Words and Automata, Lecture 3
de Bruijn cycles and Lyndon words

Dominique Perrin

1er novembre 2013

We will prove the following beautiful result.

Theorem (Fredericksen, Maiorana 1978)

Let $\ell_1 < \ell_2 < \ldots < \ell_m$ be the increasing sequence of Lyndon words of length dividing n. The word $\ell_1\ell_2\cdots\ell_m$ is a de Bruijn cycle of order n which is moreover minimal.

Example

For example, if $n = 4$ and $A = \{a, b\}$, then

$$aaaabaababababbabbbb = a\ aaab\ aab\ ab\ abb\ abbb\ b$$

is a de Bruijn cycle of order 4.

We will use the following lemma.

Lemma

Let w be a prefix of length n of a Lyndon word and let ℓ be its longest prefix in L. Then w is the n-extension of ℓ.

Let $w = a_1 \cdots a_n$ with $a_i \in A$. Set $w = \ell s$ and let v be such that $wv \in L$. Since a Lyndon word is smaller than any of its proper suffixes, we have $\ell sv < sv$ and thus $a_j \leq a_{j+r}$ for $1 \leq j \leq n - r$ with $r = |\ell|$. It is not possible to have $a_j < a_{j+r}$ since otherwise $a_1 \cdots a_{j+r}$ is a prefix of w which is a Lyndon word longer than ℓ. This implies that r is a period of w and thus the conclusion.

Proof of the Theorem

Since w has length k^n, we only need to prove that any word $w = a_1 \cdots a_n$ of length n appears as a factor of $\ell_1 \cdots \ell_m \ell_1 \ell_2$. We denote by a the first letter of the alphabet and by z the largest one. We consider the following cases.

Case a) Assume first that w is primitive and that $w = uv$ with $vu = \ell_k$ and that u is not a power of z. Set $u = pbq$ with $p \in z^*$ and b a letter $b < z$. Since vpb is a prefix of a Lyndon word, $vpz \in L$. Repeatedly use the fact that for $\ell \in L$ with $\ell < m$ one has $\ell m \in L$, We obtain that $vz^{|u|}$ is a Lyndon word. Thus $\ell_{k+1} \leq vz^{|u|}$. This implies that v is a prefix of ℓ_{k+1} and thus w is a factor of $\ell_k \ell_{k+1}$.

Case b) Assume next that $w = uv$ is primitive, that $u \in z^*$ and that $vu \in L$. We can first rule out the case where $v \in a^*$. Indeed, $z^j a^{n-j}$ is a factor of $\ell_{m-1}\ell_m\ell_1\ell_2$. Let k be the least index such that $v \leq \ell_k$. Then $\ell_k \leq vu$ and thus v is a prefix of ℓ_k. Let $v' \leq v$ be the Lyndon word such that v is a sesquipower of v'.

Case b1) If $v' \neq \ell_{k-1}$, then ℓ_{k-1} ends with u. Indeed, let v'' be the word v' with its last letter changed into a. The word visited before v' by Algorithm $\textsc{Sesquipower}(n, k)$ is, in view of Algorithm $\textsc{SesquipowerBis}(n, k)$, the word $v''z^{n-|v'|}$. Thus w is a factor of $\ell_{k-1}\ell_k$.

Case b2) Otherwise, $v' = \ell_{k-1}$. Then u is a suffix of ℓ_{k-2} and v is a prefix of $\ell_{k-1}\ell_k$. Thus w is a factor of $\ell_{k-2}\ell_{k-1}\ell_k$.

Case c) Assume finally that $w = (uv)^d$ with d dividing n and $vu = \ell_k$.
Case c1) If $u \notin z^*$ then $\ell_{k+1} \leq (vu)^{d-1} vz^{|u|}$ since the latter is a Lyndon word. Thus w is a factor of $\ell_k \ell_{k+1}$.
Case c2) Otherwise, ℓ_{k-1} ends with at least $(d-1)|w|$ letters z and $\ell_{k+1} \leq (vu)^{d-1} z^{|w|}$. Thus w is a factor of $\ell_{k-1}\ell_k\ell_{k+1}$.

We illustrate the cases in the proof for $n = 6$ and $A = \{a, b\}$.
Table 1 gives the sequence ℓ_k.

k	1	2	3	4	5	6	7	8
ℓ_k	a	$aaaaab$	$aaaabb$	$aaabab$	$aaabbb$	aab	$aababb$	$aabbab$

9	10	11	12	13	14
$aabbbb$	ab	$ababbb$	abb	$abbbbb$	b

TAB.: The Lyndon words of length dividing 6.

(a) Let $w = aaba$. Then $u = aa$, $v = ba$ and $vu = \ell_2$. We find w as a factor of $\ell_2\ell_3 = aaabaabb$.

(b1) Let $w = baaaab$. Then $u = b$ and $v = aaaab$. We find $k = 3$, $v' = v$ and w is a factor of $\ell_2\ell_3$.

(b2) Let $w = bbabab$. Then $u = bb$, $v = abab$ and vu. We find $k = 11$. We have $v' = ab$ and we find w as a factor of $\ell_9\ell_{10}\ell_{11}$.

(c1) Let $w = (aba)^2$. Then $u = a$, $v = ba$ and $k = 6$. We find w as a factor of $\ell_6\ell_7$.

(c2) Let $w = (bab)^2$. Then $u = b$, $v = ab$ and $k = 12$. We find w as a factor of $\ell_{11}\ell_{12}\ell_{13}$.