Two squares canonical factorization

Haoyue Bai, Franya Franek and William. F. Smyth

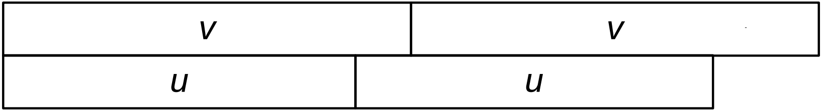
Department of Computing and Software
McMaster University, Hamilton, Ontario, Canada

Prague Stringology Conference 2014

Outline

Introduction

a configuration of two ***proportional*** squares $\boldsymbol{u}^2$ and $\boldsymbol{v}^2$



has been investigated in many different contexts:

- *Smyth et al.*: investigating three squares with intention to find a position for amortization argument for the runs conjecture a unique factorization of the type

$$\boldsymbol{u} = \boldsymbol{u_1 u_2 u_1}, \ \boldsymbol{v} = \boldsymbol{u_1 u_2 u_1 u_1 u_2} \iff \frac{3|\boldsymbol{u}|}{2} < |\boldsymbol{v}| < 2|\boldsymbol{u}|$$

- in a computational framework for computations of $\sigma_d(n)$ developed by *Deza*, *F.*, and *Jiang*: such configurations are used in *Liu*'s PhD thesis to speed up computation of certain values $\sigma_d(n)$ in the $(d, n{-}d)$ table

$\sigma_d(n)$ denotes the maximum number of distinct squares in a string of length n with d distinct symbols

- *Lam*: two rightmost squares $\boldsymbol{u}^2 \lhd \boldsymbol{v}^2$ in $\boldsymbol{x}$ have a very particular structure $\boldsymbol{u} = \boldsymbol{u_1}^{e_1}\boldsymbol{u_2}$ and $\boldsymbol{v} = \boldsymbol{u_1}^{e_1}\boldsymbol{u_2}\boldsymbol{u_1}^{e_2}$ a primitive $\boldsymbol{u_1}$ and a non-trivial proper prefix $\boldsymbol{u_2}$ of $\boldsymbol{u_1}$, and $e_1 \geq e_2 \geq 1$.

  note that two rightmost squares in $\boldsymbol{x}$ are necessarily proportional

- *Deza, F., Thierry*: two proportional squares $\boldsymbol{u}^2 \lhd \boldsymbol{v}^2$ form a *factorizable double square* if either $\boldsymbol{u}$ or $\boldsymbol{v}$ is primitive or $\boldsymbol{u}^2$ is rightmost in $\boldsymbol{v}^2$

  A factorizable double square has a *unique factorization* $(\boldsymbol{u_1}, \boldsymbol{u_2}, e_1, e_2)$: $\boldsymbol{u} = \boldsymbol{u_1}^{e_1}\boldsymbol{u_2}$ and $\boldsymbol{v} = \boldsymbol{u_1}^{e_1}\boldsymbol{u_2}\boldsymbol{u_1}^{e_2}$

where $\boldsymbol{u_1}$ is primitive, $\boldsymbol{u_2}$ a non-trivial proper prefix of $\boldsymbol{u_1}$, and $e_1 \geq e_2 \geq 1$.

moreover, there are some additional constrains to the structure of a factorizable double square

in this contribution we generalize and extend the factorization to all pairs of proportional squares starting at the same position

Basic notions

x is primitive $\iff x \neq y^p$ for any string y and any integer $p \geq 2$

Ex: *aab aab* is not primitive, while *aabaaba* is

primitive root of x: the smallest y s.t. $x = y^p$ for some integer $p \geq 1$ *(is unique and primitive)*

u^2 is primitively rooted $\iff u$ is a primitive string

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

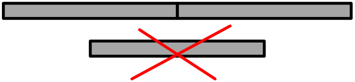
$x \lhd y \iff x$ is a proper prefix of y

A *double square* $\mathrm{DS}(\boldsymbol{u}, \boldsymbol{v})$: $\boldsymbol{u}^2 \lhd \boldsymbol{v}^2$ and $|\boldsymbol{u}| < |\boldsymbol{v}| < 2|\boldsymbol{u}|$.

*Note that in Deza, F., Thierry this would be called a **balanced double square***

Lemma (*Synchronization Principle*)

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



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.

