THE EHRENFEUCHT-SILBERGER PROBLEM

ŠTĚPÁN HOLUB AND DIRK NOWOTKA

Abstract. We consider repetitions in words and solve a longstanding open problem about the relation between the period and the length of its longest unbordered factor. A word u is called bordered if there exists a proper prefix that is also a suffix of u, otherwise it is called unbordered. In 1979 Ehrenfeucht and Silberger raised the following problem: What is the maximum length of a word w, w.r.t. the length τ of its longest unbordered factor still allowing that τ is shorter than the period π of w. We show that if w is longer than $7(\tau - 1)/3$ then $\tau = \pi$ which gives the optimal asymtotic bound.

Introduction

Combinatorial problems about repetitions lie at the core of algorithmic questions regarding strings (called words here), being it search, compression, or coding algorithms. Despite a long tradition of research many questions about the combinatorial properties of data structures as simple as words remain open. The focus of this paper is on the solution of such a question namely the problem by Ehrenfeucht and Silberger which had been open for about three decades.

When repetitions in words of symbols are considered then two notions are central: the *period*, which gives the least amount by which a word has to be shifted in order to overlap with itself, and the shortest *border*, which denotes the least (nontrivial) overlap of a word with itself. Both notions are related in several ways, for example, the length of the shortest border of a word w is not larger than the period of w, and hence, the period of an unbordered word is its length. Moreover, a shortest border itself is always unbordered. Deeper dependencies between the period of a word and its unbordered factors have been investigated and exploited in applications for decades; see also the references to related work below.

Let us recall the problem by Ehrenfeucht and Silberger. Let w be a (finite) word of length $|w|$, let $\tau(w)$ denote the length of the largest unbordered factor of w, and let $\pi(w)$ denote the period of w. Certainly, $\tau(w) \leq \pi(w)$ since the period of a factor of w cannot be larger than the period of w itself. Moreover, it is well-known that $\tau(w) = \pi(w)$ when $|w| \geq 2\pi(w)$. So, the interesting cases are those where $|w| < 2\pi(w)$. Actually, the interesting cases are also the most common ones. By far most words have a period that is longer than one half of their length. When such words are considered, a bound on $|w|$, enforcing $\tau(w) = \pi(w)$, that depends on $\tau(w)$ becomes more interesting than one depending on $\pi(w)$.

The problem by Ehrenfeucht and Silberger asks about a bound of $|w|$ depending on $\tau(w)$ such that $\tau(w) = \pi(w)$ is enforced. In this paper we establish the following

---

Date: February 2009.

The work on this article has been supported by the research project MSM 0021620839.

fact for all finite words w:

$$\text{If} \quad |w| > \frac{7}{3}(\tau(w) - 1) \quad \text{then} \quad \tau(w) = \pi(w) \ .$$

This bound on the length of w is asymtotically tight (see the following example).

Previous Work. Ehrenfeucht and Silberger raised the problem described above in [7]. They conjectured that $|w| \geq 2\,\tau(w)$ implies $\tau(w) = \pi(w)$. That conjecture was falsified shortly thereafter by Assous and Pouzet [1] by the following example:

$$w = a^n ba^{n+1} ba^n ba^{n+2} ba^n ba^{n+1} ba^n$$

where $n \geq 1$ and $\tau(w) = 3n+6$ and $\pi(w) = 4n+7$ and $|w| = 7n+10$, that is, $\tau(w) < \pi(w)$ and $|w| = 7/3\,\tau(w) - 4 > 2\tau(w)$. Assous and Pouzet in turn conjectured that $3\tau(w)$ is the bound on the length of w for establishing $\tau(w) = \pi(w)$. Duval [5] did the next step towards answering the conjecture. He established that $|w| \geq 4\,\tau(w) - 6$ implies $\tau(w) = \pi(w)$ and conjectures that, if w possesses an unbordered prefix of length $\tau(w)$, then $|w| \geq 2\,\tau(w)$ implies $\tau(w) = \pi(w)$. Note that a positive answer to Duval's conjecture yields the bound $3\,\tau(w)$ for the general question. Despite some partial results [12, 6, 8] towards a solution Duval's conjecture was only solved in 2004 [9, 10] with a new proof given in [11]. The proof of (the extended version of) Duval's conjecture lowered the bound for Ehrenfeucht and Silberger's problem to $3\,\tau(w) - 2$ as conjectured by Assous and Pouzet [1]. However, there remained a gap of $\tau(w)/3$ between that bound and the largest known example which is given above. The bound of $7\tau(w)/3$ has been conjectured in [9, 10]. This conjecture is proved in this paper, and the problem by Ehrenfeucht and Silberger is finally solved.

Other Related Work. The result related most closely to the problem by Ehrenfeucht and Silberger is the so called critical factorization theorem (CFT).

