(Trying to do a)
Counting of distinct repetitions in words[1]

One World Combinatorics

Robert Mercaș

08.02.2021

Loughborough University

[1]Szilárd Zsolt Fazekas and RM

1. Preliminaries
2. Basics
3. Results
4. Final Remarks

Basics

[2] https://cs.uwaterloo.ca/∼shallit/repetitions.html

Basics

hotshots[2]

[2] https://cs.uwaterloo.ca/∼shallit/repetitions.html

Basics

hotshots[2] shshsh

[2]https://cs.uwaterloo.ca/∼shallit/repetitions.html

Basics

hotshots[2] shshsh tratratratra

[2] https://cs.uwaterloo.ca/∼shallit/repetitions.html

Basics

hotshots[2] shshsh tratratratra

kth power: $x^k = x \cdot x \cdot \dots \cdot x$, where x is some word

[2] https://cs.uwaterloo.ca/~shallit/repetitions.html

Counting squares

Number of kth powers

Counting squares

Number of kth powers: $\mathcal{O}(n^2)$

$$aaa \cdots a$$

Counting squares

Number of kth powers: $\mathcal{O}(n^2)$

$$aaa \cdots a$$

Number of kth powers with primitive root

Counting squares

Number of kth powers: $\mathcal{O}(n^2)$

$$aaa \cdots a$$

Number of kth powers with primitive root:
$< n \log n$ [Bannai et al., 2020]

Fibonacci word: $0110100110010110 \cdots$

Counting squares

Number of kth powers: $\mathcal{O}(n^2)$

$$aaa \cdots a$$

Number of kth powers with primitive root:
$< n \log n$ [Bannai et al., 2020]

Fibonacci word: $0110100110010110 \cdots$

Distinct powers: if the root is the same do not count it again

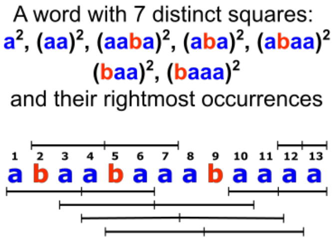
Counting squares

Number of kth powers: $\mathcal{O}(n^2)$

$$aaa \cdots a$$

Number of kth powers with primitive root: $< n \log n$ [Bannai et al., 2020]

Fibonacci word: $0110100110010110 \cdots$



Distinct powers: if the root is the same do not count it again

Words rich in distinct squares

$abaabaaabaabaaabaaaab\cdots$ [Fraenkel and Simpson, 1998]

Words rich in distinct squares

$abaabaaabaabaaabaaaab\cdots$ [Fraenkel and Simpson, 1998]

$ab\ a^2b\ a^3b\ a^2b\ a^3b\ a^4b\ a^3b\ a^4b\ a^5b\ \cdots$

Words rich in distinct squares

$abaabaaabaabaaabaaaab\cdots$ [Fraenkel and Simpson, 1998]

$ab\ a^2b\ a^3b\ a^2b\ a^3b\ a^4b\ a^3b\ a^4b\ a^5b\ \cdots$

squares: unary, $a^kba^j\ a^kba^j$, $a^kba^{k+1}b\ a^kba^{k+1}b$

Words rich in distinct squares

$$abaabaaabaabaaabaaaab\cdots$$ [Fraenkel and Simpson, 1998]

$$ab\ a^2b\ a^3b\ a^2b\ a^3b\ a^4b\ a^3b\ a^4b\ a^5b\ \cdots$$

squares: unary, $a^kba^j\ a^kba^j$, $a^kba^{k+1}b\ a^kba^{k+1}b$

$$abaabaaabaaaab\cdots$$ [Jonoska et al., 2014]

Words rich in distinct squares

$abaabaaabaabaaabaaaab\cdots$ [Fraenkel and Simpson, 1998]

$ab\ a^2b\ a^3b\ a^2b\ a^3b\ a^4b\ a^3b\ a^4b\ a^5b\ \cdots$

squares: unary, $a^kba^j\ a^kba^j$, $a^kba^{k+1}b\ a^kba^{k+1}b$

$abaabaaabaaaab\cdots$ [Jonoska et al., 2014]

$ab\ a^2b\ a^3b\ a^4b\ a^5b\ \cdots$

Words rich in distinct squares

$abaabaaabaabaaabaaaab\cdots$ [Fraenkel and Simpson, 1998]

$ab\ a^2b\ a^3b\ a^2b\ a^3b\ a^4b\ a^3b\ a^4b\ a^5b\ \cdots$

squares: unary, $a^kba^j\ a^kba^j$, $a^kba^{k+1}b\ a^kba^{k+1}b$

$abaabaaabaaaab\cdots$ [Jonoska et al., 2014]

$ab\ a^2b\ a^3b\ a^4b\ a^5b\ \cdots$

squares: unary, $a^kba^j\ a^kba^j$

Words rich in distinct squares

$$abaabaaabaabaaabaaaab \cdots$$ [Fraenkel and Simpson, 1998]

$$ab\ a^2b\ a^3b\ a^2b\ a^3b\ a^4b\ a^3b\ a^4b\ a^5b\ \cdots$$

squares: unary, $a^kba^j\ a^kba^j$, $a^kba^{k+1}b\ a^kba^{k+1}b$

$$abaabaaabaaaab \cdots$$ [Jonoska et al., 2014]

$$ab\ a^2b\ a^3b\ a^4b\ a^5b\ \cdots$$

squares: unary, $a^kba^j\ a^kba^j$

lower bound: $\frac{2k-1}{2k+2}n$, k=number of b's

Upper bounds

Lemma (Three square prefixes, [Crochemore and Rytter, 1995])

If u^2 prefix of v^2, v^2 prefix of w^2, and u primitive then $|u| + |v| \leq |w|$.

Upper bounds

Lemma (Three square prefixes, [Crochemore and Rytter, 1995])

If u^2 prefix of v^2, v^2 prefix of w^2, and u primitive then $|u| + |v| \leq |w|$.

Less than $2n$ primitively-rooted squares [Fraenkel and Simpson, 1998]

Upper bounds

Lemma (Three square prefixes, [Crochemore and Rytter, 1995])

If u^2 prefix of v^2, v^2 prefix of w^2, and u primitive then $|u| + |v| \leq |w|$.

Less than $2n$ primitively-rooted squares [Fraenkel and Simpson, 1998]

Direct proofs [Hickerson, 2003],[Ilie, 2005],[Lam, 2013]

Upper bounds

Lemma (Three square prefixes, [Crochemore and Rytter, 1995])

If u^2 prefix of v^2, v^2 prefix of w^2, and u primitive then $|u| + |v| \leq |w|$.

Less than $2n$ primitively-rooted squares [Fraenkel and Simpson, 1998]

Direct proofs [Hickerson, 2003],[Ilie, 2005],[Lam, 2013]

Best bound: $\frac{11n}{6}$ [Deza et al., 2015]

Upper bounds

Lemma (Three square prefixes, [Crochemore and Rytter, 1995])

If u^2 prefix of v^2, v^2 prefix of w^2, and u primitive then $|u| + |v| \leq |w|$.

Less than $2n$ primitively-rooted squares [Fraenkel and Simpson, 1998]

Direct proofs [Hickerson, 2003],[Ilie, 2005],[Lam, 2013]

Best bound: $\frac{11n}{6}$ [Deza et al., 2015]

Potential: $\frac{3n}{2}$ [Thierry, 2020]

Upper bounds

Lemma (Three square prefixes, [Crochemore and Rytter, 1995])

If u^2 prefix of v^2, v^2 prefix of w^2, and u primitive then $|u| + |v| \leq |w|$.

Less than $2n$ primitively-rooted squares [Fraenkel and Simpson, 1998]

Direct proofs [Hickerson, 2003],[Ilie, 2005],[Lam, 2013]

Best bound: $\frac{11n}{6}$ [Deza et al., 2015]

Potential: $\frac{3n}{2}$ [Thierry, 2020]

Bounds for binary alphabet suffice [Manea and Seki, 2015]

Upper bounds

Lemma (Three square prefixes, [Crochemore and Rytter, 1995])

If u^2 prefix of v^2, v^2 prefix of w^2, and u primitive then $|u| + |v| \leq |w|$.

Less than $2n$ primitively-rooted squares [Fraenkel and Simpson, 1998]

Direct proofs [Hickerson, 2003],[Ilie, 2005],[Lam, 2013]

Best bound: $\frac{11n}{6}$ [Deza et al., 2015]

Potential: $\frac{3n}{2}$ [Thierry, 2020]

Bounds for binary alphabet suffice [Manea and Seki, 2015]

Circular squares [Amit and Gawrychowski, 2017]: $>1.25n$ and $<3.14n$

Runs

Non-expanding repetitions of power at least 2.

