Periodicity of Circular Words

László Hegedüs

Department of Computer Science,
Faculty of Informatics, University of Debrecen,
Debrecen, Hungary

19 September, 2013

[0] The work is supported by the TÁMOP 4.2.2/C-11/1/KONV-2012-0001 and 4.2.2/B-10/1-2010-0024 projects. The projects are implemented through the New Hungary Development Plan, co-financed by the European Social Fund and the European Regional Development Fund.

Motivation

- Investigate classes of words instead of just one word.
- We can derive some information about all elements of the class from the properties of some representative.
- How to generalise notions to equivalence classes? (Circular/cyclic words in our case.)

Preliminaries

- Basic notions, like *alphabet, word, factor, prefix, suffix* are assumed to be known.
- Let $w = w_1 \dots w_n$ (for some $w_1, \dots, w_n \in \Sigma$) and $p \in \mathbb{N}$. Then $w^{\frac{p}{n}} = w^{\lfloor \frac{p}{n} \rfloor} w'$, where $w' = w_1 \dots w_{(p \mod n)}$.

Example

$(aababa)^{\frac{15}{6}} = (aababa)^2 aab = aababaaababaaab$

- A word $w \in \Sigma^+$ is *primitive* if there does not exist $v \in \Sigma^*$ such that $w = v^p$ where p is a positive integer greater than one.
- Words x and y are *conjugates* if there exist words $u, v \in \Sigma^*$ such that $x = uv$ and $y = vu$.

Example

Consider the words $abbaab$ and $baabab$. Chosing $u = ab$ and $v = baab$ we get that $abbaab$ and $baabab$ are conjugates.

Preliminaries

- The positive integer p is a *period* of a word $w = w_1 \dots w_n$ (over some alphabet Σ) if $w_i = w_{i+p}$ for all $i = 1, \dots, n-p$.

Example

The word *acbacbac* has periods 3, 6 and 8.

- The word $u \in \Sigma^*$ is a *border* of $w \in \Sigma^*$ if there exist $x, y \in \Sigma^*$ such that $w = ux = yu$.

Example

The word *abababa* has (non-trivial) borders *a*, *aba* and *ababa*.

Properties of periods

Lemma

A word $w \in \Sigma^*$ *has some border* u *if and only if* $|w| - |u|$ *is a period of* w.

Theorem (Fine and Wilf's theorem for words)

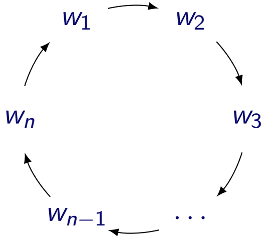
Let w *be a non-empty word of length* n. *If* w *has two distinct periods* p *and* q *such that* $p + q - \gcd(p, q) \leq n$, *then* $\gcd(p, q)$ *is also a period of* w.

Circular words

- A *circular word* is obtained from a finite non-empty word $w = w_1 w_2 w_3 \dots w_{n-1} w_n$ by joining it at the two ends (i.e., the beginning and the end of w).

- The word w is a/the *base* of the circular word w_o.

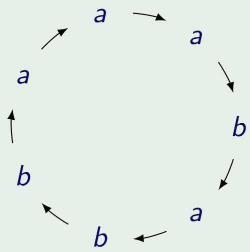
$$w = w_1 w_2 w_3 \dots w_{n-1} w_n \Longrightarrow$$



$$= \quad w_o$$

Example

$aababba \implies$



$= (aababba)_o$

We can start "reading" the circular word at any position.

$$(aababba)_o = \{aababba, ababbaa, babbaaa, abbaaab, bbaaaba, baaabab, aaababb\}$$

- Thus, $w_o = \{v \mid v \text{ is a conjugate of } w\}$.

Circular words (some conventions)

Let $w \in \Sigma^*$ and consider the circular word w_o.

- The circular word w_o is defined to be a set, but we will still refer to it as a word for convenience.
- The word $v \in \Sigma^*$ is a *factor* of w_o if v is a factor of some $u \in w_o$.
- Obviously, no circular word has any prefixes or suffixes.
- We will denote the *length* of w_o with $|w_o|$ and $|w_o| = |w|$. (Note that $|w_o|$ is not the cardinality of the set w_o.)

The shift operator

- Let $w \in \Sigma^+$ and k a non-negative integer. Then $w^{(k)}$ denotes the word that we get from w by *shifting* it by k positions to the left.

Example

$(aabcabc)^{(4)} = abcaabc$

- Thus, another definition of w_o is $w_o = \{w^{(k)} \mid k = 1, \ldots, |w|\}$.

