Walking on Words

Ian Pratt-Hartmann

Department of Computer Science, Manchester University, UK
Instytut Informatyki, Uniwersytet Opolski, Polska
email: ian.pratt@manchester.ac.uk

CPM 2024
Fukuoka, Japan
24.6.24

Outline

c	b	a	d	e	f	g	h

c	b	a	d	e	f	g	h

a

c	b	a	d	e	f	g	h

$a \quad b$

c	b	a	d	e	f	g	h

$a \quad b \quad c$

c	b	a	d	e	f	g	h

$a \quad b \quad c \quad b$

c	b	a	d	e	f	g	h

$a \quad b \quad c \quad b \quad a$

c	b	a	d	e	f	g	h

$a \quad b \quad c \quad b \quad a \quad a$

c	b	a	d	e	f	g	h

$a \quad b \quad c \quad b \quad a \quad a \quad a$

c	b	a	d	e	f	g	h

$a \quad b \quad c \quad b \quad a \quad a \quad a \quad d$

c	b	a	d	e	f	g	h

$a \quad b \quad c \quad b \quad a \quad a \quad a \quad d \quad e$

c	b	a	d	e	f	g	h

$a \quad b \quad c \quad b \quad a \quad a \quad a \quad d \quad e \quad f$

c	b	a	d	e	f	g	h

$a \quad b \quad c \quad b \quad a \quad a \quad a \quad d \quad e \quad f \quad e$

c	b	a	d	e	f	g	h

$a \quad b \quad c \quad b \quad a \quad a \quad a \quad d \quad e \quad f \quad e \quad d$

c	b	a	d	e	f	g	h

$a \quad b \quad c \quad b \quad a \quad a \quad a \quad d \quad e \quad f \quad e \quad d \quad e$

c	b	a	d	e	f	g	h

$a \quad b \quad c \quad b \quad a \quad a \quad a \quad d \quad e \quad f \quad e \quad d \quad e \quad f$

c	b	a	d	e	f	g	h

$a \quad b \quad c \quad b \quad a \quad a \quad a \quad d \quad e \quad f \quad e \quad d \quad e \quad f \quad g$

c	b	a	d	e	f	g	h

$a \quad b \quad c \quad b \quad a \quad a \quad a \quad d \quad e \quad f \quad e \quad d \quad e \quad f \quad g \quad h$

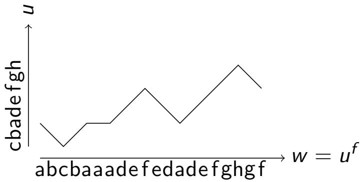
c	b	a	d	e	f	g	h

$a \quad b \quad c \quad b \quad a \quad a \quad a \quad d \quad e \quad f \quad e \quad d \quad e \quad f \quad g \quad h \quad g$

c	b	a	d	e	f	g	h

$a \quad b \quad c \quad b \quad a \quad a \quad a \quad d \quad e \quad f \quad e \quad d \quad e \quad f \quad g \quad h \quad g \quad f$

- Let $u = a_1 \cdots a_n$ be a word ($n \geq 0$). If $f : [1, m] \to [1, n]$ is a function (for some $m \geq 0$), write $u^f = a_{f(1)} \cdots a_{f(m)}$.
- We call f a walk if it is a surjection satisfying $|f(i+1) - f(i)| \leq 1$ for all i ($1 \leq i < m$).
- If f is a walk and $w = u^f$, we say u generates w.



- Given a word w, what words u might have generated it?

- Every word u generates itself and its reversal.
- All other words generated by u are strictly longer than u.
- Say u is primitive if it is not generated by any shorter word.

$$\begin{array}{lllll} \epsilon, & abc, & abcdca, & abcdedca, & \ldots \\ abbc, & abacde, & abcdcbcde, & abcdcb, & \ldots \end{array}$$

- A word u is a primitive generator of a word w if u is primitive and u generates w.
- Generation is transitive: if u generates v and v generates w, then u generates w. Hence, every word w has some primitive generator u (and in fact $\tilde{u}$ as well).

Outline

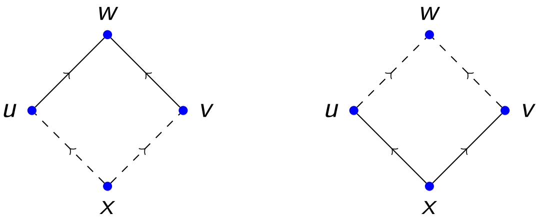
- Principal result: primitive generators are unique up to reversal:

  Theorem

  If u and v are primitive generators of w, then $u = v$ or $u = \tilde{v}$.

