Repetitions in Words: Classical and Recent Results

Narad Rampersad

Department of Mathematics
University of Liège

The most basic repetition

- A square is a non-empty word of the form xx (like $\texttt{bonbon}$).
- A word is squarefree if it contains no square as a factor.
- Any long word over 2 symbols contains squares.
- What if we use 3 symbols?

Infinitely many words avoiding squares

Theorem (Thue 1906)

There are infinitely many squarefree words over 3 symbols.

Generating squarefree words

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

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

- These words are squarefree.

The Thue–Morse word

- Iterate the morphism $0 \to 01$; $1 \to 10$:

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

- The limit word is the Thue–Morse word $\mathbf{t}$.

Words avoiding cubes

- Thue 1906: $\mathbf{t}$ contains no cube.
- A cube is a non-empty word of the form xxx (like `shshsh`).
- A word is cubefree if it contains no cube as a factor.

The Fibonacci word

- Iterate the morphism $0 \to 01;\ 1 \to 0$:

$$0 \to 01 \to 010 \to 01001 \to 01001010 \to \cdots$$

- The limit word is the Fibonacci word $\mathbf{f}$.

Repetitions in the Fibonacci word

- Kolotov 1981: $\mathbf{f}$ contains no 5-power ($xxxxx$).
  - used to construct an algebra of minimal growth with certain nilpotency properties
- Karhumäki 1983: $\mathbf{f}$ contains no 4-power ($xxxx$).
- Mignosi and Pirillo 1992: $\mathbf{f}$ contains no repetition of exponent greater than $2 + \varphi = 3.61803399 \cdots$.

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?
- avoidable: there is an infinite word that avoids the pattern.

Doubled patterns

- A doubled pattern: 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).

Current results on avoiding long patterns

Theorem (Bell and Goh 2007; R. 2009)

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.

- k-avoidable: there is an infinite word over a k-letter alphabet that avoids the pattern.

The technique

- A combinatorial lemma of Golod and Shafarevich (1964).
- Originally used to construct counterexamples to the General Burnside Problem and Kurosh's Problem (ring-theoretic analogue).

General Burnside Problem

If G is a finitely generated group and every element of G has finite order, then must G be finite?

Optimality of the patterns result

- The Zimin patterns:

$$Z_1 = x, \quad Z_2 = xyx, \quad Z_3 = xyxzxyx, \quad \ldots$$

- Z_k contains k distinct variables, has length $2^k - 1$, and is unavoidable.

Avoiding binary patterns over a binary alphabet

Theorem (Roth 1992)

Any pattern over $\{x, y\}$ of length at least 6 is 2-avoidable.

Fractional repetitions

- We denote squares by $xx = x^2$ and cubes by $xxx = x^3$.
- What would $x^{7/4}$ or $x^{8/5}$ mean?
- $\texttt{ingoing} = x^{7/4}$ for $x = \texttt{ingo}$
- $\texttt{outshout} = x^{8/5}$ for $x = \texttt{outsh}$
- 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?
- If $k > 7/4$, then k-powers are avoidable over a 3-letter alphabet (Dejean 1972).
- repetition threshold:

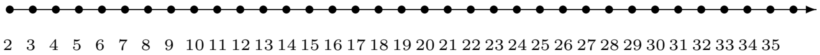
$$\mathsf{RT}(n) = \inf \{k \in \mathbb{Q} : \text{there is an infinite word over an } n\text{-letter alphabet that avoids } k\text{-powers}\}$$

Dejean's Conjecture

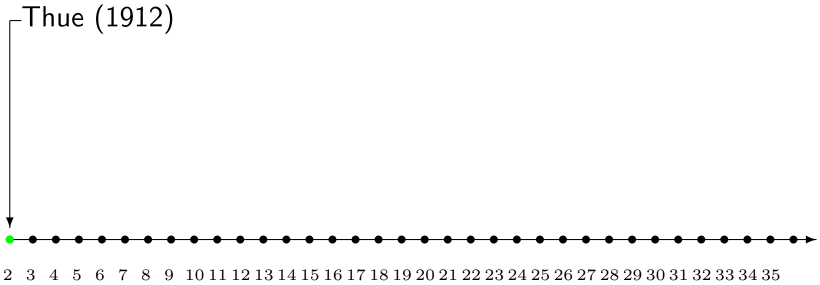
Dejean's Conjecture (1972)

