Maximum number of distinct and nonequivalent nonstandard 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

DLT 2014, 2014–08–27

Definition of square

Square is a factor xy, such that $x = y$.
For example: *aba aba* is a square.

Definition of square

Square is a factor xy, such that $x = y$.
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 xy, such that $x = y$.
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)

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

What about non-standard equalities?

Definition of $\approx$-square

For binary relation $\approx$, the $\approx$-*square* is a factor xy, such that $x \approx y$.

What about non-standard equalities?

Definition of $\approx$-square

For binary relation $\approx$, the $\approx$-*square* is a factor xy, such that $x \approx y$.

Some candidates for $\approx$ relation

- Abelian equality,
- order preserving matching,
- parametrized matching.

Candidates for $\approx$

$\approx_{ab}$ – Abelian

$x \approx_{ab} y$ if each character of the alphabet occurs the same number of times in x and y.
In other words y is an anagram of x.

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 $\approx$

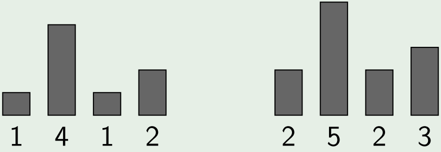
$\approx_{op}$ – order preserving

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

$$x[i] \leq x[j] \text{ iff } y[i] \leq y[j]$$

Example

$1412 \approx_{op} 2523$,



Candidates for $\approx$

$\approx_{param}$ – parametrized

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

$$x[i] = x[j] \text{ iff } y[i] = y[j].$$

Example

$$1412 \approx_{param} 2123$$

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

Maximal number of distinct squares

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 $\approx$),

For all "normal" relations $\approx$:

$$SQ_{\approx}(n) \geq SQ'_{\approx}(n)$$

Abelian squares

Some examples of Abelian squares

$$u = 01001\ 11000$$

$$v = 00110\ 01001$$

u, v are:

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

Abelian squares

Theorem

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

Abelian squares

Theorem

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

Proof.

Take word:

$$w_k = 0^k 1 0^k 1 0^{2k}$$

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

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

for $k \le a + b \le 2k$.
Note that $SQ'_{\mathrm{Abel}}(w_k) = \Theta(n)$. □

Abelian squares

Theorem

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

Abelian squares

Theorem

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

Proof.

Take a word:

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

Since $|\mathbf{w}_k| = \Theta(k^2)$ we have to show that it contains at least $\Theta(k^3 / \log n)$ different Abelian squares. □

Abelian squares, proof continued

Definition of $Sums_{i,j}$

Let

$$\text{Sums}(a, b) = |\{i \otimes j \ : a \leq i \leq j \leq b\}|.$$

where $i \otimes j \ = \ \sum_{t=i}^{j} \ t = (i + j)(j - i + 1)/2.$

Example

$\text{Sums}(2, 5) \ = \ \{2, 3, 4, 5, 7, 9, 12, 14\}.$

since $7 = 3 \otimes 4, \quad 9 = 2 \otimes 4 = 4 \otimes 5, \quad 12 = 3 \otimes 5, \quad 14 = 2 \otimes 5$

Abelian squares, proof continued

Definition of $Sums_{i,j}$

Let

$$\text{Sums}(a,b) = |\{i \otimes j \ : a \leq i \leq j \leq b\}|.$$

where $i \otimes j \ = \ \sum_{t=i}^{j} \ t = (i+j)(j-i+1)/2.$

Example

$\text{Sums}(2,5) \ = \ \{2,3,4,5,7,9,12,14\}.$

since $7 = 3 \otimes 4, \quad 9 = 2 \otimes 4 = 4 \otimes 5, \quad 12 = 3 \otimes 5, \quad 14 = 2 \otimes 5$

Bounds on $Sums_{i,j}$

This set is interesting since, it is quite dense:

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

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

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

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

- $x \approx_{ab} y$,
- there are exactly p characters 0 in x, and in y,
- there are exactly q characters 1 in x, and in y.

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

We will also use:

$$\mathbf{w}_{p,q} = \sum_{i=p}^{q} 0^i 1^i = 0^p 1^p 0^{p+1} 1^{p+1} \ \ldots \ 0^q 1^q$$

Abelian squares, proof continued

Lemma. Balanced Abelian squares – $(p, p)_{ab}$-squares

For any $p \in Sums_{\lceil 3k/4 \rceil, k}$ the $(p, p)_{ab}$-square occurs in $\mathbf{w}_k$.

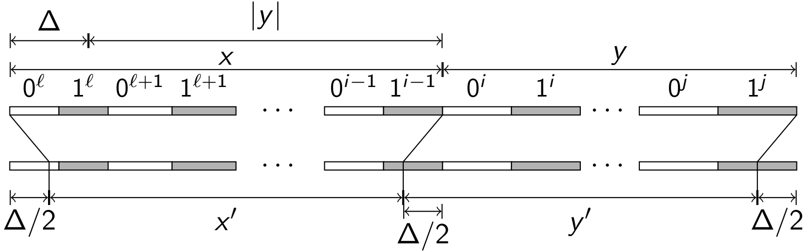
This lemma gives $\Theta(k^2/\log k)$ different Abelian squares in word $\mathbf{w}_k$ of length $\Theta(k^2)$.

Proof.

