Avoiding repetitions in words I

Narad Rampersad

Department of Mathematics and Statistics
University of Winnipeg

Repetitions in words

- What kinds of repetitions can/cannot be avoided in words (sequences)?
- e.g., the word

$$ab\color{magenta}{aa}\color{black}{bbaba}\color{magenta}{abab}$$

  contains several repetitions
- but in the word

$$abcbacbcabcba$$

  the same sequence of symbols never repeats twice in succession

Types of repetitions

- a square is a non-empty word of the form xx (like $\texttt{couscous}$)
- a word is squarefree if it contains no square
- a cube is a non-empty word xxx
- a t-power is a non-empty word x^t (x repeated t times)
- any long word over 2 symbols contains squares
- Over 3 symbols?

Thue's work

Theorem (Thue 1906)

There is an infinite squarefree word over 3 symbols.

Subsequent work

- Thue's result was rediscovered many times
- e.g., by Arshon (1937); Morse and Hedlund (1940)
- a systematic study of avoidable repetitions was begun by Bean, Ehrenfeucht, and McNulty (1979)

Morphisms

- typical construction of squarefree words: find a map that produces a longer squarefree word from a shorter squarefree word
- e.g., the map (morphism) f that sends $a \to abcab$; $b \to acabcb$; $c \to acbcacb$
- $f(acb) = abcab\,acbcacb\,acabcb$ is squarefree
- if this morphism preserves squarefreeness we can generate an infinite word by iteration

Formal definition of morphism

- let Σ, Δ be finite alphabets
- a morphism is a map $h : \Sigma^* \to \Delta^*$, satisfying $h(xy) = h(x)h(y)$ for all $x, y \in \Sigma^*$
- we usually specify a morphism by giving the values $h(a)$ for all $a \in \Sigma$.
- e.g., $h : \{a, b, c\}^* \to \{a, b, c\}^*$ maps

$$\begin{aligned} a &\to abc \\ b &\to ac \\ c &\to b \end{aligned}$$

Generating sequences with morphisms

- a morphism $h : \Sigma^* \to \Sigma^*$ such that $h(a) = ax$ for some $a \in \Sigma$ and $x \in \Sigma^*$ with $h^i(x) \neq \epsilon$ for all i is prolongable on a
- if we iterate h we get

$$\begin{aligned} h(a) &= ax \\ h^2(a) &= a\,x\,h(x) \\ &\vdots \\ h^\omega(a) &= a\,x\,h(x)\,h^2(x)\,h^3(x)\cdots . \end{aligned}$$

- $h^\omega(a)$ is a purely morphic word

Iterating a morphism

The morphism $a \to abc, b \to ac, c \to b$ given earlier is prolongable on a, so iteration gives

$$a \to abc \to abcacb \to \cdots \to abcacbabcbacabcacb \cdots$$

Preserving squarefreeness

- What conditions on a morphism guarantee that it preserves squarefreeness?
- we say a morphism is infix if no image of a letter appears inside the image of another letter
- $a \to abc$; $b \to ac$; $c \to b$ is not infix

A sufficient condition for infix morphisms

Theorem (Thue 1912; Bean et. al. 1979)

Let $f : A^* \to B^*$ be a morphism. If f is infix and $f(w)$ is squarefree whenever w is a squarefree word of length at most 3, then f preserves squarefreeness in general.

Proof of Thue's criterion

Claim: Let $w = w_0 w_1 \cdots w_n$ be a squarefree word, where each $w_i \in A$. If $f(w) = xf(a)y$ for some $x, y \in A^*$ and $a \in A$, then there exists j such that $x = f(w_1 \cdots w_{j-1})$, $a = w_j$, and $y = f(w_{j+1} \cdots w_n)$.

