One more proof of Thue's theorem on nonrepetitive sequences

Jakub Kozik

Jagiellonian University

17th CLA

repetition/sqare

nonsense lemma

Mississippi

Thue '06:

There exists an infinite square-free word over 3-letter alphabet.

Proof 1

Consider substitution s :

$a \to abcab$ $\qquad$ $b \to acabcb$

$c \to acbcac$

If w is square-free

then $s(w)$ is square free.

Proof 2 – Thue sequence

Thue–Morse sequence:

$0 \to 01$

$1 \to 10$

$$\begin{cases} t_0 = 0 \\ t_{2n} = t_n \\ t_{2n+1} = 1 - t_n \end{cases}$$

0
01
0110
01101001
0110100110010110...
...

avoids cubes (i.e. www)

Thue sequence

01101000110010110...

2 1 0 2 0 1 2 ...

avoids squares

Entropy compression argument

(following Moser–Tardos '10)

|3|2|3|1|1|2|3|2|3|3|1 →

3 →

3 →

2 →

3 →

2 →

1 →

1 →

3 →

2 →

3 →

M

a

a c

a c c

a c

a c b

a c b c

a c b c b

a c b

a c b a

a c b a a

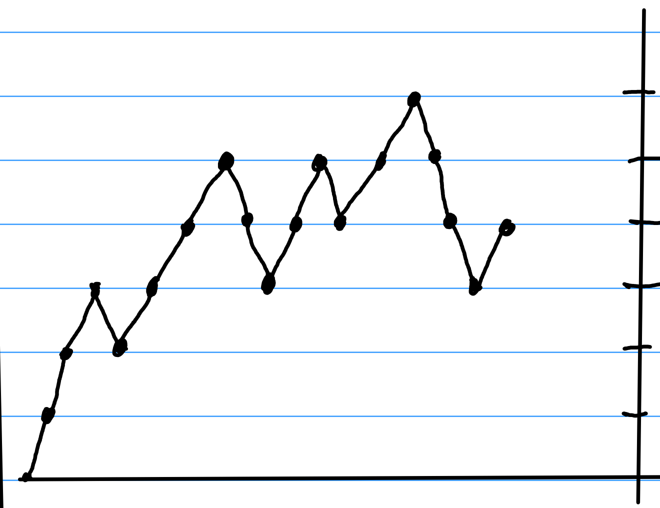
a c b a

a c b a c

a c b a c b

a c b

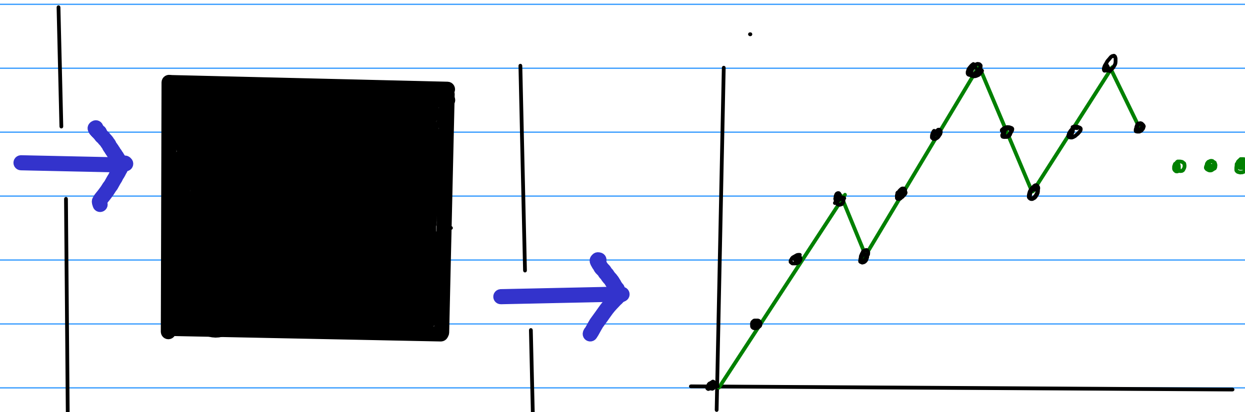
a c b c



Entropy compression argument

(following Moser–Tardos '10)

