Avoiding repetitions in words II

Narad Rampersad

Department of Mathematics and Statistics
University of Winnipeg

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.

Subsequences

- $\mathbf{w} = w_0 w_1 w_2 \cdots$
- subsequence: a word $w_{i_0} w_{i_1} \cdots$, where $0 \leq i_0 < i_1 < \cdots$.
- arithmetic subsequence of difference j: a word $w_i w_{i+j} w_{i+2j} \cdots$, where $i \geq 0$ and $j \geq 1$.

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)

Let p be a prime and let m be a non-negative integer. There exists an infinite word over a finite alphabet that avoids $(1 + 1/p^m)$-powers in arithmetic progressions of all differences, except those differences that are a multiple of p.

Squares in arithmetic progressions

Corollary

There exists an infinite word over a 4-letter alphabet that contains no squares in any arithmetic progression of odd difference.

The construction

Let $q = p^{m+1}$. We construct an infinite word with the desired properties over the alphabet

$$\Sigma = \{n : 0 < n < 2q^2 \text{ and } q \nmid n\}.$$

Define $\mathbf{w} = a_1 a_2 \cdots$ as follows. For $n \geq 1$, write $n = q^t n'$, where $q \nmid n'$, and define

$$a_n = \begin{cases} n' \bmod q^2, & \text{if } t = 0; \\ q^2 + (n' \bmod q^2), & \text{if } t > 0. \end{cases}$$

An example of the construction

Take $p = 2$ and $m = 0$ (so that $q = 2$). Then $\Sigma = \{1, 3, 5, 7\}$ and, writing $n = 2^t n'$,

$$a_n = \begin{cases} n' \bmod 4, & \text{if } n \text{ is odd;} \\ 4 + (n' \bmod 4), & \text{if } n \text{ is even.} \end{cases}$$

It follows that

$$\mathbf{w} = 153517351537173515 3\cdots$$

contains no squares in arithmetic progressions of odd difference.

Proof of the construction

Recall: $\mathbf{w} = a_1 a_2 \cdots$ and $q = p^{m+1}$. For $n \geq 1$, write $n = q^t n'$, where $q \nmid n'$. Then

$$a_n = \begin{cases} n' \bmod q^2, & \text{if } t = 0; \\ q^2 + (n' \bmod q^2), & \text{if } t > 0. \end{cases}$$

Suppose $\mathbf{w}$ contains a $(1 + 1/p^m)$-power in an arithmetic progression of difference k, where k is not a multiple of p:

$$a_i a_{i+k} \cdots a_{i+(s-1)k} = a_{i+rk} a_{i+(r+1)k} \cdots a_{i+(r+s-1)k}$$

for some integers i, r, s satisfying $s/r \geq 1/p^m$.

Proof of the construction

If $a_i = a_{i+rk} > q^2$ then q divides both i and $i + rk$ and hence divides rk.

If $a_i = a_{i+rk} < q^2$, then $i \bmod q^2 = (i + rk) \bmod q^2$, so that q^2 divides rk.

In either case, since p does not divide k, it must be the case that q divides r.

So we write $r = q^\ell r'$ for some positive integers ℓ, r' with r' not divisible by q.

Proof of the construction

Recall that $s/r \geq 1/p^m$ and $q = p^{m+1}$, so that

$$\begin{aligned} s &\geq q^\ell r'/p^m \\ &= pq^{\ell-1}r' \\ &\geq pq^{\ell-1}. \end{aligned}$$

Thus $\{i, i+k, \ldots, i+(s-1)k\}$ forms a complete set of residue classes modulo $pq^{\ell-1}$.

There exists $j \in \{i, i+k, \ldots, i+(s-1)k\}$ such that

$$j \equiv q^{\ell-1} \pmod{pq^{\ell-1}}.$$

Proof of the construction

Write

$$\begin{aligned} j &= apq^{\ell-1} + q^{\ell-1} \\ &= q^{\ell-1}(ap+1), \end{aligned}$$

for some non-negative integer a.