In *abaabaaabaaaa* we have $(aba)^{7/3}$ and $(aaba)^{10/4}$

Runs

Non-expanding repetitions of power at least 2.

In *abaabaaabaaaa* we have $(aba)^{7/3}$ and $(aaba)^{10/4}$

Introduced [Kolpakov and Kucherov, 1999]

Runs

Non-expanding repetitions of power at least 2.

In *abaabaaabaaaa* we have $(aba)^{7/3}$ and $(aaba)^{10/4}$

Introduced [Kolpakov and Kucherov, 1999]

Best known upper bound
$183/193 \approx 0.9482n$ [Bannai et al., 2014],[Fischer et al., 2015],[Holub, 2017]

Runs

Non-expanding repetitions of power at least 2.

In *abaabaaabaaaa* we have $(aba)^{7/3}$ and $(aaba)^{10/4}$

Introduced [Kolpakov and Kucherov, 1999]

Best known upper bound
$183/193 \approx 0.9482n$ [Bannai et al., 2014],[Fischer et al., 2015],[Holub, 2017]

Best known lower bound is $0.944575712n$ [Simpson, 2010]

Runs

Non-expanding repetitions of power at least 2.

In *abaabaaabaaaa* we have $(aba)^{7/3}$ and $(aaba)^{10/4}$

Introduced [Kolpakov and Kucherov, 1999]

Best known upper bound
$183/193 \approx 0.9482n$ [Bannai et al., 2014],[Fischer et al., 2015],[Holub, 2017]

Best known lower bound is $0.944575712n$ [Simpson, 2010]

Strategy: Lyndon roots (of sorts)

Existing results on higher powers ($k > 2$) and strategies

Existing results on higher powers ($k > 2$) and strategies

Less than $\frac{n}{k-2}$ [Crochemore et al., 2010]
Between $\frac{n}{2} - 2\sqrt{n}$ and $\frac{4n}{5}$ [Kubica et al., 2009]

Existing results on higher powers ($k > 2$) and strategies

Less than $\frac{n}{k-2}$ [Crochemore et al., 2010]
Between $\frac{n}{2} - 2\sqrt{n}$ and $\frac{4n}{5}$ [Kubica et al., 2009]

Strategy: (primitive) roots, or very techical

Suffix mountains

Suffix mountains

Idea: separate the squares

Suffix mountains

Idea: separate the squares

Squares in one group share a common prefix

Suffix mountains

Idea: separate the squares

Squares in one group share a common prefix

Show that the number of squares in a group is no more than the number of occurrences of the common prefix

Suffix mountains

Idea: separate the squares

Squares in one group share a common prefix

Show that the number of squares in a group is no more than the number of occurrences of the common prefix

Lemma (Synchronization)

There are only 2 occurrences of a primitive word u in u^2.

Suffix mountains

Idea: separate the squares

Squares in one group share a common prefix

Show that the number of squares in a group is no more than the number of occurrences of the common prefix

Lemma (Synchronization)

There are only 2 occurrences of a primitive word u in u^2.

Lemma (Fine and Wilf)

If p and q are periods of a word with length at least $p + q - \gcd(p, q)$, then $\gcd(p, q)$ is also a period.

Suffix Array

order the suffixes lexicographically

a	a	b	a	b	a
1	2	3	4	5	6

Suffix Array

order the suffixes lexicographically

$$\begin{array}{cccccc} a & a & b & a & b & a \\ 1 & 2 & 3 & 4 & 5 & 6 \end{array}$$

$(a < b)$

Suffix Array

order the suffixes lexicographically

a	a	b	a	b	a
1	2	3	4	5	6

$(a < b)$ $a, aababa, aba, ababa, ba, baba$

Suffix Array

order the suffixes lexicographically

$$\begin{array}{cccccc} a & a & b & a & b & a \\ 1 & 2 & 3 & 4 & 5 & 6 \end{array}$$

$(a < b)$ $\qquad a, aababa, aba, ababa, ba, baba$

put the starting positions into an array

$$6, 1, 4, 2, 5, 3$$

Suffix Array

order the suffixes lexicographically

$$\begin{array}{cccccc} a & a & b & a & b & a \\ 1 & 2 & 3 & 4 & 5 & 6 \end{array}$$

$(a < b)$ $\qquad a, aababa, aba, ababa, ba, baba$

put the starting positions into an array

$$6, 1, 4, 2, 5, 3$$

useful for indexing text

Why 'suffix mountains'

clust(u) - all suffixes which start with u

Why 'suffix mountains'

clust(u) - all suffixes which start with u

$$\begin{array}{ccccccccc} w = & a & a & b & b & a & b & a & a \\ & 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \end{array}$$

Why 'suffix mountains'

clust(u) - all suffixes which start with u

$$\begin{array}{ccccccccc} w = & a & a & b & b & a & b & a & a \\ & 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \end{array}$$

$$\begin{aligned} \mathbf{SA}(w) &= [8,7,1,5,2,6,4,3] \\ \mathbf{clust}(a) &= [8,7,1,5,2] \\ \mathbf{clust}(ab) &= [5,2] \\ \mathbf{clust}(aa) &= [7,1] \end{aligned}$$

Why 'suffix mountains'

clust(u) - all suffixes which start with u

$$\begin{array}{lcccccccc} w = & a & a & b & b & a & b & a & a \\ & 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \end{array}$$

$$\begin{aligned} \mathbf{SA}(w) &= [8,7,1,5,2,6,4,3] \\ \mathbf{clust}(a) &= [8,7,1,5,2] \\ \mathbf{clust}(ab) &= [5,2] \\ \mathbf{clust}(aa) &= [7,1] \end{aligned}$$

$[8, \quad 7, \quad 1\ , \quad 5, \quad 2\ , \quad 6, \quad 4, \quad 3]$

Why 'suffix mountains'

clust(u) - all suffixes which start with u

$$\begin{array}{lcccccccc} w = & a & a & b & b & a & b & a & a \\ & 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \end{array}$$

$$\begin{aligned} \mathbf{SA}(w) &= [8,7,1,5,2,6,4,3] \\ \mathbf{clust}(a) &= [8,7,1,5,2] \\ \mathbf{clust}(ab) &= [5,2] \\ \mathbf{clust}(aa) &= [7,1] \end{aligned}$$

$$[8,\ 7,\ 1,\ 5,\ 2]$$
$$[8,\ 7,\ 1,\ 5,\ 2,\ 6,\ 4,\ 3]$$

Why 'suffix mountains'

clust(u) - all suffixes which start with u

$$\begin{array}{lcccccccc} w = & a & a & b & b & a & b & a & a \\ & 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \end{array}$$

$$\begin{aligned} \mathbf{SA}(w) &= [8,7,1,5,2,6,4,3] \\ \mathbf{clust}(a) &= [8,7,1,5,2] \\ \mathbf{clust}(ab) &= [5,2] \\ \mathbf{clust}(aa) &= [7,1] \end{aligned}$$

$$\begin{array}{llllllll} & [7, & 1] & & & & & \\ [8, & 7, & 1, & 5, & 2] & & & \\ [8, & 7, & 1, & 5, & 2, & 6, & 4, & 3] \end{array}$$

Why 'suffix mountains'

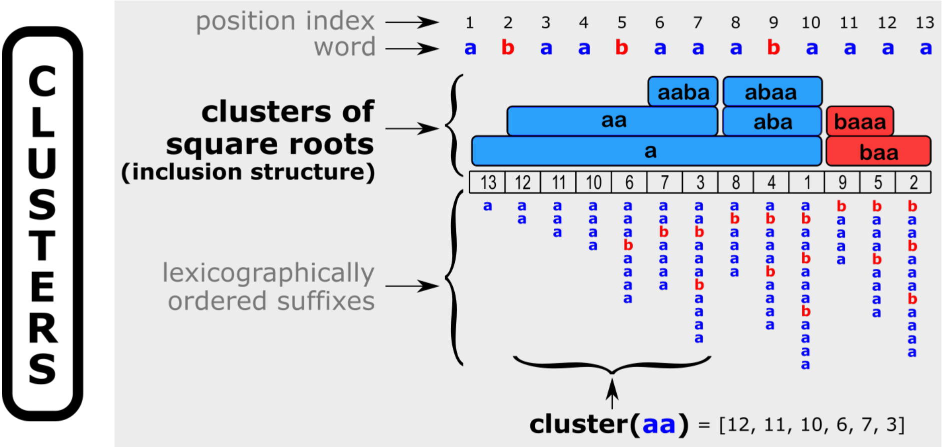
clust(u) - all suffixes which start with u

$$\begin{array}{lcccccccc} w = & a & a & b & b & a & b & a & a \\ & 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \end{array}$$

$$\begin{aligned} \mathbf{SA}(w) &= [8,7,1,5,2,6,4,3] \\ \mathbf{clust}(a) &= [8,7,1,5,2] \\ \mathbf{clust}(ab) &= [5,2] \\ \mathbf{clust}(aa) &= [7,1] \end{aligned}$$

$$\begin{array}{llllllll} & [7, & 1] & [5, & 2] & & & \\ [8, & 7, & 1, & 5, & 2] & & & \\ [8, & 7, & 1, & 5, & 2, & 6, & 4, & 3] \end{array}$$

Mountain example

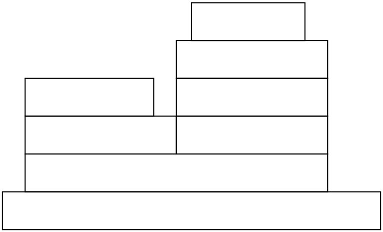


Suffix mountains

We consider suffix mountains formed of words which are roots of squares.