Proof of Thue's criterion

- suppose the contrary: $f(a)$ is a factor of $f(w_j w_{j+1})$ for some j and $f(a)$ crosses the boundary between $f(w_j)$ and $f(w_{j+1})$
- write $f(w_j) = pq$, $f(w_{j+1}) = rs$, and $f(a) = qr$
- $f(aw_ja) = qrpqqr$ contains a square, so aw_ja must contain a square
- i.e., $a = w_j$ and similarly $a = w_{j+1}$
- so $w_j = w_{j+1}$, contradicting the squarefreeness of w

Proof of Thue's criterion

- Let $w = w_0 w_1 \cdots w_n$ be a shortest squarefree word such that $f(w)$ contains a square.
- write $f(w) = xyyz$ and $f(w_i) = W_i$
- i.e., $f(w) = W_0 W_1 \cdots W_n = xyyz$
- we may suppose that x is a proper prefix of W_0 and z is a proper suffix of W_n
- write $W_0 = xW_0''$ and $W_n = W_n'z$
- then $yy = W_0'' W_1 \cdots W_{n-1} W_n'$

Proof of Thue's criterion

- then $y = W_0''W_1 \cdots W_j' = W_j''W_{j+1} \cdots W_n'$, where $W_j = W_j'W_j''$
- we must have $n \geq 3$
- if $j = 0$ then W_1W_2 is a factor of W_0, contradicting the infix property of f
- similarly, we cannot have $j = n$

Proof of Thue's criterion

- by earlier claim,

$$W_0'' = W_j'', W_1 = W_{j+1}, \ldots, W_j' = W_n'.$$

so

$$f(w_0 w_j w_n) = xW_0''W_j'W_j''W_n'z = xW_0''W_j'W_0''W_j'z$$

contains the square $W_0''W_j'W_0''W_j'$

- then $w_0 w_j w_n$ contains a square

Proof of Thue's criterion

- without loss of generality, suppose that $w_0 = w_j$
- from $W_1 = W_{j+1}, \ldots, W_{j-1} = W_{2j-1}$ we have $w_1 = w_{j+1}, \ldots, w_{j-1} = w_{2j-1}$
- then $w = w_0 w_1 \cdots w_{j-1} w_0 w_1 \cdots w_{j-1} w_n$ contains a square, a contradiction
- this completes the proof

Generating squarefree words

- the map $a \to abcab$; $b \to acabcb$; $c \to acbcacb$ satisfies the conditions of the theorem
- so it preserves squarefreeness
- if we iterate it we get squarefree words:

$$a \to abcab \to abcabacabcbacbcacbabcabacabcb$$

- so there is an infinite squarefree word over $\{a, b, c\}$

A general criterion

Theorem (Crochemore 1982)

Let $f : A^* \to B^*$ be a morphism. Then f preserves squarefreeness if and only if it preserves squarefreeness on words of length at most

$$\max\left\{3, 1 + \left\lceil \frac{M(f) - 3}{m(f)} \right\rceil\right\},$$

where $M(f) = \max_{a \in A} |f(a)|$ and $m(f) = \min_{a \in A} |f(a)|$.

Consequences

- we have an algorithm to decide if a morphism is squarefree
- simply test if it is squarefree on words of a certain length (the bound in the theorem)
- What about t-powers?
- Recall: a square looks like xx; a t-power looks like $xx \cdots xx$ (t-times)

A criterion for t-power-freeness

Theorem (Richomme and Wlazinski 2007)

Let $t \geq 3$ and let $f : A^* \to B^*$ be a uniform morphism. There exists a finite set $T \subseteq A^*$ such that f preserves t-power-freeness if and only if $f(T)$ consists of t-power-free words.

(uniform means the lengths of the images, $|f(a)|$, are the same for all $a \in A$)

The general case

Open problem

Is there an algorithm to determine if an arbitrary morphism is t-power-free?

Changing the problem slightly

- our initial goal was to generate long t-power-free words
- a morphism that preserves t-power-freeness can accomplish this
- but some morphisms can generate long t-power-free words without preserving t-power-freeness in general

An non-squarefree morphism

- consider f defined by

$$a \to abc \qquad b \to ac \qquad c \to b$$

- iterates are squarefree:

$$a \to abc \to abcacb \to abcacbabcbac \to \cdots$$

- but $f(aba) = ab\textcolor{magenta}{caca}bc$ is not

Fixed points

- suppose f generates an infinite word $\mathbf{x}$ by iteration
- we write $\mathbf{x} = f(\mathbf{x})$ and call $\mathbf{x}$ a fixed point of f
- Can we determine if $\mathbf{x}$ is t-power-free?

Deciding if a fixed point is t-power-free

Theorem (Mignosi and Séébold 1993)

There is an algorithm to decide the following problem:

> *Given* $t \geq 2$ *and a morphism* f *with fixed point* $\mathbf{x}$*, is* $\mathbf{x}$ t*-power-free?*

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 $w = x^r$ for some rational r, then w is a r-power.
- An r^+-power is a word x^s where $s > r$.
- What fractional powers can be avoided on a given alphabet?