We also have

$$\begin{aligned} j + rk &= q^{\ell-1}(ap+1) + q^{\ell} r' k \\ &= q^{\ell-1}(ap+1+qr'k). \end{aligned}$$

Proof of the construction

Also $a_j = a_{j+rk}$, and so from the definition of $\mathbf{w}$ we have

$$ap + 1 \equiv ap + 1 + qr'k \pmod{q^2},$$

so that $qr'k \equiv 0 \pmod{q^2}$.

This implies $r'k \equiv 0 \pmod{q}$. However, p does not divide k, and q does not divide r', so this congruence cannot be satisfied. This contradiction completes the proof.

The paperfolding word

- again take $p = 2$ and $m = 0$
- then $\mathbf{w} = 15351735153717351 53\cdots$
- apply the map $1, 5 \to 0$, $3, 7 \to 1$ to get
- $\mathbf{f} = 0010011000110110001\cdots$
- this is the ordinary paperfolding word

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 $|w| \in \{1, 3, 5\}$.

2-dimensional words

- 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 $x_t = w_{i_1+j_1 t, i_2+j_2 t}$.

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 4-letter alphabet A that avoid squares in all arithmetic progressions of odd difference.
- Define $\mathbf{w}$ over the alphabet $A \times A$ by $w_{m,n} = (u_m, v_n)$.

Lines through the 2D-word

- Consider an arbitrary line

$$\begin{aligned} \mathbf{x} &= (w_{i_1+j_1t,i_2+j_2t})_{t\geq 0}, \\ &= (\,(u_{i_1+j_1t}, v_{i_2+j_2t})\,)_{t\geq 0}, \end{aligned}$$

  for some i_1, i_2, j_1, j_2, with $\gcd(j_1, j_2) = 1$.
- Without loss of generality, we may assume j_1 is odd.
- Then $(u_{i_1+j_1t})_{t\geq 0}$ is an arithmetic subsequence of odd difference of $\mathbf{u}$ and hence is squarefree.
- $\mathbf{x}$ is therefore also squarefree.

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

The adjacency matrix of a morphism

Given a morphism $\varphi : \Sigma^* \to \Sigma^*$ for some finite set $\Sigma = \{a_1, a_2, \ldots, a_d\}$, we define the adjacency matrix $M = M_\varphi$ as follows:

$$M = (m_{i,j})_{1 \le i,j \le d}$$

where $m_{i,j}$ is the number of occurrences of a_i in $\varphi(a_j)$, i.e., $m_{i,j} = |\varphi(a_j)|_{a_i}$.

An example

$$\varphi : a \to ab$$
$$b \to cc$$
$$c \to bb.$$

$$M_\varphi = \begin{array}{c} \\ a \\ b \\ c \end{array} \begin{array}{c} \begin{array}{ccc} a & b & c \end{array} \\ \begin{pmatrix} 1 & 0 & 0 \\ 1 & 0 & 2 \\ 0 & 2 & 0 \end{pmatrix} \end{array}$$

Properties of M

- Define $\psi : \Sigma^* \to \mathbb{Z}^d$ by

$$\psi(w) = [|w|_{a_1}, |w|_{a_2}, \ldots, |w|_{a_d}]^T.$$

- Then

$$\psi(\varphi(w)) = M_\varphi \psi(w).$$

- By induction $M_\varphi^n = M_{\varphi^n}$, and hence

$$\psi(\varphi^n(w)) = M_\varphi^n \psi(w).$$

Dekking's construction

- Define a map:

$$a \to aaab, \quad b \to abb.$$

- The limit of the sequence

$$a \to aaab \to aaabaaabaaababb \to \cdots$$

contains no abelian 4-power.