Suffix mountains

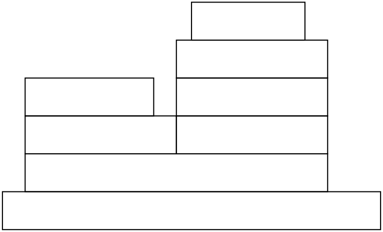
We consider suffix mountains formed of words which are roots of squares.



Suffix mountains

We consider suffix mountains formed of words which are roots of squares.

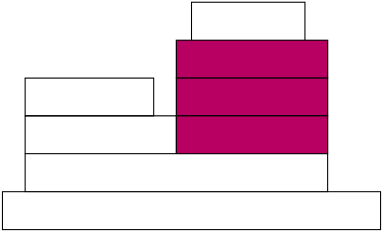
We want to show that a cluster is at least as large as the number of clusters on top of it



Suffix mountains

We consider suffix mountains formed of words which are roots of squares.

We want to show that a cluster is at least as large as the number of clusters on top of it

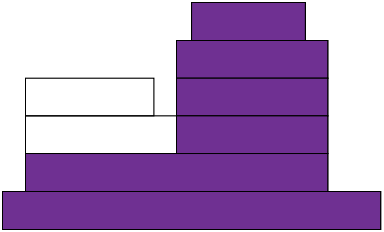


block - consecutive clusters of same size

Suffix mountains

We consider suffix mountains formed of words which are roots of squares.

We want to show that a cluster is at least as large as the number of clusters on top of it

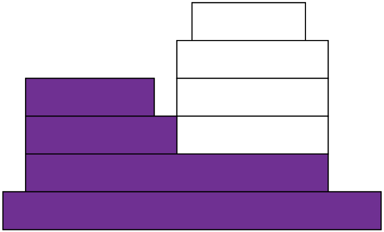


block - consecutive clusters of same size
chain - consecutive clusters

Suffix mountains

We consider suffix mountains formed of words which are roots of squares.

We want to show that a cluster is at least as large as the number of clusters on top of it



block - consecutive clusters of same size
chain - consecutive clusters

Suffix mountains

- Formed of roots of squares

Suffix mountains

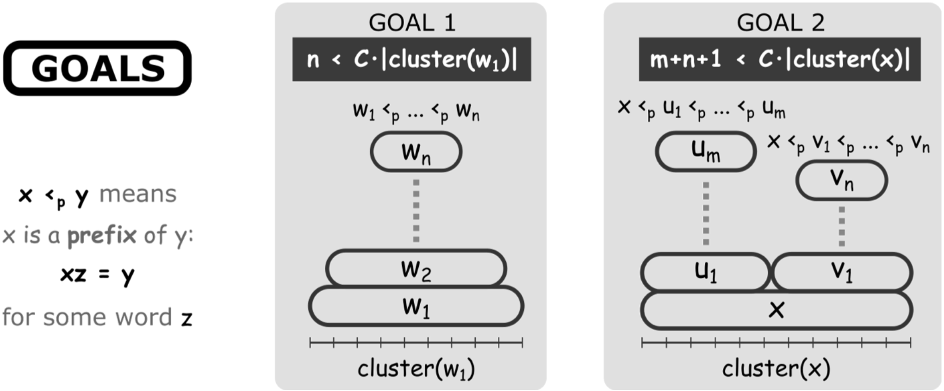
- Formed of roots of squares
- Show: a cluster is bigger than the number of clusters on top of it

Suffix mountains

- Formed of roots of squares
- Show: a cluster is bigger than the number of clusters on top of it
- If true, then there are less squares in a suffix mountain than the size of the base cluster $\implies$ bound n

Suffix mountains

- Formed of roots of squares
- Show: a cluster is bigger than the number of clusters on top of it
- If true, then there are less squares in a suffix mountain than the size of the base cluster $\Longrightarrow$ bound n



Obsevations

- The set of suffixes with a common prefix are contiguous in the suffix array

Obsevations

- The set of suffixes with a common prefix are contiguous in the suffix array
- If a kth power u^k is a factor of a word, then $\mathbf{clust}(u) \geq k$

Obsevations

- The set of suffixes with a common prefix are contiguous in the suffix array
- If a kth power u^k is a factor of a word, then $\mathbf{clust}(u) \geq k$

For any two clusters $\mathbf{clust}(u) = [i_1..j_1]$ and $\mathbf{clust}(v) = [i_2..j_2]$,

Obsevations

- The set of suffixes with a common prefix are contiguous in the suffix array
- If a kth power u^k is a factor of a word, then $\mathbf{clust}(u) \geq k$

For any two clusters $\mathbf{clust}(u) = [i_1..j_1]$ and $\mathbf{clust}(v) = [i_2..j_2]$,

- if u is a prefix of v, then $\mathbf{clust}(v) \subseteq \mathbf{clust}(u)$, and vice versa;

Obsevations

- The set of suffixes with a common prefix are contiguous in the suffix array
- If a kth power u^k is a factor of a word, then $\mathbf{clust}(u) \geq k$

For any two clusters $\mathbf{clust}(u) = [i_1..j_1]$ and $\mathbf{clust}(v) = [i_2..j_2]$,

- if u is a prefix of v, then $\mathbf{clust}(v) \subseteq \mathbf{clust}(u)$, and vice versa;
- if u and v are incomparable with respect to the prefix order, then $j_1 < i_2$ or $j_2 < i_1$, that is, the clusters do not overlap;

Obsevations

- The set of suffixes with a common prefix are contiguous in the suffix array
- If a kth power u^k is a factor of a word, then $\mathbf{clust}(u) \geq k$

For any two clusters $\mathbf{clust}(u) = [i_1..j_1]$ and $\mathbf{clust}(v) = [i_2..j_2]$,

- if u is a prefix of v, then $\mathbf{clust}(v) \subseteq \mathbf{clust}(u)$, and vice versa;
- if u and v are incomparable with respect to the prefix order, then $j_1 < i_2$ or $j_2 < i_1$, that is, the clusters do not overlap;
- if $\mathbf{clust}(u) \cap \mathbf{clust}(v) \neq \emptyset$, then either u is prefix of v or v is prefix of u

Blocks

Lemma

If $u^k \neq v^k$ with $k > 1$ are factors of w and $\mathbf{clust}(u) = \mathbf{clust}(v)$, then either u or v is primitive.

Blocks

Lemma

If $u^k \neq v^k$ with $k > 1$ are factors of w and $\textbf{clust}(u) = \textbf{clust}(v)$, then either u or v is primitive.

1. *if $u = t^n$ with $n > 1$ and primitive t, then $v = t^n t'$ for $\varepsilon \neq t' \leq_p t$;*
2. *if $v = t^n$ with $n > 1$ and primitive t, then $u = t^{n-1} t'$ for $\varepsilon \neq t' \leq_p t$.*

Blocks

Lemma

If $u^k \neq v^k$ *with* $k > 1$ *are factors of* w *and* $\mathbf{clust}(u) = \mathbf{clust}(v)$, *then either* u *or* v *is primitive.*

1. *if* $u = t^n$ *with* $n > 1$ *and primitive* t, *then* $v = t^n t'$ *for* $\varepsilon \neq t' \leq_p t$;
2. *if* $v = t^n$ *with* $n > 1$ *and primitive* t, *then* $u = t^{n-1} t'$ *for* $\varepsilon \neq t' \leq_p t$.

Lemma

Let $u^k \neq v^k$ *with* $k > 1$ *be factors of* w, *where* $\mathbf{clust}(u) = \mathbf{clust}(v)$ *and* $u \leq_p v$. *If their corresponding rightmost occurrences start at positions* u_s *and* v_s, *respectively, then* $|u_s - v_s| \geq |u|$.

Blocks

Lemma

If $u^k \neq v^k$ with $k > 1$ are factors of w and $\mathbf{clust}(u) = \mathbf{clust}(v)$, then either u or v is primitive.

