# Critical exponent of infinite balanced words via the Pell number system

Aseem Baranwal, Jeffrey Shallit

School of Computer Science
University of Waterloo
Canada

Email : shallit@uwaterloo.ca
Web : https://cs.uwaterloo.ca/~shallit/

# Repetitions

## Repetitions

A *square* is a nonempty word of the form $xx$, such as the English word `hotshots`.

An *n'th power*, for $n$ a natural number, is a nonempty word of the form $x^n = \overbrace{xx \cdots x}^{n}$.

## Periods

A positive integer $p$ is the *period* of a (finite or infinite) word $x$ if $x[i] = x[i+p]$ for all indices $i$ that are meaningful.

For example, `alfalfa` has period 3.

The smallest period is called *the period*.

# Exponents

## Fractional exponent

A word $x$ is an $e$'th power (or has *exponent* $e$) if $e = |x|/p$, where $p$ is the (smallest) period. For example, `alfalfa` is a $7/3$-power.

## Critical exponent

The *critical exponent* of a word $w$ is the supremum of the exponents of all nonempty factors of $x$. For example, `Mississippi` has critical exponent $7/3$, arising from the factor `ississi`.

# Repetition threshold

**Repetition threshold**

The *repetition threshold* of a set of infinite words $S$ is the infimum, over all words $w \in S$, of the critical exponent of $w$.

- It measures the *largest unavoidable repetition*.
- For $S$ the set of all infinite words over a $k$-letter alphabet, Dejean gave a famous conjecture about this threshold in 1972.
- Finally proven by Currie and Rampersad, and independently by Rao, in 2011.

# Another example of repetition threshold

Consider the set of bi-infinite odd palindromes $O_k$ over a $k$-letter alphabet.

Results from JOS (2016):

- The repetition threshold for $O_2$ is $7/3$.
- The repetition threshold for $O_3$ is $7/4$.
- The repetition threshold for $O_k$, $k \geq 4$, is $3/2$.

# Balanced words

- A (finite or infinite) word $w$ is *balanced* if, for each letter $a$, the number of occurrences of $a$ in two identical-length finite factors of $w$ is almost the same.
- More precisely, we demand that $||x|_a - |y|_a| \leq 1$ for all letters $a$ and all identical-length factors $x, y$ of $w$.
- For a two-letter alphabet, the infinite balanced words are well understood.
    - They coincide with the *Sturmian words*, words of the form

      $$(\lfloor (n+1)\alpha + \beta \rfloor - \lfloor n\alpha + \beta \rfloor)_{n \geq 1}$$

      for real $\alpha, \beta$, $0 \leq \alpha, \beta < 1$.

# Balanced words over larger alphabets

For larger alphabets, there is a characterization of infinite balanced words due to Hubert and Graham, independently:

- Formed by taking an infinite balanced word over a 2-letter alphabet, and replacing the 0's that occur with $x^\omega = xxx\cdots$ and the 1's that occur with $y^\omega = yyy\cdots$, where $x, y$ are constant-gap words over disjoint alphabets.
- A word $x$ is called *constant-gap* if, for each letter $a$, the distance between two consecutive occurrences of $a$ in $x^\omega$ is a constant (depending on the letter).
- Example: 0102 is a constant-gap word (successive 0's occur at distance 2, 1's and 2's occur at distance 4).
- But 0120 is not constant-gap.

# Example of infinite balanced word over a 5-letter alphabet

Example: start with the Sturmian word with $\alpha = \sqrt{2} - 1$ and $\beta = 0$:

$$0\ 1\ 0\ 1\ 0\ 0\ 1\ 0\ 1\ 0\ 0\ 1\ 0\ 1\ 0\ 1\ 0\ 0\ 1\ 0\ 1\ \cdots$$

Substitute $(0102)^\omega$ for the 0's and $(34)^\omega$ for the 1's to get:

$$\begin{array}{cccccccccccccccc} 0 & 1 & 0 & 1 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 1 & 0 & 1 & 0 & \cdots \\ \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \cdots \\ 0 & 3 & 1 & 4 & 0 & 2 & 3 & 0 & 4 & 1 & 0 & 3 & 2 & 4 & 0 & \cdots \end{array}$$

This is a balanced word over a 5-letter alphabet. Call it $\mathbf{x}_5$.

# Critical exponents of infinite balanced words

Imposing balance is a strong condition.

There are only polynomially-many balanced words of length $n$ over a $k$-letter alphabet.

We could consider the repetition threshold of the class of infinite balanced words.

# Previous work on balanced words

- For alphabet size 2, the repetition threshold is $(5+\sqrt{5})/2 = 3.61803\cdots$ (Damanik & Lenz; Justin & Pirillo).
- Rampersad-JOS-Vandomme proved that the repetition threshold for alphabet size 3 is $2+\sqrt{2}/2 = 2.7071\cdots$.
- They also proved that the repetition threshold for alphabet size 4 is $(5+\sqrt{5})/4 = 1.809\cdots$.
- They conjectured that for $k \geq 5$, the repetition threshold for the infinite balanced words over a $k$-letter alphabet is $(k-2)/(k-3)$.