What is the CFT? Let $w = uv$ be a factorization of a word w into u and v. The local period of w at the point $|u|$ is the length q of the shortest square centered at $|u|$. More formally, let x be the shortest word such that x is a prefix of vy and a suffix of zu for some y and z, then $q = |x|$. It is straightforward to see that q is not larger than the period of w. The factorization uv is called critical if q quals the period of w. The CFT states that a critical factorization exists for every nonempty word w, and moreover, a critical factorization uv can always be found such that $|u|$ is shorter than the period of w. The CFT was conjectured first by Schützenberger [13], proved by Césari and Vincent [2], and brought into its current form by Duval [4]. Crochemore and Perrin [3] found a new elegant proof of the CFT using lexicographic orders, and realized a direct application of the theorem in a new string-matching algorithm.

How does the CFT relate to the problem by Ehrenfeucht and Silberger? Observe that the shortest square x^2 centered at some point in w is always such that x is unbordered. If x results from a critical factorization and occurs in w, then w contains an unbordered factor of the length of its period. Therefore, it follows from the CFT that $|w| > 2\pi(w) - 2$ implies $\tau(w) = \pi(w)$. This bound is asymptotically optimal. In this paper, we establish the asymptotically optimal bound on $|w|$ enforcing the equality $\tau(w) = \pi(w)$ in terms of $\tau(w)$ instead of $\pi(w)$. This rounds off the long lasting research effort on the mutual relationship between the two basic properties of a word w, that is $\tau(w)$ and $\pi(w)$.

1. Notation and Basic Facts

Let us fix a finite set A of letters, called alphabet, for the rest of this paper. Let A^* denote the monoid of all finite words over A including the *empty word* denoted by ε. Let $w = uv \in A^*$. Then $u^{-1}w = v$ and $wv^{-1} = u$. In general, we denote variables over A by a, b, c, and d and variables over A^* are usually denoted by f, g, h, r through z, and α through δ, and ξ including their subscripted and primed versions. The letters i through q are to range over the set of nonnegative integers.

Let $w = a_1a_2\cdots a_n$. The word $a_na_{n-1}\cdots a_1$ is called the *reversal* of w denoted by $\overline{w}$. We denote the length n of w by $|w|$, in particular $|\varepsilon| = 0$. Let $0 \le i \le n$. Then $u = a_1a_2\cdots a_i$ is called a *prefix* of w, denoted by $u \le_{\mathrm{p}} w$, and $v = a_{i+1}a_{i+2}\cdots a_n$ is called a *suffix* of w, denoted by $v \le_{\mathrm{s}} w$. A prefix or suffix is called *proper* when $0 < i < n$. The *longest common prefix* w of two words u and v is denoted by $u \wedge_{\mathrm{p}} v$ and is defined by $w = u$, if $u \le_{\mathrm{p}} v$, or $w = v$, if $v \le_{\mathrm{p}} u$, or $wa \le_{\mathrm{p}} u$ and $wb \le_{\mathrm{p}} v$ for some different letters a and b. The *longest common suffix* of u and v, denoted $u \wedge_{\mathrm{s}} v$, is definied similarly, as one would expect. An integer $1 \le p \le n$ is a *period* of w if $a_i = a_{i+p}$ for all $1 \le i \le n - p$. The smallest period of w is called *the period* of w, denoted by $\pi(w)$. A nonempty word u is called a *border* of a word w, if $w = uy = zu$ for some nonempty words y and z. We call w *bordered*, if it has a border, otherwise w is called *unbordered*. Let $\tau(w)$ denote the maximum length of unbordered factors of w, and $\tau_2(w)$ denote the maximum length of unbordered factors occurring at least twice in w. We have that

$$\tau(w) \le \pi(w) \ .$$

Indeed, let $u = b_1b_2\cdots b_{\tau(w)}$ be an unbordered factor of w. If $\tau(w) > \pi(w)$ then $b_i = b_{i+\pi(w)}$ for all $1 \le i \le \tau(w) - \pi(w)$ and $b_1b_2\cdots b_{\tau(w)-\pi(w)}$ is a border of u; a contradiction.

Let $\lhd$ be a total order on A. Then $\lhd$ extends to a *lexicographic order*, also denoted by $\lhd$, on A^* with $u \lhd v$ if either $u \le_{\mathrm{p}} v$ or $xa \le_{\mathrm{p}} u$ and $xb \le_{\mathrm{p}} v$ and $a \lhd b$. Let $\overline{\lhd}$ denote a lexicographic order on the reversals, that is, $u \mathrel{\overline{\lhd}} v$ if $\overline{u} \lhd \overline{v}$. Let $\lhd^a$ and $\lhd_b$ denote lexicographic orders where the maximum letter a or the minimum letter b is fixed in the respective orders on A. A $\lhd$-*maximum prefix* (*suffix*) α of a word w is defined as a prefix (suffix) of w such that $v \mathrel{\overline{\lhd}} \alpha$ ($v \lhd \alpha$) for all $v \le_{\mathrm{p}} w$ ($v \le_{\mathrm{s}} w$).

The notions of maximum pre- and suffix are symmetric. It is general practice that facts involving the maximum ends of words are mostly formulated for maximum suffixes. The analogue version involving maximum prefixes is tacitly assumed.