Proposition (Properties of the shift operator)

1. $w^{(0)} = w^{(|w|)} = w$
2. $(w^{(k)})^{(\ell)} = w^{(k+\ell)}$
3. $w^{(k)} = (w^{(\ell)})^{(k-\ell)}$ $(\textit{if } \ell \leq k)$
4. $w^{(k)} = w^{(k-|w|)}$ $(\textit{if } k > |w|)$
5. $w^{(-k)} = w^{(|w|-k)}$

Weak- and strong periods of circular words

- The positive integer p is a *strong period* of some circular word w_o if p is a period of all $v \in w_o$.
- The positive integer p is a *weak period* of some circular word w_o if p is a period of at least one $v \in w_o$.

Example

Consider the circular word $(acacbacb)_o$.
Its elements with their periods:

$$\begin{array}{ll} acacbacb - 8 & cbacbaca - 8 \\ bacacbac - 5, 8 & acbacacb - 5, 8 \\ cbacacba - 5, 8 & cacbacba - 8 \\ acbacbac - 3, 6, 8 & bacbacac - 8 \end{array}$$

It has weak periods $3, 5, 6, 8$ and its only strong period is 8.

Strong periods of circular words

Proposition (Properties of strong periods)

1. *The word* $w \in \Sigma^*$ *is primitive if and only if the only strong period of* w_o *is* $|w|$.
2. *All strong periods of* $w_o \in \Sigma^*$ *are in the form* $p, 2p, \ldots, mp$, *where* $m = \frac{|w|}{p}$ *and* p *divides* $|w|$.
3. *If* $|w|$ *is prime and there exists some strong period* $p < |w|$ *of* w_o, *then* w *is a unary word.*

- Too much restriction in the case of full words.
- Probably more interesting if we consider partial words. (future)

Weak periods of circular words

Theorem (Fine and Wilf's theorem does not apply to weak periods)

There are infinitely many circular words with the following property: w_o has two distinct weak periods p and q such that $p + q - \gcd(p, q) \leq |w_o|$ and $\gcd(p, q)$ is not a weak period of w_o.

Proof.

- Let $w = ab^{p-1}ab^{p-2}$. Clearly, w has period p.
- Notice, that $w^{(2)} = b^{p-2}ab^{p-2}ab$ has period $p - 1$.
- $p + (p - 1) - 1 \leq |w|$, since $|w| = 2p - 1$
- Thus w_o has p and $p - 1$ as weak periods and it is not a unary word.

$\square$

Weak periods of circular words

Lemma

If w_o has weak period $p \leq \frac{|w|}{2}$, then $|w| - p$ is also a weak period of w_o.

Lemma

An arbitrary word w of length at least 2 over $\{a, b\}$ has periods p and $p - 1$ if and only if w_o has aa or bb as factors.

Lemma (The relation between periods and squares)

Let $w_o \in \Sigma^+$. Then p is a weak period of w_o if and only if there exists some $x \in \Sigma^+$ with $|x| = |w| - p$, such that xx is a factor of ww.

Algorithm

```
function CONSTRUCTWORD(n, p, q)
    for i ← 0 to n − 1 do
        W[i + 1] ← (i mod p) + 1
                        ▷ Now W is an array, containing the sequence (1, ..., p)^(n/p).
                ▷ Denotes, that w = (x_1 ... x_p)^(n/p) for some letters x_i (i = 1, ..., p).
    ℓ ← 1
    while ℓ ≤ n do
        eqs_ℓ[p][p] ← I_{p×p}            ▷ I_{p×p} is the p by p identity matrix.
                                         ▷ eqs_ℓ[i][j] = 1 denotes that x_i = x_j
        for i ← 1 to n − q do
            X ← W[((i + ℓ − 1) mod n) + 1]
            Y ← W[((i + ℓ + q − 1) mod n) + 1]
            if eqs_ℓ[X][Y] = 0 then
                eqs_ℓ[X][Y] ← 1
                eqs_ℓ[Y][X] ← 1
        for i ← 1 to p do                ▷ Because equality is transitive.
            for j ← 1 to p do
                if eqs[i][j] = 1 then
                    for k ← 1 to p do
                        if eqs[i][k] = 1 then
                            eqs[j][k] ← 1
        ℓ ← ℓ + 1
    return eqs_1, ..., eqs_n
```

An example

- We are looking for a circular word of length 9 which has weak periods 5 and 2.
- Run the above algorithm with parameters $n = 9, p = 5$ and $q = 2$.
- Since $n = 9$, the result consists of nine matrices, namely $eqs_1, eqs_2 \dots eqs_9$.
- The ones that do not represent unary words are the following two:

$$eqs_4 = \begin{pmatrix} 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 & 1 \end{pmatrix}, eqs_5 = \begin{pmatrix} 1 & 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 \end{pmatrix}.$$

Let us use, for example, eqs_4 now.

$$eqs_4 = \begin{pmatrix} 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 & 1 \end{pmatrix}.$$

The following equalities have to be satisfied by $w = x_1x_2x_3x_4x_5x_1x_2x_3x_4$:

$$x_1 = x_3 \quad \text{and} \quad x_2 = x_4 = x_5.$$

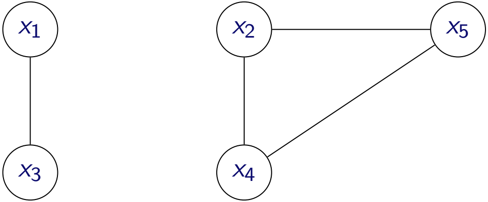


Figure: Graph representation of eqs_4.

