Repetitions in Words

Narad Rampersad

Department of Mathematics
University of Liège

Words avoiding squares

- A square is a word of the form xx (like bonbon).
- A word is squarefree if it contains no square as a factor.

Squarefree words using 3 symbols (Thue 1906)

Iterate the substitution $0 \to 012$; $1 \to 02$; $2 \to 1$:

$$0 \to 012 \to 012021 \to 012021012102 \to \cdots$$

These words are squarefree.

Words avoiding cubes

- A cube is a word of the form xxx (like `shshsh`).
- A word is cubefree if it contains no cube as a factor.

Cubefree words using 2 symbols (Thue 1906)

Iterate the substitution $0 \to 01$; $1 \to 10$:

$$0 \to 01 \to 0110 \to 01101001 \to 0110100110010110 \to \cdots$$

These words are cubefree.

Patterns

- Squares (xx) and cubes (xxx) are patterns with one variable.
- Patterns can have several variables.
- 01122011 is an instance of the pattern $xyyx$.
- Given a pattern, is it avoidable over a finite alphabet?

Doubled patterns

- A doubled pattern is one in which every variable occurs at least twice (like *xyzyxz*).
- Any doubled pattern is avoidable (Bean, Ehrenfeucht, McNulty; Zimin 1979).
- Any doubled pattern is avoidable over a 4-letter alphabet (Bell and Goh 2007).

Avoiding long patterns

- Consequence: a pattern with k variables and length at least 2^k is avoidable on a 4-letter alphabet.
- A pattern with k variables and length at least $200 \cdot 5^k$ is avoidable on a 2-letter alphabet (Cassaigne and Roth).
- Using the method of Bell and Goh, we can improve this result.

Current results on avoiding long patterns

Theorem

Let p be a pattern containing k distinct variables.

(a) If p has length at least 2^k then p is 4-avoidable.

(b) If p has length at least 3^k then p is 3-avoidable.

(c) If p has length at least 4^k then p is 2-avoidable.

Open problem: improve the bounds in (b) and (c).

The technique

- We use a special case of a theorem of Golod and Shafarevich (1964).
- Let S be a set of words over an d-letter alphabet, each of length at least 2.
- Suppose S has at most r_i words of length i for $i \geq 2$.

A power series criterion

Theorem

If the power series expansion of

$$G(z) := \left(1 - dz + \sum_{i \geq 2} r_i z^i\right)^{-1}$$

has non-negative coefficients, then there are least $[z^n]G(z)$ words of length n over a d-letter alphabet that contain no word of S as a factor.

Fractional repetitions

- We denote squares by $xx = x^2$ and cubes by $xxx = x^3$.
- What would $x^{7/4}$ or $x^{8/3}$ mean?
- $0111011 = x^{7/4}$ for $x = 0111$
- $00100100 = x^{8/3}$ for $x = 001$
- If $w = x^k$ for some rational k, then w is a k-power.

Avoiding fractional repetitions

- What fractional powers can be avoided on a given alphabet?
- Dejean (1972) showed that if $k > 7/4$, then k-powers are avoidable over a 3-letter alphabet.
- repetition threshold:

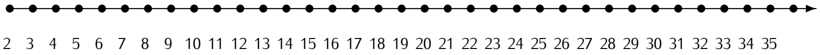
$$RT(n) = \inf \{k \in \mathbb{Q} : \text{there is an infinite word over an } n\text{-letter alphabet that avoids } k\text{-powers}\}$$

Dejean's Conjecture

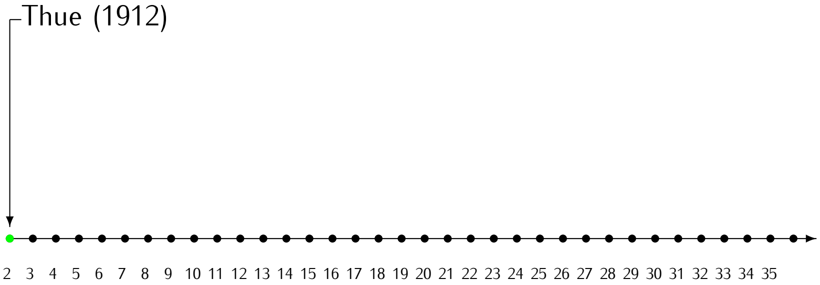
Dejean's Conjecture (1972)

$$RT(n) = \begin{cases} 2, & n = 2 \\ 7/4, & n = 3 \\ 7/5, & n = 4 \\ n/(n-1), & n \geq 5. \end{cases}$$