  The proof has a 'geometrical' character, and is elementary.
- As a consequence, every word featuring more than one letter has exactly two primitive generators, of the form u and $\tilde{u}$.
- Say that u and v are primitive conjugates if they have the same primitive generators.

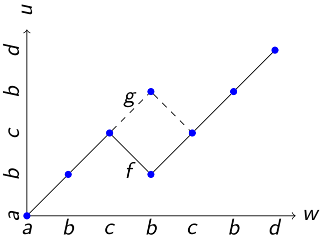
- An immediate corollary is confluence: if u and v generate w, there exists an x that generates u and v.
- It is also possible to prove an analogous amalgamation result.



- Thus, the following are equivalent: (i) u and v are primitive conjugates; (ii) u and v generate a common word; (iii) u and v are generated by a common word.
- However, primitive conjugacy classes are not in general lattices under the relation of generation.

Outline

- Primitive generators are unique, but generating walks are not.



- Say that a word u is perfect if $u^f = u^g \Rightarrow f = g$. It is trivial to show:

Theorem

Let u be a word. Then u is perfect if and only if it contains no non-trivial palindrome as a factor.

- If $u[i,j]$ is a non-trivial palindrome, we call $\langle i,j\rangle$ a defect of u.
- Denote the set of defects of u by Δ_u.
- If $|u| = n$, then Δ_u is a binary relation on $[1,n]$. Denote its equivalence closure by Δ_u^*.

Theorem

If u is primitive, and f, g walks on u of length m, then $u^f = u^g$ if and only if $\langle f(i), g(i)\rangle \in \Delta_u^*$ for all i $(1 \leq i \leq m)$.

Thus, whether $u^f = u^g$ depends only on the positions and lengths of the non-trivial palindromes in u.

Outline

- The motivation comes from first-order logic.
- If $\mathcal{L}$ is a fragment of first-order logic, $Sat(\mathcal{L})$ is the following problem:

  Given: a sentence φ of $\mathcal{L}$;
  Output: Y if φ has a model, N otherwise.
- We seek fragments $\mathcal{L}$ for which $Sat(\mathcal{L})$ is decidable.
- One such fragment is the adjacent fragment, e.g.:

$$\forall x_1 x_2 x_3 \exists x_4 \forall x_5 (p(x_1 x_2 x_3 x_4 x_3 x_4 x_5) \rightarrow p(x_1 x_2 x_3 x_2 x_3 x_4 x_5)).$$

- We require the above theorem on primitive generators to establish the decidability of $Sat(\mathcal{L})$ in this case.

- Consider the alphabet $[1, k] = \{1, \ldots, k\}$ and the endomorphism $\sigma : [1, k]^* \to [1, k]^*$ defined by

$$\sigma(i) = \begin{cases} 1 \cdot (i+1) & \text{if } i < k \\ 1 & \text{if } i = k. \end{cases}$$

- This defines a sequence of words

$$\{\alpha_n^{(k)}\}_{n \geq 1} = 1, \quad \sigma(1), \quad \sigma(\sigma(1)), \quad \sigma(\sigma(\sigma(1))), \quad \ldots$$

- With $k = 2$, we obtain the Fibonacci words

$$\{\alpha_n^{(2)}\}_{n \geq 1} = 1, \quad 12, \quad 121, \quad 12112, \quad \ldots$$

- With $k = 3$, we obtain the tribonacci words

$$\{\alpha_n^{(3)}\}_{n \geq 1} = 1, \quad 12, \quad 1213, \quad 1213121, \quad \ldots$$

- It is simple to show that, for all k and all $n > k$,

$$\alpha_n^{(k)} = \alpha_{n-1}^{(k)} \alpha_{n-2}^{(k)} \cdots \alpha_{n-k}^{(k)}.$$

- All terms of the k-bonacci sequence (from the kth) have the same primitive generator.

Theorem

For all $k \geq 2$, there exists a word γ_k such that, for all $n \geq k$, γ_k is the primitive generator of $\alpha_n^{(k)}$.

Outline

- Let w be a non-empty word and u a primitive generator of w. Define the primitive incompressibility of w to be $|u|/|w|$.
- We ask, for a fixed alphabet A, what is the expected primitive incompressibility of words over A of length n?
- For $|A| \geq 4$ we observe the following pattern: