How many double squares can a string contain?

F. Franek, joint work with A. Deza and A. Thierry

Advanced Optimization Laboratory
Department of Computing and Software
McMaster University, Hamilton, Ontario, Canada

Séminaire d'algorithmique et de complexité
du plateau de Saclay
February, 2014

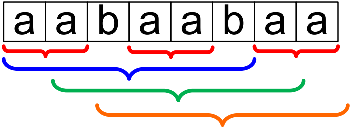
Outline

Motivation and background

We are dealing with finite strings over finite alphabets. There is no particular requirement about the order of the alphabet.

What is the *maximum number of distinct squares problem* ?

We are counting types of squares rather than their occurrences.



has 6 occurrences of squares, but only 4 distinct squares, *aa*, *aabaab*, *abaaba*, and *baabaa*.

A trivial bound: the number of all occurrences of primitively rooted squares in a string of length n is bounded by $O(n \log n)$ (*Crochemore 1978*) and the number of distinct non-primitively rooted squares is $O(n)$ (*Kubica et al. 2013*)

Could it be $O(n)$? And if so, what would be the constant?

Why this is not simple? In a string of length n, $O(\log n)$ squares can start at the same position!

a	a	b	a	a	b	a	a	b	a	a	b

It is easy to compute it for short strings, so why induction cannot be used?

a	a	b	a	a	b

+

a	a	b	a	a	b

a	a	b	a	a	b	a	a	b	a	a	b

Concatenation does both “destroys" existing types through multiple-occurrences and “creates" new types. Of course, same holds true for the reverse process - partitioning of strings.

Theorem (*Fraenkel-Simpson*, 1998)

There are at most $2n$ distinct squares in a string of length n.

Count only the **rightmost** *occurrences. Fraenkel-Simpson showed that if there are three rightmost squares uu, vv, and ww starting at the same position so that $|u| < |v| < |w|$, then ww contains a farther copy of uu, based on Crochemore-Rytter (1995) Lemma showing that in such a case, $|w| \geq |u| + |v|$.*

Fraenkel-Simpson hypothesized that the number of distinct squares should be bounded by n, i.e.

$$\sigma(n) \leq n$$

where $\sigma(n) = \max\{\, s(x) \;:\; x \textit{ is a string of length } n \,\}$.

Fraenkel-Simpson gave an infinite sequence of strings $\{x_n\}_{n=1}^{\infty}$ so that $|x_n| \nearrow \infty$ and

$$\frac{s(x_n)}{|x_n|} \nearrow 1$$

where $s(x) = \textit{number of distinct squares in } x$.

- In 2005 *Ilie* provided a simpler proof of *Fraenkel-Simpson*'s Theorem and in 2007 presented an asymptotic upper bound of $2n-\theta(\log n)$.

- In 2011 *Deza-F.* proposed a d-step approach to the problem and conjectured that $\sigma_d(n) \leq n - d$, where $\sigma_d(n) = \max\{\, s(x) \;:\; x \textit{ is a string of length } n \textit{ with } d \textit{ distinct symbols} \,\}$.

Basic notions and tools

Definition

non-trivial power of a string x is a concatenation of m copies of x denotes as x^m; x^2 is a *square*, x^3 a *cube*.

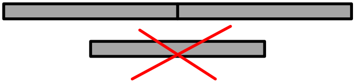
A string x is *primitive* if $x \neq y^n$ for any y and any $n \geq 2$.

primitive root of x is the smallest primitive y so that $x = y^n$.

x and y are *conjugates* if $x = uv$ and $y = vu$ for some u, v.

Lemma (*Synchronization principle*)

Given a primitive string x, a proper suffix y of x, a proper prefix z of x, and $m \geq 0$, there are exactly m occurrences of x in $yx^m z$.



Lemma (*Common factor lemma*)

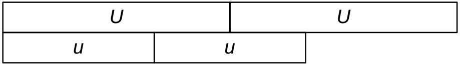
For any strings x and y, if a non-trivial power of x and a non-trivial power of y have a common factor of length $|x|+|y|$, then the primitive roots of x and y are conjugates. In particular, if x and y are primitive, then x and y are conjugates.

Double squares

- *Fraenkel-Simpson*: only two rightmost squares can start at the same position. Thus, only one rightmost square or two rightmost squares may start at any position.

- *Lam* (2009 – *unpublished*) tried bounding the number of *double squares* and hence bound the number of distinct squares. His approach is based on a taxonomy of all possible configurations of two double squares yielding a bound of $\frac{94}{48}n \approx 1.98n$.

A configuration of two squares



has been investigated in many different contexts:

- *Smyth et. al.*: with intention to find a position for amortization argument for runs conjecture.
- in computational framework by *Deza-F.-Jiang*: such configurations are used in *Liu*'s Ph.D. thesis to speed up computation of $\sigma_d(n)$.
- *Lam*: two rightmost squares have a unique structure.

Lemma

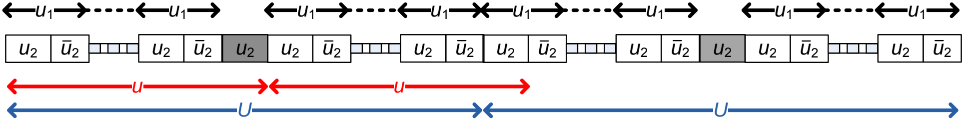
Let uu and UU be two squares in a string x starting at the same position with $|u| < |U|$ such that either

(b) *both uu and UU are rightmost occurrences, or*

(a) $|U| < |uu|$ *and either uu or UU is primitively rooted.*

Then $|u| < |U| < |uu| < |UU|$ and there is a unique primitive string u_1, a unique proper prefix u_2 of u_1, and unique integers e_1 and e_2 satisfying $1 \le e_2 \le e_1$ such that $\boldsymbol{u} = \boldsymbol{u_1}^{e_1}\boldsymbol{u_2}$ and $\boldsymbol{U} = \boldsymbol{u_1}^{e_1}\boldsymbol{u_2}\boldsymbol{u_1}^{e_2}$; i.e. uu and UU form a double square.

$$u_1^{\mathcal{U}(1)}\, u_2\, u_1^{\mathcal{U}(2)}\, u_1^{\mathcal{U}(1)}\, u_2\, u_1^{\mathcal{U}(2)}$$



Thus, only strings of length at least 10 may contain a double square: $|UU| = 2((\mathcal{U}(1)+\mathcal{U}(2))|u_1|+|u_2|) \geq 2((1+1)2+1) = 10$.

Cyclic shift (rotation) to the right is controlled by

$$lcp(u_1, \overline{u}_1)$$

while cyclic shift to the left is controlled by

$$lcs(u_1, \overline{u}_1)$$

lcp = largest common prefix
lcs = largest common suffix

$$u_1 = aaabaa,\ u_2 = aaab,\ \overline{u}_2 = aa,\ \mathcal{U}(1) = \mathcal{U}(2) = 2$$

aaabaaaaabaaaaabaaabaaaaabaaaaabaaaaabaaaaabaaabaaaaabaaaaa

[ ] [ ) ( ])

aaabaaaaabaaaaabaaabaaaaabaaaaabaaaaabaaaaabaaabaaaaabaa . . .

[ ] [ ) ( ])

. aabaaaaabaaaaabaaabaaaaabaaaaabaaaaabaaaaabaaabaaaaabaaa . .

[ ] [ ) ( ])

. . abaaaaabaaaaabaaabaaaaabaaaaabaaaaabaaaaabaaabaaaaabaaaa .

[ ] [ ) ( ])

. . . baaaaabaaaaabaaabaaaaabaaaaabaaaaabaaaaabaaabaaaaabaaaaa

$u_1 = aaabaa$, $u_2 = aaab$, $\overline{u}_2 = aa$, $\mathcal{U}(1) = 2$, and $\mathcal{U}(2) = 1$.

aaabaaaaabaaaaabaaabaaaaabaaaaabaaaaabaaabaaaaa

[ ][ )( ])

aaabaaaaabaaaaabaaabaaaaabaaaaabaaaaabaaabaa . . .

[ ][ )( ])

. **aabaaaaabaaaaabaaabaaaaabaaaaabaaaaabaaabaaa** . .

[ ][ )( ])

. . **abaaaaabaaaaabaaabaaaaabaaaaabaaaaabaaabaaaa** .

[ ][ )( ])

. . . **baaaaabaaaaabaaabaaaaabaaaaabaaaaabaaabaaaaa**

Definition

For a double square $\mathcal{U}$, $\overline{v}vv\overline{v}$ where $|\overline{v}| = |\overline{u}_2|$ and $|v| = |u_2|$ is an *inversion factor*

$$\mathcal{U} = u_1{}^{\mathcal{U}(1)} u_2 u_1{}^{\mathcal{U}(2)+\mathcal{U}(1)} u_2 u_1{}^{\mathcal{U}(2)} =$$

$$u_1{}^{(\mathcal{U}(1)-1)} u_2 \underbrace{\overline{u}_2 u_2 u_2 \overline{u}_2}_{N_1} u_1{}^{\mathcal{U}(2)+\mathcal{U}(1)-2} u_2 \underbrace{\overline{u}_2 u_2 u_2 \overline{u}_2}_{N_2} u_1{}^{(\mathcal{U}(2)-1)}$$

N_1 N_2

natural inversion factors

A cyclic shift of an inversion factor is an inversion factor, also controlled by $lcp(u_1, \overline{u}_1)$ and $lcs(u_1, \overline{u}_1)$.

```
                      L1                                      L2
[                          ] [          ) (                ]                      )
aaabaaaaabaaaaabaaaaabaaaaabaaabaaaaabaaaaabaaaaabaaaaabaaaaabaaaaabaaabaaaaabaa
aaabaaaaabaaaaabaaaaabaaaabaaabaaaabaaaaabaaaaabaaaaabaaaaabaaaabaaabaaaabaa
aaabaaaaabaaaaabaaaaabaaaaabaaabaaaaabaaaaabaaaaabaaaaabaaaaabaaaaabaaabaaaaabaa
aaabaaaaabaaaaabaaaaabaaaaabaaabaaaaabaaaaabaaaaabaaaaabaaaaabaaaaabaaabaaaaabaa
                         R1                                       R2
```

All inversion factors are cyclic shifts of the natural ones:

Lemma (*Inversion factor lemma*)

Given a double square $\mathcal{U}$, there is an inversion factor of $\mathcal{U}$ within the string UU starting at position i $\iff$ $i \in [L_1, R_1] \cup [L_2, R_2]$.

Inversion factor lemma for distinct squares

Theorem (*Fraenkel-Simpson, Ilie*)

At most two rightmost squares can start at the same position.

Let us assume that 3 rightmost squares uu, UU, vv start at the same position.
By item (c) of Inversion factor lemma, uu and UU form a double square $\mathcal{U}$: $u = u_1{}^{\mathcal{U}(1)}u_2$ and $U = u_1{}^{\mathcal{U}(1)}u_2u_1{}^{\mathcal{U}(2)}$.
Since the first v contains an inversion factor, the second v must also contain an inversion factor.

Cont. on the next slide

Cont. from the previous slide

If the inversion factor in the second v were from $[L_2, R_2]$, then $|v| = |U|$, a contradiction.
Hence v must not contain an inversion factor from $[L_2, R_2]$ and so $u_1{}^{\mathcal{U}(1)} u_2 u_1{}^{\mathcal{U}(1)+\mathcal{U}(2)-1} u_2$ must be a prefix of v.
Therefore vv contains another copy of $u_1{}^{\mathcal{U}(1)} u_2 u_1{}^{\mathcal{U}(1)} u_2 = uu$, a contradiction.

Fundamental Lemma:

Lemma

Let x be a string starting with a double square $\mathcal{U}$. Let $\mathcal{V}$ be a double square with $\mathfrak{s}(\mathcal{U}) < \mathfrak{s}(\mathcal{V})$, then either