...|0|0|2|1|0|1|2



input sequences

3^N

output sequences (Logs)

$o(4^N)$

mapping is almost injective

Nonrepetitive sequences - entropy compression

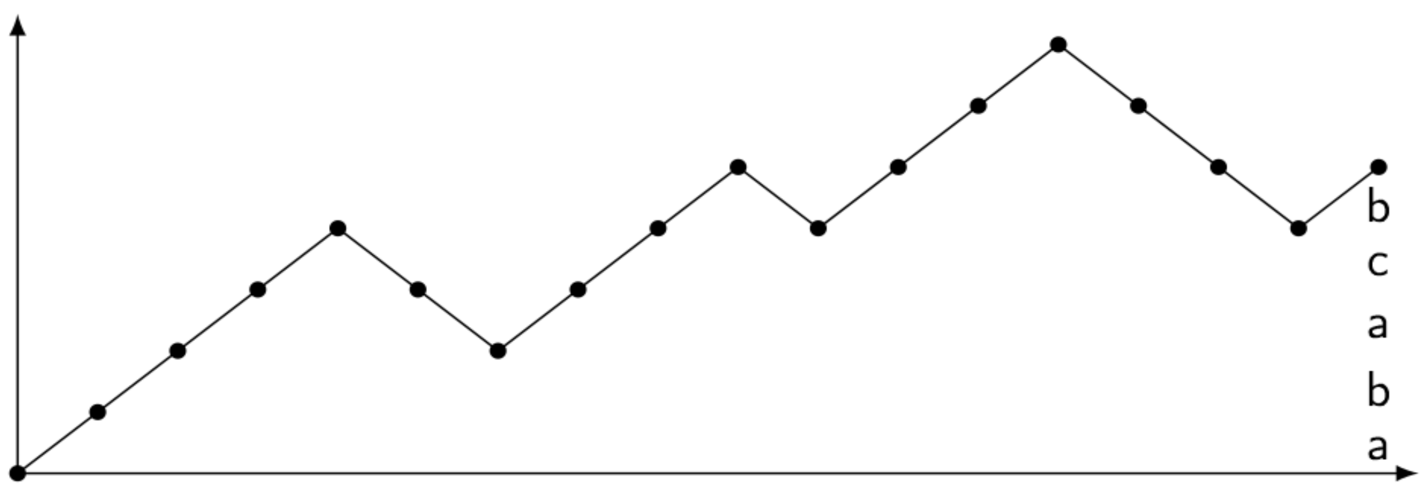
(Grytczuk, JK, Micek 2010)

Algorithm (suppose it never stops)

$i := 1$
while $i \leqslant n$ **do**
  $s_i :=$ random symbol
  if $s_1, \ldots, s_i$ is nonrepetitive **then**
    $i := i + 1$
  else
    there is exactly one repetition, say of size h
    $i := i - h + 1$

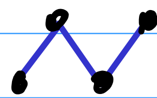
Counting (N steps)

# of lattice paths	$o(4^N)$
# of final words	$o(1)$
Total logs	$o(4^N)$
Total random seq.	C^N

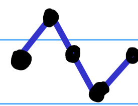


Constraining input sequence

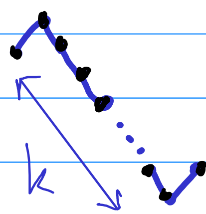
- no repeated letters on the input

  → no "" on the output

- no squares of length 4 on the input

  → no "" on the output

- no squares of length 2k on the input

  → no "