Note that both x and y must repeat at least twice

these lemmas are really a folklore, but we included proofs as we did not know of a published proof of the Common Factor Lemma

A simple corollary of Common Factor Lemma:

Corollary (Uniqueness Lemma)

$\boldsymbol{x}$ and $\boldsymbol{y}$ primitive strings, $p, q \geq 1$:

(a) $\boldsymbol{x}^p = \boldsymbol{y}^q \;\Rightarrow\; \boldsymbol{x} = \boldsymbol{y} \;\&\; p = q$

(b) $p, q \geq 2, \quad \boldsymbol{x_1} \lhd \boldsymbol{x}, \quad \boldsymbol{y_1} \lhd \boldsymbol{y}$

$$\boldsymbol{x}^p \boldsymbol{x_1} = \boldsymbol{y}^q \boldsymbol{y_1} \;\Rightarrow\; \boldsymbol{x} = \boldsymbol{y} \;\&\; \boldsymbol{x_1} = \boldsymbol{y_1} \;\&\; p = q$$

(a) $x^p = y^q$

- $p = 1$

  then $\boldsymbol{x} = \boldsymbol{y}^q$, $\boldsymbol{x}$ primitive $\Rightarrow$ $q = 1$ and $\boldsymbol{x} = \boldsymbol{y}$

- $p, q \geq 2$

  $\boldsymbol{x}^p$ and $\boldsymbol{y}^q$ have a common factor of length $\geq |\boldsymbol{x}|+|\boldsymbol{y}|$, by the Common Factor Lemma $\boldsymbol{x} \sim \boldsymbol{y}$, hence $\boldsymbol{x} = \boldsymbol{y}$

(b) $\boldsymbol{x}^p\boldsymbol{x_1} = \boldsymbol{y}^q\boldsymbol{y_1}$, $p, q \geq 2$

$\boldsymbol{x}^p\boldsymbol{x_1} = \boldsymbol{y}^q\boldsymbol{y_1}$ have a common factor of length $|\boldsymbol{x}|+|\boldsymbol{y}|$, hence $x = y$

the requirement $p, q \geq 2$ is essential – for instance:

$\boldsymbol{x} = aabb$, $\boldsymbol{x_1} = aa$ and $p = 2$:

$$\boldsymbol{x}^2\boldsymbol{x_1} = aabbaabbaa$$

$\boldsymbol{y} = aabbaabba$, $\boldsymbol{y_1} = a$ and $q = 1$:

$$\boldsymbol{y}^1\boldsymbol{y_1} = aabbaabbaa$$

$$\boldsymbol{x}^2\boldsymbol{x_1} = \boldsymbol{y}^1\boldsymbol{y_1}$$

Main results

Lemma (Two Squares Factorization Lemma)

$\forall\ \mathrm{DS}(\boldsymbol{u}, \boldsymbol{v})\ \exists$ *unique* $\boldsymbol{u_1}, \boldsymbol{u_2}, e_1, e_2$ *such that*
$\boldsymbol{u} = \boldsymbol{u_1}^{e_1}\boldsymbol{u_2}$ *and* $\boldsymbol{v} = \boldsymbol{u_1}^{e_1}\boldsymbol{u_2}\boldsymbol{u_1}^{e_2}$

$\boldsymbol{u_1}$ *primitive*

$\boldsymbol{u_2}$ *is a possibly trivial proper prefix of* $\boldsymbol{u_1}$

$e_1 \geq e_2 \geq 1$

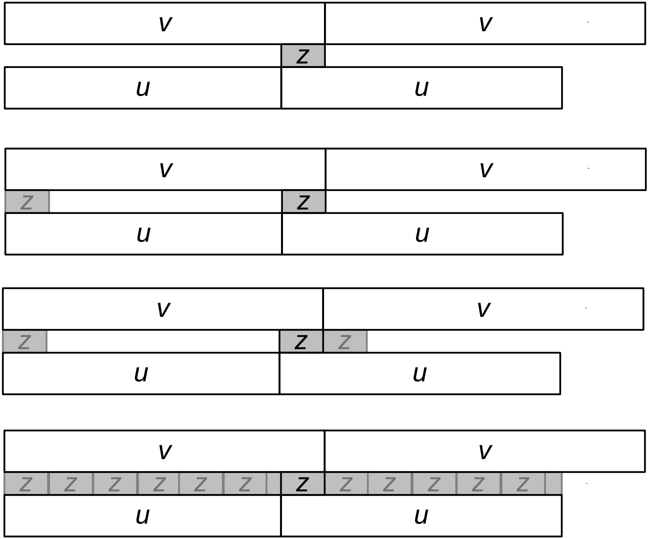
Moreover,

(a) $|\boldsymbol{u_2}| = 0 \ \Rightarrow\ e_1 > e_2 \geq 1$

(b) $|\boldsymbol{u_2}| > 0 \ \Rightarrow\ \boldsymbol{v}$ *primitive*

(c) $|\boldsymbol{u_2}| > 0\ \&\ e_1 \geq 2 \ \Rightarrow\ \boldsymbol{u}$ *primitive*

