An explicit condition for boundedly supermultiplicative subshifts

Vuong Bui (joint work with Matthieu Rosenfeld)

work done during postdoctoral fellowship in Montpellier

One World Combinatorics on Words Seminar
March 2025

Problem setting

We fix

- an alphabet Σ
- a set of forbidden factors $\mathcal{F} \subseteq \Sigma^+$

We care about

- the language $\mathcal{L}$ of words avoiding $\mathcal{F}$ (that is, every $w \in \mathcal{L}$ does not contain any factor from $\mathcal{F}$)

We do not consider

- finite $\mathcal{F}$ since the language can be recognized by an automaton (rich theory there).

In this talk, we mostly consider $\mathcal{F}$ containing

- squares
- the results of other sets are briefly mentioned

THREE (similar) ideas will be presented. (The first already known, the two latter extend the first.)

Submultiplicativity (and an upper bound)

For any set $\mathcal{F}$ of forbidden factors, the number C_n of words avoiding $\mathcal{F}$ with length n is *submultiplicative*:

$$C_{\ell+m} \leq C_\ell C_m.$$

By Fekete's lemma, the limit of the growth rate exists and is:

$$\alpha = \lim_{n\to\infty} \sqrt[n]{C_n} = \inf \sqrt[n]{C_n}.$$

In other words, we have an upper bound on α: For every n,

$$\alpha \leq \sqrt[n]{C_n}.$$

A second way to upper bound the growth rate α

- Any *finite subset* of the forbidden factors induces an *superset* of the language, which can be formulated by an *automaton*.
- The larger the subset, the closer the upper bound is, and it converges to α.

How to get a lower bound $\alpha \geq ??$

Do almost the same as the upper bound:

- take a finite subset $\mathcal{F}'$ of forbidden factors
- consider the corresponding automaton

and

- (roughly) upper bound the number of words that are $\mathcal{F}'$-free but not $\mathcal{F}$-free by the corresponding number of $\mathcal{F}'$-free words.
- subtract these upper bounds to get a lower bound on the number of $\mathcal{F}$-free words.

Case study with $\mathcal{F}$ containing all squares[1]

- suppose the alphabet Σ has 4 letters
- let C_n denote the number of square-free words of length n

Suppose $C_m \geq 2C_{m-1}$ for every $m \leq n$, we prove

$$C_{n+1} \geq 2C_n.$$

If v is square-free and $c \in \Sigma$ but vc is not square-free, then

$$vc = uyy$$

for some square yy, say of period $p = |y|$. Hence,

$$\begin{aligned} C_{n+1} &\geq C_n|\Sigma| - \sum_{p\geq 1} C_{n+1-p} \\ &\geq 4C_n - \sum_{p\geq 1} \frac{C_n}{2^{p-1}} \\ &= C_n\left(4 - (2^0 + 2^{-1} + 2^{-2} + \dots)\right) \\ &\geq 2C_n. \end{aligned}$$

[1] Rosenfeld's example

Improve the bound?

In total, we want to conclude the existence of β so that

$$C_{n+1} \geq \beta C_n$$

for every n.
We can integrate an automata (more context of suffix) into counting to improve β, or simply increase $|\Sigma|$ if we just want to prove some proof of concept.
Note that $C_{n+1} \geq \beta C_n$ is different and stronger than

$$C_n \geq \beta^n.$$

We will see how they make a difference later.

A second way to obtain lower bounds on α, in terms of C_n itself instead of an automata

- Submultiplicativity gives an upper bound on α in terms of C_n.
- What if we can prove a form of supermultiplicativity, say by start with

$$C_{\ell+m} \geq C_\ell C_m - \#\text{ invalid concatenations?}$$

Pick a square-free word u of length ℓ and a square-free word v of length m. If $u \cdot v$ is not square-free, then

$$u \cdot v = xy \odot yz,$$

where $|xy| \leq \ell$ or $|xy| > \ell$ (depending on the relative positions of $\cdot$ to $\odot$). Hence,

$$C_{\ell+m} \geq C_\ell C_m - \sum_{p \geq 1} \left(\sum_{i=0}^{p-1} C_{\ell-i} C_{m-(p-i)} + \sum_{i=1}^{p-1} C_{\ell-i} C_{m-(p-i)} \right).$$