1. *if $u = t^n$ with $n > 1$ and primitive t, then $v = t^n t'$ for $\varepsilon \neq t' \leq_p t$;*
2. *if $v = t^n$ with $n > 1$ and primitive t, then $u = t^{n-1} t'$ for $\varepsilon \neq t' \leq_p t$.*

Lemma

Let $u^k \neq v^k$ with $k > 1$ be factors of w, where $\mathbf{clust}(u) = \mathbf{clust}(v)$ and $u \leq_p v$. If their corresponding rightmost occurrences start at positions u_s and v_s, respectively, then $|u_s - v_s| \geq |u|$.

Corollary

Let $u_1^k, \ldots, u_n^k$ be squares in w such that $\mathbf{clust}_w(u_1) = \cdots = \mathbf{clust}_w(u_n)$. Then, $|\mathbf{clust}_w(u_1)| > (k-1)n$.

Connections

How are the squares connected, and why clusters?!

Connections

How are the squares connected, and why clusters?!

Why might [Three square prefixes] not be enough?

Connections

How are the squares connected, and why clusters?!

Why might [Three square prefixes] not be enough?

Our approach:

Connections

How are the squares connected, and why clusters?!

Why might [Three square prefixes] not be enough?

Our approach:
We will try to count not the whole squares, but some really long prefixes.

Connections

How are the squares connected, and why clusters?!

Why might [Three square prefixes] not be enough?

Our approach:
We will try to count not the whole squares, but some really long prefixes.

Idea was also mentioned/used in [Jonoska et al., 2014, Lemma 2]
and [Bannai et al., 2014].)

Representatives (of squares)

For $x \leq_p u$, the *x-representative* (x-rep) of u^2 is the longest prefix of u^2 which ends in x.

x-rep is of length at least $|u| + |x|$

Representatives (of squares)

For $x \leq_p u$, the *x-representative* (x-rep) of u^2 is the longest prefix of u^2 which ends in x.

x-rep is of length at least $|u| + |x|$

In the word $w = abaabcabaabab$ we have the square

$$u = (aba)^2 \text{ starting at position 7.}$$

Representatives (of squares)

For $x \leq_p u$, the *x-representative* (x-rep) of u^2 is the longest prefix of u^2 which ends in x.

x-rep is of length at least $|u| + |x|$

In the word $w = abaabcabaabab$ we have the square

$$u = (aba)^2 \text{ starting at position 7.}$$

a-rep of u^2 is $abaaba = u^2$,

Representatives (of squares)

For $x \leq_p u$, the *x-representative* (x-rep) of u^2 is the longest prefix of u^2 which ends in x.

x-rep is of length at least $|u| + |x|$

In the word $w = abaabcabaabab$ we have the square

$$u = (aba)^2 \text{ starting at position 7.}$$

a-rep of u^2 is $abaaba = u^2$,

ab-rep of u^2 is $abaab$.

Anchors

Let x-rep of square u^2 be $uu'x$. Let u_s be its starting position and u_m be

$$u_m = u_s + |u|$$

Anchors

Let x-rep of square u^2 be $uu'x$. Let u_s be its starting position and u_m be

$$u_m = u_s + |u|$$

The x-anchor of u^2 in w is

$$\Psi(u^2, x) = u_s + |uu'|$$

Anchors

Let x-rep of square u^2 be $uu'x$. Let u_s be its starting position and u_m be

$$u_m = u_s + |u|$$

The x-anchor of u^2 in w is

$$\Psi(u^2, x) = u_s + |uu'|$$

$w = abaabcabaabab$

The a-rep of u^2 is $abaaba = u^2$, first occurring at 7,

$$\Psi(u^2, a) = 7 + 5 = 12.$$

Anchors

Let x-rep of square u^2 be $uu'x$. Let u_s be its starting position and u_m be

$$u_m = u_s + |u|$$

The x-anchor of u^2 in w is

$$\Psi(u^2, x) = u_s + |uu'|$$

$w = abaabcabaabab$

The a-rep of u^2 is $abaaba = u^2$, first occurring at 7,

$$\Psi(u^2, a) = 7 + 5 = 12.$$

The ab-rep of u^2 is $abaab$, first occurring at 1,

$$\Psi(u^2, ab) = 1 + 3 = 4.$$

Collision Lemma

Lemma

Let w be an arbitrary word with two square factors u^2, v^2 such that $u <_p v$, and let $x \leq_p u$ be a common prefix of theirs.

Collision Lemma

Lemma

Let w be an arbitrary word with two square factors u^2, v^2 such that $u <_p v$, and let $x \leq_p u$ be a common prefix of theirs. If $\mathbf{\Psi}_w(u^2, x) = \mathbf{\Psi}_w(v^2, x)$, then $u = t^k$ for some primitive word t with $|t| < |x|$ and $k \geq 2$.

Collision Lemma

Lemma

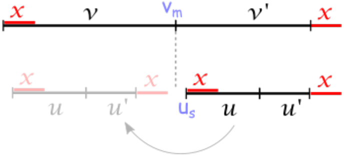
Let w be an arbitrary word with two square factors u^2, v^2 such that $u <_p v$, and let $x \leq_p u$ be a common prefix of theirs. If $\mathbf{\Psi}_w(u^2, x) = \mathbf{\Psi}_w(v^2, x)$, then $u = t^k$ for some primitive word t with $|t| < |x|$ and $k \geq 2$. Moreover, $tu'x \leq_p v$, where $u'x$ is the longest prefix of u bordered by x.

Collision Lemma

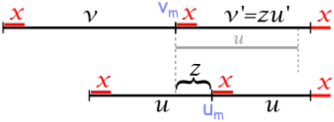
Lemma

Let w be an arbitrary word with two square factors u^2, v^2 such that $u <_p v$, and let $x \leq_p u$ be a common prefix of theirs. If $\mathbf{\Psi}_w(u^2, x) = \mathbf{\Psi}_w(v^2, x)$, then $u = t^k$ for some primitive word t with $|t| < |x|$ and $k \geq 2$. Moreover, $tu'x \leq_p v$, where $u'x$ is the longest prefix of u bordered by x.

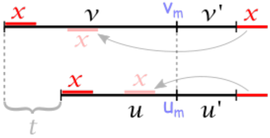
(1)



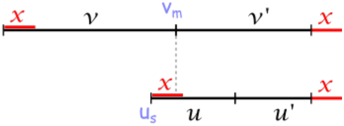
(3a)



(2)



(3b)



Chain chain chain

Corollary

Let $u_1^2, \ldots, u_n^2$ and $v_1^2, \ldots, v_n^2$ be squares in a word w with their roots from the same chain, and x a common prefix of theirs.

Chain chain chain

Corollary

Let $u_1^2, \ldots, u_n^2$ *and* $v_1^2, \ldots, v_n^2$ *be squares in a word* w *with their roots from the same chain, and* x *a common prefix of theirs. If* $\boldsymbol{\Psi}_w(u_i^2, x) = \boldsymbol{\Psi}_w(v_i^2, x)$*, for all* $i \in \{1, \ldots, n\}$*,*

Chain chain chain

Corollary

Let $u_1^2, \ldots, u_n^2$ and $v_1^2, \ldots, v_n^2$ be squares in a word w with their roots from the same chain, and x a common prefix of theirs. If $\mathbf{\Psi}_w(u_i^2, x) = \mathbf{\Psi}_w(v_i^2, x)$, for all $i \in \{1, \ldots, n\}$, then there exists a primitive word t shorter than x, such that $u_i = t^{k_i}$ with $k_i \geq 2$, for all $i \in \{1, \ldots, n\}$.

Chain chain chain

Corollary

Let $u_1^2, \ldots, u_n^2$ and $v_1^2, \ldots, v_n^2$ be squares in a word w with their roots from the same chain, and x a common prefix of theirs. If $\mathbf{\Psi}_w(u_i^2, x) = \mathbf{\Psi}_w(v_i^2, x)$, for all $i \in \{1, \ldots, n\}$, then there exists a primitive word t shorter than x, such that $u_i = t^{k_i}$ with $k_i \geq 2$, for all $i \in \{1, \ldots, n\}$.

Proof.

If the x-anchor of some u_i^2 and v_i^2 coincide, there is some primitive t_i with $|t_i| < |x|$ such that $u_i = t_i^{k_i}$ with $k_i \geq 2$ and $t_i x$ is a prefix of v_i.

Chain chain chain

Corollary

Let $u_1^2, \ldots, u_n^2$ and $v_1^2, \ldots, v_n^2$ be squares in a word w with their roots from the same chain, and x a common prefix of theirs. If $\mathbf{\Psi}_w(u_i^2, x) = \mathbf{\Psi}_w(v_i^2, x)$, for all $i \in \{1, \ldots, n\}$, then there exists a primitive word t shorter than x, such that $u_i = t^{k_i}$ with $k_i \geq 2$, for all $i \in \{1, \ldots, n\}$.