McMaster University


v
v
z
u
u
v
v
z
z
u
u
v
v
z
z
z
u
u
v
v
z z z z z z z z z z z z
u
u

$\exists\ k \geq 1$ s.t. $\boldsymbol{u} = \boldsymbol{z}^k \boldsymbol{z'}$ for some proper prefix $\boldsymbol{z'}$ of $\boldsymbol{z}$

$\boldsymbol{u_1}$ primitive root of $\boldsymbol{z}$

$\boldsymbol{z} = \boldsymbol{u_1}^{e_2}$ for some $e_2 \geq 1$

for some $e_1 \geq e_2 k$ and some prefix $\boldsymbol{u_2}$ of $\boldsymbol{u_1}$,

$\boldsymbol{u} = \boldsymbol{u_1}^{e_1} \boldsymbol{u_2}$ and $\boldsymbol{v} = \boldsymbol{uz} = \boldsymbol{u_1}^{e_1} \boldsymbol{u_2} \boldsymbol{u_1}^{e_2}$

now the uniqueness and other properties:

(i) $|\boldsymbol{u_2}| = 0$

$\boldsymbol{u} = \boldsymbol{u_1}^{e_1}$ and $\boldsymbol{v} = \boldsymbol{u_1}^{e_1+e_2} \Rightarrow \boldsymbol{v} = \boldsymbol{u_1}^{2(e_1+e_2)}$

$|\boldsymbol{v}| < 2|\boldsymbol{u}|$ and $e_1 \geq e_2 \Rightarrow e_1 > e_2$

uniqueness of $\boldsymbol{u_1}$ is a consequence of Uniqueness Lemma (a)

(ii) $|\boldsymbol{u_2}| > 0$

assume $(\boldsymbol{w_1}, \boldsymbol{w_2}, f_1, f_2)$

$\boldsymbol{u} = \boldsymbol{u_1}^{e_1}\boldsymbol{u_2} = \boldsymbol{w_1}^{f_1}\boldsymbol{w_2}$ & $\boldsymbol{v} = \boldsymbol{u_1}^{e_1}\boldsymbol{u_2}\boldsymbol{u_1}^{e_2} = \boldsymbol{w_1}^{f_1}\boldsymbol{w_2}\boldsymbol{w_1}^{f_2}$

$$e_1 = f_1 = 1 \quad \Rightarrow \quad \boldsymbol{v} = \boldsymbol{u}\boldsymbol{u_1} = \boldsymbol{u}\boldsymbol{w_1} \quad \Rightarrow \quad \boldsymbol{u} = \boldsymbol{v}$$

WLOG assume that $f_1 > e_1 = 1$

$\boldsymbol{u} = \boldsymbol{u_1}\boldsymbol{u_2} = \boldsymbol{w_1}^{f_1}\boldsymbol{w_2}$ & $\boldsymbol{v} = \boldsymbol{u_1}\boldsymbol{u_2}\boldsymbol{u_1} = \boldsymbol{w_1}^{f_1}\boldsymbol{w_2}\boldsymbol{w_1}^{f_2} \Rightarrow$
$\boldsymbol{u_1} = \boldsymbol{w_1}^{f_2}$

$\boldsymbol{u_1}$ primitive forces $f_2 = 1$ and $\boldsymbol{u_1} = \boldsymbol{w_1}$

$\boldsymbol{u_1}\boldsymbol{u_2} = \boldsymbol{w_1}^{f_1}\boldsymbol{w_2} = \boldsymbol{u_1}^{f_1}\boldsymbol{w_2}$, implies that $f_1 = 1$
a contradiction

show that $\boldsymbol{v}$ is primitive

suppose the contrary: $\boldsymbol{v} = \boldsymbol{w}^k, k \geq 2$

$|\boldsymbol{w}| \leq \frac{|\boldsymbol{v}|}{2} \leq |\boldsymbol{u_1}^{e_1}| + |\boldsymbol{u_2}|$

$\boldsymbol{w}^{2k} = \boldsymbol{v}^2 = \boldsymbol{u_1}^{e_1}\boldsymbol{u_2}\boldsymbol{u_1}^{e_1+e_2}\boldsymbol{u_2}\boldsymbol{u_1}^{e_2}$

