Unbordered words and (local) periods

We say that the word w is *bordered* if there exists a nonempty word $u \neq w$ that is both a prefix and a suffix of w. Such a word u is called a *border* of w. If u is itself bordered with a border z, then z is also a border of w. This implies that the shortest border of a bordered word is not bordered. If a border u of w is longer than $|w|/2$, then u overlaps itself, hence it is bordered. This yields that each bordered word is of the form uvu, $u \neq \varepsilon$. Moreover, it is easy to see that $|w| - |u|$ is a period of w. If we denote the least period of w as $\pi(w)$, we have that $\pi(w) = |w| - |u|$ holds for a bordered word w with the longest border u. Thus a word w is *unbordered* (that is, not bordered) if and only if $\pi(w)$ is maximal possible, namely $|w|$.

Lyndon words. A primitive word is not an integer power of a shorter word. On the other hand, any bordered word is a rational power of a shorter word, e.g. $aabaaba = (aab)^{\frac{7}{3}}$. We shall show that each primitive word is conjugate of an unbordered word. There can be more unbordered conjugates but one of them is always so called *Lyndon word*, which is defined as the primitive word which is the minimal element of its conjugacy class (that is the equivalence class of mutually conjugate words) with respect to some lexicographic order $\lhd$.

Recall that all words conjugate with a primitive word are themselves primitive. The primitivity implies that the Lyndon word is obtained by a unique conjugation: if uv is primitive, then $uv \neq vu$.

We observe several properties of lexicographic orders. Such an order is given by a linear order on letters. If $\lhd$ and $\blacktriangleleft$ are lexicographic orders given by inverse orders on letters (that is, $a \lhd b$ iff $b \blacktriangleleft a$), then for two prefix incomparable words we also have $u \lhd v$ iff $v \blacktriangleleft u$. However, if u is a prefix of v, them the orders behave in the same way: $u \lhd v$ and $u \blacktriangleleft v$. Further, for each z, we have $u \lhd v$ iff $zu \lhd zv$. If u and v are prefix incomparable (in particular, if they are of the same length), we also have $u \lhd v$ iff $uz \lhd vz'$ for each z and z'.

Theorem. Each Lyndon word is unbordered.

Proof. By contradiction. Let $w = uvu$ be a Lyndon word. Minimality implies $uvu \lhd vuu$, hence $uv \lhd vu$, hence also $uuv \lhd uvu$, which contradicts the minimality of w. □

Lyndon words can be characterized as "selfminimal":

Theorem. A word w is a Lyndon word if and only if it is (strictly) smaller than any of its proper suffixes.

Proof. Let w be a Lyndon word, and let u be a prefix of of w and v a suffix of w such that $|u| = |v|$. The word w is unbordered, therefore $u \neq v$ and $u \lhd v$, hence also $w \lhd v$.

Assume now that w is not a Lyndon word. If w is not primitive, then it is lexicographically larger than its primitive root. If $w = uv$ is primitive and vu is the Lyndon conjugate of w, then $vu \lhd uv$, hence also $v \lhd uv$. □

Theorem. Let u and v be Lyndon words (with respect to $\lhd$). If $u \lhd v$ and $u \neq v$, then uv is a Lyndon word.

Proof. We first show $uv \lhd v$. If $v = uv'$, then $uv \lhd uv' = v$ follows from $v \lhd v'$. If u is not a prefix of v, then $uv \lhd v$ follows from $u \lhd v$.

Let now $z \neq v$ be a proper suffix of uv. If $uv = uv'z$, then $uv \lhd v \lhd z$. If $z = z'v$, then $u \lhd z'$, thus $uv \lhd z'v$.

We have shown that uv is smaller than all its proper suffixes, hence it is a Lyndon word. □

The factorization $w = uv$ of w is called *standard* (with respect to $\lhd$), if v is the smallest proper suffix of w. Obviously, every word of length at least two admits the standard factorization. It is also clear from the definition that v is the longest proper suffix of w which is a Lyndon word (it is smaller than any of its suffixes, and any longer suffix has a smaller proper suffix, namely v).

We now have another characterization of Lyndon words.

Theorem. A word w is a Lyndon word if and only if w is a letter or $w = uv$ for some Lyndon words u and v such that $u \neq v$ and $u \lhd v$.

Proof. The "if" part follows from the previous theorem.

