An Extension of the Lyndon Schützenberger Result to Pseudoperiodic Words

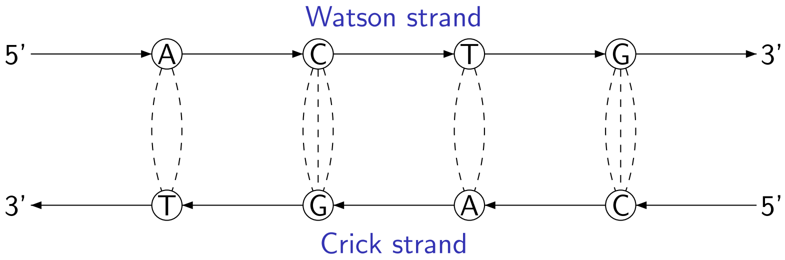
Elena Czeizler[1] Eugen Czeizler[1]
Lila Kari[2] Shinnosuke Seki[2]

[1]Department of IT, Åbo Akademi University

[2]Department of Computer Science, University of Western Ontario

June 30th, 2009

Informational equivalence between Watson and Crick strands



The Crick strand CAGT is obtained from the Watson strand ACTG by the antimorphic involution τ defined as $\tau(\mathtt{A}) = \mathtt{T}$, $\tau(\mathtt{T}) = \mathtt{A}$, $\tau(\mathtt{C}) = \mathtt{G}$, $\tau(\mathtt{G}) = \mathtt{C}$.

Observation

Two WK-complementary strands are equivalent w.r.t. the information they encode.

A mathematical model of Watson-Crick complementarity

Definition

A mapping $\theta : \Sigma^* \rightarrow \Sigma^*$ is called an antimorphic involution if

1. for any $x, y \in \Sigma^*$, $\theta(xy) = \theta(y)\theta(x)$ (antimorphism), and
2. θ^2 is the identity (involutive).

Actually,

$$\tau(\mathtt{ACTG}) = \tau(\mathtt{G})\tau(\mathtt{ACT}) = \cdots = \tau(\mathtt{G})\tau(\mathtt{T})\tau(\mathtt{C})\tau(\mathtt{A}) = \mathtt{CAGT}.$$

Observation

For an arbitrary antimorphic involution θ and a word $u \in \Sigma^*$, u and $\theta(u)$ are informationally equivalent.

θ-primitivity

A word $u \in \Sigma^+$ is primitive if for any $t \in \Sigma^+$, $u \in t^+$ implies $u = t$.

Definition ([CKS08])

A word $u \in \Sigma^+$ is said to be θ-primitive if for any $t \in \Sigma^+$, $u \in t\{t, \theta(t)\}^*$ implies $u = t$.

Definition ([CKS08])

The θ-primitive root of $u \in \Sigma^+$ is a θ-primitive word $t \in \Sigma^+$ s.t. $u \in t\{t, \theta(t)\}^*$.

The uniqueness of θ-primitive root is guaranteed.

An extended Fine and Wilf theorem

Let $u, v \in \Sigma^+$ and θ be an antimorphic involution.

Theorem ([CKS08])

If there exist $\alpha \in \{u, \theta(u)\}^*$ *and* $\beta \in \{v, \theta(v)\}^*$ *which share a prefix of length* $\mathrm{lcm}(|u|, |v|)$, *then* $u, v \in \{t, \theta(t)\}^+$ *for some* θ*-primitive word* $t \in \Sigma^+$.

Theorem ([CKS08])

If there exist $\alpha \in \{u, \theta(u)\}^*$ *and* $\beta \in \{v, \theta(v)\}^*$ *which share a prefix of length* $2\max(|u|, |v|) + \min(|u|, |v|) - \gcd(|u|, |v|)$, *then* $u, v \in \{t, \theta(t)\}^+$ *for some* θ*-primitive word* $t \in \Sigma^+$.

Lyndon-Schützenberger equation

For words $u, v, w \in \Sigma^*$, an equation of the form

$$u^\ell = v^n w^m$$

is called the Lyndon-Schützenberger equation originally proposed in [LySc62].

Theorem

If $\ell, n, m \geq 2$, then the above equation implies $u, v, w \in t^+$ for some primitive word $t \in \Sigma^+$.

Its concise proof is available at, e.g., [Lot83, HaNo04].

An extended Lyndon Schützenberger equation

We extend the LS-equation as follows:

$$u_1 \cdots u_\ell = v_1 \cdots v_n w_m \cdots w_1,$$

for $u, v, w \in \Sigma^+$ and $\ell, n, m \geq 2$, where $u_1, \ldots, u_\ell \in \{u, \theta(u)\}$, $v_1, \ldots, v_n \in \{v, \theta(v)\}$, and $w_1, \ldots, w_m \in \{w, \theta(w)\}$.

Problem

Find conditions on ℓ, n, m under which the exLS equation implies $u, v, w \in \{t, \theta(t)\}^+$ for some θ-primitive word $t \in \Sigma^+$.

When ℓ, n, m guarantees the existence of such t, we say that (ℓ, n, m) imposes θ-periodicity.