Proof.

If the x-anchor of some u_i^2 and v_i^2 coincide, there is some primitive t_i with $|t_i| < |x|$ such that $u_i = t_i^{k_i}$ with $k_i \geq 2$ and $t_i x$ is a prefix of v_i.
Since the roots form a prefix chain, we get that the words $t_i x$ also form a prefix chain (either $t_i x \leq_p t_j x$ or $t_j x \leq_p t_i x$)..

Chain chain chain

Corollary

Let $u_1^2, \ldots, u_n^2$ and $v_1^2, \ldots, v_n^2$ be squares in a word w with their roots from the same chain, and x a common prefix of theirs. If $\mathbf{\Psi}_w(u_i^2, x) = \mathbf{\Psi}_w(v_i^2, x)$, for all $i \in \{1, \ldots, n\}$, then there exists a primitive word t shorter than x, such that $u_i = t^{k_i}$ with $k_i \geq 2$, for all $i \in \{1, \ldots, n\}$.

Proof.

If the x-anchor of some u_i^2 and v_i^2 coincide, there is some primitive t_i with $|t_i| < |x|$ such that $u_i = t_i^{k_i}$ with $k_i \geq 2$ and $t_i x$ is a prefix of v_i.
Since the roots form a prefix chain, we get that the words $t_i x$ also form a prefix chain (either $t_i x \leq_p t_j x$ or $t_j x \leq_p t_i x$)..
Since x prefix of all the squares, we get that $t_i x$ has periods $|t_i|, |t_j|$.

Chain chain chain

Corollary

Let $u_1^2, \ldots, u_n^2$ and $v_1^2, \ldots, v_n^2$ be squares in a word w with their roots from the same chain, and x a common prefix of theirs. If $\boldsymbol{\Psi}_w(u_i^2, x) = \boldsymbol{\Psi}_w(v_i^2, x)$, for all $i \in \{1, \ldots, n\}$, then there exists a primitive word t shorter than x, such that $u_i = t^{k_i}$ with $k_i \geq 2$, for all $i \in \{1, \ldots, n\}$.

Proof.

If the x-anchor of some u_i^2 and v_i^2 coincide, there is some primitive t_i with $|t_i| < |x|$ such that $u_i = t_i^{k_i}$ with $k_i \geq 2$ and $t_i x$ is a prefix of v_i.
Since the roots form a prefix chain, we get that the words $t_i x$ also form a prefix chain (either $t_i x \leq_p t_j x$ or $t_j x \leq_p t_i x$)..
Since x prefix of all the squares, we get that $t_i x$ has periods $|t_i|, |t_j|$.

Since $|t_i x| > |t_i| + |t_j| > |t_i| + |t_j| - \gcd(|t_i|, |t_j|)$, by [Fine and Wilf] we have that t_i and t_j have a common primitive root t, so $t_i = t_j = t$. □

Main Result

Theorem

For all words w and squares $u_1^2, \ldots, u_n^2$ in w with $u_1 <_p \cdots <_p u_n$:

$$|\mathbf{clust}_w(u_1)| \geq n + 1.$$

Proof idea

- in decreasing order of length, assign to u_i the position $\mathbf{\Psi}_w(u_i^2, x)$ (if not previously assigned)

Proof idea

- in decreasing order of length, assign to u_i the position $\boldsymbol{\Psi}_w(u_i^2, x)$ (if not previously assigned)
- if no collision, done. Otherwise, for largest such square [CL] $u = t^k$

Proof idea

- in decreasing order of length, assign to u_i the position $\mathbf{\Psi}_w(u_i^2, x)$ (if not previously assigned)
- if no collision, done. Otherwise, for largest such square [CL] $u = t^k$
- we get squares $t^{k_1}, \ldots, t^{k_m}$ with $1 < k_1 < \cdots < k_m = k$

Proof idea

- in decreasing order of length, assign to u_i the position $\boldsymbol{\Psi}_w(u_i^2, x)$ (if not previously assigned)
- if no collision, done. Otherwise, for largest such square [CL] $u = t^k$
- we get squares $t^{k_1}, \ldots, t^{k_m}$ with $1 < k_1 < \cdots < k_m = k$
- if $x = t^\ell t'$, then $|x| \leq |t^{k_1}|$ since $x = u_1$, so $k_1 > \ell$ and $k_m \geq m + \ell$.

Proof idea

- in decreasing order of length, assign to u_i the position $\mathbf{\Psi}_w(u_i^2, x)$ (if not previously assigned)
- if no collision, done. Otherwise, for largest such square [CL] $u = t^k$
- we get squares $t^{k_1}, \ldots, t^{k_m}$ with $1 < k_1 < \cdots < k_m = k$
- if $x = t^\ell t'$, then $|x| \leq |t^{k_1}|$ since $x = u_1$, so $k_1 > \ell$ and $k_m \geq m + \ell$.
- By [CL] we know that $t^{k_m} t' \leq_p u_n$, so $t^{m+\ell} t' \leq_p u_n$.

Proof idea

- in decreasing order of length, assign to u_i the position $\mathbf{\Psi}_w(u_i^2, x)$ (if not previously assigned)
- if no collision, done. Otherwise, for largest such square [CL] $u = t^k$
- we get squares $t^{k_1}, \ldots, t^{k_m}$ with $1 < k_1 < \cdots < k_m = k$
- if $x = t^\ell t'$, then $|x| \leq |t^{k_1}|$ since $x = u_1$, so $k_1 > \ell$ and $k_m \geq m + \ell$.
- By [CL] we know that $t^{k_m} t' \leq_p u_n$, so $t^{m+\ell} t' \leq_p u_n$.
- if s_i is leftmost position where $t^i x$ occurs in w, assign the position $p_i = s_i + i \cdot |t|$ to the square $(t^{k_i})^2$.

Proof idea

- in decreasing order of length, assign to u_i the position $\mathbf{\Psi}_w(u_i^2, x)$ (if not previously assigned)
- if no collision, done. Otherwise, for largest such square [CL] $u = t^k$
- we get squares $t^{k_1}, \ldots, t^{k_m}$ with $1 < k_1 < \cdots < k_m = k$
- if $x = t^\ell t'$, then $|x| \leq |t^{k_1}|$ since $x = u_1$, so $k_1 > \ell$ and $k_m \geq m + \ell$.
- By [CL] we know that $t^{k_m} t' \leq_p u_n$, so $t^{m+\ell} t' \leq_p u_n$.
- if s_i is leftmost position where $t^i x$ occurs in w, assign the position $p_i = s_i + i \cdot |t|$ to the square $(t^{k_i})^2$.
- possible collisions: $v \in \{u_1, \ldots, u_n\} \setminus \{t^{k_1}, \ldots, t^{k_m}\}$ such that $\mathbf{\Psi}_w(v^2, x) = p_i = s_i + i \cdot |t|$.

Proof idea

- in decreasing order of length, assign to u_i the position $\mathbf{\Psi}_w(u_i^2, x)$ (if not previously assigned)
- if no collision, done. Otherwise, for largest such square [CL] $u = t^k$
- we get squares $t^{k_1}, \dots, t^{k_m}$ with $1 < k_1 < \cdots < k_m = k$
- if $x = t^\ell t'$, then $|x| \leq |t^{k_1}|$ since $x = u_1$, so $k_1 > \ell$ and $k_m \geq m + \ell$.
- By [CL] we know that $t^{k_m} t' \leq_p u_n$, so $t^{m+\ell} t' \leq_p u_n$.
- if s_i is leftmost position where $t^i x$ occurs in w, assign the position $p_i = s_i + i \cdot |t|$ to the square $(t^{k_i})^2$.
- possible collisions: $v \in \{u_1, \dots, u_n\} \setminus \{t^{k_1}, \dots, t^{k_m}\}$ such that $\mathbf{\Psi}_w(v^2, x) = p_i = s_i + i \cdot |t|$.
- derive contradictions on v, which is either

Proof idea

- in decreasing order of length, assign to u_i the position $\mathbf{\Psi}_w(u_i^2, x)$ (if not previously assigned)
- if no collision, done. Otherwise, for largest such square [CL] $u = t^k$
- we get squares $t^{k_1}, \ldots, t^{k_m}$ with $1 < k_1 < \cdots < k_m = k$
- if $x = t^\ell t'$, then $|x| \leq |t^{k_1}|$ since $x = u_1$, so $k_1 > \ell$ and $k_m \geq m + \ell$.
- By [CL] we know that $t^{k_m} t' \leq_p u_n$, so $t^{m+\ell} t' \leq_p u_n$.
- if s_i is leftmost position where $t^i x$ occurs in w, assign the position $p_i = s_i + i \cdot |t|$ to the square $(t^{k_i})^2$.
- possible collisions: $v \in \{u_1, \ldots, u_n\} \setminus \{t^{k_1}, \ldots, t^{k_m}\}$ such that $\mathbf{\Psi}_w(v^2, x) = p_i = s_i + i \cdot |t|$.
- derive contradictions on v, which is either (1) a power of t,

