Square-free words

Students: Siddharth Berera, Delanna Do, Alycia Doucette, Bridget Duah, Bill Feng, Mordechai Goldberger, Andrés Gómez-Colunga, Luke Hammer, Ziqi He, Joey Lakerdas-Gayle, Amanda Lamphere, Mary Olivia Liebig, Mauditra Matin, Jacob Micheletti, Adil Oryspayev, Dan Roebuck, Sara Salazar, Noam Scully, Shiyao Shen, Wangsheng Song, Thomas Sottosanti, Juliet Whidden
Mentors: John López, Eric Rowland

$\frac{1}{3}$ **Mathematics Seminar**
Hofstra University, 2021–09–08

Combinatorics on words

Axel Thue (1863–1922)

Are there long words that don't contain repetitions?

Definition

A word on a set Σ is a sequence of elements from Σ.

Example: $\Sigma = \{0, 1\}$ $\quad w = 0110$ We call Σ the "alphabet".

Definition

A square is a word of the form xx.

$\texttt{couscous} = (\texttt{cous})^2$ $\quad \texttt{hotshots} = (\texttt{hots})^2$ $\quad 0101 = (01)^2$

Are there arbitrarily long square-free words on the alphabet $\{0, 1\}$?

010X

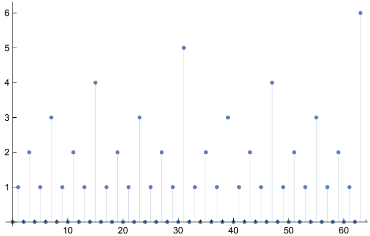
Theorem (Thue 1906)

There exist arbitrarily long square-free words on the alphabet $\{0, 1, 2\}$.

Guay-Paquet & Shallit 2009: What is the lexicographically least infinite square-free word on $\mathbb{N} = \{0, 1, 2, \ldots\}$?

$$0102010301020104\cdots$$

This is known as the ruler sequence.



Let ρ be the morphism that replaces each letter n with $0\,(n+1)$.

$$
\begin{aligned}
\rho(0) &= 01 \\
\rho^2(0) &= \rho(01) = 0102 \\
\rho^3(0) &= \rho(0102) = 01020103 \\
&\vdots \\
\rho^\infty(0) &= 010201030102010 4\cdots
\end{aligned}
$$

Theorem (Guay-Paquet–Shallit 2009)

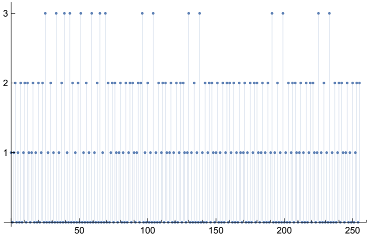
The lexicographically least square-free word on $\mathbb{N}$ is $\rho^\infty(0)$.

Varying the prefix

But this description is not robust!

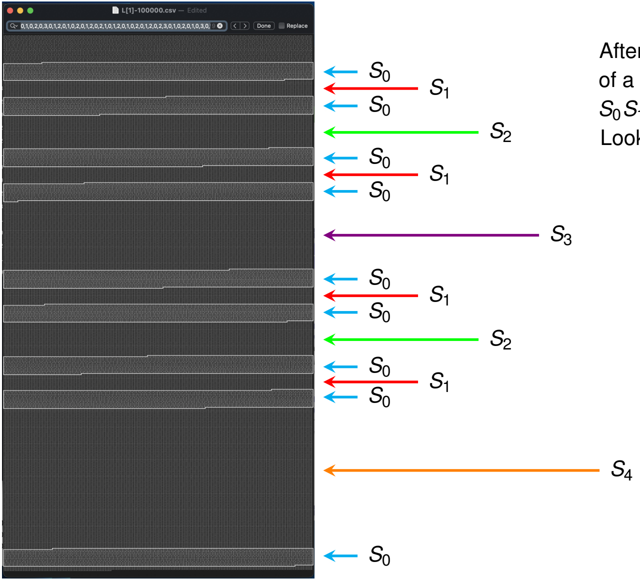
What is the lex. least infinite square-free word on $\mathbb{N}$ beginning with 1?

$$101201020120210120102012030102 01 \cdots$$



Very different word! Call it $L(1)$.
What is its structure? Is it generated by a morphism?

Structure of $L(1)$



First 100000 letters of $L(1)$

After some prefix, $L(1)$ consists of a word of the form $S_0 S_1 S_0 S_2 S_0 S_1 S_0 S_3 \cdots$. Look familiar?

The word S_{n+1} is defined in terms of S_n.

Conjecture

$L(1) = P_1\,\alpha(\rho^\infty(0))$, where P_1 is a 5177-letter prefix and $\alpha(n) = S_n$.

Proof: in progress.

Theorem

There exists a sequence of words T_n, defined recursively, such that $L(2) = 2\,\gamma(\rho^\infty(0))$, where $\gamma(n) = T_n$.

For $n \geq 3$, all $L(n)$ seem to have the same tail!

Conjecture

$L(n) = P_n\,\rho(\alpha(\rho^\infty(0)))$ for all $n \geq 3$, for some prefix P_n and $\alpha(n) = S_n$.

Longer prefixes

What if $|w| \geq 2$?

Theorem

If w can be written as $w = ps$ where p consists of two or more letters all ≥ 3, then $L(w) = p[:-1]L(p[-1]s)$.

Ex. $L(53) = 5L(3)$.

Theorem

$L(\rho(w)) = \rho(L(w))$ for all words w.

Ex. $L(040805) = \rho(L(374)) = \rho(37L(4))$.

Open questions:

- What does $L(w)$ look like in general?
- Are there only finitely many different tails that arise in $L(w)$?