Cases when (ℓ, n, m) does not impose θ-periodicity

Proposition

If one of ℓ, n, m is 2, then (ℓ, n, m) does not impose θ-periodicity.

Example

Let θ be the mirror image on $\{a, b\}$. Let $u = a^k b^2 a^{2k}$, $v = (a^{2k} b^2 a^k)^\ell a^{2k} b^2$, and $w = a^2$ for some $k, \ell \geq 1$. Then $\theta(u)^{\ell+1} u^{\ell+1} = v^2 w^k$.
Actually, when $\ell = k = 2$, then

$$\begin{aligned} \theta(u)^2 u^2 &= (a^4 b^2 a^2)^2 (a^2 b^2 a^4)^2 \\ &= a^4 b^2 a^6 b^2 a^4 b^2 a^6 b^2 a^4 \\ &= (a^4 b^2 a^6 b^2)^2 (a^2)^2 = v^2 w^2 \end{aligned}$$

Observation

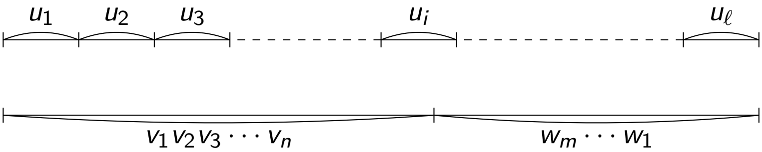
$\ell, n, m \geq 3$ must hold for (ℓ, n, m) to impose θ-periodicity.

An application of exFW theorem I

Many cases can be solved by applying the exFW theorem.

Proposition

$(\geq 6, \geq 3, \geq 3)$ *imposes θ-periodicity.*



Due to the symmetric roles of $v_1 \cdots v_n$ and $w_m \cdots w_1$ in this equation, one can assume that $|v_1 \cdots v_n| \geq |w_1 \cdots w_m|$, that is, $|v_1 \cdots v_n| \geq \frac{1}{2}|u_1 \cdots u_\ell|$.

An application of exFW theorem II

Proof.

Since $\ell \geq 6$, $|v_1 \cdots v_n| \geq \frac{1}{2}|u_1 \cdots u_\ell| \geq |u_1 u_2 u_3|$. Also note that $v_1 \cdots v_n$ is a prefix of $u_1 \cdots u_\ell$. These mean that $u_1 \cdots u_\ell$ and $v_1 \cdots v_n$ share a prefix of length

$$\max(3|u|, 3|v|) \geq 2\max(|u|, |v|) + \min(|u|, |v|).$$

Thus the exFW theorem implies that $u, v \in \{t, \theta(t)\}^+$ for some θ-primitive word $t \in \Sigma^+$. With this, the exLS equation gives $w_1 \cdots w_m \in \{t, \theta(t)\}^+$. Since t is θ-primitive, we can conclude that $w \in \{t, \theta(t)\}^+$. □

An application of exFW theorem III

This proof technique works for the cases $(5, \geq 5, \geq 5)$.

Proposition

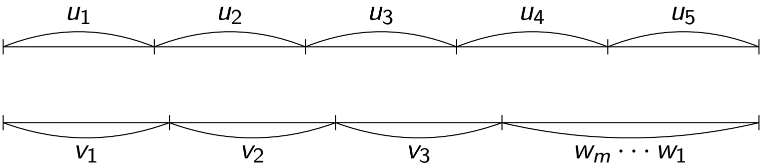
$(5, \geq 5, \geq 5)$ *imposes* θ*-periodicity.*

The remaining cases $(5, 3 \text{ or } 4, \geq 3)$ require combinatorial arguments.

Combinatorial argument on $(5, 3, \geq 3)$ I

Proposition

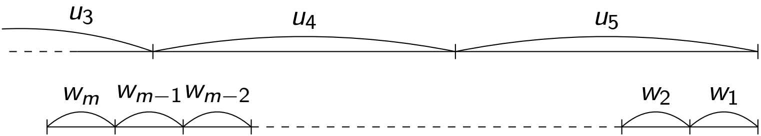
$(5, 3$ *or* $4, \geq 3)$ *impose* θ*-periodicity.*



If the border between $v_1 v_2 v_3$ and $w_m \cdots w_1$ is on anything but u_3, then the exFW theorem is still applicable.

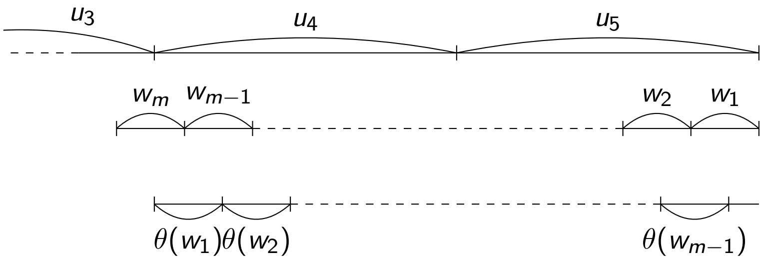
Combinatorial argument on $(5, 3, \geq 3)$ II

Even when the border is on u_3, when $|u_4 u_5| \leq |w_{m-1} \cdots w_1|$, then the exFW theorem is applicable.