The following remarks state some facts about maximum suffixes which are folklore. They are included in this paper to make it self-contained.

Remark 1.1. *Let w be a bordered word. The shortest border u of w is unbordered, and $w = uzu$. The longest border of w has length equal to $|w| - \pi(w)$.*

Indeed, if u is a border of w, then each border of u is also a border of w. Therefore u is unbordered, and it does not overlap with itself. If v is a border of w then $|w|-|v|$ is a period of w. Conversely, the prefix of w of length $|w| - \pi(w)$ is a border of w.

Remark 1.2. *Any maximum suffix of a word w occurs only once in w and is longer than $|w| - \pi(w)$.*

Indeed, let α be the $\lhd$-maximum suffix of w for some order $\lhd$. Then $w = x\alpha y$ and $\alpha \lhd \alpha y$ implies $y = \varepsilon$ by the maximality of α. If $w = uv\alpha$ with $|v| = \pi(w)$, then $u\alpha \le_{\mathrm{p}} w$ gives a contradiction again.

Remark 1.3. *Let α be its own maximum suffix w.r.t. some order $\lhd$, and let x be a prefix of α of length $\pi(\alpha)$. Then x is unbordered.*

Indeed, suppose on the contrary that x is bordered, that is, $x = ghg$ for some nonempty g. Let $\alpha = xy$. We have $gy \lhd \alpha$, by assumption, which implies $y \lhd hgy$. Note that gy is not a prefix of α otherwise $|gh| < |x|$ is a period of α contradicting the choice of x. Hence, $za \leq_{\mathrm{p}} y$ and $zb \leq_{\mathrm{p}} hgy$ for some different letters a and b with $a \lhd b$. But, $y \leq_{\mathrm{p}} \alpha$, since $|x| = \pi(w)$, implies $za \leq_{\mathrm{p}} \alpha$ which contradicts the maximality of α (since $za \leq_{\mathrm{p}} \alpha \lhd zb \leq_{\mathrm{p}} hgy$).

Let an integer q with $0 \leq q < |w|$ be called *point* in w. A nonempty word x is called a *repetition word* at point q if $w = uv$ with $|u| = q$ and there exist words y and z such that $x \leq_{\mathrm{s}} yu$ and $x \leq_{\mathrm{p}} vz$. Let $\pi(w,q)$ denote the length of the shortest repetition word at point q in w. We call $\pi(w,q)$ the *local period* at point q in w. Note that the repetition word of length $\pi(w,q)$ at point q is necessarily unbordered and $\pi(w,q) \leq \pi(w)$. A factorization $w = uv$, with $u, v \neq \varepsilon$ and $|u| = q$, is called *critical*, if $\pi(w,q) = \pi(w)$, and if this holds, then q is called a *critical point*. Let $\lhd$ be an order on A and $\blacktriangleleft$ be its inverse. Then the shorter of the $\lhd$-maximum suffix and the $\blacktriangleleft$-maximum suffix of some word w is called a *critical suffix* of w. Similarly, we define a *critical prefix* of w by the shorter of the two maximum prefixes resulting from some order and its inverse. This notation is justified by the following formulation of the so called critical factorization theorem (CFT) [3] which relates maximum suffixes and critical points.

Theorem 1.4 (CFT)**.** *Let w be a nonempty word and γ be a critical suffix of w. Then $|w| - |\gamma|$ is a critical point.*

Remark 1.5. *Let rs be an unbordered word where $|r|$ is a critical point. Then s and r do not overlap and sr is unbordered with $|s|$ as a critical point.*

Let us highlight the following definitions. They are not standard but will be central in the proof of Theorem 2.1.

Definition 1.6. *Let words g and w be given. The longest prefix g shorter than g that is also a suffix of w will be called the g-*suffix *of w.*

*The number $|ws^{-1}|$, where s is the g-suffix of w, is called the g-*period *of w, denoted by $\pi_g(w)$.*

*The shortest prefix w' of w satisfying $\pi_g(w') = \pi_g(w)$ is called the g-*critical prefix *of w.*

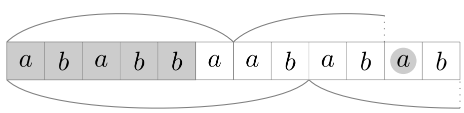
Remark 1.7. *Note that zd, where d is a letter, is the g-critical prefix of w if and only if zd is the longest prefix of w satisfying $\pi_g(z) < \pi_g(zd)$.*

Example 1.8. *Consider $w = ababbaababab$ of length 12 and $g = ababb$. The g-suffix of w is $abab$, whence $\pi_g(w) = 8$. The g-critical prefix of w is $ababbaababa$ of length 11, since*

$$\pi_g(ababbaababa) = 8\,,$$

and

$$\pi_g(ababbaabab) = 6\,.$$



Note that, by definition, the g-suffix of w can be empty, but it cannot be equal to g. For example, the abb-suffix of $aabb$ is empty. Therefore, the abb-critical prefix of $aabb$ is $aabb$ itself.