A second way to obtain lower bounds on α ... (cont.)

$$C_{\ell+m} \geq C_\ell C_m - \sum_{p\geq 1}\left(\sum_{i=0}^{p-1} C_{\ell-i} C_{m-(p-i)} + \sum_{i=1}^{p-1} C_{\ell-i} C_{m-(p-i)}\right)$$

Remember

$$C_{n+1} \geq \beta C_n,$$

which means

$$C_{\ell-i} C_{m-(p-i)} \leq \frac{C_\ell}{\beta^i}\frac{C_m}{\beta^{p-i}} = \beta^{-p} C_\ell C_m.$$

In total,

$$C_{\ell+m} \geq C_\ell C_m \left(1 - \sum_{p\geq 1}\left(\sum_{i=0}^{p-1} \beta^{-p} + \sum_{i=1}^{p-1} \beta^{-p}\right)\right) = f(\beta) C_\ell C_m,$$

where

$$f(\beta) = 1 - \sum_{p\geq 1}(2p-1)\beta^{-p}.$$

A second way to obtain lower bounds on α... (cont.)

$$C_{\ell+m} \geq \left(1 - \sum_{p\geq 1}(2p-1)\beta^{-p}\right) C_\ell C_m = f(\beta) C_\ell C_m.$$

$|\Sigma|$ large enough
$\implies$ β large enough $\implies$ $f(\beta) > 0$ $\implies$ C_n supermultiplicative.
For $|\Sigma| \geq 5$, let $\beta = \frac{|\Sigma|+\sqrt{|\Sigma|^2-4|\Sigma|}}{2}$. Then C_n is supermultiplicative with

$$C_{\ell+m} \geq \left(1 - \frac{1+\beta}{(\beta-1)^2}\right) C_\ell C_m.$$

A bit of summarization

- The first problem.
  Concatenation of a square-free word and a letter: A square may appear at the end only, with the conclusion

$$C_{n+1} \geq \beta C_n.$$

- The second problem.
  Concatenation of a square-free word and a square-free word: A square may appear at the boundary between two words only, with the conclusion

$$C_{\ell+m} \geq \text{const}\, C_\ell C_m.$$

Both conclusions are stronger than $C_n \geq \beta^n$ (up to constants). Note that the second problem uses the first conclusion. The second conclusion is "qualitatively" stronger than the first conclusion.

- What is the third? Maybe a word gets concatenated with itself? What is the conclusion? Which previous conclusions does it use?

Circular words

- A circular word is a word written on a circle (instead of writing from left to right).
- A word may be a square-free but may not be square-free when being written on a circle, e.g., "abcdefa", or more complicated with a square of period 2 as in "abacdefb".
- Such a square-free word w must be written as

$$w = ABC$$

  with CA being a square. (Imagine $ww = ABCABC$ contains a smaller square, relatable to minimal squares.)
- Square-free circular words are included in square-free words, but Shur conjectured both grow at the same exponential rate.

Same growth rate for square-free and circular square-free words when $|\Sigma| \geq 5$

- If w is square-free but not circular square free then $w = ABC$ with
$$C \cdot A = y \odot y.$$
- Suppose $|y| = i$. We have $\leq 2i - 1$ options for the position of $\odot$ in CA. Taking these two, we can recover w from the word obtained by deleting one of the two y from w.
- Let R_n be the number of circular square-free words:
$$R_n \geq C_n - \sum_{i \geq 1} \sum_{j=1}^{2i-1} C_{n-i}.$$
- It follows from $C_{n+1} \geq \beta C_n$ that
$$R_n \geq C_n \left(1 - \sum_{i \geq 1} (2i - 1)\beta^{-i} \right).$$

Same growth rate for square-free and circular square-free words when $|\Sigma| \geq 5$ (cont.)

