A stronger square conjecture on binary words

Nataša Jonoska[1], Florin Manea[2], and Shinnosuke Seki[3]

[1] Department of Mathematics and Statistics, University of South Florida, USA.
[2] Institute für Informatik, Christian-Albrechts-Universität zu Kiel, Germany.
[3] Department of Information and Computer Science, Aalto University, Finland.

Talk at SOFSEM 2014, January 27th.

Square packing

What is square?

Square

A square is a word (sequence of letters) of the form xx.

Example

Squares	Non-squares
aa	aba
abab	bbab
baabbaab	baabbaaa
abaaababaaab	aaaaaaa
⋮	⋮

Square packing
Squares on a word

Counting rule

Don't count the same square twice or more times!

What squares does the following word (length 16) contain?

abbaaaabaaaaabaa

Square packing
Squares on a word

Counting rule

Don't count the same square twice or more times!

What squares does the following word (length 16) contain?

abbaaaabaaaaabaa

- aa (occurs 8 times but count is just 1)

Square packing
Squares on a word

Counting rule

Don't count the same square twice or more times!

What squares does the following word (length 16) contain?

abbaaaabaaaaabaa

- aa (occurs 8 times but count is just 1)
- aaaa

Square packing
Squares on a word

Counting rule

Don't count the same square twice or more times!

What squares does the following word (length 16) contain?

abbaaaabaaaaabaa

- aa (occurs 8 times but count is just 1)
- aaaa
- aaaabaaaaaba

Square packing
Squares on a word

Counting rule

Don't count the same square twice or more times!

What squares does the following word (length 16) contain?

abbaaaabaaaaabaa

- aa (occurs 8 times but count is just 1)
- aaaa
- aaaabaaaaaba
- aaabaaaaabaa

Square packing

Squares on a word

Counting rule

Don't count the same square twice or more times!

What squares does the following word (length 16) contain?

abbaaaabaaaaabaa

- aa (occurs 8 times but count is just 1)
- aaaa
- aaaabaaaaaba
- aaabaaaaabaa
- bb

Square packing

Squares on a word

Counting rule

Don't count the same square twice or more times!

What squares does the following word (length 16) contain?

ab<u>baaaabaaaa</u>abaa

- aa (occurs 8 times but count is just 1)
- aaaa
- aaaabaaaaaba
- aaabaaaaabaa
- bb
- baaaabaaaa

The count is 6

Square packing
Counting squares: example

Exercise

How many squares does the following word (length 20) contain?

$$a^{20} = \text{aaaaaaaaaaaaaaaaaaaa}$$

Square packing

Counting squares: example

Exercise

How many squares does the following word (length 20) contain?

$$\mathrm{a}^{20} = \mathrm{aaaaaaaaaaaaaaaaaaaa}$$

Answer

10 ($a^2, a^4, a^6, a^8, a^{10}, a^{12}, a^{14}, a^{16}, a^{18}, a^{20}$).

Square packing

Counting squares: example

Exercise

How many squares does the following word (length 20) contain?

$$\mathtt{a}^{20} = \mathtt{aaaaaaaaaaaaaaaaaaaa}$$

Answer

10 ($a^2, a^4, a^6, a^8, a^{10}, a^{12}, a^{14}, a^{16}, a^{18}, a^{20}$).

Proposition

The unary word of length n, that is, a^n, contains exactly $\lfloor n/2 \rfloor$ squares.

Square packing

Counting squares: example

Exercise

How many squares does the following word (length 20) contain?

abaabaaabaabaaabaaaa

Square packing

Counting squares: example

Exercise

How many squares does the following word (length 20) contain?

abaabaaabaabaaabaaaa

Answer

13 (a^2, $(aa)^2$, $(aab)^2$, $(aba)^2$, $(baa)^2$, $(aaba)^2$, $(abaa)^2$, $(baaa)^2$, $(aabaaab)^2$, $(abaaaba)^2$, $(abaabaa)^2$, $(baaabaa)^2$, $(baabaaa)^2$).