2. Solution of the Ehrenfeucht-Silberger Problem

This entire section is devoted to the proof of the main result of this paper: the solution of the Ehrenfeucht-Silberger problem.

Theorem 2.1. *Let $w \in A^*$. If $|w| > \frac{7}{3}(\tau(w) - 1)$ then $\tau(w) = \pi(w)$.*

We identify two particular unbordered factors w and show that the assumption of the theorem, namely that these factors are strictly smaller than $\frac{3}{7}|w| + 1$, leads either to a contradiction or to $\tau(w) = \pi(w)$.

Note that the claim holds trivially if every letter in w occurs only once because $\tau(w) = \pi(w) = |w|$ holds in that case. Let

$$w = v'uzuv$$

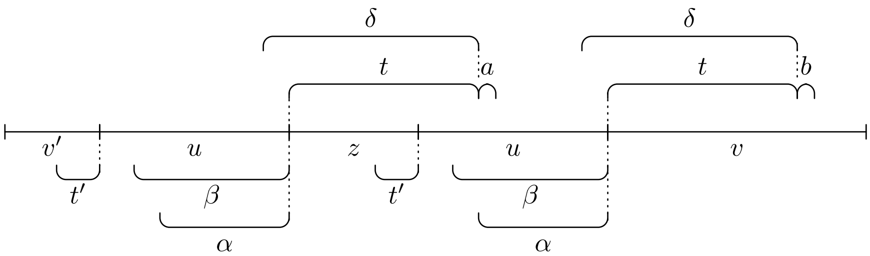
such that $|u| = \tau_2(w)$ and z is of maximum length. It is clear that such a factorization exists whenever a letter occurs more than once in w. Based on such a factorization of w we fix some more notation for the rest of this proof. Let

$$t = v \wedge_{\mathrm{p}} zu \quad \text{and} \quad t' = v' \wedge_{\mathrm{s}} uz \ .$$

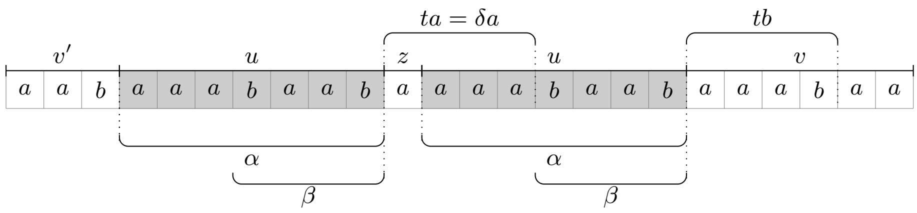
If $t \neq v$, then let

- ta be a prefix of zu and tb be a prefix of v with $a \neq b$,
- δa be the $\lhd^a$-maximum suffix of $t'uta$ for some fixed order $\lhd^a$ such that a is the maximum in A,
- α be the $\lhd^a$-maximum suffix of $t'u$, and
- β be the $\blacktriangleleft_a$-maximum suffix of $t'u$ where $\blacktriangleleft_a$ is the inverse order of $\lhd^a$.

The notation introduced so far is exemplified by the following figure where we assume that $t \neq v$ and $t' \neq v'$ and $|t'| < |z| < |t| < |\delta| < |\alpha t|$ and $|\alpha| < |\beta| < |u|$.



The example of long words where the period exceeds the length of the longest unbordered factors by Assous and Pouzet (see page 2) turns out to highlight the most interesting cases of this proof. We therefore use it as a running example throughout this section. The notation introduced above applied to a word of Assous and Pouzet is illustrated by the following figure. In this case t' is empty.



We can suppose w.l.o.g. that v' is as short as possible. This in particular implies the following claim.

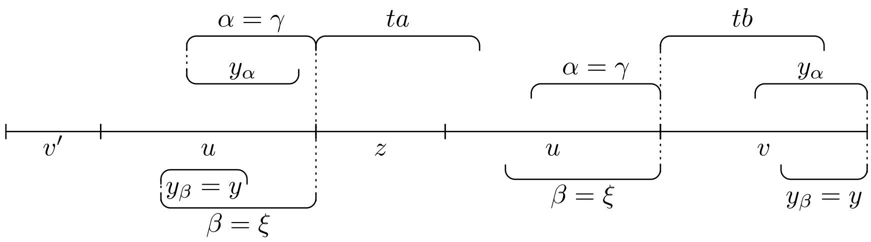
Claim 2.2.

