Repetitions in Words—Part II

Narad Rampersad

Department of Mathematics and Statistics
University of Winnipeg

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 $s > r$.

Avoiding fractional repetitions

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

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

A more precise question

- Let $\alpha > 1$ be a real number.
- Suppose that for every $\epsilon > 0$ there exists an infinite word avoiding $(\alpha + \epsilon)$-powers.
- Is there a *single* word that avoids α^+-powers?
- We use a compactness argument from topology.

Some topology

A topological space T consists of a set X together with a collection $\mathcal{S}$ of subsets of X (the open sets) such that

(a) $\emptyset$ and X are both in $\mathcal{S}$;

(b) The union of any collection of sets in $\mathcal{S}$ is again in $\mathcal{S}$;

(c) The intersection of any finite collection of sets in $\mathcal{S}$ is again in $\mathcal{S}$.

The complements of the open sets are the closed sets.

Compactness

- An open cover of a set Y is a collection of open sets $\mathcal{O} \subseteq \mathcal{S}$ such that $Y \subseteq \bigcup_{O \in \mathcal{O}} O$.
- A topological space $T = (X, \mathcal{S})$ is compact if every open cover of X has a finite subcover.
- Equivalently, if $\mathcal{C}$ is a collection of closed sets such that every finite intersection of sets from $\mathcal{C}$ is nonempty, then the intersection of all sets in $\mathcal{C}$ is also nonempty.

A topology on infinite words

- There is a natural topology on Σ^ω, the space of one-sided infinite words over a finite alphabet Σ.
- The open sets have the form $L\Sigma^\omega$, where $L \subseteq \Sigma^*$ is any language of finite words.
- This topological space is compact.

Applying the compactness argument

- Let β be a real number.
- Let $W_k(\beta)$ denote the set of all infinite words over $\Sigma_k = \{0, 1, \ldots, k-1\}$ avoiding β-powers.
- $W_k(\beta)$ is closed: it is the complement of the open set $L\Sigma^\omega$, where L is the language of all finite words containing a β-power.

Applying the compactness argument

- Suppose $W_k(\alpha+\epsilon) \neq \emptyset$ for all ϵ.
- If $\alpha \leq \beta$, then $W_k(\alpha) \subseteq W_k(\beta)$.
- The intersection of any finite number of the $W_k(\alpha+\epsilon)$ equals $W_k(\alpha+\epsilon')$, where ϵ' is the smallest of the ϵ, and is therefore nonempty.
- By compactness $W = \bigcap_{\epsilon>0} W_k(\alpha+\epsilon)$ is nonempty.
- Any word $\mathbf{w} \in W$ is α^+-power-free.

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

$$\begin{aligned} h(0) &= 012021201210212021 0 \\ h(1) &= 120102012021020102 1 \\ h(2) &= 201210120102101210 2 \end{aligned}$$

- h is a $(7/4)^+$-power-free morphism
- it 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:

$$12345163 2415$$

Pansiot encoding:

$$0101101.$$

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

The Pansiot encoding

Example (k=6)

Word:

$$12345163 2415$$

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:

$$123451632415$$

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:

$$123451632415$$

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 $0 \to 101101$; $1 \to 10$:

$$1 \to 10 \to 10101101 \to 1010110110101101101010110110 \to \cdots$$

- Decode:

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

A map into the symmetric group

- Moulin Ollagnier proved the conjecture for $5 \leq k \leq 11$.
- His observation: a word $w = a_1 a_2 \cdots a_{k-1}$ containing no repeated letter can be associated with a permutation:

$$\begin{pmatrix} 1 & 2 & 3 & \cdots & k-1 & k \\ a_1 & a_2 & a_3 & \cdots & a_{k-1} & b \end{pmatrix}$$

- b is the unique letter that does not occur in w.

A map into the symmetric group

- Moving from one $(k-1)$-letter block to the next $(k-1)$-letter block by a "0" in the Pansiot encoding corresponds to multiplication on the right by

$$\sigma_0 = \begin{pmatrix} 1 & 2 & 3 & \cdots & k-1 & k \\ 2 & 3 & 4 & \cdots & 1 & k \end{pmatrix}.$$

- Moving from one block to the next by a "1" corresponds to multiplication on the right by

$$\sigma_1 = \begin{pmatrix} 1 & 2 & 3 & \cdots & k-1 & k \\ 2 & 3 & 4 & \cdots & k & 1 \end{pmatrix}.$$

A map into the symmetric group

Example (k=6)

Word:

$$\color{red}{12345}\color{black}{1632415}$$

Pansiot encoding:

$$0101101$$

Permutation:

$$\begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 1 & 2 & 3 & 4 & 5 & 6 \end{pmatrix}$$

A map into the symmetric group

Example (k=6)

Word:

$$1\textcolor{red}{23451}632415$$

Pansiot encoding:

$$0101101$$

Permutation:

$$\sigma_0 = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 2 & 3 & 4 & 5 & 1 & 6 \end{pmatrix}$$

A map into the symmetric group

Example (k=6)