Dekking's method

- the idea is to map letters to elements of $\mathbb{Z}/n\mathbb{Z}$ for some n
- abelian repetitions correspond to certain arithmetic progressions in $\mathbb{Z}/n\mathbb{Z}$
- show no such arithmetic progressions exist

Some definitions

- Let $\varphi : \Sigma^* \to \Sigma^*$ be a morphism.
- Call the words $\varphi(a)$, for $a \in \Sigma$, blocks.
- If $\varphi(a) = vv'$, $v' \neq \epsilon$, then v is a left subblock and v' a right subblock.
- Let G be a finite abelian group (written additively).
- $A \subseteq G$ is progression-free of order n if for all $a \in A$, $a, a+g, a+2g, \ldots, a+(n-1)g \in A$ implies $g = 0$.

φ-injectivity

- Let $f : \Sigma^* \to G$ be a morphism: i.e., $f(\epsilon) = 0$ and $f(a_1 a_2 \cdots a_i) = \sum_{1 \le j \le i} f(a_j)$.
- Call f a weight function.
- Let $v_1 v'_1, v_2 v'_2, \ldots, v_n v'_n$ be blocks.
- f is φ-injective if

$$f(v_1) = f(v_2) = \cdots = f(v_n)$$

implies either $v_1 = v_2 = \cdots = v_n$ or $v'_1 = v'_2 = \cdots = v'_n$.

The main criterion

Suppose that

(a) The adjacency matrix of φ is invertible.

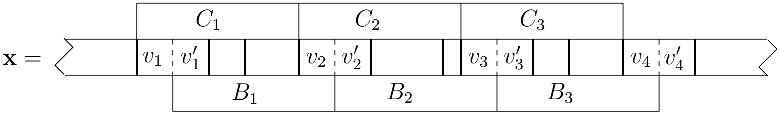
(b) $f(\varphi(a)) = 0$ for all $a \in \Sigma$;

(c) the set $A = \{f(v) : \ v$ a left subblock of $\varphi\}$ is progression-free of order $n+1$;

(d) f is φ-injective.

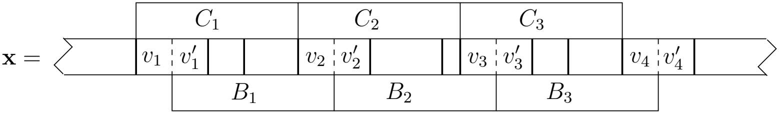
If φ is prolongable on a, and $\varphi^\omega(a)$ avoids "short" abelian n-powers, then $\varphi^\omega(a)$ is abelian n-power-free.

Proof

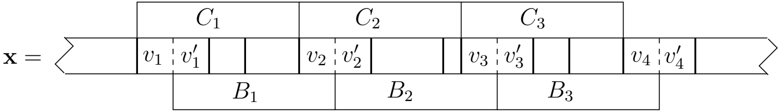


Let $\mathbf{x} = \varphi^\omega(a)$. Suppose $B_1 B_2 \cdots B_n$ is an abelian n-power in $\mathbf{x}$ with $|B_i|$ minimal.

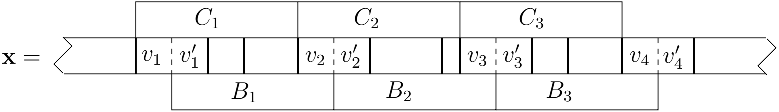
Suppose the B_i are not "short": i.e., $|B_i| > \max_{a\in\Sigma} |\varphi(a)|$.



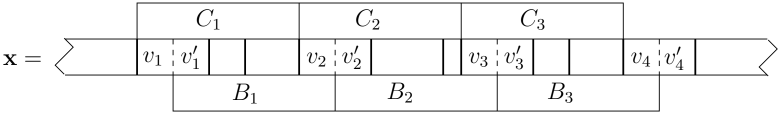
- Since the B_i's contain the same numbers of each letter, we have $f(B_1) = f(B_2) = \cdots = f(B_n)$.
- By hypothesis $f(\varphi(a)) = 0$ for every $a \in \Sigma$.
- Hence $f(B_i) = f(v_i') + f(v_{i+1})$.