$$|\alpha| \le |u| \quad \textit{and} \quad |\beta| \le |u| \ . \tag{1}$$

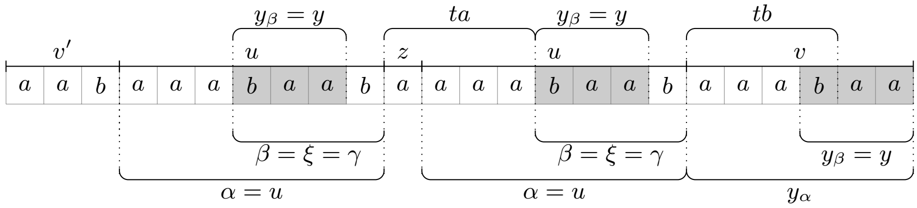
Proof. If α is longer than u, then the prefix $\hat{u}$ of α of length $\pi(\alpha)$ is unbordered by Remark 1.3. It is of length at least $|u|$, otherwise u is bordered. From $|u| = \tau_2(w)$ follows $|\hat{u}| = |u|$ since $\hat{u}$ occurs at least twice in w. We have a factorization $\hat{v}'\hat{u}\hat{z}\hat{u}\hat{v}$ of w where $\hat{v}' = v'u\alpha^{-1}$ and $|\hat{z}| = |z|$ and $\hat{v} = \hat{u}^{-1}\alpha v$; contradicting the minimality of $|v'|$. □ □

2.1. **The First Factor.** In this subsection we describe, using the factorization introduced above, a particular factor of w, which is likely to be unbordered and long; see the factor $uzuv_0d$ in the proof of Claim 2.4 below. The basic assumption of our proof, namely that there are no too long unbordered factors, will yield important additional restrictions on w.

Let γ denote the shorter of α and β, and let y_α and y_β denote the α- and β-suffix of uv for the rest of this proof. Moreover, let y be the shorter of y_α and y_β and let ξ be either α or β so that $y = y_\xi$. The following figure shall illustrate the considered setting by an example where $|\alpha| < |\beta|$ and $|y_\alpha| > |y_\beta|$, that is, we have $\gamma = \alpha$, $y = y_\beta$ and $\xi = \beta$.



The same situation for our running example is depicted next.



We prove the following property of γ first.

Claim 2.3. *If $v_0\gamma$ is a prefix of γv with $v_0 \neq \varepsilon$, then $uzu\gamma^{-1}v_0\gamma$ is unborderd.*

Proof. Suppose on the contrary that $uzu\gamma^{-1}v_0\gamma$ has a shortest border h. Note that h is, like every shortest border of a factor in w, not longer than $|u| = \tau_2(w)$. In fact $|h| < |u|$ since $|h| = |u|$ contradicts the maximality of $|z|$. If $|\gamma| < |h| < |u|$ then γ occurs more than once in u contradicting Remark 1.2. And finally, if $|h| \leq |\gamma|$ then u is bordered by h since then $h \leq_{\mathrm{s}} \gamma \leq_{\mathrm{s}} u$; a contradiction which concludes the proof. □ □

We shall now consider the ξ-critical prefix of w in order to prove the following inequalities.

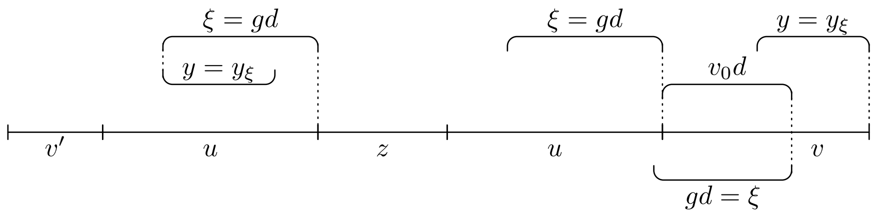
Claim 2.4.

$$|v'| < |u| \quad and \quad |v| < |u| \quad and \quad |v| \leq |ty|.$$

Proof. Within this proof suppose w.l.o.g. that $|v'| \leq |v|$. Note, the assumption that v' is as short as possible does not harm generality.

The claim is trivial if $|y| \geq |v|$. We therefore suppose that the ξ-critical prefix of w can be written as $v'uzuv_0d$, where d is a letter. We let g denote the ξ-suffix of $v'uzuv_0$.

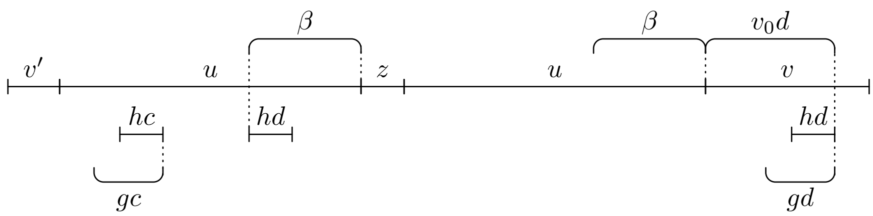
Assume first that $gd = \xi$ as illustrated in the next figure.



Then the word $uzuv_0d$ is unbordered, by Claim 2.3. From $|v| < |v_0d\xi|$ we obtain $|uzuv_0d| > |zuv|$. Therefore $|uzuv_0d| \geq |w|/2+1 > \frac{3}{7}|w|+1$; a contradiction. This implies that gc is a prefix of ξ with $c \neq d$.

