On the maximum number of non-standard squares in a word

Tomasz Kociumaka[1] Jakub Radoszewski[1]
Wojciech Rytter [1,2] **Tomasz Waleń**[1]

[1]Faculty of Mathematics, Informatics and Mechanics,
University of Warsaw, Warsaw, Poland
[kociumaka,jrad,rytter,walen]@mimuw.edu.pl

[2]Faculty of Mathematics and Computer Science,
Copernicus University, Toruń, Poland

Definition of square

Square is a factor $w_1 w_2$, such that $w_1 = w_2$.
For example: *aba aba* is a square.

Definition of square

Square is a factor $w_1 w_2$, such that $w_1 = w_2$.
For example: *aba aba* is a square.

Maximal number of distinct squares $SQ(n)$

The $SQ(n)$ denotes the maximal number of distinct squares in a word of length n.

Definition of square

Square is a factor $w_1 w_2$, such that $w_1 = w_2$.
For example: *aba aba* is a square.

Maximal number of distinct squares $SQ(n)$

The $SQ(n)$ denotes the maximal number of distinct squares in a word of length n.

Theorem (Ilie, 2007)

$$SQ(n) \leq 2n - O(\log n).$$

What about non-standard equalities?

Definition of $\approx$-square

For some binary relation $\approx$, the $\approx$-*square* is a factor $w_1 w_2$, such that $w_1 \approx w_2$.

What about non-standard equalities?

Definition of $\approx$-square

For some binary relation $\approx$, the $\approx$-*square* is a factor $w_1 w_2$, such that $w_1 \approx w_2$.

What about maximal number of distinct squares?

First we should precise what does it mean *distinct*:

- $SQ_{\approx}(n)$ denotes the maximal number of distinct factors (in a sense of $=$ relation) that are $\approx$-squares in a word of length n,
- $SQ'_{\approx}(n)$ denotes the maximal number of distinct factors (in a sense of $\approx$ relation) that are $\approx$-squares in a word of length n (valid for transitive $\approx$),

Candidates for $\approx$

$\approx_{ab}$ – abelian

$w_1 \approx_{ab} w_2$ if each character of the alphabet occurs the same number of times in w_1 and w_2.

Example

$$1321 \approx_{ab} 1213,$$

Abelian squares were first studied by Erdös [1961], who posed a question on the smallest alphabet size for which there exists an infinite abelian-square-free word.

Candidates for $\approx$

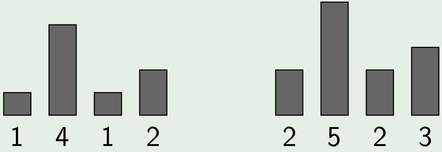
$\approx_{op}$ – order preserving

$w_1 \approx_{op} w_2$ if for all $1 \leq i,j \leq |w_1| = |w_2|$,

$$w_1[i] \leq w_1[j] \text{ iff } w_2[i] \leq w_2[j]$$

Example

$$1412 \approx_{op} 2523,$$



Candidates for $\approx$

$\approx_{param}$ – parametrized

(similar to $\approx_{op}$),
$w_1 \approx_{param} w_2$ if for all $1 \leq i, j \leq |w_1| = |w_2|$,

$$w_1[i] = w_1[j] \text{ iff } w_2[i] = w_2[j].$$

Example

$$1412 \approx_{param} 2123$$

Parameterized equality has been proposed by Baker [JCSS, 1995].

Abelian squares

Some examples of Abelian squares

$$u = 01001\ 11000$$

$$v = 00110\ 01001$$

u, v are different in sense of definition of SQ_{ab} (since $u \neq v$), but they are equivalent in sense of definition of SQ'_{ab} (since $u \approx_{ab} v$).

Abelian squares

Theorem

$$SQ_{ab}(n) = \Theta(n^2)$$

Abelian squares

Theorem

$$SQ_{ab}(n) = \Theta(n^2)$$

Proof.