(a) $\mathfrak{s}(\mathcal{V}) < R_1(\mathcal{U})$, *in which case either*

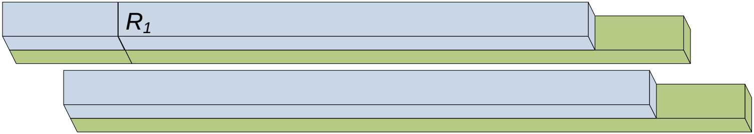
- (a_1) $\mathcal{V}$ *is an* α-*mate of* $\mathcal{U}$ (cyclic shift), *or*
- (a_2) $\mathcal{V}$ *is a* β-*mate of* $\mathcal{U}$ (cyclic shift of U to V), *or*
- (a_3) $\mathcal{V}$ *is a* γ-*mate of* $\mathcal{U}$ (cyclic shift of U to v), *or*
- (a_4) $\mathcal{V}$ *is a* δ-*mate of* $\mathcal{U}$ (big tail),

or

(b) $R_1(\mathcal{U}) \leq \mathfrak{s}(\mathcal{V})$, *then*

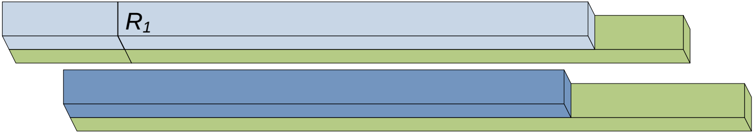
- (b_1) $\mathcal{V}$ *is a* ε-*mate of* $\mathcal{U}$ (big gap).

α-mate (cyclic shift):




```
[                   ][     )(           ]              )
aaabaaaaabaaaaabaaabaaaaabaaaaabaaaaabaaabaa...
 [                   ][     )(           ]              )
.aabaaaaabaaaaabaaabaaaaabaaaaabaaaaabaaabaa..
```

β-mate (cyclic shift of U to V)



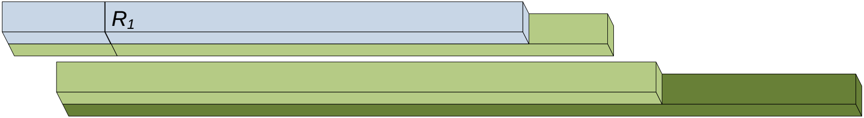
[ ] [ ) ( ])

baaaaabaaaaabaaaaabaaaaabaaaaabaaabaaaaabaaaaabaaaaabaaaaabaaaaabaaaaabaaabaaaaa. . . .

[ ] [ ) ( ])

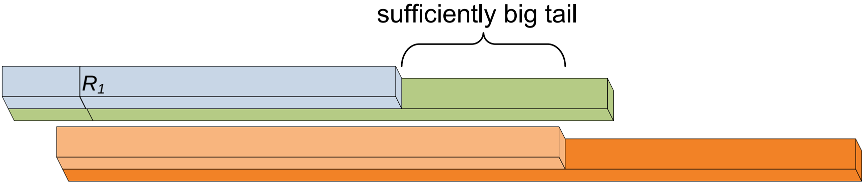
. . . aaabaaaaabaaaaabaaaaabaaaaaaabaaabaaaaabaaaaabaaaaabaaaaabaaaaaaabaaabaaaaabaa.

γ-mate (cyclic shift of U to v)



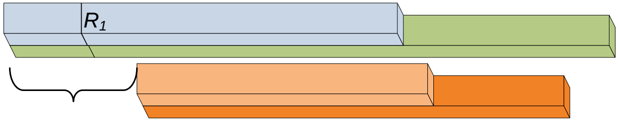

```
[                    ][  )(              ]        )
aabaabaabaabaabaaabaabaabaabaabaabaaab
        [                         ][            )(           ]                           )
        aabaabaabaaabaabaabaabaabaabaabaaabaabaabaaabaabaabaabaabaaba
```

δ-mate (big tail)




```
[              ][ )(           ]     )
aabaabaabaabaabaaabaabaabaabaabaabaaab
      [                           ][                  )(         ]                                     )
      aabaabaabaaabaabaabaabaabaabaaaabaabaabaaabaabaabaabaabaabaaabaabaabaabaabaaaabaabaabaaabaabaab
```

ε-mate (big gap)