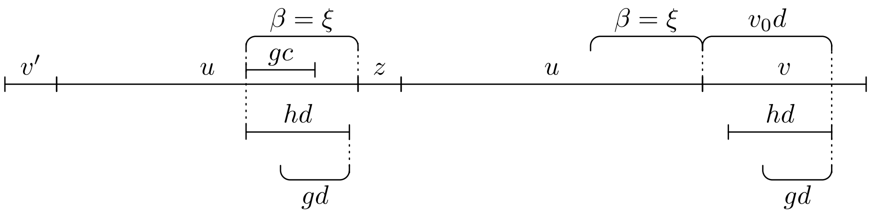
Suppose $c \lhd^a d$ and consider βzuv_0d. Since $|\beta zuv_0d| > |zuv|$, it must be bordered, as above. Let hd be its shortest border. We proceed by a case distinction on the length of h.

Suppose $|h| \leq |g|$ as illustrated in the next figure.



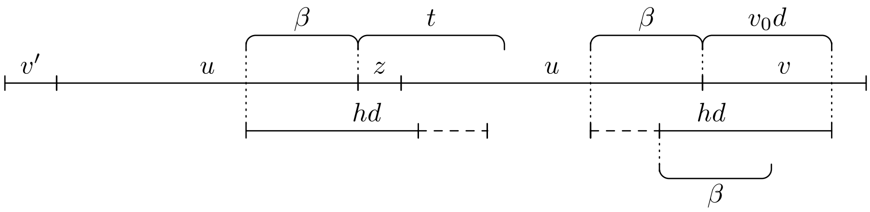
Then hd is a prefix of β and the occurrence of $hc \leq_{\mathrm{s}} gc$ in ξ, and hence also in β, contradicts the maximality of β since $hd \blacktriangleleft_a hc$.

Suppose $|g| < |h| < |\beta|$ as illustrated in the next figure.



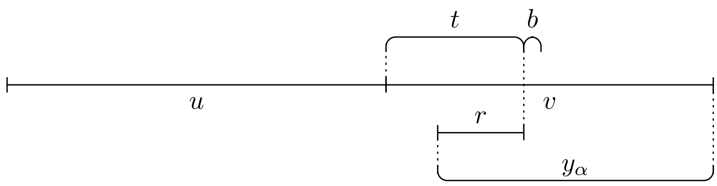
Then gd occurs in u and $\xi = \beta$ since $gd \blacktriangleleft_a gc$. Therefore h contradicts the assumption that g is the ξ-suffix of $v'uzuv_0$.

It remains that $|h| \geq |\beta|$ which implies $\beta \leq_{\mathrm{p}} h$ as illustrated next.



The choice of u implies $|h| < |u|$, whence either $h = \beta v_0$ or the word $uzuv_0h^{-1}\beta$ is unbordered, by Claim 2.3. From $|h| < |u|$ we have $|uzuv_0h^{-1}\beta| > |zuv| > |w|/2 + 1$. Therefore, $h = \beta v_0$, which implies $v_0d \leq_{\mathrm{p}} t$, and $|v| \leq |ty|$. The remaining inequalities follow from $|\beta v| \leq |\beta v_0 dy| = |hdy| < |u\beta|$, where the last inequality uses $|hd| \leq |u|$ and $|y| \leq |y_\beta| < |\beta|$. The possibility $d \triangleleft^a c$ is similar considering αzuv_0d. $\square$ $\square$

Suppose that $v \neq t$. Recall that tb is a prefix of v, and ta a prefix of zu. From $|v| \leq |ty|$, and from $|y| \leq |y_\alpha|$ we deduce that utb and y_α have in uv an overlap rb where r is nonempty. In other words, r is a suffix of ut such that $uv = utr^{-1}y_\alpha$.

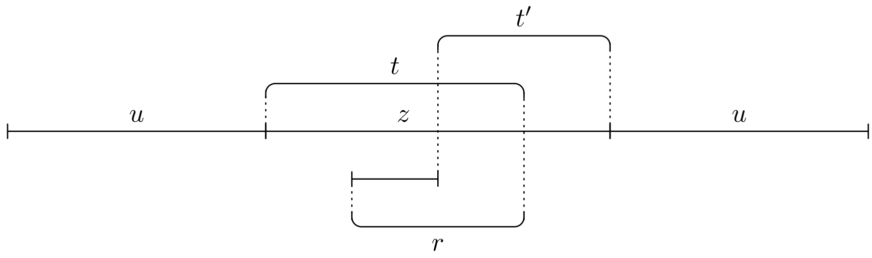


Since $|y_\alpha| < |\alpha|$, we have

$$|t| > |v| - |\alpha| + |r| \ . \tag{2}$$

The word rb is a prefix of α, and ra is a suffix of uta, which is a prefix of uzu. The maximality of α implies that ra is not a factor of $t'u$, and thus

$$|r| > |t| + |t'| - |z| \ . \tag{3}$$



2.2. **The Second Factor.** Let us now turn our attention to the word δ. In particular, we consider the factor $\delta t^{-1}zuv_\delta d$ as defined below in the proof of Claim 2.6.

The following claim points out that every factor of $t'uv$ is strictly less than δa w.r.t. $\lhd^a$. In particular, δa does not occur in $t'uv$.