$\boldsymbol{w}^{2k}$ and $\boldsymbol{u_1}^{e_1+e_2}\boldsymbol{u_2}$ have a common factor $\boldsymbol{u_1}^{e_1+e_2}\boldsymbol{u_2}$ of length $(|\boldsymbol{u_1}^{e_1}|+|\boldsymbol{u_2}|)+|\boldsymbol{u_1}^{e_2}| \geq |\boldsymbol{w}|+|\boldsymbol{u_1}|$

apply Common Factor Lemma to conclude that $\boldsymbol{w} \sim \boldsymbol{u_1}$, thus $\boldsymbol{w} = \boldsymbol{u_1}$

primitive string $\boldsymbol{u_1} = \boldsymbol{u_2}\overline{\boldsymbol{u}}_2$ aligns with $\boldsymbol{u_2}\boldsymbol{u_1}$, and so $\overline{\boldsymbol{u}}_2$ is a prefix of $\boldsymbol{u_1}$, in contradiction to Synchronization Principle

let $e_2 \geq 2$ show that $\boldsymbol{u}$ is primitive
suppose the contrary: $\boldsymbol{u} = \boldsymbol{w}^k$, $k \geq 2$

Hence $|\boldsymbol{w}| \leq \frac{|\boldsymbol{u}|}{2} = \frac{(|\boldsymbol{u_1}^{e_1}|+|\boldsymbol{u_2}|)}{2} < |\boldsymbol{u_1}^{e_1-1}|+|\boldsymbol{u_2}|$

$e_2 \geq 1$ and $e_2 \geq 2 \quad \Rightarrow \quad e_1+e_2 \geq 3$

$\boldsymbol{u_1}^{e_1}\boldsymbol{u_2} \lhd \boldsymbol{u}^2 = \boldsymbol{w}^{2k}$

so $\boldsymbol{w}^{2k}$ and $\boldsymbol{u_1}^{e_1+e_2}$ have a common factor $\boldsymbol{u_1}^{e_1}\boldsymbol{u_2}$

since $|\boldsymbol{u_1}^{e_1}\boldsymbol{u_2}| \geq |\boldsymbol{v}|+|\boldsymbol{u_1}|$, applying Common Factor Lemma,
$\boldsymbol{u_1} = \boldsymbol{w}$

this in turn implies $\boldsymbol{u} = \boldsymbol{u_1}^{e_1}\boldsymbol{u_2} = \boldsymbol{u_1}^k$, impossible since
$0 < |\boldsymbol{u_2}| < |\boldsymbol{u_1}|$

observations:

$|\boldsymbol{u_2}| > 0$ if any one of the following conditions holds:

(a) $\boldsymbol{v}$ is primitive

(b) $\boldsymbol{u}$ is primitive

(c) $\boldsymbol{u}^2$ is rightmost in $\boldsymbol{v}^2$

moreover:

(d) $|\boldsymbol{u_2}| > 0 \iff \boldsymbol{v}$ is primitive

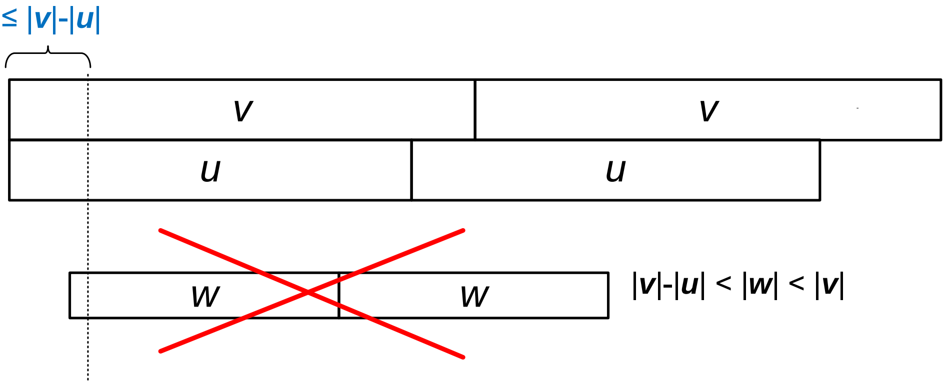
Applications

we concluded the paper with a comment and a sketch of how the canonical factorization could be applied to New Periodicity Lemma:

Lemma (2006, Fan, Puglisi, Smyth, and Turpin)