Word:

$$12\color{red}{34516}\color{black}32415$$

Pansiot encoding:

$$0101101$$

Permutation:

$$\sigma_0\color{red}{\sigma_1}\color{black} = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 3 & 4 & 5 & 1 & 6 & 2 \end{pmatrix}$$

A map into the symmetric group

Example (k=6)

Word:

$$123451632415$$

Pansiot encoding:

$$0101101$$

Permutation:

$$\sigma_0\sigma_1\sigma_0 = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 4 & 5 & 1 & 6 & 3 & 2 \end{pmatrix}$$

A map into the symmetric group

Example (k=6)

Word:

$$1234\textcolor{red}{51632}415$$

Pansiot encoding:

$$0101101$$

Permutation:

$$\sigma_0\sigma_1\sigma_0\textcolor{red}{\sigma_1} = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 5 & 1 & 6 & 3 & 2 & 4 \end{pmatrix}$$

A map into the symmetric group

Example (k=6)

Word:

$$123451632415$$

Pansiot encoding:

$$0101101$$

Permutation:

$$\sigma_0\sigma_1\sigma_0\sigma_1\sigma_1 = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 1 & 6 & 3 & 2 & 4 & 5 \end{pmatrix}$$

A map into the symmetric group

Example (k=6)

Word:

$$123451\textcolor{red}{63241}5$$

Pansiot encoding:

$$0101101$$

Permutation:

$$\sigma_0\sigma_1\sigma_0\sigma_1\sigma_1\textcolor{red}{\sigma_0} = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 6 & 3 & 2 & 4 & 1 & 5 \end{pmatrix}$$

A map into the symmetric group

Example (k=6)

Word:

$$1234516\textcolor{red}{32415}$$

Pansiot encoding:

$$0101101$$

Permutation:

$$\sigma_0\sigma_1\sigma_0\sigma_1\sigma_1\sigma_0\textcolor{red}{\sigma_1} = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 3 & 2 & 4 & 1 & 5 & 6 \end{pmatrix}$$

A map into the symmetric group

Example (k=6)

Word:

$$123451632415$$

Pansiot encoding:

$$0101101$$

Permutation:

$$\sigma_0\sigma_1\sigma_0\sigma_1\sigma_1\sigma_0\sigma_1 = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 3 & 2 & 4 & 1 & 5 & 6 \end{pmatrix}$$

A map into the symmetric group

- Define map ψ from the binary Pansiot codewords to the symmetric group S_k by

$$\begin{aligned} 0 &\to \sigma_0 \\ 1 &\to \sigma_1, \end{aligned}$$

and if $y = y_0 y_1 \cdots y_\ell$ is a word over $\{0, 1\}$, then

$$y \to \sigma_{y_0} \sigma_{y_1} \cdots \sigma_{y_\ell}.$$

Kernel repetitions

- Alphabet size k
- w a word over an k-letter alphabet
- x the binary Pansiot encoding of w
- Write $x = pe$ with e also a prefix of x; p non-empty.
- Call p the period and e the excess.
- If $|e| \geq k - 1$ and $\psi(p)$ is the identity permutation, x is a kernel repetition.
- w then has exponent $(|pe| + k - 1)/|p|$.

Kernel repetitions

Example (k=4)

$$\begin{array}{rrcl} \text{Word:} & w & = & 12341341234 13 \\ \text{Pansiot encoding:} & x & = & \underbrace{1100011}_{p}\underbrace{110}_{e} \\ \text{Permutation:} & \psi(p) & = & \begin{pmatrix} 1 & 2 & 3 & 4 \\ 1 & 2 & 3 & 4 \end{pmatrix} \end{array}$$

x is a kernel repetition; w has exponent

$$(|pe| + k - 1)/|p| = (10 + 4 - 1)/7 = 13/7.$$

Moulin Ollagnier's approach

- Generate an infinite Pansiot encoding $\mathbf{x}$ by iterating a binary morphism f.
- $\mathbf{x}$ encodes a word $\mathbf{w}$ over an n-letter alphabet.
- $\mathbf{x}$ must not contain a kernel repetition $x = pe$ with $(|pe| + k - 1)/|p| > RT(k)$.

The algebraic condition

- f maps $0 \to f(0)$; $1 \to f(1)$.
- algebraic condition for f: for some permutation τ,

$$\psi(f(0)) = \tau^{-1} \cdot \psi(0) \cdot \tau, \quad \psi(f(1)) = \tau^{-1} \cdot \psi(1) \cdot \tau.$$

- Ensures that f maps kernel repetitions to kernel repetitions
- Long kernel repetitions are the images under f of shorter kernel repetitions (more or less).

Checking the candidate word

- Check finitely many kernel repetitions in $\mathbf{x}$: verify none have $(|pe| + k - 1)/|p| > RT(k)$.
- Check that $\mathbf{w}$ does not contain other forbidden repetitions that do not arise from kernel repetitions in $\mathbf{x}$.
- These have length at most $(k-1)^2$—only finitely many to check.