Take word:

$$w(k) = 0^k 1 0^k 1 0^{2k}$$

it contains $\Theta(k^2)$ *ab*-squares of form:

$$0^a 1 0^b \quad 0^{k-b} 1 0^{a+2b-k}$$

for $k \leq a + b \leq 2k$. $\square$

Abelian squares

Theorem

$$SQ'_{ab}(n) = \Omega(n^{1.5}/\log n)$$

Abelian squares

Theorem

$$SQ'_{ab}(n) = \Omega(n^{1.5} / \log n)$$

Proof.

Take a word:

$$w(k) = \sum_{i=1}^{k} 0^i 1^i = 01\ 0011\ 000111\ \ldots\ 0^k 1^k$$

$\square$

Abelian squares, proof continued

Definition of $Sums_{i,j}$

Let

$$Sums_{i,j} = \{s_{p,q} : i \leq p \leq q \leq j\}$$

where $s_{p,q} = \sum_{k=p}^{q} k$.

Abelian squares, proof continued

Definition of $Sums_{i,j}$

Let

$$Sums_{i,j} = \{s_{p,q} : i \leq p \leq q \leq j\}$$

where $s_{p,q} = \sum_{k=p}^{q} k$.

Bounds on $Sums_{i,j}$

$$|Sums_{i,j}| = \Omega(|j-i|^2/\log j)$$

Abelian squares, notion of $(p, q)_{ab}$-squares

$(p, q)_{ab}$-square for $\Sigma = \{0, 1\}$

$w_1 w_2$ is $(p, q)_{ab}$-square if:

- $w_1 \approx_{ab} w_2$,
- there are exactly p characters 0 in w_1, w_2,
- there are exactly q characters 1 in w_1, w_2.

01001 11000, 00110 01001 are $(2, 3)_{ab}$-squares.

Abelian squares, proof continued

Lemma. $(t,t)_{ab}$-squares

For any $t = s_{p,q} \in Sums_{3k/4,k}$ the $(t,t)_{ab}$-square occurs in $w(k)$.

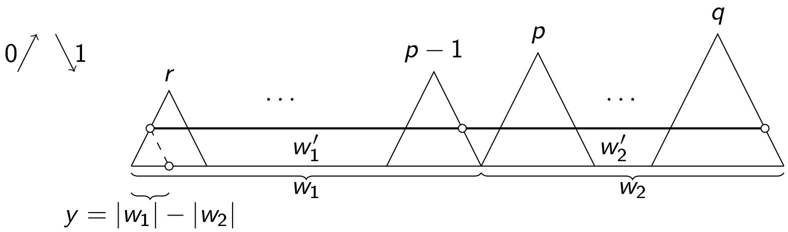
Proof.

Let $w_2 = 0^p 1^p \dots 0^q 1^q$, and $w_1 = 0^r 1^r \dots 0^{p-1} 1^{p-1}$ (where r is maximal such that $s_{r,p-1} \geq s_{p,q}$).

Let $y = |w_1| - |w_2|$:

- if $y = 0$ we are done,
- if not we adjust $w_1 w_2$ into $w_1' w_2'$:
  w_1' cutting away from w_1 first $y/2$ zeros and last $y/2$ ones,
  w_2' cut away from w_2 last $y/2$ ones and add $y/2$ zeros at the beginning.

□



Abelian squares, proof continued

Lemma. $(t, t+\delta)_{ab}$-squares

For any $t = s_{p,q} \in Sums_{3k/4,k}$ the $w(k)$ contains at least $k/4$ different $(t, s)_{ab}$-squares.

Proof.

Locate occurrence of $(t, t)_{ab}$-square and slightly extend it (if $s > t$), or shrink it (if $s < t$).
We can do use procedure for at least $k/4$ values of s. □

Abelian squares, proof continued

Finally

After combining previous lemmas we have

$$|Sums_{3k/4,k}| \cdot k/4 = \Theta(k^3/\log n)$$

different squares within word of length $\Theta(k^2)$, and this gives required bound $\Omega(n^2/\log n)$.

Order preserving squares

Theorem

For unbounded alphabet $SQ_{op}(n) = \Theta(n^2)$

Proof.

Take word:

$$w(k) = 123\ldots k$$

Every factor of $w(k)$ of even length is an order-preserving square. □

Order preserving squares

Theorem

For alphabet of constant size $SQ_{op}(n) = \Theta(n)$

Order preserving squares

Theorem

For alphabet of constant size $SQ_{op}(n) = \Theta(n)$

Proof.