Square packing
Counting squares: example

Exercise

How many squares does the following word (length 20) contain?

abaabaaabaabaaabaaaa

Answer

13 (a^2, $(aa)^2$, $(aab)^2$, $(aba)^2$, $(baa)^2$, $(aaba)^2$, $(abaa)^2$, $(baaa)^2$, $(aabaaab)^2$, $(abaaaba)^2$, $(abaabaa)^2$, $(baaabaa)^2$, $(baabaaa)^2$).

Thus, binary words outperform unary words in packing squares.

Square packing
Quiz-setting

Square packing

Find a word that contains many squares relative to its length.

Upper bounds

Linear bounds; proved and conjectured

Theorem ([Fraenkel & Simpson 1998])

A word of length n contains at most $2n$ squares.

Idea of simple proof [Ilie 2005]

Ignore all but the rightmost occurrence of each square, and we can see that, at each position, at most 2 rightmost occurrences can start. □

Upper bounds

Linear bounds; proved and conjectured

Theorem ([Fraenkel & Simpson 1998])

A word of length n contains at most $2n$ squares.

Idea of simple proof [Ilie 2005]

Ignore all but the rightmost occurrence of each square, and we can see that, at each position, at most 2 rightmost occurrences can start. □

Square conjecture

A word of length n contains at most n squares.

Upper bounds

Conjecture

Unsuccessful attempt

If we could show that "at each position of a word, at most one rightmost occurrence could start," then the conjecture would be proved.

Counterexample

Consider *abaababaab*. The rightmost occurrences of *abaaba* and *abaababaab* start both on position 1.

Square-dense words
Examples

Known square-dense words include

Square-dense words
Examples

Known square-dense words include

- the following word by Fraenkel and Simpson:

$$\begin{aligned} w_{\mathrm{fs}} &= abaabaaabaabaaabaaaabaaabaaaabaaaaab\cdots \\ &= a^1ba^2ba^3ba^2ba^3ba^4ba^3ba^4ba^5b\cdots \end{aligned}$$

Square-dense words
Examples

Known square-dense words include

- the following word by Fraenkel and Simpson:

$$\begin{aligned} w_{\mathrm{fs}} &= abaabaaabaabaaabaaaabaaabaaaabaaaaab\cdots \\ &= a^1ba^2ba^3ba^2ba^3ba^4ba^3ba^4ba^5b\cdots \end{aligned}$$

- Let us propose a "simpler" one as:

$$w_{\mathrm{jms}} = abaabaaabaaaabaaaaab\cdots = a^1ba^2ba^3ba^4ba^5b\cdots .$$

Square-dense words
Examples

Known square-dense words include

- the following word by Fraenkel and Simpson:

$$\begin{aligned} w_{\mathrm{fs}} &= abaabaaabaabaaabaaaabaaabaaaabaaaaab\cdots \\ &= a^1ba^2ba^3ba^2ba^3ba^4ba^3ba^4ba^5b\cdots \end{aligned}$$

- Let us propose a "simpler" one as:

$$w_{\mathrm{jms}} = abaabaaabaaaabaaaaab\cdots = a^1ba^2ba^3ba^4ba^5b\cdots .$$

Note that both of these words are binary.

Square-densest words
Candidates are binary

Conjecture

Square-densest words are binary.

binary words with at most 1 b is square-sparse

Recall that the word a^n contains exactly $\lfloor n/2 \rfloor$ squares. Consider a word $a^i b a^{n-i-1}$ with 1 b. The b cannot make contribution to any square, so replacing it with a does not destroy any square. Hence, it contains at most $\lfloor n/2 \rfloor$ squares.

Stronger conjecture

Stronger conjecture on binary words

Let us strengthen the square conjecture by the # of occurrences of b's as follows.

Stronger conjecture on binary words

For $k \geq 2$, a binary word w of length n with k b's contains at most $\frac{2k-1}{2k+2}n$ squares.

Stronger conjecture

Asymptotic tightness

Theorem

Both of the words w_{fs} and w_{jms} achieve the tighter bound asymptotically.

Inductive approach to the stronger conjecture
Preliminaries

A binary word w with k b's can be represented as

$$w = a^{i_1} b a^{i_2} b \cdots a^{i_k} b a^{i_{k+1}}$$

for some coefficients $i_1, \ldots, i_{k+1} \geq 0$.