Avoiding cubes on the binary alphabet

- x^2 cannot be avoided on a binary alphabet
- What about x^3?
- iterate the morphism $0 \to 001$, $1 \to 011$:

$$0 \to 001 \to 001011011 \to \cdots \to 001001011001011011\cdots$$

- this word avoids x^3 (exercise!)

Fractional repetitions on the binary alphabet

Theorem (Thue 1912)

There is an infinite word over 2 symbols that contains no 2^+-power (i.e., no s-power where $s > 2$).

The Thue–Morse word

- the word Thue constructed is obtained by iterating $0 \to 01$, $1 \to 10$:

$$0110100110010110100101100110100 1\cdots$$

- this is the Thue–Morse word

Fractional repetitions on the ternary alphabet

- x^2 is avoidable on a 3-letter alphabet
- Can repetitions with smaller exponent be avoided?
- There is an infinite word over 3 symbols that contains no $(7/4)^+$-power (i.e., no s-power where $s > 7/4$) (Dejean 1972).

Repetition threshold

- 2^+-powers are avoidable on 2 letters
- $(7/4)^+$-powers are avoidable on 3 letters
- What about larger alphabets?
- repetition threshold:

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

Dejean's Conjecture (1972)

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

The ternary alphabet

- Dejean proved that $RT(3) = 7/4$ using the morphism

$$\begin{aligned} h(0) &= 0120212012102120210 \\ h(1) &= 1201020120210201021 \\ h(2) &= 2012101201021012102 \end{aligned}$$

- h maps $(7/4)^+$-power-free words to $(7/4)^+$-power-free words
- by iterating h on 0, we obtain an infinite word with the desired property

Morphic constructions for larger alphabets

- Can a similar construction exist for larger alphabets?
- Brandenburg (1983): No.
- For each integer $k \geq 2$, define

$$\alpha_k = \begin{cases} 7/4, & \text{if } k = 3; \\ 7/5, & \text{if } k = 4; \\ \frac{k}{k-1}, & \text{if } k \neq 3, 4. \end{cases}$$

- Dejean's Conjecture is that $RT(k) = \alpha_k$.

No α_k^+-power-free morphisms

Theorem

Let Σ_k be an alphabet of size $k \geq 4$. There exists no growing α_k^+-power-free morphism from Σ_k to Σ_k.

growing morphism refers to a morphism h such that $h(a) \neq \epsilon$ for all $a \in \Sigma$ and $|h(a)| > 1$ for at least one letter $a \in \Sigma$

Implications of Brandenburg's result

- We cannot hope to prove Dejean's Conjecture by producing α_k^+-free morphisms.
- It could be the case that there exist morphisms h that are not α_k^+-free but still generate an infinite α_k^+-free word by iteration.
- Still, this is strong evidence that a new idea is needed in order to attack Dejean's Conjecture for larger alphabets.
- new idea provided by Pansiot

Pansiot's approach

- Alphabet size k
- A word of length at least $k+2$ must contain a factor with exponent at least $k/(k-1)$.
- If a word avoids $(k/(k-1))^+$-powers, every block of length $k-1$ consists of $k-1$ different letters.

The Pansiot encoding

- The letter following a block y of length $k-1$ is either
  - the first letter of y; or
  - the unique letter that does not occur in y.
- Pansiot encoding: encode first case with a 0; second case with a 1.
- Can uniquely reconstruct the original word from the Pansiot encoding.

The Pansiot encoding

Example (k=6)

Word:

$$123451632415$$

Pansiot encoding:

$$0101101.$$

We reconstruct the original word from the prefix 12345 and the code 0101101.

The Pansiot encoding

Example (k=6)

Word:

$$123451632415$$

Pansiot encoding:

$$0101101.$$

We reconstruct the original word from the prefix 12345 and the code 0101101.

The Pansiot encoding

Example (k=6)

Word:

$$123451632415$$

Pansiot encoding:

$$0101101.$$

We reconstruct the original word from the prefix 12345 and the code 0101101.