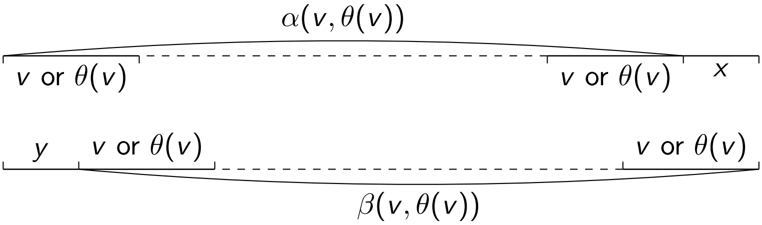


Combinatorial argument on $(5, 3, \geq 3)$ III

The case when the border is on u_3 and $u_5 = \theta(u_4)$ is illustrated.



Combinatorial argument on $(5, 3, \geq 3)$ IV



For a θ-primitive word v, we consider the overlap of the form

$$\alpha(v, \theta(v)) \cdot x = y \cdot \beta(v, \theta(v)),$$

where $\alpha(v, \theta(v)), \beta(v, \theta(v)) \in \{v, \theta(v)\}^+$, $x, y \in \Sigma^+$, and $|x|, |y| < |v|$.

Combinatorial argument on $(5, 3, \geq 3)$ V

Theorem

All possible overlaps of the above-mentioned form are given in the following table (modulo a substitution of v by $\theta(v)$) together with the characterization of their sets of solutions.

Equation	Solution
$v^i x = y\theta(v)^i, i \geq 1$	$v = yp, x = \theta(y), p = \theta(p)$, and whenever $i \geq 2$, $y = \theta(y)$
$vx = yv$	$v = (pq)^{j+1}p$, $x = qp$, $y = pq$ for some $p, q \in \Sigma^+$, $j \geq 0$
$v\theta(v)x = yv\theta(v)$,	$v = (pq)^{j+1}p$, $x = \theta(pq)$, $y = pq$, with $j \geq 0$, $qp = \theta(qp)$
$v^{i+1}x = y\theta(v)^i v, i \geq 1$	$v = r(tr)^{n+m}r(tr)^n$, $x = (tr)^m r(tr)^n$, $y = r(tr)^{n+m}$
$v\theta(v)^i x = yv^{i+1}$, $i \geq 1$	$v = (rt)^n r(rt)^{m+n}r$, $y = (rt)^n r(rt)^m$, $x = (rt)^{m+n}r$
$v\theta(v)^i x = yv^i\theta(v)$, $i \geq 2$	$v = (rt)^n r(rt)^{m+n}r$, $y = (rt)^n r(rt)^m$, $x = (tr)^m r(tr)^n$

Combinatorial argument on $(5, 3, \geq 3)$ VI

Theorem

$(5, \geq 3, \geq 3)$ *imposes* θ*-periodicity.*

Summary on the exLS equation

ℓ	n	m	θ-periodicity	How to prove
≥ 6	≥ 3	≥ 3	YES	exFW theorem
5	≥ 5	≥ 5	YES	exFW theorem
5	4	≥ 4	YES	combinatorial arguments
5	3	≥ 3	YES	
4	≥ 3	≥ 3	OPEN	?
3	≥ 3	≥ 3	OPEN	
2			NO	examples
	2		NO	
		2	NO	

Future directions

1. Solving the open cases of exLS equation;
2. Concise proof technique;
3. Further extensions of exLS equation; e.g., to n words of

$$u_1 \cdots u_\ell \in \{v_1, \theta(v_1)\}^{k_1}\{v_2, \theta(v_2)\}^{k_2} \cdots \{v_n, \theta(v_n)\}^{k_n}$$

References I

[ChKa97] Christian Choffrut and Juhani Karhumäki.
Combinatorics of words.
In Crzegorz Rozenberg and Arto Salomaa, editors, *Handbook of Formal Languages*, vol.1, pp.320-438. Springer-Verlag, Berlin-Heidelberg-NewYork, 1997.

[CKS08] E. Czeizler, L. Kari, and S. Seki.
On a special class of primitive words.
In *Proc. MFCS 2008*, LNCS 5162, pp. 265-277, Berlin-Heidelberg, 2008, Springer.

[FiWi65] N. J. Fine and H. S. Wilf.
Uniqueness theorem for periodic functions.
Proc. American Mathematical Society, 16:109-114, 1965.

References II

[HaNo04] T. Harju and D. Nowotka.
The equation $x^i = y^j z^k$ in a free semigroup.
Semigroup Forum, 68:488-490, 2004.

[Lot83] M. Lothaire.
Combinatorics on Words.
Encyclopedia of Mathematics and its Applications, 17, Addison-Wesley Publishing Co., 1983.

[LySc62] R. C. Lyndon and M. P. Schützenberger.
The equation $a^M = b^N c^P$ in a free group.
Michigan Mathematical Journal, 9:289-298, 1962.