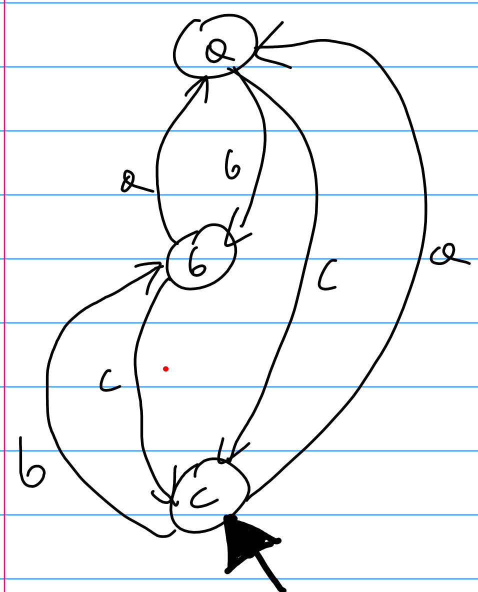
" on the output

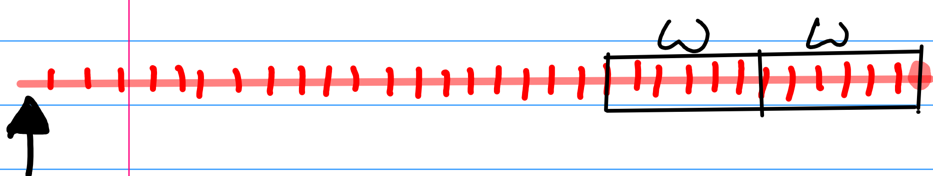
Obs. 1

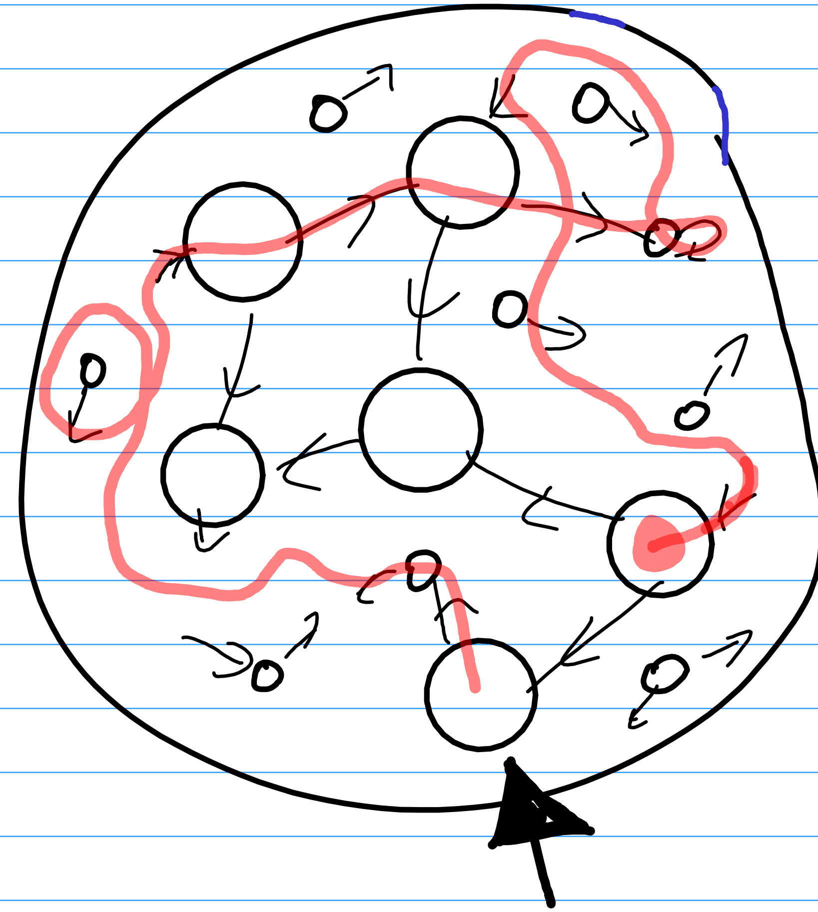
Language of words that avoid squares of lengths up to k is REGULAR.

any constant

How retractions affect input automaton?





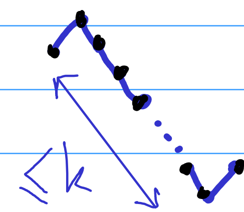


Obs. Retractions preserve the state of the automaton.