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

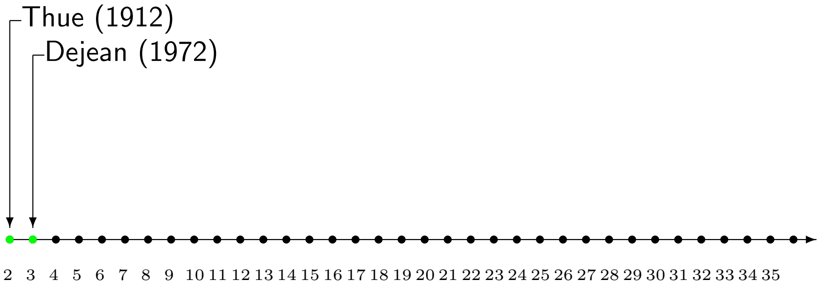
History of the conjecture



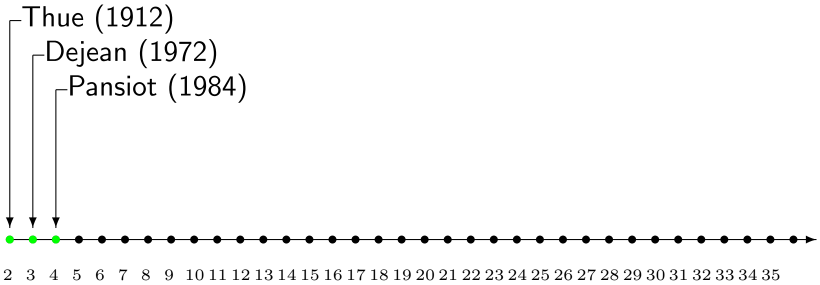
History of the conjecture



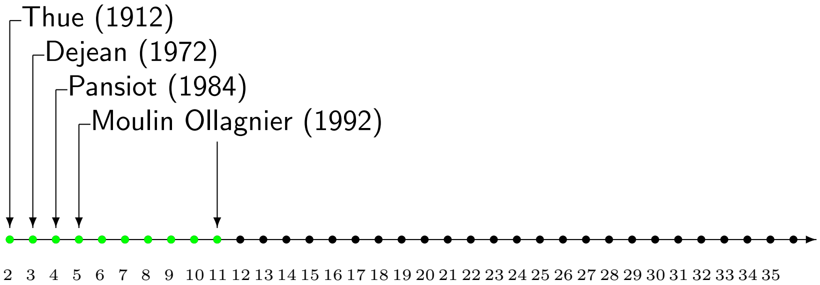
History of the conjecture



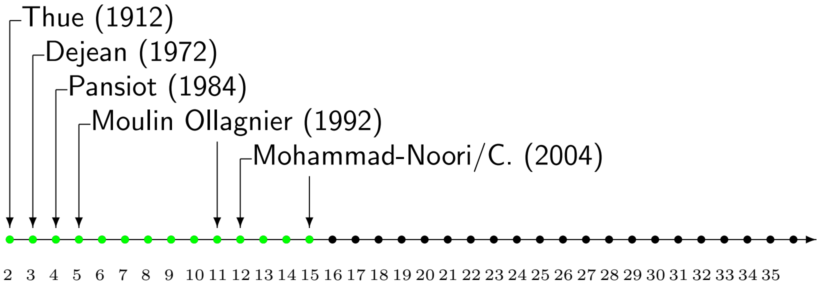
History of the conjecture



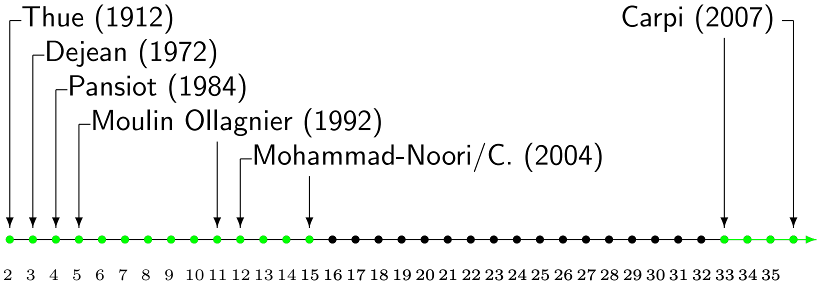
History of the conjecture



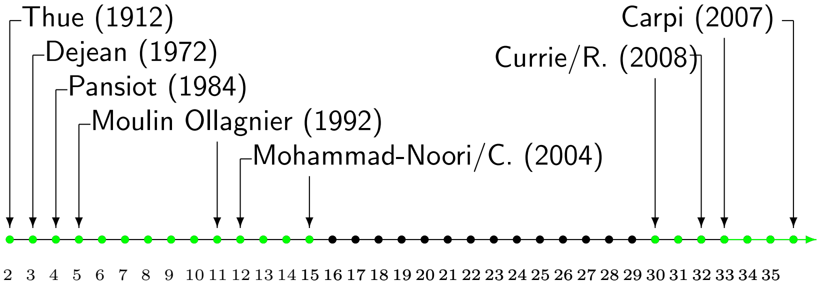
History of the conjecture



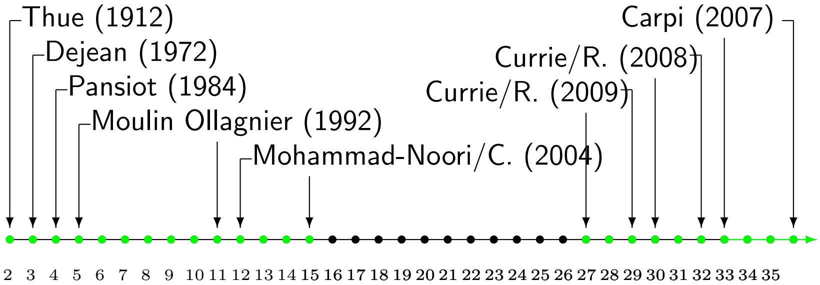
History of the conjecture



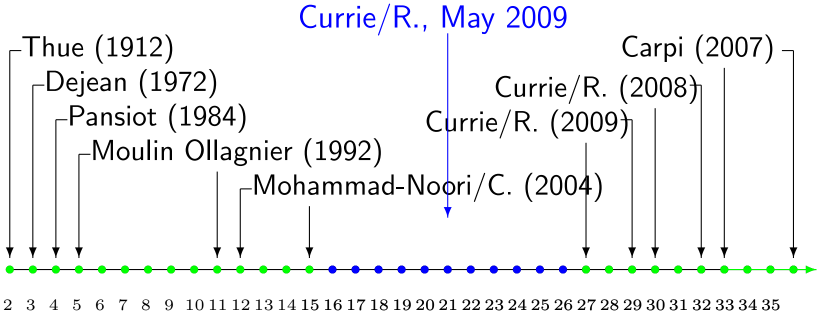
History of the conjecture



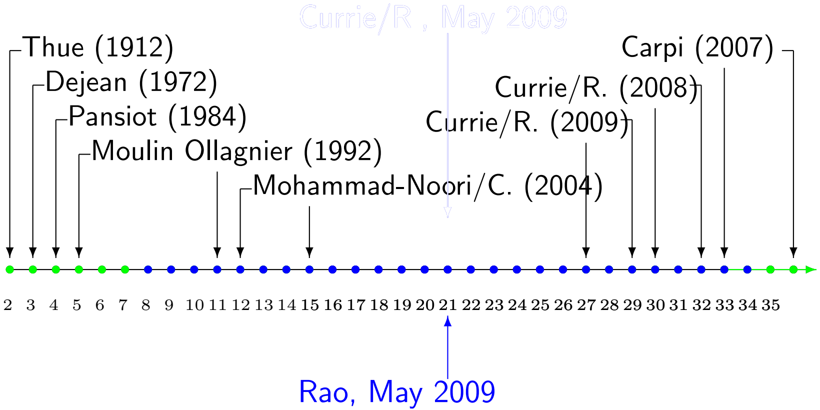
History of the conjecture



History of the conjecture



History of the conjecture



A highly non-repetitive word

Theorem (Beck 1981)

For any $\epsilon > 0$, there is some N_ϵ and an infinite binary word such that any two identical factors of length $n > N_\epsilon$ are at distance $> (2-\epsilon)^n$.

- Proof is non-constructive—uses the probabilistic method (Lovász Local Lemma).
- No constructive proof known (but see Carpi and D'Alonzo 2009).

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?
- Stronger than avoiding $7/4^+$-powers.

Avoiding approximate repetitions over 3 letters

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.