Let $\boldsymbol{x} = \mathrm{DS}(\boldsymbol{u}, \boldsymbol{v})$, where we require that $\boldsymbol{u}^2$ be regular and that $\boldsymbol{v}$ be primitive. There is no square $\boldsymbol{w}^2$ starting at position i, $1 \leq i < |\boldsymbol{v}| - |\boldsymbol{u}|$ with $|\boldsymbol{v}| - |\boldsymbol{u}| < |\boldsymbol{w}| < |\boldsymbol{v}|$ except possibly $|\boldsymbol{w}| = |\boldsymbol{u}|$.

v^2 primitive, u^2 regular



we can report that since the final submission to PSC2014, we were able to prove using the canonical factorization an extended NPL:

Theorem

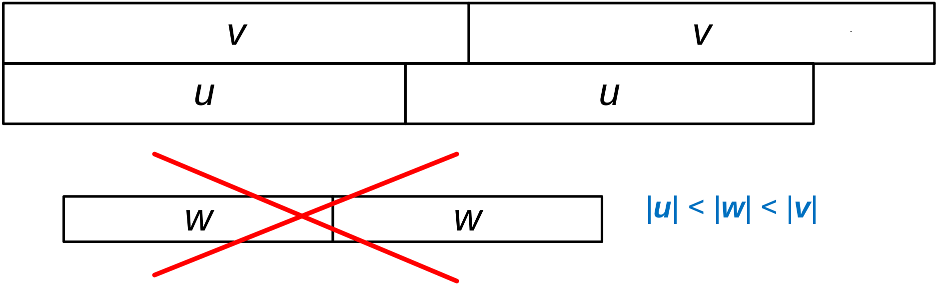
Consider a double square $\mathrm{DS}(\boldsymbol{u}, \boldsymbol{v})$ *and let* $\boldsymbol{u'}$ *be a suffix of* $\boldsymbol{u}$ *so that* $\boldsymbol{v} = \boldsymbol{uu'}$*. Let* $\boldsymbol{w}^2$ *be any square that is a factor of* $\boldsymbol{v}^2$*. Then exactly one of the following mutually exclusive cases holds:*

(a) $\boldsymbol{w} = \boldsymbol{v}$, *or*

(b) $|\boldsymbol{w}| < |\boldsymbol{u}|$, *or*

(c) $|\boldsymbol{u}| \leq |\boldsymbol{w}| < |\boldsymbol{v}|$ *and the primitive root of* $\boldsymbol{w}$ *is a conjugate of the primitive root of* $\boldsymbol{u'}$*.*

McMaster University



$|u| < |w| < |v|$

Lemma (*Crochemore-Rytter* (1995), *Fraenkel-Simpson* (1998))

Let $\boldsymbol{u}^2 \lhd \boldsymbol{v}^2 \lhd \boldsymbol{w}^2$ and let $\boldsymbol{u}$ be primitive, then $|\boldsymbol{u}|+|\boldsymbol{v}| \leq |\boldsymbol{w}|$.

we can also report that since the final submission to PSC2014, we were able to prove using the canonical factorization a generalization of the above lemma:

Theorem (*Bai*, *Deza*, and *F.*)

Let $\boldsymbol{u}^2 \lhd \boldsymbol{v}^2 \lhd \boldsymbol{w}^2$. Then either

(*a*) $|\boldsymbol{u}|+|\boldsymbol{v}| \leq |\boldsymbol{w}|$

or *(inclusive or)*

(*b*) *$\boldsymbol{u}$, $\boldsymbol{v}$, and $\boldsymbol{w}$ have the same primitive root*

THANK YOU

M. Crochemore and W. Rytter
Squares, cubes, and time-space efficient string searching
Algorithmica, 1995

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

A. Deza, F. Franek, and A. Thierry
How many double squares can a string contain?
to appear in Discrete Applied Mathematics, 2014

A. Deza, F. Franek, and M. Jiang
A computational framework for determining square-maximal strings
Proceedings of the Prague Stringology Conference 2012

K. Fan, S.J. Puglisi, W. F. Smyth, and A.Turpin
A new periodicity lemma
SIAM J. Discrete Math., 2006

A.S. Fraenkel and J. Simpson
How many squares can a string contain?
Journal of Combinatorial Theory, Series A, 1998

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

E. Kopylova and W.F. Smyth
The three squares lemma revisited
Journal of Discrete Algorithms, 2012

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
PhD thesis, Dept. of Computing and Software, McMaster University, 2013