- When $|\Sigma|$ large enough, we have β large enough, hence

$$R_n \geq \text{const}\, C_n.$$

- However, it is obvious that

$$C_n \geq R_n,$$

  which means both C_n, R_n grow at the same rate.
  This partially solves Schur's conjecture, which asks for every alphabet size.
- Remark: A suffix and a prefix of the same word are concatenated (instead of different words). This suggests a relatable observation: The two seem to be independent enough in long words.

Some results for the general set of forbidden factors $\mathcal{F}$ (first)[2]

Theorem

Suppose there exists $\beta > 1$ *such that*

$$|\Sigma| \geq \beta + \sum_{f \in \mathcal{F}} \beta^{1-|f|},$$

then for all $n \geq 0$*, the number* C_n *of* $\mathcal{F}$*-free words of length* n *satisfies*

$$C_{n+1} \geq \beta C_n.$$

[2] This is not exactly new.

Some results for the general set of forbidden factors $\mathcal{F}$ (second)

Theorem

Suppose there exists $\beta > 1$ *such that*

$$|\Sigma| > \beta + \sum_{f \in F} \beta^{1-|f|},$$

then for every ℓ, m,

$$C_{\ell+m} \geq \gamma C_\ell C_m,$$

where

$$\gamma = 1 - \sum_{f \in \mathcal{F}} (|f| - 1)\beta^{-|f|} > 0.$$

Some results for the general set of forbidden factors $\mathcal{F}$ (third)

Theorem

Suppose there exists $\beta > 1$ such that $C_{n+1} \geq \beta C_n$ for all $n \geq 0$, and such that

$$\gamma = 1 - \sum_{f \in \mathcal{F}} (|f| - 1)\beta^{-|f|} > 0,$$

then

$$C_n \geq R_n \geq \gamma C_n.$$

Some problems we have not solved (at least yet)

- How to manage the states of the automaton when concatenating the boundaries of two words in the proof of the supermultiplicativity? (A success would probably reduce $|\Sigma| \geq 5$ to something smaller!)
- How to explain the coincidence that the two conditions

$$\beta + \sum_{f \in \mathcal{F}} \beta^{1-|f|} \leq |\Sigma|,$$

$$1 - \sum_{f \in \mathcal{F}} (|f| - 1) x^{-|f|} > 0,$$

one is the derivative of the other? It is not so clear in the construction of the proof.

Some problems we have not solved (cont.)

- The technique always fails when the number of forbidden factors of length n in $\mathcal{F}$ grows at the same rate as C_n. For example: the language of self-avoiding walks with $\mathcal{F}$ being self-avoiding polygons.
  Meanwhile, the problem of square-free words seems to be relatively easy for the reason that the growth of C_n is the square of the growth of $\mathcal{F}$ being the minimal squares (at least for $|\Sigma| \geq 5$).

Some cousin (but not quite related) problems

- Some number, say the number of $\mathcal{F}$-free words, is *naturally* submultiplicative. Proving the other direction of supermultiplicativity is usually harder. For example: $\|A^n\|$ for a matrix A is obviously submultiplicative but proving some form of supermultiplicativity may be tricky. No one knows yet how to prove something similarly good for self-avoiding walks.
- Some numbers are however naturally supermultiplicative (also with one-line proofs). Proving the submultiplicativity often asks for something else.
  For example, the submultiplicativity of the number of $P(n)$ of polyominoes on the square lattice can be proved using: For every n, there exists some ℓ so that $(n-1)/4 \leq \ell \leq n/2$ and

$$P(n) \leq n^3 P(\ell) P(n-\ell).$$

Some takeaways

- We can bound growth rates by both automaton and the number of words C_n itself. Just subtract things appropriately. They seem dual. Good bounds obtained by either ask for computational power.
- Concatenating a letter to the end can be extended to concatenating another word to the end, and also concatenating the word itself to the end. They all work similarly with different consequences.

THANK YOU!