Let $w_1 w_2$ is a $\approx_{op}$-square, there are three posibilities:

- case (a): $\Sigma(w_1) = \Sigma(w_2)$, so $w_1 = w_2$ and $w_1 w_2$ is regular square,
- case (b): $\Sigma(w_1) \cap \Sigma(w_2) = \emptyset$,
- case (c): $\Sigma(w_1) \cap \Sigma(w_2) \neq \emptyset$,

$\square$

Order preserving squares, $|\Sigma| = O(1)$

$\approx_{op}$-squares such that $\Sigma(w_1) = \Sigma(w_2)$

In this case $w_1 w_2$ is a regular square. Since there could be at most $2n$ of regular squares, there could be only $O(n)$ of such $\approx_{op}$-squares.

Order preserving squares, $|\Sigma| = O(1)$

$\approx_{op}$-squares such that $\Sigma(w_1) = \Sigma(w_2)$

In this case $w_1 w_2$ is a regular square. Since there could be at most $2n$ of regular squares, there could be only $O(n)$ of such $\approx_{op}$-squares.

$\approx_{op}$-squares such that $\Sigma(w_1) \cap \Sigma(w_2) = \emptyset$

We can change each letter from $\Sigma(w_2)$ to a letter from $\Sigma(w_1)$ and look for regular squares in such word. Once again there could be only $O(n)$ of such squares.

Order preserving squares, $|\Sigma| = O(1)$

$\approx_{op}$-squares such that $\Sigma(w_1) \cap \Sigma(w_2) \neq \emptyset$

Let $\Sigma_1 = \Sigma(w_1)$, $\Sigma_2 = \Sigma(w_2)$. We can prove that at each position there could be only a single occurrence of $\approx_{op}$-square $w_1' w_2'$ such that $\Sigma(w_1') = \Sigma_1$ and $\Sigma(w_2') = \Sigma_2$.

Order preserving squares, $|\Sigma| = O(1)$

$\approx_{op}$-squares such that $\Sigma(w_1) \cap \Sigma(w_2) \neq \emptyset$

Let $\Sigma_1 = \Sigma(w_1)$, $\Sigma_2 = \Sigma(w_2)$. We can prove that at each position there could be only a single occurrence of $\approx_{op}$-square $w_1' w_2'$ such that $\Sigma(w_1') = \Sigma_1$ and $\Sigma(w_2') = \Sigma_2$.

And finally

For $|\Sigma| = O(1)$, there are $O(1)$ possible of choices for $(\Sigma(w_1), \Sigma(w_2))$, and for each choice we have at most $O(n)$ different $\approx_{op}$-squares. $SQ_{op}(n) = O(n)$.

Summary

In this talk:

- $SQ_{ab}(n) = \Theta(n^2)$
- $SQ'_{ab}(n) = \Omega(n^{1.5}/\log n)$
- $SQ_{op}(n) = \Theta(n^2)$ for unbounded Σ,
- $SQ_{op}(n) = \Theta(n)$ for constant size Σ.

We also have:

- $SQ_{op}(n) = \Omega(nk)$ for $|\Sigma| = k$,
- $SQ_{param}(n) = \Theta(n^2)$ for unbounded Σ,
- $SQ_{param}(n) = \Theta(n)$ for $\Sigma = \{0, 1\}$.

Algorithms?

- for abelian: we can enumerate all $\approx_{ab}$-squares in $O(n^2/\log^2 n + Occ)$ for constant size alphabet,
- for order-preserving: we can enumerate all $\approx_{op}$-squares in $O(n \log n + Occ)$ time for constant size alphabet.

Words avoiding order preserving squares

Theorem

There exists inifite word over alphabet $\Sigma = \{0, 1, 2\}$ that avoid $\approx_{op}$-squares of length at least 4. (since it is impossible to avoid squares of length 2).

Proof.

Take any square free word (i.e. Thue-Morse word) over alphabet $\{0, 1, 2\}$.
Apply morphism:

$$h(x) = x1 \text{ for } x \in \{0, 1, 2\}$$

$\square$

Thank you for your attention!