi(wz) > \pi(w)$ for each extension wz with $z \leq_{\mathrm{p}} w$ and $|z| \geq 2$, by Theorem 2. The shortest extension increasing the period is for $z = aa$, that is, $w.aa = aaabaaaa$ with $\pi(waa) = 5$.

However, we have $\pi(wz) < |w|$ and the corresponding conjugate $w' = abaaaa$ of w is bordered. In this example, we need an extension $z = aaa$ of length 3 in order to obtain $\pi(wz) = |w|$.

References

[1] Berstel, J., Karhumäki, J.: Combinatorics on words — A tutorial. Bull. EATCS **79**, 178–229 (2003)

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

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

[4] Crochemore, M., Rytter, W.: Squares, cubes, and time-space efficient string searching. Algorithmica **13**(5), 405–425 (1995)

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

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

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

[8] Lothaire, M.: Combinatorics on Words, *Encyclopedia of Mathematics*, vol. 17. Addison-Wesley, Reading, MA (1983. Reprinted in the Cambridge Mathematical Library, Cambridge Univ. Press, 1997)

[9] Lothaire, M.: Algebraic Combinatorics on Words, *Encyclopedia of Mathematics and its Applications*, vol. 90. Cambridge University Press, Cambridge, United Kingdom (2002)

[10] Lothaire, M.: Algorithmic Combinatorics on Words. Cambridge University Press, Cambridge, United Kingdom (2005)

[11] Morita, H., van Wijngaarden, A.J., Vinck, A.J.H.: On the construction of maximal prefix-synchronized codes. IEEE Trans. Inform. Theory **42**, 2158–2166 (1996)

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

Turku
Centre *for*
Computer
Science

Lemminkäisenkatu 14 A, 20520 Turku, Finland | 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 Sciences

ISBN 952-12-1661-1
ISSN 1239-1891