The Pansiot encoding

Example (k=6)

Word:

$$12345163 2415$$

Pansiot encoding:

$$0101101.$$

We reconstruct the original word from the prefix 12345 and the code 0101101.

The Pansiot encoding

Example (k=6)

Word:

$$123451632415$$

Pansiot encoding:

$$0101101.$$

We reconstruct the original word from the prefix 12345 and the code 0101101.

The Pansiot encoding

Example (k=6)

Word:

$$1234516324 15$$

Pansiot encoding:

$$0101101.$$

We reconstruct the original word from the prefix 12345 and the code 0101101.

The Pansiot encoding

Example (k=6)

Word:

$$123451632415$$

Pansiot encoding:

$$0101101.$$

We reconstruct the original word from the prefix 12345 and the code 0101101.

The Pansiot encoding

Example (k=6)

Word:

$$123451632415$$

Pansiot encoding:

$$0101101.$$

We reconstruct the original word from the prefix 12345 and the code 0101101.

The Pansiot encoding

Example (k=6)

Word:

$$123451632415$$

Pansiot encoding:

$$0101101.$$

We reconstruct the original word from the prefix 12345 and the code 0101101.

Constructing the Pansiot encoding

- Proving Dejean's conjecture for $k = 4$: need an infinite $(7/5)^+$-power-free word $\mathbf{w}$
- Instead, find the binary Pansiot encoding of $\mathbf{w}$
- Binary encoding: iterate $0 \to 101101$; $1 \to 10$:

$$1 \to 10 \to 10101101 \to 10101101101011011010110110 \to \cdots$$

- Decode:

$$\mathbf{w} = 12342143241342314321\cdots$$

The final resolution of the conjecture

- Combined work of: Dejean (1972), Pansiot (1984), Moulin Ollagnier (1992), Currie and Mohammad-Noori (2007), Carpi (2007), Currie and Rampersad (2009), Rao (2009)
- Major breakthrough: Carpi's proof of the conjecture for $k \geq 33$

Words with irrational critical exponent

- the critical exponent of a word $\mathbf{w}$ is the quantity

$$\sup\{r \in \mathbb{Q} : \mathbf{w} \text{ contains an } r\text{-power}\}$$

- the words constructed to verify Dejean's conjecture have rational critical exponent
- 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 $\mathbf{f}$
- $\mathbf{f}$ has critical exponent $2 + \varphi = 3.61803399\cdots$

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).

Long patterns

Theorem (Ochem and Pinlou 2012; Blanchet-Sadri and Woodhouse 2012)

Let p be a pattern containing k distinct variables.

(a) If p has length at least 2^k then p is 3-avoidable.

(b) If p has length at least $3 \cdot 2^{k-1}$ 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.

Showing the unavoidability of the Zimin patterns

- by induction on k
- Let Σ be an alphabet of size s.
- Clearly Z_1 is unavoidable on Σ.
- Suppose Z_k is unavoidable on Σ.
- Then there exists N such that every word of length N contains an instance of Z_k.
- There are s^N such words.

Showing the unavoidability of the Zimin patterns

- Let $w \in \Sigma^*$ be a word of length $M = s^N(N+1) + N$:

$$w = x_0 a_0 x_1 a_1 \cdots x_{s^N-1} a_{s^N-1} x_{s^N},$$

  where for $0 \leq i \leq s^N$, $|x_i| = N$ and $|a_i| = 1$.
- There exists $i < j$ such that $x_i = x_j$.
- By the induction hypothesis x_i contains an instance of Z_k.

Showing the unavoidability of the Zimin patterns

- Write $x_i = x'yx''$, where y is an instance of Z_k, so

$$yx''a_ix_{i+1}a_{i+1}\cdots x_{j-1}a_{j-1}x'y$$

  begins and ends with an instance of Z_k.
- It is therefore an instance of Z_{k+1}.
- So any word of length M over Σ contains an instance of Z_{k+1}.
- Σ was arbitrary, so Z_{k+1} is unavoidable.

Existence of long unavoidable patterns

- so there exist patterns of length $2^k - 1$ with k variables that are unavoidable
- a similar argument shows that there exist patterns of length $3 \cdot 2^{k-1} - 1$ with k variables that are unavoidable over the binary alphabet

Summary

We have seen:

- integer powers
- fractional powers
- patterns

There are many more types of repetitions whose avoidability/unavoidability has been studied.

The End