Square-free reducts of words

Combinatorics on Words Seminar

Szymon Stankiewicz

joint work with Jarosław Grytczuk

March 15, 2021

Definitions

We will say, that word W can be **squre reduced in one step** to word U (denoted by $W \to U$) iff exist words A, B, C such that $W = ABBC$ and $U = ABC$.

Word W can be **square reduced** to word U (denoted by $W \rightsquigarrow U$) iff $W = U$ or if we can find a sequence of one step square reductions starting with word W and ending with word U. It's obvious that relation $\rightsquigarrow$ is a transitive and reflexive closure of relation $\to$.

Word U is a **reduct** of word W iff $W \rightsquigarrow U$ and U is square-free.

We will denote the set of all reducts of word W by $R(W)$ and size of this set by $r(W)$. Let $f_k(n)$ be the maximal value of $r(W)$ over all words of length n over alphabet of size k.

Example 1 - $(ab)^5$

- $ababababab \to abababab$
- $ababababab \to ababab$
- $ababababab \to ababab$

- $ababababab \rightsquigarrow ababababab$
- $ababababab \rightsquigarrow abababab$
- $ababababab \rightsquigarrow ababab$
- $ababababab \rightsquigarrow abab$
- $ababababab \rightsquigarrow ab$

ab is the only reduct of $ababababab$.

Example 2 - *abcbabcbc*

$$abcbabcbc \qquad\qquad abcbabcbc$$

$$\downarrow \qquad\qquad \downarrow$$

$$abcbabc \qquad\qquad abcbc$$

$$\downarrow$$

$$abc$$

abcbabcbc has two reducts

Binary words

Proposition

Every binary word W satisfies $r(W) = 1$.

Sketch of proof:

- Only six square-free binary words exist: 0, 1, 01, 10, 010, 101.
- If $W \rightsquigarrow U$, then first letter, last letter and set of letters of W is the same as for U.

Ternary words

Theorem

For every integer $k \geq 1$, there exists a ternary word W with $r(W) \geq k$.

$$A = \text{abacabcbacabacbabc}$$

$$B = \text{abacabcbacbcacbabc}$$

$$C = \text{abacbcacbacabcbabc}$$

$$D = \text{abacabcbabcbabacabacacbcacbabcbababc}$$

Ternary words

Lemma

Morphism $\varphi : \mathrm{a} \mapsto A,\ \mathrm{b} \mapsto B,\ \mathrm{c} \mapsto C$ is square-free.

By the result of Crochemore, to prove that a morphism is square-free it suffices to check its images on the set of square-free words of length at most 3.

Corollary

Morphism $\varphi' : \mathrm{a} \mapsto B,\ \mathrm{b} \mapsto A,\ \mathrm{c} \mapsto C$ is square-free.

Ternary words

Lemma

$D \rightsquigarrow A$ and $D \rightsquigarrow B$.

$D \rightsquigarrow A$	$D \rightsquigarrow B$
$abacabcbabcbabacabacacbcacbabcbababc$	$abacabcbabcbabacabacacbcacbabcbababc$
$abacabcbabacabacacbcacbabcbababc$	$abacabcbabacabacacbcacbabcbababc$
$abacabcbacabacacbcacbabcbababc$	$abacabcbacabacacbcacbabcbababc$
$abacabcbacabacacbabcbababc$	$abacabcbacacbcacbabcbababc$
$abacabcbacabacbabcbababc$	$abacabcbacbcacbabcbababc$
$abacabcbacabacbababc$	$abacabcbacbcacbababc$
$abacabcbacabacbabc$	$abacabcbacbcacbabc$

Ternary words

We can define new morphism $\psi : \mathrm{a} \mapsto D,\ \mathrm{b} \mapsto D,\ \mathrm{c} \mapsto C$. Since $D \rightsquigarrow A$ and $D \rightsquigarrow B$ we know, that for each ternary word W $\psi(W) \rightsquigarrow \varphi(W)$ and $\psi(W) \rightsquigarrow \varphi'(W)$.