The "only if" part is obtained using the standard factorization $w = uv$ of w. We have $u \lhd uv \lhd v$, hence $u \lhd v$ and $u \neq v$. It remains to show that u is Lyndon. If u is letter, we are done. Otherwise, let u_1u_2 be the standard factorization of u. Then $v \lhd u_2$, since otherwise u_2v is a Lyndon word, contradicting the definition of v. Therefore, for any proper suffix u' of u, we have $u \lhd v \lhd u_2 \lhd u'$, and u is a Lyndon word. □

Previous facts culminate in the following theorem that defines the *Lyndon factorization* of any word with respect to given lexicographic order $\lhd$.

Theorem. Each word w can be uniquely written as a product of a non-increasing sequence of Lyndon words (with respect to $\lhd$).

Proof. Consider the factorization $w = u_1u_2\cdots u_k$ such that $(u_1u_2\cdots u_{j-1}) \cdot u_j$ is the standard factorization of $u_1u_2\cdots u_j$ for each $j = 2, 3, \dots, k$ and u_1 is a Lyndon word. In other words, we take from w successively longest Lyndon suffixes.

Clearly, the factorization is well defined, and all u_i are Lyndon words. If $u_i \lhd u_{i+1}$ and $u_i \neq u_{i+1}$, then also $u_i \neq u_{i+1}$ is a Lyndon word, and we obtain a contradiction with the maximality of u_{i+1}. The sequence of factors is therefore non-increasing.

Let $w = v_1v_2\dots v_m = z_1z_2\cdots z_\ell$ be two factorizations satisfying the assumptions. WLOG let $|v_1| \leq |z_1|$, and let $z_1 = v_1v_2\cdots v_jr$, where r is a nonempty prefix of v_{j+1}. Then

$$z_1 \lhd r \lhd v_{j+1} \lhd v_1 \lhd z_1,$$

hence $j = 0$, and $z_1 = v_1 = r$. The proof is concluded by induction on the length of w. □

Local period and the Critical Factorization.

Definition. The *local period* of a word w at the position $k \in \{0, 1, \dots, |w|\}$ is the length of the shortest nonempty word x that is suffix comparable with u and prefix comparable with v, where $w = uv$ and $k = |u|$.

Informally, the local period is the length of x of the shortest square x^2 that is centered at the position k. Observe that such an x is always unbordered (otherwise it can be replaced by its border).

It is easy to see that the local period is at any position at most $\pi(w)$. If the local period at the position k is equal $\pi(w)$, we call k a *critical position* and the factorization $w = uv$, with $|u| = k$ or, more precisely, the pair (u, v), is called a *critical factorization* of w.

Theorem (Critical Factorization Theorem). Every nonempty word w admits a critical factorization.

Proof. Choose two inverse lexicographic orders $\lhd$ and $\blacktriangleleft$. Let α be the maximum suffix of w w.r.t. $\lhd$ and let β be the maximum suffix of w w.r.t. $\blacktriangleleft$. If w has the period one, then the claim is obvious. Otherwise, we have $\alpha \neq \beta$, and WLOG we can assume $|\alpha| < |\beta|$.

We show that $k = |w| - |\alpha|$ is a critical point. Let x^2 be the shortest square centered at the position k. Assume, that $|x|$ is not a period of w; in particular, w is not a factor of x^ω.

1. If x is both a prefix of α and a suffix of $w\alpha^{-1}$, then $x\alpha$ is a suffix of w, hence $x\alpha \lhd \alpha$. Then $\alpha \lhd x^{-1}\alpha$, which contradicts the maximality of α.

2. If α is a prefix of x and x is a suffix of $w\alpha^{-1}$, then α is a prefix of $x\alpha$, hence $\alpha \lhd x\alpha$, again a contradiction.

3. Let finally x be a prefix of α and $w\alpha^{-1}$ be a suffix of x. Let $x^i x'$ be the longest prefix of α, with the period $|x|$. Then there are two distinct letters c and d such that $x'c$ and $x^i x'd$ prefixes of α. The maximality of α implies $d \lhd c$. Let $v = \beta\alpha^{-1}$. Since v is a suffix of x, we conclude that $vx'd$ is a factor of w. However, that contradicts the maximality of β because $vx'c$ is a prefix of β and $vx'c \blacktriangleleft vx'd$. $\square$