Searching by computer

- Moulin Ollagnier found by computer search binary morphisms to generate $\mathbf{x}$ for $5 \leq k \leq 11$.
- For $k = 5$:

$$
\begin{aligned}
0 &\rightarrow 010101101101010110110 \\
1 &\rightarrow 101010101101101101101.
\end{aligned}
$$

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 $k \geq 33$

A quantitative version of Dejean's Theorem

Conjecture (Shur)

Let ρ_k be the real number such that the number of $RT(k)^+$-free words of length n over a k-letter alphabet grows like $(\rho_k)^n$. Then ρ_k tends to a limit $\hat{\alpha} \approx 1.242$ as k tends to infinity.

A highly non-repetitive word

Theorem (Beck 1981)

For any $\epsilon > 0$, there exist N_ϵ and an infinite binary word $\mathbf{w}$ such that any two identical factors of $\mathbf{w}$ of length $n > N_\epsilon$ are separated by a distance at least $(2-\epsilon)^n$.

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

The probabilistic method

- we want to show the existence of an object (word) avoiding certain "bad" events (here, repetitions)
- choose a word at random and show that with positive probability, it avoids repetitions
- this would be easy if the presence of repetitions were independent events
- but repetitions can overlap
- we use the Lovász local lemma

A dependency graph

Given a set S of probability events, we construct a dependency digraph $D = (S, E)$, where the event X is mutually independent of the events $\{Y \ : (X, Y) \notin E\}$.

The Lovász local lemma

Let $A_1, A_2, \ldots, A_t$ be events in a probability space, with a dependency digraph $D = (S, E)$. Suppose there exist real numbers $x_1, x_2, \ldots, x_t$ with $0 \leq x_i < 1$ for $1 \leq i \leq t$ such that

$$\Pr(A_i) \leq x_i \prod_{(i,j) \in E} (1 - x_j)$$

for $1 \leq i \leq t$. Then the probability that none of the events $A_1, A_2, \ldots, A_t$ occur is at least

$$\prod_{1 \leq i \leq t} (1 - x_i).$$

Probabilistic argument for squarefree words

We use this method to prove the existence of an infinite squarefree word over a finite alphabet.

Let $A_{i,r}$ be the event that there exists a square of length $2r$ beginning at position i of a word of length n, i.e., that

$$a_i a_{i+1} \cdots a_{i+r-1} = a_{i+r} a_{i+r+1} \cdots a_{i+2r-1}.$$

Then the event $A_{i,r}$ is mutually independent of the set of all events $A_{j,s}$ when $i + 2r - 1 < j$ or $i > j + 2s - 1$.

In the dependency digraph, (i, r) is connected to (j, s) by an edge in each direction if $i + 2r - 1 \geq j$ and $i \leq j + 2s - 1$.

As in the statement of the lemma, we now associate a real number $x_{i,r}$ with each event $A_{i,r}$. We then have

$$\begin{aligned}\prod_{((i,r),(j,s))\in E}(1-x_{j,s}) &= \prod_{\substack{i-2s+1\leq j\leq i+2r-1\\ 0\leq j\leq n-2s\\ 1\leq s\leq n/2}}(1-x_{j,s})\\ &\geq \prod_{s\geq 1}(1-x_{j,s})^{2r+2s-1}.\end{aligned}$$

Take logs to get

$$\sum_{((i,r),(j,s))\in E} \log(1 - x_{j,s}) \geq \sum_{s\geq 1} (2r + 2s + 1)\log(1 - x_{j,s}).$$

Now we choose the $x_{j,s}$. This is somewhat of a black art: choosing $x_{j,s} = \alpha^{-s}$ for some α often works.

Suppose that we can find real numbers $c < -1$ and $\alpha < 1$ such that $\log(1-\alpha) \geq c\alpha$. Then we set $x_{j,s} = \alpha^{-s}$ and we have

$$\begin{aligned}
& \sum_{s \geq 1} (2r + 2s - 1) \log(1 - x_{j,s}) \\
\geq\ & \sum_{s \geq 1} (2r + 2s - 1) c\alpha^{-s} \\
=\ & (2r-1)c \sum_{s \geq 1} \alpha^{-s} + 2c \sum_{s \geq 1} s\alpha^{-s} \\
=\ & \frac{(2r-1)c}{\alpha - 1} + \frac{2c\alpha}{(\alpha - 1)^2}.
\end{aligned}$$

Now if our events take place over an alphabet of size k, then $\Pr(A_{i,r}) = k^{-r}$, so if

$$\log \Pr(A_{i,r}) = -r \log k \leq -r \log \alpha + \frac{(2r-1)c}{\alpha - 1} + \frac{2c\alpha}{(\alpha - 1)^2},$$

the conditions of the local lemma will be satisfied.

We conclude that with positive probability none of the events $A_{i,r}$ occur; i.e., there exists a squarefree word of length n over an alphabet of size k.

If $\alpha = 6.23$, $c = -1.091$, and $k \geq 13$, the inequality is satisified and we have our result.

The End