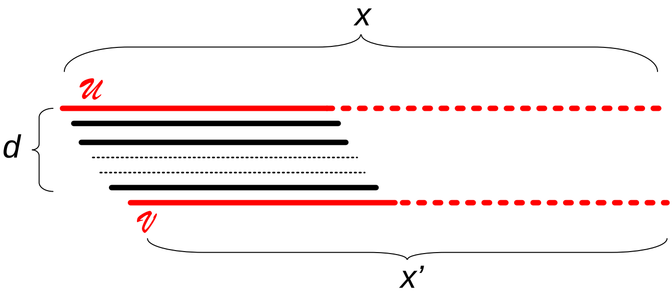
sufficiently big gap

```
[              ][ )(           ]     )
aabaabaabaabaabaaabaabaabaabaabaabaaab
                [                 ][    )(           ]           )
                aabaabaabaabaabaabaaabaabaabaabaabaabaabaabaaabaab
```

An upper bound for the number of double squares

We show by induction a bound $\delta(x) \leq \frac{5}{6}|x| - \frac{1}{3}|u|$, where uu is the shorter square of the leftmost double square of x.

The fundamental lemma basically says that either the gap $G(\mathcal{U},\mathcal{V})$ is "big" or the tail $T(\mathcal{U},\mathcal{V})$ is "big" (for δ-mate and ε-mate), or it is case of α-mate, β-mate, or γ-mate.



Lemma (Gap-Tail lemma)

$\delta(x') \leq \frac{5}{6}|x'| - \frac{1}{3}|v|$ *implies*

$$\delta(x) \leq \frac{5}{6}|x| - \frac{1}{3}|u| + d - \frac{1}{2}|G(\mathcal{U},\mathcal{V})| - \frac{1}{3}|T(\mathcal{U},\mathcal{V})|$$

We deal with α-mates, β-mates, and γ-mates separately.

It is possible as they form families, either a pure α-family, or $\alpha+\beta$-family, or $\alpha+\beta+\gamma$-family.

$\mathcal{U}$-family consists only of α-mates

Illustration of α-family with $\mathcal{U}(1) = \mathcal{U}(2)$

```
aaabaaaaabaaaaabaaabaaaaabaaaaabaaaaabaaaaabaaabaaaaabaaaaa

[                    ] [               ) (   ]                               )
aaabaaaaabaaaaabaaabaaaaabaaaaabaaaaabaaaaabaaabaaaaabaa...
 [                    ] [               ) (   ]                               )
.aabaaaaabaaaaabaaabaaaaabaaaaabaaaaabaaaaabaaabaaaaabaaa..
  [                    ] [               ) (   ]                               )
..abaaaaabaaaaabaaabaaaaabaaaaabaaaaabaaaaabaaabaaaaabaaaa.
   [                    ] [               ) (   ]                               )
...baaaaabaaaaabaaabaaaaabaaaaabaaaaabaaaaabaaabaaaaabaaaaa
```

Illustration of α-family with $\mathcal{U}(1) > \mathcal{U}(2)$

aaabaaaaabaaaaabaaabaaaaabaaaaabaaaaabaaabaaaaa

[ ][ )( ])

aaabaaaaabaaaaabaaabaaaaaabaaaaabaaaaabaaabaa...

[ ][ )( ])

.**aabaaaaabaaaaabaaabaaaaabaaaaabaaaaabaaabaaa**..

[ ][ )( ])

..**abaaaaabaaaaabaaabaaaaabaaaaabaaaaabaaabaaaa**.

[ ][ )( ])

...**baaaaabaaaaabaaabaaaaabaaaaabaaaaabaaabaaaaa**

It is easy to estimate the size of α-family, as it is controlled by $lcp(u_1, \overline{u}_1)$ and $lcp(y, u_2)$ where y is x without UU: the size $\leq |u_1|$.

- Either there are no other double squares, and then it can be shown directly that the bound holds, or
- There is a $\mathcal{V}$ underneath, and we can use induction using the Gap-Tail lemma. $\mathcal{V}$ must be either γ-mate, or δ-mate, or ε-mate, and the Gap-Tail lemma can be applied to propagate the bound.