Proof idea

- in decreasing order of length, assign to u_i the position $\mathbf{\Psi}_w(u_i^2, x)$ (if not previously assigned)
- if no collision, done. Otherwise, for largest such square [CL] $u = t^k$
- we get squares $t^{k_1}, \ldots, t^{k_m}$ with $1 < k_1 < \cdots < k_m = k$
- if $x = t^\ell t'$, then $|x| \leq |t^{k_1}|$ since $x = u_1$, so $k_1 > \ell$ and $k_m \geq m + \ell$.
- By [CL] we know that $t^{k_m} t' \leq_p u_n$, so $t^{m+\ell} t' \leq_p u_n$.
- if s_i is leftmost position where $t^i x$ occurs in w, assign the position $p_i = s_i + i \cdot |t|$ to the square $(t^{k_i})^2$.
- possible collisions: $v \in \{u_1, \ldots, u_n\} \setminus \{t^{k_1}, \ldots, t^{k_m}\}$ such that $\mathbf{\Psi}_w(v^2, x) = p_i = s_i + i \cdot |t|$.
- derive contradictions on v, which is either (1) a power of t, (2) some other prefix of a power of t or

Proof idea

- in decreasing order of length, assign to u_i the position $\mathbf{\Psi}_w(u_i^2, x)$ (if not previously assigned)
- if no collision, done. Otherwise, for largest such square [CL] $u = t^k$
- we get squares $t^{k_1}, \ldots, t^{k_m}$ with $1 < k_1 < \cdots < k_m = k$
- if $x = t^\ell t'$, then $|x| \leq |t^{k_1}|$ since $x = u_1$, so $k_1 > \ell$ and $k_m \geq m + \ell$.
- By [CL] we know that $t^{k_m} t' \leq_p u_n$, so $t^{m+\ell} t' \leq_p u_n$.
- if s_i is leftmost position where $t^i x$ occurs in w, assign the position $p_i = s_i + i \cdot |t|$ to the square $(t^{k_i})^2$.
- possible collisions: $v \in \{u_1, \ldots, u_n\} \setminus \{t^{k_1}, \ldots, t^{k_m}\}$ such that $\mathbf{\Psi}_w(v^2, x) = p_i = s_i + i \cdot |t|$.
- derive contradictions on v, which is either (1) a power of t, (2) some other prefix of a power of t or (3) neither, thus has ut' as a prefix $\square$

Doesn't work for multiple peaks

Chains $x_1 <_p \cdots <_p x_k$, $u_1 <_p \cdots <_p u_m$ and $v_1 <_p \cdots <_p v_n$, where $x_k <_p u_1$ and $x_k <_p v_1$, but u_1 and v_1 are incomparable by $<_p$.

Doesn't work for multiple peaks

Chains $x_1 <_p \cdots <_p x_k$, $u_1 <_p \cdots <_p u_m$ and $v_1 <_p \cdots <_p v_n$, where $x_k <_p u_1$ and $x_k <_p v_1$, but u_1 and v_1 are incomparable by $<_p$.

[Theorem]: we know that $|\mathbf{clust}(u_1)| \geq m + 1$ and $|\mathbf{clust}(v_1)| \geq n + 1$, so $|\mathbf{clust}(x_k)| \geq m + n + 2$.

Doesn't work for multiple peaks

Chains $x_1 <_p \cdots <_p x_k$, $u_1 <_p \cdots <_p u_m$ and $v_1 <_p \cdots <_p v_n$, where $x_k <_p u_1$ and $x_k <_p v_1$, but u_1 and v_1 are incomparable by $<_p$.

[Theorem]: we know that $|\mathbf{clust}(u_1)| \geq m+1$ and $|\mathbf{clust}(v_1)| \geq n+1$, so $|\mathbf{clust}(x_k)| \geq m+n+2$.

For x_i's we cannot use previous argument, since $\mathbf{\Psi}_w(u_j^2, x_i) = \mathbf{\Psi}_w(v_\ell^2, x_i)$ is possible without either u_j or v_ℓ being non-primitive.

Doesn't work for multiple peaks

Chains $x_1 <_p \cdots <_p x_k$, $u_1 <_p \cdots <_p u_m$ and $v_1 <_p \cdots <_p v_n$, where $x_k <_p u_1$ and $x_k <_p v_1$, but u_1 and v_1 are incomparable by $<_p$.

[Theorem]: we know that $|\mathbf{clust}(u_1)| \geq m + 1$ and $|\mathbf{clust}(v_1)| \geq n + 1$, so $|\mathbf{clust}(x_k)| \geq m + n + 2$.

For x_i's we cannot use previous argument, since $\mathbf{\Psi}_w(u_j^2, x_i) = \mathbf{\Psi}_w(v_\ell^2, x_i)$ is possible without either u_j or v_ℓ being non-primitive.

Take incomparable $u_j = yzzyz$ and $v_\ell = zyz$ $(y, z$ bordered by $x_i)$.

Doesn't work for multiple peaks

Chains $x_1 <_p \cdots <_p x_k$, $u_1 <_p \cdots <_p u_m$ and $v_1 <_p \cdots <_p v_n$, where $x_k <_p u_1$ and $x_k <_p v_1$, but u_1 and v_1 are incomparable by $<_p$.

[Theorem]: we know that $|\mathbf{clust}(u_1)| \geq m+1$ and $|\mathbf{clust}(v_1)| \geq n+1$, so $|\mathbf{clust}(x_k)| \geq m+n+2$.

For x_i's we cannot use previous argument, since $\mathbf{\Psi}_w(u_j^2, x_i) = \mathbf{\Psi}_w(v_\ell^2, x_i)$ is possible without either u_j or v_ℓ being non-primitive.

Take incomparable $u_j = yzzyz$ and $v_\ell = zyz$ $(y, z$ bordered by $x_i)$.

Then $\mathbf{\Psi}_w((yzzyz)^2, x_i) = |w| - |x_i| + 1 = \mathbf{\Psi}_w((zyz)^2, x_i)$ in:

$$w = yzzyzyzzyz$$

Doesn't work for multiple peaks

Chains $x_1 <_p \cdots <_p x_k$, $u_1 <_p \cdots <_p u_m$ and $v_1 <_p \cdots <_p v_n$, where $x_k <_p u_1$ and $x_k <_p v_1$, but u_1 and v_1 are incomparable by $<_p$.

[Theorem]: we know that $|\mathbf{clust}(u_1)| \geq m+1$ and $|\mathbf{clust}(v_1)| \geq n+1$, so $|\mathbf{clust}(x_k)| \geq m+n+2$.

For x_i's we cannot use previous argument, since $\mathbf{\Psi}_w(u_j^2, x_i) = \mathbf{\Psi}_w(v_\ell^2, x_i)$ is possible without either u_j or v_ℓ being non-primitive.

Take incomparable $u_j = yzzyz$ and $v_\ell = zyz$ (y, z bordered by x_i).

Then $\mathbf{\Psi}_w((yzzyz)^2, x_i) = |w| - |x_i| + 1 = \mathbf{\Psi}_w((zyz)^2, x_i)$ in:

$$w = yzzyzyzzyz$$

In such a case u_j and v_ℓ have a special structure resembling the reverses of the FS double squares analysed in [Deza et al., 2015]

Doesn't work for multiple peaks

Chains $x_1 <_p \cdots <_p x_k$, $u_1 <_p \cdots <_p u_m$ and $v_1 <_p \cdots <_p v_n$, where $x_k <_p u_1$ and $x_k <_p v_1$, but u_1 and v_1 are incomparable by $<_p$.

[Theorem]: we know that $|\mathbf{clust}(u_1)| \geq m+1$ and $|\mathbf{clust}(v_1)| \geq n+1$, so $|\mathbf{clust}(x_k)| \geq m+n+2$.

For x_i's we cannot use previous argument, since $\boldsymbol{\Psi}_w(u_j^2, x_i) = \boldsymbol{\Psi}_w(v_\ell^2, x_i)$ is possible without either u_j or v_ℓ being non-primitive.

Take incomparable $u_j = yzzyz$ and $v_\ell = zyz$ $(y, z$ bordered by $x_i)$.

Then $\boldsymbol{\Psi}_w((yzzyz)^2, x_i) = |w| - |x_i| + 1 = \boldsymbol{\Psi}_w((zyz)^2, x_i)$ in:

$$w = yzzyzyzzyz$$

In such a case u_j and v_ℓ have a special structure resembling the reverses of the FS double squares analysed in [Deza et al., 2015]

A refinement of the anchor positions and assignment strategy might work.

