Non-constructive methods in combinatorics on words

Narad Rampersad

Department of Mathematics
University of Liège

The Thue–Morse word

$$\mathbf{t} = 01101001100101101001011001101001\cdots$$

- Let t_i denote the i-th symbol of $\mathbf{t}$.
- t_i is the number of 1's mod 2 in the binary expansion of i.
- Studied by Thue in 1906, Morse in 1921, and many others later.

Combinatorial properties of the T–M word

- $\mathbf{t}$ contains no overlap (Thue 1912).
- An overlap is a factor of the form xxx (like `shshsh`) or $xyxyx$ (like `entente`).
- $\mathbf{t}$ is the lexicographically largest infinite binary word starting with 0 that avoids overlaps (Berstel 1994).

Patterns

- xxx and $xyxyx$ are patterns (Bean, Ehrenfeucht, McNulty; Zimin 1979).
- x, y, etc., are variables.
- Which patterns are avoidable?
- How many symbols are required to avoid a pattern?
- xx can be avoided using 3 symbols (Thue 1906).

Avoiding a specified set of words

- We apply a special case of a result 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.

Avoiding the pattern xx

Proposition

For $n \geq 0$ there are at least 5^n words of length n over an alphabet of size 7 that avoid the pattern xx.

- Let S be the set of squares over an alphabet of size 7.
- For $n \geq 1$ the set S contains 7^n squares of length $2n$.

Applying the power series criterion

- Define

$$\begin{aligned} G(z) &:= \left(1 - 7z + \sum_{i\geq 1} 7^i z^{2i}\right)^{-1} \\ &= \left(1 - 7z + \frac{7z^2}{1-7z^2}\right)^{-1} \\ &= 1 + 7z + 42z^2 + 245z^3 + 1372z^4 + 7546z^5 + \cdots . \end{aligned}$$

- It is easy to show that $[z^n]G(z) \geq 5^n$ for $n \geq 0$.

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.

Avoiding a finite set of words

- Let S be a set of words of length n.
- We want an infinite word that contains no element of S as a factor.
- If S contains very few words (as a function of n), this is probably easy.
- We would like to show that even when S is quite large, this is still possible.

The probabilistic method

- Let $A_1, \ldots, A_n$ be events in a probability space.
- A graph $G = (V, E)$ is a dependency graph if $V = \{1, \ldots, n\}$ and for all i, A_i is mutually independent of all the A_j's for which there is no edge $\{i, j\} \in E$.

The Lovász Local Lemma

Lovász Local Lemma (symmetric version)

Let $G = (V, E)$ be a dependency graph for events $A_1, \ldots, A_n$. Suppose that the maximum degree of G is d and that there is a real number p for which $\Pr[A_i] \leq p$ for all $i = 1, \ldots, n$. If $4pd \leq 1$, then

$$\Pr[\cap \overline{A_i}] \geq (1 - 2p)^n > 0.$$

Looking at a random word

- Fix an alphabet with k symbols.
- We consider a random word and use the local lemma to show that with positive probability the word avoids S.
- Let t be an arbitrary positive integer and let w be a random word of length t.

Calculating probabilities

- For $i = 1, \ldots, t$, A_i is the event that w contains an occurrence of a word from S at position i.
- Then $\Pr[A_i] = |S|/k^n$.
- We may take $d = 2n - 1$: there are at most $2n - 1$ overlapping pairs of occurrences of factors of length n in w.

Applying the local lemma

- For
$$|S| \le \frac{k^n}{4(2n-1)},$$
we have
$$4(|S|/k^n)(2n-1) \le 1.$$
- By the local lemma, with positive probability w contains no occurrence of any word of S.

The result summarized

Theorem

Fix an alphabet A of size $k \geq 2$. Let S be a set of words of length $n \geq 1$ over A. If

$$|S| \leq \frac{k^n}{4(2n-1)},$$

then there is an infinite word over A that avoids S.

- We pass from finite to infinite words by a standard compactness argument.
- One can obtain a stronger result using other methods.

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 $> (2-\epsilon)^n$.

- One of the first uses of the local lemma in combinatorics on words.
- No constructive proof known.

Nonrepetitive colourings of the real line

Theorem (Rote, see Grytczuk and Śliwa 2003)

There exists a colouring f of $\mathbb{R}$, $f : \mathbb{R} \to \{0, 1\}$, such that no two line segments are coloured alike with respect to translations. Formally, for every $\epsilon > 0$ and every pair of real numbers $x < y$, there exists $0 \leq t < \epsilon$ such that $f(x + t) \neq f(y + t)$.

We specify the colouring

- Define $f(x) = 0$ if $\log|x|$ is rational and $f(x) = 1$ otherwise.
- Consider two points $0 \leq x < y$.
- If $f(x) = f(y)$, then let $x + t_1 = e^{q_1}$, where $0 \leq t_1 < \epsilon$ and q_1 is rational.
- Now $f(x + t_1) = 0$.
- If $f(x + t_1) = f(y + t_1) = 0$, then $y + t_1 = e^{q_2}$ for some rational number $q_2 \neq q_1$.

We derive a contradiction

- Let $x + t_2 = e^{q_3}$, where $t_1 < t_2 < \epsilon$ and q_3 is rational.
- If again $f(y + t_2) = 0$, then $y + t_2 = e^{q_4}$ for some rational number q_4.
- Now $x - y = e^{q_1} - e^{q_2} = e^{q_3} - e^{q_4}$, where the q_i's are all distinct rational integers.
- Algebraic powers of e are linearly independent over the algebraic numbers (Lindemann–Weierstrass 1885)!

Measurable colourings of the real line

Theorem (Alon, Grytczuk, Lasoń, Michałek 2009)

There is a 5-coloring of the real line such that no pair of intervals (not necessarily adjacent) has the same measure of every color.

- Proof uses topological methods
- Is 5 optimal?

The End