$\mathcal{U}$-family consists of α-mates and β-mates

Illustration of α+β-family

```
aaabaaaaabaaaaabaaaaabaaaaabaaaaabaaabaaaaabaaaaabaaaaabaaaaabaaaaabaaaaabaaabaaaaabaaaaaaabaaaaa

[                              ][    )(                      ]          )
aaabaaaaabaaaaabaaaaabaaaaabaaaaabaaabaaaaabaaaaabaaaaabaaaaabaaaaabaaaaabaaabaa..............α-segment starts
 [                              ][    )(                      ]          )
.aabaaaaabaaaaabaaaaabaaaaabaaaabaaabaaaaabaaaaabaaaaabaaaaabaaaaabaaaabaaabaaaa.............
  [                              ][    )(                      ]          )
..abaaaaabaaaaabaaaaabaaaaabaaaaabaaabaaaaabaaaaabaaaaabaaaaabaaaaabaaaaabaaabaaaa............
   [                              ][    )(                      ]          )
...baaaaabaaaaabaaaaabaaaaabaaaaabaaabaaaaabaaaaabaaaaabaaaaabaaaaabaaaaabaaabaaaaa...........
     [                       ][          )(              ]                      )
.....aaabaaaaabaaaaabaaaaabaaaaabaaabaaaaabaaaaabaaaaabaaaaabaaaaabaaaaabaaabaaaaaabaa.........β-segment starts
      [                       ][          )(              ]                      )
......aabaaaaabaaaaabaaaaabaaaabaaabaaaaabaaaaabaaaaabaaaaabaaaaabaaaabaaabaaaaaabaaa........
       [                       ][          )(              ]                      )
.......abaaaaabaaaaabaaaaabaaaaabaaabaaaaabaaaaabaaaaabaaaaabaaaaabaaaaabaaabaaaaaabaaaa.......
        [                       ][          )(              ]                      )
........baaaaabaaaaabaaaaabaaaaabaaabaaaaabaaaaabaaaaabaaaaabaaaaabaaaaabaaabaaaaaabaaaaa......
          [                ][                )(  ]                                 )
..........aaabaaaaabaaaaabaaaaabaaabaaaaabaaaaabaaaaabaaaaabaaaaabaaaaabaaabaaaaaabaaaaabaa...β-segment starts
           [                ][                )(  ]                                 )
...........aabaaaaabaaaaabaaaabaaabaaaaabaaaaabaaaaabaaaaabaaaaabaaaabaaabaaaaaabaaaaabaaa..
            [                ][                )(  ]                                 )
............abaaaaabaaaaabaaaaabaaabaaaaabaaaaabaaaaabaaaaabaaaaabaaaaabaaabaaaaaabaaaaabaaaa.
             [                ][                )(  ]                                 )
.............baaaaabaaaaabaaaaabaaabaaaaabaaaaabaaaaabaaaaabaaaaabaaaaabaaabaaaaaabaaaaabaaaaa
```

It is more complicated to estimate the size of a $\alpha+\beta$-family:

$$\leq \begin{cases} \lceil \frac{\mathcal{U}(1)-\mathcal{U}(2)}{2} \rceil |u_1| & \text{if } \mathcal{U}(2) = 1 \\ \frac{\mathcal{U}(1)-\mathcal{U}(2)}{2} |u_1| & \text{if } \mathcal{U}(2) > 1 \end{cases}$$

- Either there are no other double squares, and then it can be shown directly that the bound holds, or
- There is a $\mathcal{V}$ underneath, and we can use induction using the Gap-Tail lemma. $\mathcal{V}$ must be either δ-mate, or ε-mate, and the Gap-Tail lemma can be applied to propagate the bound. (*Special care needed for ε-mate case and super-ε-mate must be put in play !*)