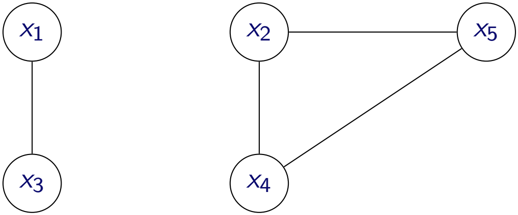


Figure: Graph representation of eqs_4.

Choosing $x_1 = x_3 = a$, and $x_2 = x_4 = x_5 = b$, we have $w = ababbabab$ with period 5 and $w^{(4)} = babababab$ with period 2.

All weak periods of w_o are $2, 4, 5, 6, 7, 8, 9$.

Number of weak periods

Theorem

For given $n \geq 3$ the maximal number of weak periods of non-unary circular words over $\{a, b\}$ of length n is

$$
\begin{cases}
\lfloor \frac{n}{2} \rfloor + \lfloor \frac{n}{4} \rfloor + 1 & \text{if } n \text{ is odd,} \\
\frac{n}{2} & \text{if } n \text{ is even and } (n \mod 3) = 1 \text{ or } (n \mod 3) = 0, \\
\frac{n}{2} + 1 & \text{if } n \text{ is even and } (n \mod 3) = 2.
\end{cases}
$$

Relatively prime weak periods

Theorem

Given $w_o \in \Sigma^*$ *with weak periods* p, q *(with* $p > q \geq 2$*). Let* $m = p \bmod q$*. If* $\gcd(p, q) = 1$ *and* w_o *is not a unary word, then*

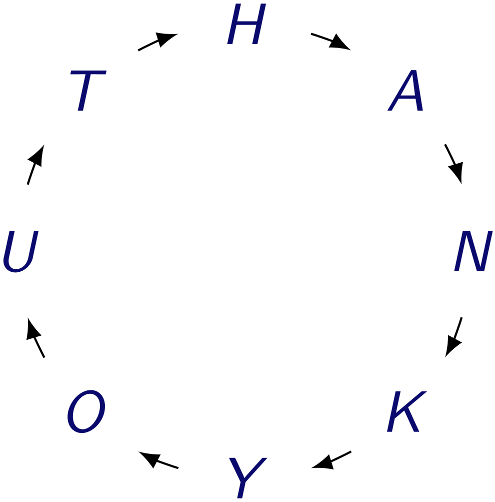
$$|w_o| \leq \begin{cases} p + q \cdot \lfloor \frac{p}{q} \rfloor & \text{if } m = 1, \\ p + q \cdot \lceil \frac{p}{q} \rceil & \text{if } m > 1. \end{cases}$$

Proof idea:

- $m = 1$. Choose a non-unary $v \in \Sigma^*$ with $|v| = q$.
- $u := v^{\frac{p}{q}} = v^{\lfloor \frac{p}{q} \rfloor} v_1$. Then $w = uv^{\lfloor \frac{p}{q} \rfloor} = v^{\lfloor \frac{p}{q} \rfloor} v_1 v^{\lfloor \frac{p}{q} \rfloor}$ has period p and $w^{(p)} = v^{\lfloor \frac{p}{q} \rfloor} u = v^{\lfloor \frac{p}{q} \rfloor} v^{\lfloor \frac{p}{q} \rfloor} v_1$ has period q.
- Thus w_o has weak periods p and q. Furthermore, $|w_o| = p + q \cdot \lfloor \frac{p}{q} \rfloor$.
- To show that any z_o of length at least $|w_o| + 1$ is unary we have to show that we cannot partition the letters of z_o into more than one equivalence class.

Some (possible) future directions

1. What about partial words? (e.g, $(ab\diamond\diamond b)_o$ has strong (weak) periods 3 and 2)
2. Define $S(w) = \{v \in \Sigma^* \mid (v, w) \in P\}$, where P is some relation. What about the weak- and strong periods of $S(w)$? What kind of relations could turn out to be interesting?
3. In 2 above, P may be a relation over several words.


H
A
N
K
Y
O
U
T