Since for any morphing χ we may prove that if $U \rightsquigarrow W$ then $\chi(U) \rightsquigarrow \chi(W)$ we may conclude that for any word S if $T \in R(S)$ then $\varphi(T), \varphi'(T) \in R(\psi(S))$.

Ternary words

Lemma

Let W be any ternary word containing one of letters a, b then $r(\psi(W)) \geq 2r(W)$.

- If $S \neq T$, then $\varphi(S) \neq \varphi(T)$ and $\varphi'(S) \neq \varphi'(T)$.
- For all S containing one of letters a, b $\varphi(S) \neq \varphi'(S)$.

So each word $S \in R(W)$ generate two distinct words $\varphi(S)$ and $\varphi'(S)$, both being elements of $R(\psi(W))$.

Ternary words

Fact

The number of square-free words of length n over a 3-letter alphabet is at least c^n, for some constant $c > 1$.

Theorem

There exists a constant $\alpha > 1$ *such that* $f_3(n) \geq \alpha^n$.

Ternary words

Consider a word $W_m = CDDDCDDD \cdots CDDD = (CDDD)^m$.

- DDD can be reduced to any of the words A, B, AB, BA, ABA, BAB.
- W_m can be reduced to any square-free word over alphabet $\{A, B, C\}$ having m letters C and starting with CA or CB. The number of such words is at least c^m.
- Length of W_m is at most $36m$.

Theorem follows for $\alpha = c^{\frac{1}{36}}$ and since c is roughly 1.3 we know that $\alpha \geq 1.0073$.

Words over four letters

Theorem

For every integer $k \geq 1$, there exists a word over a 4-letter alphabet with exactly k distinct reducts.

Words over four letters

Let us fix the 4-letter alphabet as $\{a, b, x, y\}$. Take the word $F =$ xabaxababx having exactly two reducts $P = xabx$ and $Q =$ xabaxabx $= Q'P$.

Let W_∞ be any infinite square-free word over the alphabet $\{a, b, y\}$ starting with the letter y. Let $W_1, W_2, \ldots$ be any sequence of prefixes of the word W_∞ with strictly growing lengths such that each of them ends with letter y.

Lemma

For each $i \geq 1$ word $S_i = FW_1 FW_2 \cdots FW_i$ has exactly $i + 1$ reducts.

$$R(S_i) = \{PW_i, QW_i\} \cup \{PW_1 QW_i, PW_2 QW_i, \ldots, PW_{i-1} QW_i\}.$$

Words over four letters

Lemma

Let $W = a_1 a_2 \cdots a_n$ be any square-free word, where each a_i is a single letter. Let $V = a_1^{k_1} a_2^{k_2} \cdots a_n^{k_n}$, where each k_i is a positive integer. Then every square in V is of the form x^{2k}, where $x = a_i$ for some $i \in 1, 2, ..., n$.

Words over four letters

Theorem

Let U be a word over alphabet $\{a, b, x\}$ starting and ending with the letter x. Let $\alpha = r(U)^{\frac{1}{|U|+5}}$. Then there exists a constant c such that $f_4(n) \geq c\alpha^n$, for all $n \in \mathbb{N}$

Let S be any infinite square-free word over alphabet $\{a, b, y\}$ starting with the letter y. Let T be a word obtained form S by duplicating every occurrence of the letter y in S, except the first one. Hence, the word T can be written uniquely as $T = T_1 T_2 T_3 \ldots$, where each factor T_i starts and ends with the letter y, and these are the only occurrences of this letter in T_i. Finally, let us define $V_j = UT_1 UT_2 \cdots UT_j$, for each $j \geq 1$.