Optimality

Using the lower bound construction in [Jonoska et al., 2014], we easily illustrate the extremal cases

Optimality

Using the lower bound construction in [Jonoska et al., 2014], we easily illustrate the extremal cases

- $|\mathbf{clust}_w(u_i)| = n - i + 2$, take $u_i = ab^{i-1}$ and $w = u_1 u_2 \cdots u_n u_n$.

Optimality

Using the lower bound construction in [Jonoska et al., 2014], we easily illustrate the extremal cases

- $|\textbf{clust}_w(u_i)| = n - i + 2$, take $u_i = ab^{i-1}$ and $w = u_1 u_2 \cdots u_n u_n$.
- $|\textbf{clust}_w(u_1)| = |\textbf{clust}_w(u_n)| = n + 1$ is realised by the roots $u_i = a^{n-1}ba^{i-1}$ and $w = u_1 u_2 \cdots u_n u_n$.

1. Preliminaries
2. Basics
3. Results
4. Final Remarks

Cluster sizes diversity - conjecture

Can all possible combinations of cluster sizes be realised in some w?

Cluster sizes diversity - conjecture

Can all possible combinations of cluster sizes be realised in some w?

Consider a chain of square roots $u_1 <_p \cdots <_p u_n$.

Cluster sizes diversity - conjecture

Can all possible combinations of cluster sizes be realised in some w?

Consider a chain of square roots $u_1 <_p \cdots <_p u_n$.

[Theorem]: $|\textbf{clust}(u_{i-1})| \geq |\textbf{clust}(u_i)|$ and $|\textbf{clust}(u_i)| \geq n - i + 2$

Cluster sizes diversity - conjecture

Can all possible combinations of cluster sizes be realised in some w?

Consider a chain of square roots $u_1 <_p \cdots <_p u_n$.

[Theorem]: $|\mathbf{clust}(u_{i-1})| \geq |\mathbf{clust}(u_i)|$ and $|\mathbf{clust}(u_i)| \geq n - i + 2$

- enough to consider $|\mathbf{clust}_w(u_1)| = n + 1$

Cluster sizes diversity - conjecture

Can all possible combinations of cluster sizes be realised in some w?

Consider a chain of square roots $u_1 <_p \cdots <_p u_n$.

[Theorem]: $|\mathbf{clust}(u_{i-1})| \geq |\mathbf{clust}(u_i)|$ and $|\mathbf{clust}(u_i)| \geq n - i + 2$

- enough to consider $|\mathbf{clust}_w(u_1)| = n + 1$
- denote by $d_i = |\mathbf{clust}(u_i)| - (n - i + 2)$ the amount that the length of the cluster at level i has over the minimum required for that level as a consequence of [Theorem] and let $D = \max_i\{d_i\}$

Cluster sizes diversity - conjecture

Can all possible combinations of cluster sizes be realised in some w?

Consider a chain of square roots $u_1 <_p \cdots <_p u_n$.

[Theorem]: $|\mathbf{clust}(u_{i-1})| \geq |\mathbf{clust}(u_i)|$ and $|\mathbf{clust}(u_i)| \geq n - i + 2$

- enough to consider $|\mathbf{clust}_w(u_1)| = n + 1$
- denote by $d_i = |\mathbf{clust}(u_i)| - (n - i + 2)$ the amount that the length of the cluster at level i has over the minimum required for that level as a consequence of [Theorem] and let $D = \max_i\{d_i\}$
- (conjecture) the idea from [Jonoska et al., 2014] can be modified to produce any combination of cluster sizes

Cluster sizes diversity - conjecture

Can all possible combinations of cluster sizes be realised in some w?

Consider a chain of square roots $u_1 <_p \cdots <_p u_n$.

[Theorem]: $|\textbf{clust}(u_{i-1})| \geq |\textbf{clust}(u_i)|$ and $|\textbf{clust}(u_i)| \geq n - i + 2$

- enough to consider $|\textbf{clust}_w(u_1)| = n + 1$
- denote by $d_i = |\textbf{clust}(u_i)| - (n - i + 2)$ the amount that the length of the cluster at level i has over the minimum required for that level as a consequence of [Theorem] and let $D = \max_i\{d_i\}$
- (conjecture) the idea from [Jonoska et al., 2014] can be modified to produce any combination of cluster sizes
- take words $u_i = a^D ba^{\ell_i}$, adjusting ℓ_i so u_i occurs exactly at positions $p_j = \sum_{k=1}^{j} |u_k|$ for $j \geq i - d_i$, which are the unique starting positions of squares u_j^2 in the word $w = u_1 \ldots u_n u_n$.

Cluster diversity - example

Let the clusters of $u_1, \ldots, u_6$ be of size $7, 7, 5, 5, 3, 2,$ respectively.

Cluster diversity - example

Let the clusters of $u_1, \ldots, u_6$ be of size $7, 7, 5, 5, 3, 2$, respectively.

The values $d_1, \ldots, d_6 = 0, 1, 0, 1, 0, 0$, hence $D = 1$

Cluster diversity - example

Let the clusters of $u_1, \ldots, u_6$ be of size $7, 7, 5, 5, 3, 2$, respectively.

The values $d_1, \ldots, d_6 = 0, 1, 0, 1, 0, 0$, hence $D = 1$

We can set $u_1 = ab$, $u_2 = aba$, $u_3 = aba^3$, $u_4 = aba^4$, $u_5 = aba^6$ and $u_6 = aba^8$

$$u_1 u_2 u_3 u_4 u_5 u_6 u_6 = ababaaba^3 aba^4 aba^6 aba^8 aba^8$$

Cluster diversity - example

Let the clusters of $u_1, \ldots, u_6$ be of size $7, 7, 5, 5, 3, 2,$ respectively.

The values $d_1, \ldots, d_6 = 0, 1, 0, 1, 0, 0,$ hence $D = 1$

We can set $u_1 = ab$, $u_2 = aba$, $u_3 = aba^3$, $u_4 = aba^4$, $u_5 = aba^6$ and $u_6 = aba^8$

$$u_1 u_2 u_3 u_4 u_5 u_6 u_6 = ababaaba^3 aba^4 aba^6 aba^8 aba^8$$

We expect that investigating the shortest words which realise a combination of cluster sizes could lead to improvements in both lower and upper bounds on distinct repetitions (the above are NOT).

Runs

We cannot readily apply the method used for distinct squares, because here multiple occurrences of a repetition have to be taken into account.

Runs

We cannot readily apply the method used for distinct squares, because here multiple occurrences of a repetition have to be taken into account.

Consider a run $(a_1 \cdots a_n)^{\frac{k}{n}}$ beginning at some position i in w.

Runs

We cannot readily apply the method used for distinct squares, because here multiple occurrences of a repetition have to be taken into account.

Consider a run $(a_1 \cdots a_n)^{\frac{k}{n}}$ beginning at some position i in w. The run ending square (RES) is the square $w[i+k-2n..i+k-1]$.

Runs

We cannot readily apply the method used for distinct squares, because here multiple occurrences of a repetition have to be taken into account.

Consider a run $(a_1 \cdots a_n)^{\frac{k}{n}}$ beginning at some position i in w. The run ending square (RES) is the square $w[i+k-2n..i+k-1]$.

If $w = aababaa$ and the run is $(ab)^{\frac{5}{2}}$

Runs

We cannot readily apply the method used for distinct squares, because here multiple occurrences of a repetition have to be taken into account.

Consider a run $(a_1 \cdots a_n)^{\frac{k}{n}}$ beginning at some position i in w. The run ending square (RES) is the square $w[i+k-2n..i+k-1]$.

If $w = aababaa$ and the run is $(ab)^{\frac{5}{2}}$, then the RES is $baba$.

Runs

We cannot readily apply the method used for distinct squares, because here multiple occurrences of a repetition have to be taken into account.

Consider a run $(a_1 \cdots a_n)^{\frac{k}{n}}$ beginning at some position i in w. The run ending square (RES) is the square $w[i+k-2n..i+k-1]$.

If $w = aababaa$ and the run is $(ab)^{\frac{5}{2}}$, then the RES is $baba$.

An upper bound on the number of run ending squares is implicitly an upper bound on the number of runs.