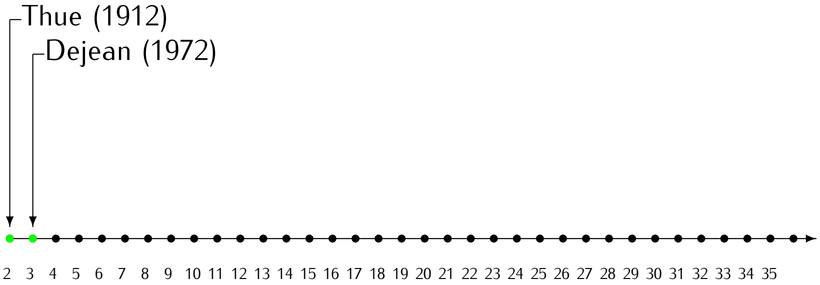
History of the conjecture



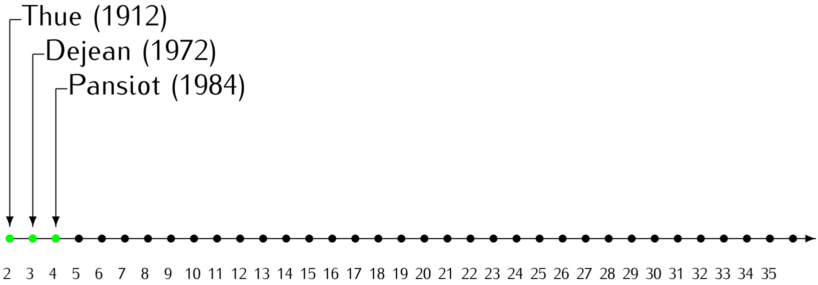
History of the conjecture



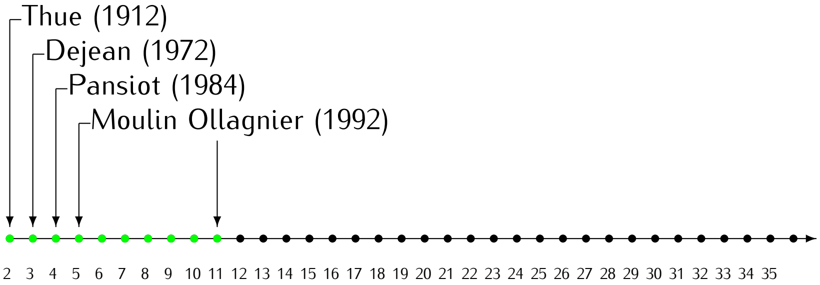
History of the conjecture



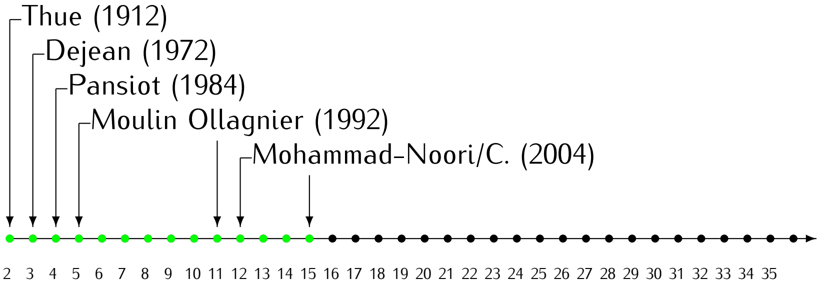
History of the conjecture



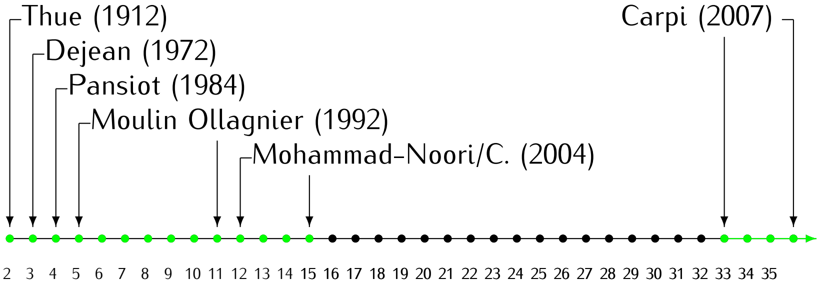
History of the conjecture



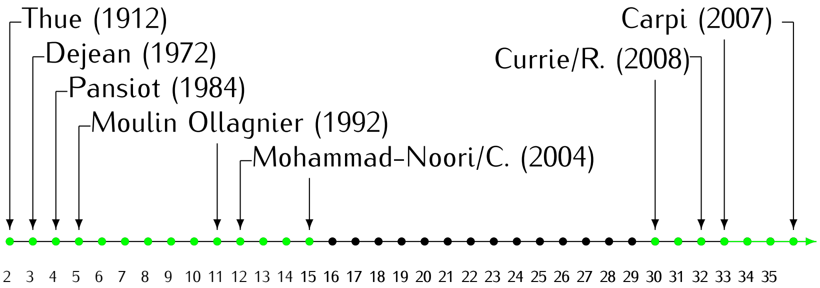
History of the conjecture



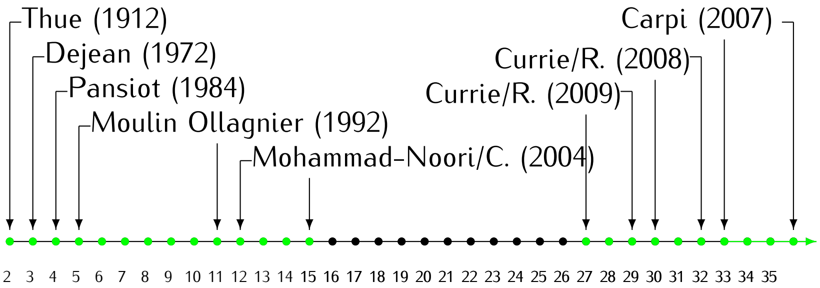
History of the conjecture



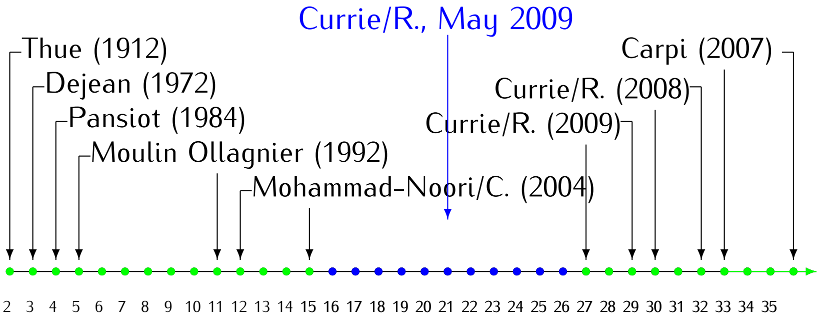
History of the conjecture



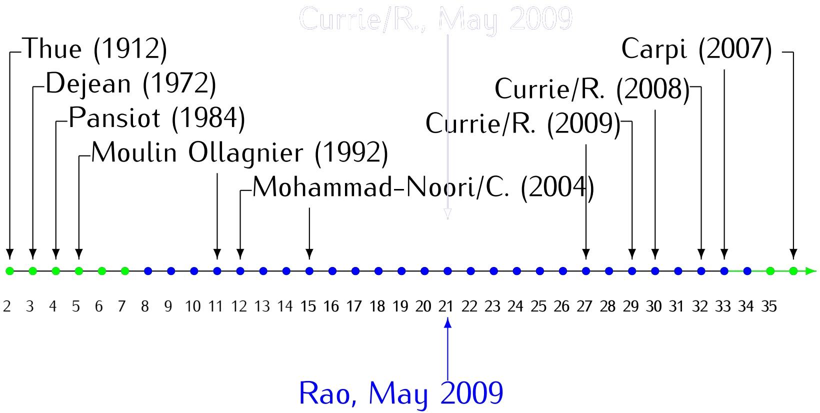
History of the conjecture



History of the conjecture



History of the conjecture



Approximate repetitions

- Instead of avoiding exact repetitions xx, we avoid "approximate" repetitions xx' where x and x' are almost equal.
- E. g. can we avoid xx' where x and x' have the same length and agree in more than 3/4 of their positions?
- This is stronger than avoiding 7/4-powers.

Avoiding approximate repetitions

Theorem (Ochem, R., Shallit 2008)

There is an infinite word $\mathbf{w}$ over $\{0, 1, 2\}$ that avoids all xx' where x and x' have the same length and agree in more than 3/4 of their positions.