Coefficient sequence

The *coefficient sequence of w* means the sequence $i_1, \ldots, i_{k+1}$.

Coefficient set

The *coefficient set* of w, denoted by $I(w)$, is the **multiset** $\{i_1, i_2, \ldots, i_k, i_{k+1}\}$.

Inductive approach to the stronger conjecture

sufficient condition inequality

Let

$I(w)[j]$	the j-th smallest element of $I(w)$
$I(w)[\max]$	the maximum element of $I(w)$
$\#\mathrm{Sq}(w)$	the number of squares on the word w

Lemma

If a word w_k of length n with k b's satisfies

$$\#\mathrm{Sq}(w_k) \leq \left\lfloor \frac{I(w_k)[\max]}{2} \right\rfloor + \sum_{j=1}^{|I(w_k)|-2} (I(w_k)[j] + 1),$$

then $\#\mathrm{Sq}(w_k) \leq \frac{2k-1}{2k+2} n$.

Inductive approach to the stronger conjecture

Induction strategy

Induction hypothesis

The inequality holds for all words with at most $k-1$ b's.

Inductive approach to the stronger conjecture
Induction strategy

Induction hypothesis

The inequality holds for all words with at most $k-1$ b's.

1. Apply an operation to a word w_{k-1} with $k-1$ b's to yield a word w_k with k b's.

Inductive approach to the stronger conjecture

Induction strategy

Induction hypothesis

The inequality holds for all words with at most $k-1$ b's.

1. Apply an operation to a word w_{k-1} with $k-1$ b's to yield a word w_k with k b's.
2. Prove that the # of squares thus created is small enough to keep the inequality valid.

Inductive approach to the stronger conjecture

Induction strategy: example

Induction hypothesis

$w_{k-1} = a^1 b a^2 b \cdots a^{k-1} b a^k$ (with $k-1$ b's) satisfies the inequality, that is, $\#\mathrm{Sq}(w_{k-1}) \leq \lfloor k/2 \rfloor + \sum_{j=1}^{k-2} (j+1)$.

Inductive approach to the stronger conjecture

Induction strategy: example

Induction hypothesis

$w_{k-1} = a^1 b a^2 b \cdots a^{k-1} b a^k$ (with $k-1$ b's) satisfies the inequality, that is, $\#\mathrm{Sq}(w_{k-1}) \leq \lfloor k/2 \rfloor + \sum_{j=1}^{k-2}(j+1)$.

1. $w_{k-1} \leftarrow ba^{k+1} \Rightarrow w_k = a^1 b a^2 b \cdots a^{k-1} b a^k b a^k a$.

Inductive approach to the stronger conjecture

Induction strategy: example

Induction hypothesis

$w_{k-1} = a^1 b a^2 b \cdots a^{k-1} b a^k$ (with $k-1$ b's) satisfies the inequality, that is, $\#\mathrm{Sq}(w_{k-1}) \leq \lfloor k/2 \rfloor + \sum_{j=1}^{k-2}(j+1)$.

1. $w_{k-1} \leftarrow ba^{k+1} \Rightarrow w_k = a^1 b a^2 b \cdots a^{k-1} b a^k b a^k a$.
2. The catenation creates $(k-1)+1$ squares, that is, $(a^{k-1}ba)^2, (a^{k-2}ba^2)^2, \ldots, (aba^{k-1})^2, (ba^k)^2$.

Inductive approach to the stronger conjecture

Induction strategy: example

Induction hypothesis

$w_{k-1} = a^1ba^2b\cdots a^{k-1}ba^k$ (with $k-1$ b's) satisfies the inequality, that is, $\#\mathrm{Sq}(w_{k-1}) \leq \lfloor k/2 \rfloor + \sum_{j=1}^{k-2}(j+1)$.