Let $p = i \otimes j$ and $\ell < i$ be the largest index s. t. $\ell \otimes (i-1) \geq p$. Take subwords $x = \mathbf{w}_{\ell,i-1}$, $y = \mathbf{w}_{i,j}$ of $\mathbf{w}_k$.

- if $|x| = |y|$, then xy is $(p,p)_{ab}$-square
- otherwise we can do some cutting and shifting of x and y. Let $\Delta = |x| - |y| > 0$. We modify x, y to obtain x', y':
  x': cut the first $\Delta/2$ zeros and the last $\Delta/2$ ones.
  y': add $\Delta/2$ ones on the left, and remove last $\Delta/2$ ones.

□



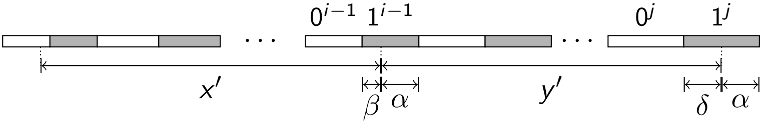
Abelian squares, proof continued

Lemma. $(p, p \pm \delta)_{ab}$-squares

For any $p = (i \otimes j) \in Sums_{\lceil 3k/4 \rceil, k}$ the $\mathbf{w}_k$ contains at least $k/4$ different $(p, p \pm \delta)_{ab}$-squares.

Proof.

Modify $(p, p)_{ab}$-square from previous lemma by slightly extending it or shrink it.
We can do that for at least $k/4$ values of δ. □



Abelian squares, proof continued

Finally

Combining previous lemmas we have

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

different Abelian squares within word $\mathbf{w}_k$ of length $\Theta(k^2)$, and this gives required bound $\Omega(n^{1.5}/\log n)$.

Order preserving squares, trivial bound on $SQ_{\mathrm{op}}(n)$

Theorem

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

Proof.

Take word:

$$w_k = 123 \dots k$$

Every factor of w_k of even length is an order-preserving square. □

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

Theorem

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

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

Theorem

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

Proof.

Let xy is a $\approx_{op}$-square, there are two possibilities:

- case (a): $\Sigma(x) = \Sigma(y)$, so $x = y$ and xy is regular square, so there could be $2n$ of such squares,
- case (b): $\Sigma(x) \neq \Sigma(y)$, we can show that there are $O(n)$ of such squares.

□

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

Lemma

Let w be a word of length n over an alphabet Σ, and let Σ_1, Σ_2 be two distinct subsets of Σ, $|\Sigma_1| = |\Sigma_2|$. Let f be a given bijection between Σ_1 and Σ_2. Then there are at most n distinct subwords of w of the form $xf(x)$, where $\mathrm{Alph}(x) = \Sigma_1$.

Example

Let

$$w = 123212313 22$$

$$\Sigma_1 = \{1, 2\}, \quad \Sigma_2 = \{1, 3\}, \quad f(1) = 1, f(2) = 3$$

The factor 212313 is of form $xf(x)$ ($x = 212$, $f(x) = 313$).

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

Proof.

Suppose a word $xf(x)$, where $\mathrm{Alph}(x) = \Sigma_1$, starts at position i.
Let $j > i$ be the first occurrence of a letter in $\Sigma_2 - \Sigma_1$, $w[j] = c$.
This letter is located in $f(x)$.
Let $k \geq i$ be the first occurrence of $f^{-1}(c)$.
Then $|x| = j - k$ and this uniquely determines the word $xf(x)$ as $w[i..i + 2(j - k) - 1]$.
So the number of such distinct subwords does not exceed n. □

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

And finally

For $|\Sigma| = O(1)$, there are $O(1)$ possible of choices for (Σ_1, Σ_2, f) with $\Sigma_1 \neq \Sigma_2$ and f being an non-decreasing bijection.
For each choice we have at most n different $\approx_{op}$-squares due to the previous lemma.

Words avoiding order preserving squares

Theorem

There exists infinite 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 τ (i.e. Thue-Morse word) over alphabet $\{0, 1, 2\}$.
Consider morphism:

$$\psi \quad : \; 0 \mapsto 10,\, 1 \mapsto 11,\, 2 \mapsto 12.$$

By case-by-case analysis we can prove that $\psi(\tau)$ avoids $\approx_{op}$-squares of length at least 4. □

Words avoiding parametrized cubes

Theorem

Let τ be the infinite Thue-Morse word. The word $\psi(\tau)$ is parameterized-cube-free.

Summary

In this talk:

- $SQ_{\mathrm{Abel}}(n) = \Theta(n^2)$
- $SQ'_{\mathrm{Abel}}(n) = \Omega(n^{1.5}/\log n)$
- $SQ_{\mathrm{op}}(n) = \Theta(n^2)$ for unbounded Σ,
- $SQ_{\mathrm{op}}(n) = \Theta(n)$ for constant size Σ,
- inifinite words avoiding op-squares, parametrized cubes.

Other results in the publication:

- $SQ'_{\mathrm{Abel}}(n, 2) = O(mn)$ where m is the number of blocks,
- $SQ_{\mathrm{op}}(n, k) = \Omega(kn)$,
- $SQ_{\mathrm{param}}(n) = \Theta(n^2)$ for unbounded Σ,
- $SQ_{\mathrm{param}}(n, 2) = \Theta(n)$.

Thank you for your attention!