To obtain $\mathbf{w}$, iterate the map

$$\begin{aligned} 0 &\rightarrow 0120212010210121021202 10 \\ 1 &\rightarrow 1201020121021202102010 21 \\ 2 &\rightarrow 2012101202102010210121 02. \end{aligned}$$

van der Waerden's Theorem

van der Waerden's Theorem

If the natural numbers are partitioned into finitely many sets, then one set contains arbitrarily large arithmetic progressions.

vdW rephrased

For any infinite word **w** over a finite alphabet A, there exists $a \in A$ such that for all $m \geq 1$, **w** contains a^m in a subsequence indexed by an arithmetic progression.

Repetitions in arithmetic progressions

Theorem (Carpi 1988)

For every integer $n \geq 2$, there exists an infinite word over a finite alphabet that contains no squares in any arithmetic progression except those whose difference is a multiple of n.

Folding a piece of paper

- Take an 8.5×11 piece of paper and fold it in half.
- Unfold and record the pattern of hills and valleys (0 for a hill and 1 for a valley).

$$0$$

- Fold twice, unfold, and record the pattern of hills and valleys.

$$0 \quad 0 \quad 1$$

- Fold three times, unfold, and record the pattern.

$$0 \quad 0 \quad 1 \quad 0 \quad 0 \quad 1 \quad 1$$

The paperfolding sequence

- Fold infinitely (!) many times, and unfold. The pattern obtained is the paperfolding sequence.

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

Theorem (Allouche and Bousquet-Mélou 1994)

For any paperfolding word $\mathbf{f}$, if ww is a non-empty subword of $\mathbf{f}$, then $|w| \in \{1, 3, 5\}$.

A modified paperfolding sequence

Take

$$\mathbf{f} = 0010011000110110\cdots$$

and replace the 0's and 1's in the even indexed positions by 2's and 3's respectively to obtain

$$\mathbf{v} = 2030213020312130\cdots.$$

Arithmetic progressions of odd difference

Theorem (Kao, R., Shallit, and Silva 2008)

Let $\mathbf{v}$ be obtained from a paperfolding word $\mathbf{f}$ as described above. Then $\mathbf{v}$ contains no squares in any arithmetic progression of odd difference.

Words in higher dimensions

- A 2-dimensional word is a 2D array of symbols.
- Formally: a map $\mathbf{w}$ from $\mathbb{N}^2$ to A.
- We write $w_{m,n}$ for $\mathbf{w}(m, n)$.
- A word $\mathbf{x}$ is a line of $\mathbf{w}$ if there exists i_1, i_2, j_1, j_2, such that $\gcd(j_1, j_2) = 1$ and for $t \geq 0$, we have $x_t = w_{i_1+j_1 t, i_2+j_2 t}$.

Avoiding repetitions in higher dimensions

Theorem (Carpi 1988)

There exists a 2-dimensional word **w** over a 16-letter alphabet such that every line of **w** is squarefree.

- Let $\mathbf{u} = u_0 u_1 u_2 \cdots$ and $\mathbf{v} = v_0 v_1 v_2 \cdots$ be infinite words over the alphabet $A = \{0, 1, 2, 3\}$ that avoid squares in all arithmetic progressions of odd difference.
- Define **w** over the alphabet $A \times A$ by $w_{m,n} = (u_m, v_n)$.

Abelian repetitions

Erdős 1961	abelian square: a word xx' such that x' is a permutation of x (like `reappear`)
Evdokimov 1968	abelian squares avoidable over 25 letters
Pleasants 1970	abelian squares avoidable over 5 letters
Justin 1972	abelian 5-powers avoidable over 2 letters
Dekking 1979	abelian 4-powers avoidable over 2 letters
	abelian cubes avoidable over 3 letters
Keränen 1992	abelian squares avoidable over 4 letters

Avoiding patterns in the Abelian sense

- Avoiding the pattern $xyyx$ in the Abelian sense means avoiding all words $xyy'x'$ where x and x' (resp. y and y') are permutations of each other.
- Open problem: characterize the patterns that are avoidable in the Abelian sense.

Theorem (Currie and Visentin 2008)

Any pattern over $\{x, y\}$ of length greater than 118 is avoidable in the Abelian sense over a 2-letter alphabet.

Summary

- Variations on Thue's problem:
  - patterns
  - fractional powers
  - approximate repetitions
  - repetitions in arithmetic progressions
  - repetitions in multi-dimensional words
  - Abelian squares and patterns
  - *non-repetitive colourings of the real line, graphs*
  - *repetitions in Sturmian words*
- Still many open problems

The End