Claim 2.5. *Let f be a factor of $t'uv$. Then $f \lhd^a \delta a$ and $f \neq \delta a$.*

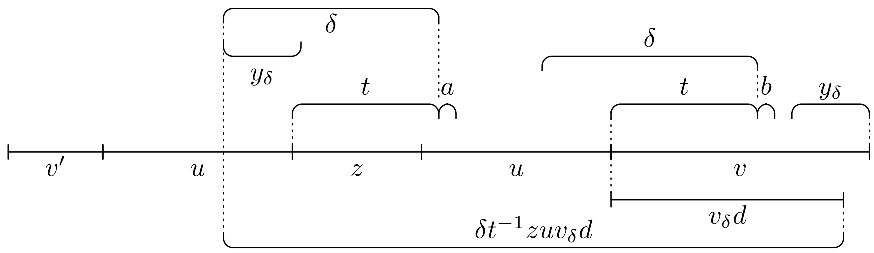
Proof. Suppose on the contrary that there exists a factor f of $t'uv$ such that $\delta a \lhd^a f$. Note that the maximality of δ is contradicted, if f occurs in $t'ut$ or y_α. Therefore, we have that there exists a prefix $f'b$ of f such that $f' \leq_{\mathrm{s}} t'ut$. But, we have $f'a \leq_{\mathrm{s}} t'uta$, and hence, $f'a \lhd^a \delta a$. The contradiction follows now from $f'b \lhd^a f'a$. □ □

Let y_δ denote the δa-suffix of w. Note that

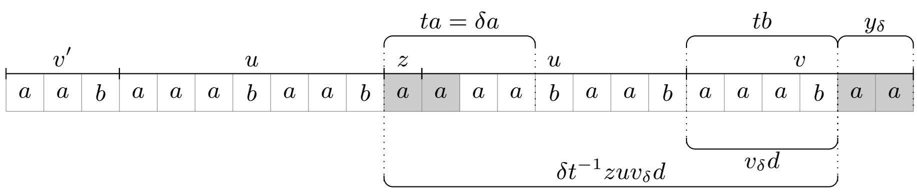
$$|y_\delta| < |v| - |t|, \tag{4}$$

since otherwise there is a suffix t_0 of $t'ut$ such $t_0 b$ is a prefix of y_δ, and $t_0 a$ is a suffix of $t'uta$ contradicting the maximality of δ.

Consider the following figure which already gives an illustration of the factor $\delta t^{-1}zuv_\delta d$ that will be defined below in the proof of Claim 2.6.



Our running example gives the following setting, with $d = b$.



Claim 2.6.

$$|\delta| - |t| + |z| + |u| + |v| - |y_\delta| < \frac{3}{7}|w| + 1 \tag{5}$$

Proof. Let $\delta t^{-1} zuv_\delta d$ be the δa-critical prefix of $v'uzuv$ and let g denote the δa-suffix of $\delta t^{-1} zuv_\delta$. This implies that $|v_\delta d| \geq |v| - |y_\delta|$ (in particular, $|v_\delta| \geq |t|$) and that gc is a prefix of δa and $c \neq d$. Note that $d \lhd^a c$, since all factors of uv are less than δa w.r.t. $\lhd^a$ by Claim 2.5.

Finally, we claim that $\delta t^{-1} zuv_\delta d$ is unbordered. Indeed, suppose on the contrary that there exists a shortest border hd. Since δa does not occur in uv and hd has to be shorter than u, we deduce that $|h| < |\delta|$. The maximality of g implies that $|h| \leq |g|$. But now hd is a suffix of gd whence hc is a factor of δa; a contradiction, since $hd \leq_{\mathrm{p}} \delta$ and $hd \lhd^a hc$. □ □

2.3. **Case Analysis.** In order to complete the proof we distinguish the following cases.

A Special Case. Consider first the special case where $t \neq v$ and z is empty and $u = \alpha$. It is not difficult to see that $uutb$ is unbordered.

Indeed, suppose that hb is the shortest border of $uutb$. The choice of u implies $|h| < |u|$, and since u is unbordered, we have $h \leq_{\mathrm{s}} t$. Now hb is a prefix of u and ha a factor of u; a contradiction to the maximality of α.

By $|uutb| \geq |w|/2 + 1 > \frac{3}{7}|w| + 1$ we can exclude this case.

Case 1. Let now either $t = v$ or $t' = v'$ but not both. By symmetry, we can suppose $t \neq v$ and $t' = v'$. Note that the assumption that v' is as short as possible does not harm the symmetry.

We are now going to show that the inequalities we have obtained in the previous subsections do not have a common solution. It is an exercise in the application of the simplex algorithm.