# Our main result

We show that the Rampersad-JOS-Vandomme conjecture is true for $k = 5$: the repetition threshold for infinite balanced words over a 5-letter alphabet is $3/2$.

Proof sketch: Step 1:

- Only finitely many balanced words over a 5-letter alphabet have critical exponent $< 3/2$
  - The longest is of length 44: 01203104120130410213014021031401203104120130
  - It can be found by a depth-first search of the tree of balanced words

# Our main result

Step 2:

- We construct a candidate infinite balanced word by starting with the Sturmian word $\alpha = \sqrt{2} - 1$ and $\beta = 0$ and substituting 0102 for the 0's and 34 for the 1's.
- That gives us the infinite balanced word

$$\mathbf{x}_5 = 03140230410324031042301403240130420 \cdots$$

  that we already constructed on a previous slide.
- We use a theorem-prover based on $\alpha$-Ostrowski representations to show that $\mathbf{x}_5$ has critical exponent $3/2$.

# Ostrowski representation

- Given a real number $0 \leq \alpha < 1$, we can consider its continued fraction $[0, a_1, a_2, \ldots]$.
- The truncation of the continued fraction $[0, a_1, \ldots, a_i]$ is a rational number $p_i/q_i$.
- The *Ostrowski $\alpha$-representation* of $n$ is a certain linear combination of the $q_i$.
- The $n$'th symbol of the Sturmian word with $\beta = 0$ has a simple characterization in terms of the Ostrowski $\alpha$-representation of $n$.

# Pell representation

- The continued fraction of $\alpha = \sqrt{2} - 1$ is $[0, 2, 2, \ldots]$.
- The corresponding $q_i$ are the *Pell numbers* $1, 2, 5, 12, 29, 70 \ldots$ and satisfy the linear recurrence $q_i = 2q_{i-1} + q_{i-2}$.
- The *Pell representation* of $n$ is of the form $\sum_i a_i q_i$, where $a_i \in \{0, 1, 2\}$ and furthermore $a_i = 2$ implies that $a_{i-1} = 0$.
- If an infinite word is generated by an automaton taking Pell representations as input, it is called *Pell-automatic*.

$\mathbf{x}_5$ is Pell-automatic, and here is the automaton:

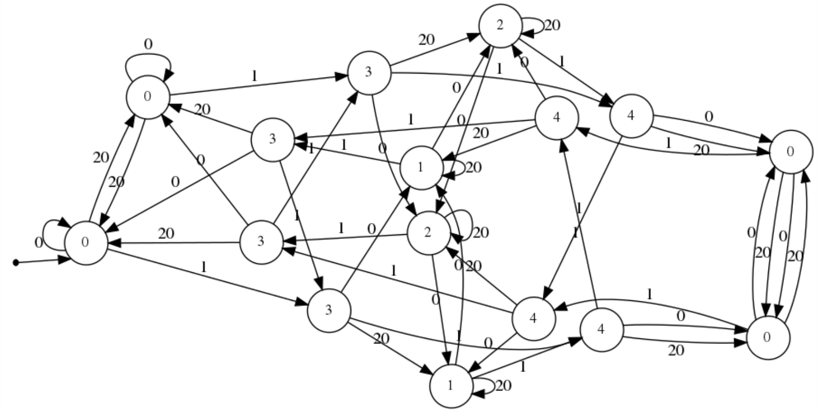

**Figure 1:** Automaton for the infinite word $\mathbf{x}_5$.

# The Walnut theorem prover

- Walnut is a theorem prover (written by Hamoon Mousavi) that can prove or disprove theorems phrased as first-order logic predicates involving automatic sequences.
- My co-author Aseem Baranwal extended Walnut to work with Pell-automatic sequences.
- The only really hard part is building the adder: an automaton that takes $x, y, z$ in parallel (represented in Pell representation) and checks if $x + y = z$.

# Using `Walnut`

- Here is a first-order logic predicate for $\mathbf{x}_5$ having a power of exponent larger than 3/2:

  $$\exists i \; \exists p \; (p \geq 1) \wedge \forall j \; (2j \leq p) \implies \mathbf{x}_5[i+j] = \mathbf{x}_5[i+j+p].$$

- When we enter this into `Walnut` it answers `false`, so $\mathbf{x}_5$ has no powers larger than 3/2.
- But $\mathbf{x}_5$ has the 3/2-power 032403, so this bound is optimal.

# Further work

- It should be possible to prove the Rampersad-JOS-Vandomme conjecture on repetition threshold for infinite balanced words for some other alphabet sizes.
- We currently have no idea how to do it in general.
- We also don't know how to classify all constant-gap words.
- It would be nice to extend `Walnut` to handle *all* Sturmian words. This is doable.
- There is a lot more to study!

# Walnut

For more about `Walnut`, visit

`https://cs.uwaterloo.ca/~shallit/walnut.html` .