Repetitions in Words—Part IV

Narad Rampersad

Department of Mathematics and Statistics
University of Winnipeg

Another method for showing exponential growth

- A special case of a theorem of Golod and Shafarevich 1964.
- Let S be a set of words over an m-letter alphabet, each of length at least 2.
- Suppose S has at most c_i words of length i for $i \geq 2$.

A power series criterion

Theorem

If the power series expansion of

$$G(x) := \left(1 - mx + \sum_{i \geq 2} c_i x^i\right)^{-1}$$

has non-negative coefficients, then there are least $[x^n]G(x)$ words of length n over a m-letter alphabet that contain no word of S as a factor.

Proof

- Let $F(x) := \sum_{i\geq 0} a_i x^i$, where a_i is the number of words of length i over an m-letter alphabet that avoid S.
- We show

$$F \geq G(x) = \left(1 - mx + \sum_{i\geq 2} c_i x^i\right)^{-1} = \sum_{i\geq 0} b_i x^i.$$

- $F \geq G$ means $a_i \geq b_i$ for all $i \geq 0$

Proof

- For $k \geq 1$, there are $m^k - a_k$ words w of length k over an m-letter alphabet that contain a word in S as a factor.
- (a) $w = w'a$, where a is a single letter and w' is a word of length $k-1$ containing a word in S as a factor
- (b) $w = xy$, where x is a word of length $k-j$ that avoids S and $y \in S$ is a word of length j.
- at most $(m^{k-1} - a_{k-1})m$ words of the form (a)
- at most $\sum_j a_{k-j}c_j$ words w of the form (b)

Proof

So

$$m^k - a_k \leq (m^{k-1} - a_{k-1})m + \sum_j a_{k-j}c_j.$$

Rearrange:

$$a_k - a_{k-1}m + \sum_j a_{k-j}c_j \geq 0, \quad k \geq 1.$$

Proof

- Define

$$\begin{aligned} H(x) &:= F(x)\left(1 - mx + \sum_{j\geq 2} c_j x^j\right) \\ &= \left(\sum_{i\geq 0} a_i x^i\right)\left(1 - mx + \sum_{j\geq 2} c_j x^j\right). \end{aligned}$$

- for $k \geq 1$, we have

$$[x^k]H(x) = a_k - a_{k-1}m + \sum_j a_{k-j}c_j.$$

Proof

- we have shown $a_k - a_{k-1}m + \sum_j a_{k-j}c_j \geq 0$
- so $[x^k]H(x) \geq 0$ for $k \geq 1$.
- Since $[x^0]H(x) = 1$, the inequality $H \geq 1$ holds and $H - 1$ has non-negative coefficients.
- Then $F = HG = (H-1)G + G \geq G$, as required.

Enumeration of squarefree words

- With almost no work, we can show that there are at least 5^n squarefree words of length n over an alphabet of size 7.
- 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 $n \geq 0$.

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?

Exponential growth of words avoiding patterns

Theorem (Bell and Goh 2007)

Let $k \geq 2$ and $m \geq 4$ be integers with $(k, m) \neq (2, 4)$. Let p be a pattern containing k distinct variables, each occurring at least twice in p. Then for $n \geq 0$, there are at least λ^n words of length n over an m-letter alphabet that avoid the pattern p, where

$$\lambda = \lambda(k, m) := m \left(1 + \frac{1}{(m-2)^k}\right)^{-1}.$$

Some special cases

Corollary

Let p be a pattern in which every variable occurs at least twice. There is an infinite word over a 4-letter alphabet that avoids p.

Some special cases

Corollary

All patterns with k variables and length at least 2^k are avoidable over a 4-letter alphabet.

Proof of Theorem

Lemma

Let $k \geq 1$ be a integer and let p be a pattern over the set of variables $\Delta = \{x_1, \ldots, x_k\}$. Suppose that for $1 \leq i \leq k$, the variable x_i occurs $a_i \geq 1$ times in p. Let $m \geq 2$ be an integer and let Σ be an m-letter alphabet. Then for $n \geq 1$, the number of words of length n over Σ that are instances of the pattern p is at most $[x^n]C(x)$, where

$$C(x) := \sum_{i_1 \geq 1} \cdots \sum_{i_k \geq 1} m^{i_1 + \cdots + i_k} x^{a_1 i_1 + \cdots + a_k i_k}.$$

- Let $k \geq 2$ and $m \geq 4$ be integers with $(k, m) \neq (2, 4)$.
- Let
$$\lambda = \lambda(k, m) := m \left(1 + \frac{1}{(m-2)^k}\right)^{-1}.$$
- We have $\lambda \geq m - 1/2$.

Let $a_1, \ldots, a_k$ be integers, each at least 2. Let

$$C(x) := \sum_{i_1 \geq 1} \cdots \sum_{i_k \geq 1} m^{i_1 + \cdots + i_k} x^{a_1 i_1 + \cdots + a_k i_k},$$

and let

$$B(x) := \sum_{i \geq 0} b_i x^i = (1 - mx + C(x))^{-1}.$$

To prove the theorem, we show $b_n \geq \lambda b_{n-1}$ for all $n \geq 0$.