Words over four letters

- $r(V_j) = r(U)^j$.
- $|V_j| \leq j(|U| + 5)$
- $r(V_j) = r(U)^j \geq (r(U)^{\frac{1}{|U|+5}})^{|V_j|} = \alpha^{|V_j|}$
- $|V_{i+1}| - |V| \leq |U| + 5$, therefore we can take $c = \alpha^{-(|U|+5)}$

One may check that the word $U = xabaxababxbabx$ satisfies $r(U) = 4$ and $|U| = 14$. Hence, in the above theorem we may take $\alpha = 4^{\frac{1}{19}} \approx 1.075$.

Posets of square reductions

Fact

For any alphabet Σ pair $(\Sigma^*, \rightsquigarrow)$ forms a poset.

Theorem

Poset $S = (\{a, b\}^*, \rightsquigarrow)$ *is universal.*

Posets of square reductions

Lemma

Let $A = a^{\alpha_1} b a^{\alpha_2} b \cdots a^{\alpha_n}$ and $B = a^{\beta_1} b a^{\beta_2} b \cdots a^{\beta_n}$. $A \rightsquigarrow B$ iff $\alpha_i \geq \beta_i$ for all $i \in 1, 2, ..., n$.

Let P be a poset and R be it's realizer of size n. For element p of poset P let $\alpha_i(p)$ be the positions of element p in i-th linear order of R.

We will map element p to $\phi(p) = a^{\alpha_1(p)} b a^{\alpha_2(p)} b \cdots a^{\alpha_n(p)}$. It's easy to see that for any element p, q of poset P $\phi(p) \rightsquigarrow \phi(q)$ iff $\alpha_i(p) \geq \alpha_i(q)$ for all $i \in 1, 2, ..., n$, which means that $p \geq q$ in poset P.

Posets of square reductions

Let G_k be a graph created by removing directions of edges from poset $([k]^*, \rightsquigarrow)$.

Theorem

G_3 has finitely many connected components.

Posets of square reductions

$$
\begin{aligned}
X_1 &= abcabac & \qquad Y_1 &= abcbac \\
X_2 &= abcacba & \qquad Y_2 &= abcba \\
X_3 &= abcbabc & \qquad Y_3 &= abc \\
X_4 &= abcbacab & \qquad Y_4 &= abcab \\
X_5 &= abcbacb & \qquad Y_5 &= abcacb
\end{aligned}
$$

$$
\begin{aligned}
S_1 &= abcbabcbcacbcacabacabcbacabcabacacbcabacac \\
S_2 &= abcbabcbcacbcabacbcabcbacbcabcacbcbabcba \\
S_3 &= abcbabcbcacbcacabacabcbabcbc \\
S_4 &= abcbabcbcacbcacabacabcbacabcabacacbcacbacab \\
S_5 &= abcbabcbcacbcabacbcabcbacbcabcacbabcacbcb
\end{aligned}
$$

Posets of square reductions

- Each square-free ternary word of leangth at least 9 contains one of the words X (up to alphabet permutation).
- For each $i \in 1, ..., 5$, $S_i \rightsquigarrow X_i$ and $S_i \rightsquigarrow Y_i$, therefore X_i and Y_i are in the same connected component.
- If $S = AX_iB$ then S is in the same connected component as AY_iB.
- For each $i \in 1, ..., 5$, $|X_i| > |Y_i|$.

Since each connected component contains square-free word of length at most 8, then G_3 has finitely many connected components.

Open problems

Question

Is there a ternary word W with $r(W) = 80$?

Other missing values up to 120: 95, 97, 101, 102, 104, 105, 107, 117, 119.

Conjecture

There exist infinitely many positive integers m such that no ternary word have exactly m distinct reducts.

Open problems

Conjecture (Fraenkel and Simpson)

Each word of length n has at most n distinct squares.

Conjecture

Each word of length n can be square reduced in one step to at most n different words.

Open problems

Conjecture

For every $k \geq 1$ G_k has finitely many connected components.

Questions?

The End