Obs. 1

Language of words over Σ without squares of size $\leq k$ is regular. (easy to count)

Obs. 2

Set of / \ paths without "

"

is even easier to count.

$$\left(\text{g.f. satisfies } f = z + z\frac{f^{k+1}}{1-f} \right)$$

Can it be true that
for large enough k
the # of input sequences
is eventually larger
than the # of logs?

Can we expect the procedure to stop?

A more direct method was initiated by Brinkhuis [7]. He considers a 25-uniform substitution f from A^* into itself defined by

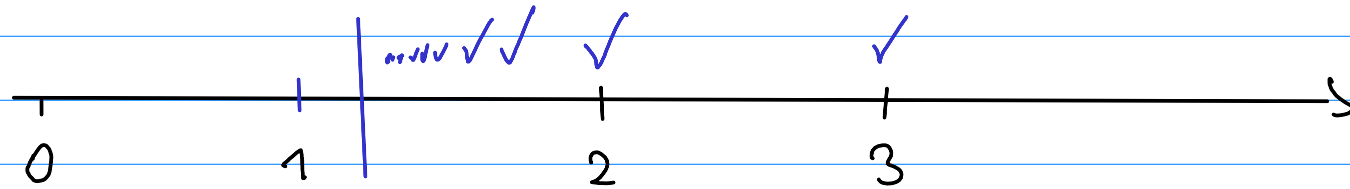
$$f: \begin{array}{rcl} 0 & \mapsto & \{U_0, V_0\} \\ 1 & \mapsto & \{U_1, V_1\} \\ 2 & \mapsto & \{U_2, V_2\} \end{array}$$

where $U_0 = x1\tilde{x}$, $V_0 = y0\tilde{y}$ and $x = 012021020102$ and $y = 012021201021$. The words $U_1, \ldots, V_2$ are obtained by applying the circular permutation $(0, 1, 2)$. He proves that f is square-free, and thus every square-free word w of length n is mapped onto 2^n square-free words of length $25n$. His bound is only $2^{n/24}$.

Better bounds by: Ekhad, Zeilberger (D.) '98, Grimm '01, Sun '03 and ...

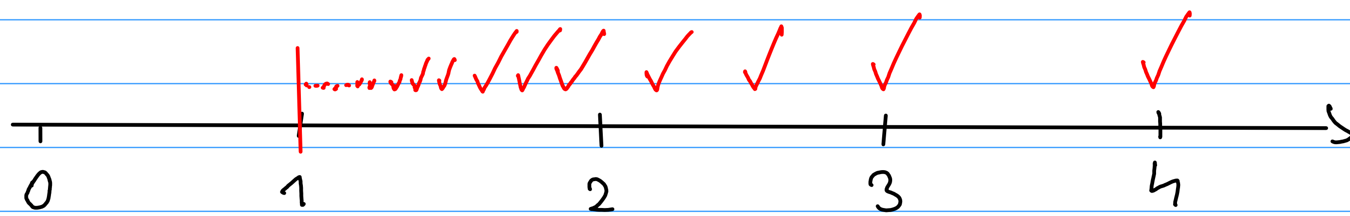
Constrained input

$$(\lambda_I)^N$$



Constrained logs

$$(\lambda_O)^N$$



... and Kolpakov '06

and Shur '09

2	4	20	[illegible]	[illegible]	[illegible]
3	2	54	1,30175907	1,3017618923	$3 \cdot 10^{-6}$
3	2^{+}	17	2,60587894	2,6058790806	$2 \cdot 10^{-7}$

(almost) entropy compression

before MT'10

Inclusion-exclusion principle

Combinatorics on words

(formal power series)

Golod/Bell, Goh

Shur's bounds

Goulden-Jackson cluster method

Probabilistic Combinatorics

(probabilistic spaces)

Lovasz Local Lemma

Shearer condition

MT-Resample

Entropy compression

E.c. with constraints

Thank

YOU!