Runs - strategy

- if $u <_p v \in \Sigma^*$, with a being their first letter

Runs - strategy

- if $u <_p v \in \Sigma^*$, with a being their first letter
- an occurrence of uu in w is a RES if it followed by $b \neq a$ or suffix of w

Runs - strategy

- if $u <_p v \in \Sigma^*$, with a being their first letter
- an occurrence of uu in w is a RES if it followed by $b \neq a$ or suffix of w
- let run ending occurrences of u^2 at positions $\{i_1 < \cdots < i_k\}$

Runs - strategy

- if $u <_p v \in \Sigma^*$, with a being their first letter
- an occurrence of uu in w is a RES if it followed by $b \neq a$ or suffix of w
- let run ending occurrences of u^2 at positions $\{i_1 < \cdots < i_k\} \subseteq \mathbf{clust}(u)$

Runs - strategy

- if $u <_p v \in \Sigma^*$, with a being their first letter
- an occurrence of uu in w is a RES if it followed by $b \neq a$ or suffix of w
- let run ending occurrences of u^2 at positions $\{i_1 < \cdots < i_k\} \subseteq \mathbf{clust}(u)$
- if $j \in \{1, \ldots, k-1\}$ then $w[i_j + |u|] \neq w[i_j + 2 \cdot |u|]$, and
  $w[i_k + |u|] \neq w[i_k + 2 \cdot |u|]$ or $i_k + 2 \cdot |u| = |w| + 1$

Runs - strategy

- if $u <_p v \in \Sigma^*$, with a being their first letter
- an occurrence of uu in w is a RES if it followed by $b \neq a$ or suffix of w
- let run ending occurrences of u^2 at positions $\{i_1 < \cdots < i_k\} \subseteq \mathbf{clust}(u)$
- if $j \in \{1, \ldots, k-1\}$ then $w[i_j + |u|] \neq w[i_j + 2 \cdot |u|]$, and
  $w[i_k + |u|] \neq w[i_k + 2 \cdot |u|]$ or $i_k + 2 \cdot |u| = |w| + 1$
- for each $j \in \{1, \ldots, k\}$, at least one of the two positions i_j and $i_j + |u|$ is not in $\mathbf{clust}(v)$, so $|\mathbf{clust}(u)| - |\mathbf{clust}(v)| \geq k$

Runs - strategy

- if $u <_p v \in \Sigma^*$, with a being their first letter
- an occurrence of uu in w is a RES if it followed by $b \neq a$ or suffix of w
- let run ending occurrences of u^2 at positions $\{i_1 < \cdots < i_k\} \subseteq \mathbf{clust}(u)$
- if $j \in \{1, \ldots, k-1\}$ then $w[i_j + |u|] \neq w[i_j + 2 \cdot |u|]$, and
  $w[i_k + |u|] \neq w[i_k + 2 \cdot |u|]$ or $i_k + 2 \cdot |u| = |w| + 1$
- for each $j \in \{1, \ldots, k\}$, at least one of the two positions i_j and $i_j + |u|$ is not in $\mathbf{clust}(v)$, so $|\mathbf{clust}(u)| - |\mathbf{clust}(v)| \geq k$
- for consecutive roots $u_1 <_p \cdots <_p u_n$, we get that $\mathbf{clust}(u_i)$ is larger than the number of all runs with run ending square u_j^2, $j \geq i$.

Runs - strategy

- if $u <_p v \in \Sigma^*$, with a being their first letter
- an occurrence of uu in w is a RES if it followed by $b \neq a$ or suffix of w
- let run ending occurrences of u^2 at positions $\{i_1 < \cdots < i_k\} \subseteq \mathbf{clust}(u)$
- if $j \in \{1, \ldots, k-1\}$ then $w[i_j + |u|] \neq w[i_j + 2 \cdot |u|]$, and
  $w[i_k + |u|] \neq w[i_k + 2 \cdot |u|]$ or $i_k + 2 \cdot |u| = |w| + 1$
- for each $j \in \{1, \ldots, k\}$, at least one of the two positions i_j and $i_j + |u|$ is not in $\mathbf{clust}(v)$, so $|\mathbf{clust}(u)| - |\mathbf{clust}(v)| \geq k$
- for consecutive roots $u_1 <_p \cdots <_p u_n$, we get that $\mathbf{clust}(u_i)$ is larger than the number of all runs with run ending square u_j^2, $j \geq i$.
- argument does not extend easily to overlapping chains of run ending squares

Runs - strategy

- if $u <_p v \in \Sigma^*$, with a being their first letter
- an occurrence of uu in w is a RES if it followed by $b \neq a$ or suffix of w
- let run ending occurrences of u^2 at positions $\{i_1 < \cdots < i_k\} \subseteq \mathbf{clust}(u)$
- if $j \in \{1, \ldots, k-1\}$ then $w[i_j + |u|] \neq w[i_j + 2 \cdot |u|]$, and $w[i_k + |u|] \neq w[i_k + 2 \cdot |u|]$ or $i_k + 2 \cdot |u| = |w| + 1$
- for each $j \in \{1, \ldots, k\}$, at least one of the two positions i_j and $i_j + |u|$ is not in $\mathbf{clust}(v)$, so $|\mathbf{clust}(u)| - |\mathbf{clust}(v)| \geq k$
- for consecutive roots $u_1 <_p \cdots <_p u_n$, we get that $\mathbf{clust}(u_i)$ is larger than the number of all runs with run ending square u_j^2, $j \geq i$.
- argument does not extend easily to overlapping chains of run ending squares (might cover cases when two chains overlap, but does not seem to work when we have 3 peeks)

Closing remarks

1. the bound for one chain is $n-1$

   (where n is the number of occurrences of the common prefix)

Closing remarks

1. the bound for one chain is $n-1$ (where n is the number of occurrences of the common prefix)

Conjecture

For any word w, any integer $\ell \geq 0$, and any set $S = \{u_1, u_2, \ldots, u_n\}$ s.t., for all $i \in \{1, \ldots, n\}$, u_i^ℓ is a factor of w and $u_1 \leq_p u_i$, we have $|S| < \frac{1}{\ell - 1}|w|_{u_1}$.

Closing remarks

1. the bound for one chain is $n-1$
   (where n is the number of occurrences of the common prefix)

Conjecture

For any word w, any integer $\ell \geq 0$, and any set $S = \{u_1, u_2, \ldots, u_n\}$ s.t., for all $i \in \{1, \ldots, n\}$, u_i^ℓ is a factor of w and $u_1 \leq_p u_i$, we have $|S| < \frac{1}{\ell-1}|w|_{u_1}$.

2. we can generalise the result for larger exponents

Closing remarks

1. the bound for one chain is $n - 1$
   (where n is the number of occurrences of the common prefix)

Conjecture

For any word w, any integer $\ell \geq 0$, and any set $S = \{u_1, u_2, \ldots, u_n\}$ s.t., for all $i \in \{1, \ldots, n\}$, u_i^ℓ is a factor of w and $u_1 \leq_p u_i$, we have $|S| < \frac{1}{\ell - 1}|w|_{u_1}$.

2. we can generalise the result for larger exponents
   (define $k-1$ representatives and anchor positions for each u^k)

Closing remarks

1. the bound for one chain is $n-1$
   (where n is the number of occurrences of the common prefix)

Conjecture

For any word w, any integer $\ell \geq 0$, and any set $S = \{u_1, u_2, \ldots, u_n\}$ s.t., for all $i \in \{1, \ldots, n\}$, u_i^ℓ is a factor of w and $u_1 \leq_p u_i$, we have $|S| < \frac{1}{\ell-1}|w|_{u_1}$.

2. we can generalise the result for larger exponents
   (define $k-1$ representatives and anchor positions for each u^k)

3. we can generalise it for single chains of run ending squares

Closing remarks

1. the bound for one chain is $n-1$
   (where n is the number of occurrences of the common prefix)

Conjecture

For any word w, any integer $\ell \geq 0$, and any set $S = \{u_1, u_2, \ldots, u_n\}$ s.t., for all $i \in \{1, \ldots, n\}$, u_i^ℓ is a factor of w and $u_1 \leq_p u_i$, we have $|S| < \frac{1}{\ell - 1}|w|_{u_1}$.

2. we can generalise the result for larger exponents
   (define $k-1$ representatives and anchor positions for each u^k)
3. we can generalise it for single chains of run ending squares
4. the big challenge: overlapping chains

Closing remarks

1. the bound for one chain is $n-1$
   (where n is the number of occurrences of the common prefix)

Conjecture

For any word w, any integer $\ell \geq 0$, and any set $S = \{u_1, u_2, \ldots, u_n\}$ s.t., for all $i \in \{1, \ldots, n\}$, u_i^ℓ is a factor of w and $u_1 \leq_p u_i$, we have $|S| < \frac{1}{\ell-1}|w|_{u_1}$.

2. we can generalise the result for larger exponents
   (define $k-1$ representatives and anchor positions for each u^k)

3. we can generalise it for single chains of run ending squares

4. the big challenge: overlapping chains
   (find a definition for anchor of u^2 which depends on anchors of all squares having u as a prefix)

Cheers

QUESTIONS