1. $w_{k-1} \leftarrow ba^{k+1} \Rightarrow w_k = a^1ba^2b\cdots a^{k-1}ba^kba^ka$.
2. The catenation creates $(k-1)+1$ squares, that is, $(a^{k-1}ba)^2, (a^{k-2}ba^2)^2, \ldots, (aba^{k-1})^2, (ba^k)^2$.
3. Thus,

$$\begin{aligned}\#\mathrm{Sq}(w_k) &\leq \lfloor (k+1)/2 \rfloor + \textstyle\sum_{j=1}^{k-2}(j+1) + (k-1)+1 \\ &= \lfloor (k+1)/2 \rfloor + \textstyle\sum_{j=1}^{k-1}(j+1)\end{aligned}$$

Inductive approach to the stronger conjecture

Theorems

1) When the coefficients are pairwise-distinct

Theorem

For $k \geq 2$, let $w_k = a^{i_0} b a^{i_1} b \cdots b a^{i_k}$ be a binary word of length n with k b's. If the coefficients $i_1, \ldots, i_{k-1}$ are pairwise-distinct, then $\#\mathrm{Sq}(w_k) \leq \frac{2k-1}{2k+2} n$.

Recall $w_{\mathrm{jms}} = a^1 b a^2 b a^3 b a^4 b \ldots$. Thus, the tighter bound holds for any subword of w_{jms}.

Inductive approach to the stronger conjecture

Theorems

2) When the coefficient sequence is sorted

Theorem

For $k \geq 2$, let $w_k = a^{i_0}ba^{i_1}b \cdots ba^{i_{k-1}}ba^{i_k}$ be a binary word of length n with k b's. If $i_1 \leq i_2 \leq \cdots \leq i_{k-1}$, then $\#\mathrm{Sq}(w_k) \leq \frac{2k-1}{2k+2}n$.

Inductive approach to the stronger conjecture
Theorems

3) Pairwise-distinct extrema

A *maximum* of the coefficient sequence $i_0, \ldots, i_k$ is its subsequence $i_\ell, i_{\ell+1}, \ldots, i_{r-1}, i_r$ such that $i_\ell < i_{\ell+1} = \cdots = i_{r-1} > i_r$. The notion of *minima* is defined analogously.

Theorem

For $k \geq 2$, let w_k be a binary word of length n with k b's. If all maxima of its coefficient sequence are pairwise-distinct, then $\#\mathrm{Sq}(w_k) \leq \frac{2k-1}{2k+2}n$.

Recall $w_{\mathrm{fs}} = a^1ba^2ba^3ba^2ba^3ba^4ba^3ba^4ba^5b\cdots$, and its maxima $i_2i_3i_2, i_3i_4i_3, i_4i_5i_4, \ldots$ are pairwise-distinct. Hence, the tighter bound holds for any subword of w_{fs}.

Inductive approach to the stronger conjecture
Theorems

4) Small # of b's

proposition

For $2 \leq k \leq 9$, any word w of length n with at most k b's satisfies $\#\mathrm{Sq}(w) \leq \frac{2k-1}{2k+2}n$.

Summary of the solved cases

The tighter bound holds if one of the following conditions is safisfied:

1. Extrema of the coefficient sequence are pairwise-distinct, including the cases when:
   - the coefficients are pairwise-distinct
   - the sequence is sorted
2. Multiplicity of coefficients is at most 4.
3. $|w|_b \leq 9$.

Open problems

- Prove the bound $\frac{2k-1}{2k+2}n$ in general,
- or Find a counterexample (a binary word of length n with k b's that contains more than $\frac{2k-1}{2k+2}n$ squares).
- Prove that square-densest words are binary.
- Generalize the bound for arbitrary alphabets.

Acknowledgements

Thank you very much for your attention!

References I

[Fraenkel & Simpson 1998]
A. S. Fraenkel and J. Simpson.
How many squares can a string contain?
Journal of Combinatorial Theory, Series A 82, pp.112-120, 1998.

[Ilie 2005]
L. Ilie.
A simple proof that a word of length n has at most $2n$ distinct squares.
Journal of Combinatorial Theory, Series A 112(1), pp.163-164, 2005.

References II

[Jonoska, Manea, S. 2013]
N. Jonoska, F. Manea, and S. Seki.
Stronger square conjecture on binary words.
SOFSEM 2014, to appear.