$\mathcal{U}$-family consists of α-mates, β-mates, and γ-mates

Illustration of $\alpha+\beta+\gamma$-family

```
              R_1
[              ][ )(           ]     )     type
aabaabaabaabaabaaabaabaabaabaabaabaaab     5  1      <--- start of α-segment
 [              ][ )(           ]     )
 abaabaabaabaabaaabaabaabaabaabaabaaaba     5  1     <--- end of α-segment
   [           ][    )(     ]           )
   aabaabaabaabaaabaabaabaabaabaabaaabaab    4  2        <--- start of β-segment
    [           ][    )(     ]           )
    abaabaabaabaaabaabaabaabaabaabaaabaaba    4  2       <--- end of β-segment
      [                 ][        )(       ]                   )
      aabaabaabaaabaabaabaabaabaabaaabaabaabaaabaabaabaabaabaaba   3  3    <--- start of γ-segment
       [                 ][        )(       ]                   )
       abaabaabaaabaabaabaabaabaabaaabaabaabaaabaabaabaabaabaabaa   3  3
        [                 ][        )(       ]                   )
        baabaabaaabaabaabaabaabaabaaabaabaabaaabaabaabaabaabaabaaa   3  3
         [                 ][        )(       ]                   )
         aabaabaaabaabaabaabaabaabaaabaabaabaaabaabaabaabaabaabaaab   2  4 not a double square
          [                 ][        )(       ]                   )
          abaabaaabaabaabaabaabaabaaabaabaabaaabaabaabaabaabaabaaaba    2  4 not a double square
           [                 ][        )(       ]                   )
           baabaaabaabaabaabaabaabaaabaabaabaaabaabaabaabaabaabaaabaa     2  4   not a double square
            [                 ][        )(       ]                   )
            aabaaabaabaabaabaabaabaaabaabaabaaabaabaabaabaabaabaaabaab     2  4   not a double square
             [                 ][        )(       ]                   )
             abaaabaabaabaabaabaabaaabaabaabaaabaabaabaabaabaabaaabaaba     1  5   not a double square
              [                 ][        )(       ]                   )
              baaabaabaabaabaabaabaaabaabaabaaabaabaabaabaabaabaaabaabaa     1  5   not a double square
              R_1
```

It is quite complex to estimate the size of a $\alpha{+}\beta{+}\gamma$-family:

$$\leq \frac{2}{3}(\mathcal{U}(1)+1)|u_1|$$

- Either there are no other double squares, and then it can be shown directly that the bound holds, or
- There is a $\mathcal{V}$ underneath, and we can use induction using the Gap-Tail lemma. $\mathcal{V}$ must be either δ-mate, or ε-mate, and the Gap-Tail lemma can be applied to propagate the bound.

Main theorems

Theorem

The number of double squares in a string of length n is bounded by $\lfloor 5n/6 \rfloor$.

Corollary

The number of distinct squares in a string of length n is bounded by $\lfloor 11n/6 \rfloor$.

- We presented a universal upper bound of $\frac{11n}{6}$ for the maximum number of distinct squares in a string of length n
- A bound of $\frac{5n}{6}$ for the maximum number of double squares
- It improves the universal bound of $2n$ by Fraenkel-Simpson
- It improves the asymptotic bound of $2n - \Theta(n)$ by Ilie
- The combinatorics of double squares is interesting on its own and possibly can be used for some other problems

THANK YOU

M. Crochemore and W. Rytter.
Squares, cubes, and time-space efficient string searching.
Algorithmica, 13:405–425, 1995.

A. Deza and F. Franek.
A *d*-step approach to the maximum number of distinct squares and runs in strings.
Discrete Applied Mathematics, 163:268–274, 2014.

A. Deza, F. Franek, and M Jiang.
A computational framework for determining square-maximal strings.
In J. Holub and J. Žďárek, editors, *Proceedings of the Prague Stringology Conference 2012*, pages 111–119, Czech Technical University in Prague, Czech Republic, 2012.

A.S. Fraenkel and J. Simpson.
How many squares can a string contain?
Journal of Combinatorial Theory, Series A, 82(1):112–120, 1998.

F. Franek, R.C.G. Fuller, J. Simpson, and W.F. Smyth.
More results on overlapping squares.
Journal of Discrete Algorithms, 17:2–8, 2012.

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):163–163, 2005.

L. Ilie.
A note on the number of squares in a word.
Theoretical Computer Science, 380(3):373–376, 2007.

E. Kopylova and W.F. Smyth.
The three squares lemma revisited.
Journal of Discrete Algorithms, 11:3–14, 2012.

M. Kubica, J. Radoszewski, W. Rytter, and T. Waleń.
On the maximum number of cubic subwords in a word.
European Journal of Combinatorics, 34:27–37, 2013.

N. H. Lam.
On the number of squares in a string.
AdvOL-Report 2013/2, McMaster University, 2013.

M. J. Liu.
Combinatorial optimization approaches to discrete problems.
Ph.D. thesis, Department of Computing and Software, McMaster University, 2013.