- Since $f(v_i v_i') = 0$, we get $f(B_i) = -f(v_i) + f(v_{i+1})$.
- Thus the $f(v_i)$ form an $(n+1)$-term arithmetic progression with difference $f(B_i)$.
- This forces $f(v_1) = f(v_2) = \cdots = f(v_{n+1})$.
- φ-injectivity forces either $v_1 = v_2 = \cdots = v_{n+1}$ or $v_1' = v_2' = \cdots = v_{n+1}'$



In the first case, we "slide" the abelian n-power to the left by $|v_1|$ symbols to get another n-power $C_1 C_2 \cdots C_n$, which is aligned with blocks of φ. In the second case we slide to the right.



- Let D_i be such that $C_i = \varphi(D_i)$.
- Since $\mathbf{x} = \varphi(\mathbf{x})$, $D_1 D_2 \cdots D_n$ is a factor of $\mathbf{x}$.
- Now $\psi(C_i) = M\psi(D_i)$, where M is the matrix of φ.
- Since M is invertible and $\psi(C_1) = \psi(C_2) = \cdots = \psi(C_n)$, we have $\psi(D_1) = \psi(D_2) = \cdots = \psi(D_n)$.
- $D_1 \cdots D_n$ is a shorter abelian n-power, contradiction.

Avoiding abelian 4-powers

- We check that the morphism φ

$$a \to aaab, \quad b \to abb$$

  verifies the criterion we just proved.
- The matrix of φ is $\begin{pmatrix} 3 & 1 \\ 1 & 2 \end{pmatrix}$, which is invertible.
- Take $G = \mathbb{Z}/5\mathbb{Z}$.
- Define $f : \{a, b\}^* \to G$ by $f(a) = 1$ and $f(b) = 2$.
- $f(aaab) = f(abb) = 0$

Avoiding abelian 4-powers

- $A = \{0, 1, 2, 3\}$ is progression free of order 5
- f is φ-injective
- $\varphi^{\omega}(a)$ has no short abelian 4-powers
- by previous criterion, $\varphi^{\omega}(a)$ avoids abelian 4-powers

Avoiding abelian cubes

Define ϑ by $\vartheta(a) = aabc$, $\vartheta(b) = bbc$, and $\vartheta(c) = acc$. The same method shows that $\vartheta^\omega(a)$ avoids abelian cubes.

Avoiding abelian squares

- Finding a construction that avoids abelian squares over 4-letters is much harder.
- Keränen found one in 1992 by intensive computer search.
- The images of the letters under his morphism each have length 85.

Avoiding patterns in the abelian sense

- an abelian instance of the pattern $xxyxyx$ is a word

$$X_1X_2Y_1X_3Y_2X_4$$

  where the X_i are all abelian equivalent and the Y_i are all abelian equivalent; e.g.,

$$0012\ 0120\ 221\ 1020\ 122\ 2100$$

- relatively little is known about avoidability of patterns in the abelian sense

Avoiding long binary patterns in the abelian sense

Theorem (Currie and Visentin 2008)

Patterns over $\{x, y\}$ of length greater than 118 are avoidable in the abelian sense on a binary alphabet.

Avoiding sum-cubes

The following avoidability result is even stronger than avoiding abelian cubes:

Theorem (Cassaigne, Currie, Schaeffer, Shallit 2011)

There exists an infinite word over a finite subset of $\mathbb{N}$ that contains no three consecutive blocks of the same length and the same sum (sum-cube).

The word is the fixed point of the morphism

$$0 \to 03 \qquad 1 \to 43 \qquad 3 \to 1 \qquad 4 \to 01.$$

Avoiding sum-squares

Open problem

Does there exist an infinite word over a finite subset of $\mathbb{N}$ that contains no two consecutive blocks of the same length and the same sum (sum-square)?

The End