The construction

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

$$
\begin{aligned}
0 &\rightarrow 012021201021012102120210 \\
1 &\rightarrow 120102012102120210201021 \\
2 &\rightarrow 201210120210201021012102.
\end{aligned}
$$

Avoiding approximate repetitions over 4 letters

Theorem (Ochem, R., Shallit 2008)

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

Unavoidable regularity

van der Waerden's Theorem

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

Unavoidable repetitions

vdW rephrased

For any infinite word $\mathbf{w}$ over a finite alphabet A, there exists $a \in A$ such that for all $m \geq 1$, $\mathbf{w}$ 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.

The Toeplitz construction

- Start with an infinite sequence of gaps, denoted ?.

$$? \quad ? \quad ? \quad ? \quad ? \quad ? \quad ? \quad ? \quad ? \quad ? \quad ? \quad ? \quad ? \quad ? \quad ? \quad \cdots$$

- Fill every other gap with alternating 0's and 1's.

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

- Repeat.

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

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

Paperfolding words

- In the limit one obtains the ordinary paperfolding word:

$$0010011000110110\cdots$$

- At each step, one may choose to fill in the gaps by either

$$0101010101\cdots$$

or

$$1010101010\cdots.$$

- Different choices result in different paperfolding words.

Repetitions in paperfolding words

Theorem (Allouche and Bousquet-Mélou 1994)

If $\mathbf{f}$ is a paperfolding word and ww is a non-empty factor of $\mathbf{f}$, then $|w| \in \{1, 3, 5\}$.

Modifying the paperfolding word

- Take

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

- 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 $\mathbf{w}$ is a 2D array of symbols.
- $w_{m,n}$: the symbol of $\mathbf{w}$ at position (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 $x_t = w_{i_1+j_1t, i_2+j_2t}$.

Avoiding repetitions in higher dimensions

Theorem (Carpi 1988)

There exists a 2-dimensional word $\mathbf{w}$ over a 16-letter alphabet such that every line of $\mathbf{w}$ is squarefree.

Constructing the 2D word

- Let $\mathbf{u} = u_0u_1u_2\cdots$ and $\mathbf{v} = v_0v_1v_2\cdots$ be infinite words over $A = \{0, 1, 2, 3\}$ that avoid squares in all arithmetic progressions of odd difference.
- Define $\mathbf{w}$ over the alphabet $A \times A$ by $w_{m,n} = (u_m, v_n)$.

Other 2D results

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

- There exists a 2-dimensional word $\mathbf{w}$ over a 4-letter alphabet, such that every line of $\mathbf{w}$ is 3^+-power-free.
- There exists a 2-dimensional word $\mathbf{w}$ over a 9-letter alphabet, such that every line of $\mathbf{w}$ is 2^+-power-free.

Non-repetitive graph colourings

- A word can be viewed as a (vertex) colouring of a path.
- What about graphs in general?
- A non-repetitive colouring of a graph G is a vertex colouring such that the sequence of colours encountered along any path yields a squarefree word (Alon, Grytczuk, Hałuszczak, and Riordan 2002).

Paths

Theorem (Thue 1906)

For any n, the path of length n has a non-repetitive 3-colouring.

Cycles

Theorem (Currie 2002)

For any n, except $n \in \{5, 7, 9, 10, 14, 17\}$, the cycle of length n has a non-repetitive 3-colouring. For the exceptional values of n, four colours are required.

Trees

Theorem

Any tree has a non-repetitive 4-colouring.

Graphs with bounded degree

Theorem (Alon et al. 2002)

There exists a constant C such that every graph with maximum degree Δ has a non-repetitive colouring using at most $C\Delta^2$ colours.

- Uses the probabilistic method.

Planar graphs

- Question (Alon et al. 2002): Does there exist a constant C such that every planar graph has a non-repetitive C-colouring?

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.
- Problem: characterize the patterns that are avoidable in the Abelian sense.

Binary patterns 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.

Two equivalent open problems

- Question (Cassaigne, Richomme, Saari, Zamboni 2010): Does there exist an infinite binary word that avoids Abelian squares at a sequence of positions with bounded gaps?
- Question (Pirillo and Varricchio 1994; Halbeisen and Hungerbühler 2000): Does there exist an infinite sequence over a finite set of integers such that no two consecutive blocks of the same length have the same sum?
- Cassaigne et al. showed that the problems are equivalent.

Summary

- Variations on Thue's problem:
  - patterns
  - fractional powers
  - approximate repetitions
  - repetitions in arithmetic progressions
  - repetitions in multi-dimensional words
  - repetitions in cycles, trees, graphs
  - Abelian squares and patterns

The End