Inequality (3) can be transformed into

$$\mathtt{L}_1 := |r| - |t| - |t'| + |z| - 1 \geq 0. \tag{6}$$

Inequalities (4) and (5) imply,

$$|\delta| + |z| + |u| < \frac{3}{7}|w|,$$

which together with $|\delta| \geq |r|$ and $|w| = |v| + |v'| + 2|u| + |z|$ yield

$$\mathtt{L}_2 := 3|v'| + 3|v| - |u| - 4|z| - 7|r| > 0. \tag{7}$$

Similarly, if we use (5) with $|\delta| \geq |y_\delta|$, we obtain

$$\mathtt{L}_3 := 7|t| - 4|v| + 3|v'| - |u| - 4|z| + 7 > 0. \tag{8}$$

One can now check that under the assumption $|v'| = |t'|$ we have

$$28\,\mathtt{L}_1 + 4\,\mathtt{L}_2 + 3\,\mathtt{L}_3 + 7\,|v'ut| = -7,$$

a contradiction.

Case 2. Suppose $t \neq v$ and $t' \neq v'$. By the special case above, we can assume

$$\mathtt{L}_0 := |u| - |\alpha| + |z| - 1 \geq 0. \tag{9}$$

Inequality (2) can be transformed into

$$\mathtt{L}_4 := |t| + |\alpha| - |r| - |v| - 1 \geq 0, \tag{10}$$

The symmetry of v and v' yields, as a mirror variant of (8), the inequality

$$\mathtt{L}_5 := 7|t'| + 3|v| - 4|v'| - |u| - 4|z| + 7 > 0. \tag{11}$$

One can now check that

$$7\,\mathrm{L}_0 + 21\,\mathrm{L}_1 + 2\,\mathrm{L}_2 + 2\,\mathrm{L}_3 + 7\,\mathrm{L}_4 + 3\,\mathrm{L}_5 = 0,$$

again a contradiction.

Now, we have already proved that if $\tau(w) < \frac{3}{7}|w| + 1$, then v is a prefix of zu, and v' is a suffix of uz. It remains to consider this one more case.

Case 3. Let $t = v$ and $t' = v'$. Then $\pi(w) \leq |uz|$. Clearly, we can suppose that $\pi(w) > |u|$, since otherwise trivially $\pi(w) = \tau(w) = |u|$. Let $w = rs$ be a critical factorization of u. Then szr is unbordered of length $\pi(w)$, unless r is a prefix, and s is a suffix of z; see Remark 1.5. Suppose the latter possibility. Now, either one of the words uz and zu is unbordered of length $\pi(w)$ or u is both prefix and suffix of z. We are therefore left with the case $w = v'u^iz'u^jv$, with $i, j \geq 2$, where u is not a suffix of uz' and not a prefix of $z'u$. Moreover, v' is a suffix of u and v is a prefix of u. The assumption $\pi(w) > |u|$ now implies that z' is nonempty. Suppose, without loss of generality, $i \leq j$.

Similarly as above, we have that either $sz'u^{j-1}r$ or $z'u^j$ is unbordered. From $|u^jz'| < \frac{3}{7}|w| + 1$ we deduce

$$|v'v| > \left(\frac{4}{3}j - i\right)|u| + \frac{4}{3}|z'| - \frac{7}{3} \geq \left(\frac{4}{3}j - i\right)|u| - 1\ . \tag{12}$$

Case 3.1: $i = j$. If v' is a suffix of uz' and v a prefix of $z'u$, then we have $\pi(w) = \tau(w) = |z'u^j|$. Otherwise we obtain from Case 1 and Case 2 an unbordered factor of $v'uz'uv$ of length at least $\frac{3}{7}|v'uz'uv| + 1$. Moreover, this factor contains u as a factor, which can be substituted with u^j to obtain an unbordered factor of w of length at least $\frac{3}{7}|v'u^jz'u^jv| + 1$.

Case 3.2: $i < j$. Since $j \geq 3$, we obtain from (12) that $|v'v| \geq 2|u| - 1$; a contradiction with Claim 2.4.

This concludes the proof of Theorem 2.1.

References

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

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

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

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

[5] 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.

[6] J.-P. Duval, T. Harju, and D. Nowotka. Unbordered factors and Lyndon words. *Discrete Math.*, 308(11):2261–2264, 2008. (submitted in 2002).

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

[8] T. Harju and D. Nowotka. Minimal Duval extensions. *Internat. J. Found. Comput. Sci.*, 15(2):349–354, 2004.

[9] T. Harju and D. Nowotka. Periodicity and unbordered words. In *STACS 2004 (Montpellier)*, volume 2996 of *Lecture Notes in Comput. Sci.*, pages 294–304, Berlin, 2004. Springer-Verlag.

[10] T. Harju and D. Nowotka. Periodicity and unbordered words: A proof of the extended Duval conjecture. *J. ACM*, 54(4), 2007.

[11] Š. Holub. A proof of the extended Duval's conjecture. *Theoret. Comput. Sci.*, 339(1):61–67, 2005.

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

[13] 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, Amsterdam, 1979. North-Holland.

Department of Algebra, Charles University of Prague, Czech Republic
E-mail address: holub@karlin.mff.cuni.cz

Institute for Formal Methods in Computer Science, Universität Stuttgart, Germany
E-mail address: nowotka@fmi.uni-stuttgart.de