- The proof is by induction on n.
- When $n = 0$, we have $b_0 = 1$ and $b_1 = m$.
- Since $m > \lambda$, the inequality $b_1 \geq \lambda b_0$ holds.
- Suppose that for all $j < n$, we have $b_j \geq \lambda b_{j-1}$.
- Since $B = (1 - mx + C)^{-1}$, we have $B(1 - mx + C) = 1$.
- Hence $[x^n]B(1 - mx + C) = 0$ for $n \geq 1$.

However,

$$B(1 - mx + C) =$$

$$\left(\sum_{i \geq 0} b_i x^i\right) \left(1 - mx + \sum_{i_1 \geq 1} \cdots \sum_{i_k \geq 1} m^{i_1 + \cdots + i_k} x^{a_1 i_1 + \cdots + a_k i_k}\right),$$

so

$$[x^n] B(1 - mx + C) =$$

$$b_n - b_{n-1} m + \sum_{i_1 \geq 1} \cdots \sum_{i_k \geq 1} m^{i_1 + \cdots + i_k} b_{n-(a_1 i_1 + \cdots + a_k i_k)} = 0.$$

Rearranging, we obtain

$$b_n = \lambda b_{n-1} + (m - \lambda) b_{n-1} - \sum_{i_1 \geq 1} \cdots \sum_{i_k \geq 1} m^{i_1 + \cdots + i_k} b_{n-(a_1 i_1 + \cdots + a_k i_k)}.$$

To show $b_n \geq \lambda b_{n-1}$ it therefore suffices to show

$$(m-\lambda)b_{n-1} - \sum_{i_1 \geq 1} \cdots \sum_{i_k \geq 1} m^{i_1+\cdots+i_k} b_{n-(a_1 i_1+\cdots+a_k i_k)} \geq 0.$$

Since $b_j \geq \lambda b_{j-1}$ for all $j < n$, we have $b_{n-i} \leq b_{n-1}/\lambda^{i-1}$ for $1 \leq i \leq n$. Hence

$$\begin{aligned}
& \sum_{i_1 \geq 1} \cdots \sum_{i_k \geq 1} m^{i_1+\cdots+i_k} b_{n-(a_1 i_1+\cdots+a_k i_k)} \\
\leq \; & \sum_{i_1 \geq 1} \cdots \sum_{i_k \geq 1} m^{i_1+\cdots+i_k} \frac{\lambda b_{n-1}}{\lambda^{a_1 i_1+\cdots+a_k i_k}} \\
= \; & \lambda b_{n-1} \sum_{i_1 \geq 1} \cdots \sum_{i_k \geq 1} \frac{m^{i_1+\cdots+i_k}}{\lambda^{a_1 i_1+\cdots+a_k i_k}} \\
= \; & \lambda b_{n-1} \sum_{i_1 \geq 1} \frac{m^{i_1}}{\lambda^{a_1 i_1}} \cdots \sum_{i_k \geq 1} \frac{m^{i_k}}{\lambda^{a_k i_k}}.
\end{aligned}$$

Since $a_i \geq 2$ for $1 \leq i \leq k$, we have

$$\begin{aligned} & \lambda b_{n-1} \sum_{i_1 \geq 1} \frac{m^{i_1}}{\lambda^{a_1 i_1}} \cdots \sum_{i_k \geq 1} \frac{m^{i_k}}{\lambda^{a_k i_k}} \\ \leq \; & \lambda b_{n-1} \sum_{i_1 \geq 1} \frac{m^{i_1}}{\lambda^{2 i_1}} \cdots \sum_{i_k \geq 1} \frac{m^{i_k}}{\lambda^{2 i_k}} \\ = \; & \lambda b_{n-1} \left(\sum_{i \geq 1} \frac{m^i}{\lambda^{2i}} \right)^k . \end{aligned}$$

Since $\lambda \geq m - 1/2$, we have $m/\lambda^2 \leq m/(m-1/2)^2 < 1$. Thus

$$\lambda b_{n-1} \left(\sum_{i \geq 1} \frac{m^i}{\lambda^{2i}} \right)^k = \lambda b_{n-1} \left(\frac{m/\lambda^2}{1 - m/\lambda^2} \right)^k = \lambda b_{n-1} \left(\frac{m}{\lambda^2 - m} \right)^k .$$

We have thus shown

$$\sum_{i_1 \geq 1} \cdots \sum_{i_k \geq 1} m^{i_1+\cdots+i_k} b_{n-(a_1 i_1+\cdots+a_k i_k)} \leq \lambda b_{n-1} \left(\frac{m}{\lambda^2 - m} \right)^k .$$

We are trying to show

$$(m - \lambda) b_{n-1} - \sum_{i_1 \geq 1} \cdots \sum_{i_k \geq 1} m^{i_1+\cdots+i_k} b_{n-(a_1 i_1+\cdots+a_k i_k)} \geq 0.$$

Clearly, it now suffices to show

$$m - \lambda \geq \lambda \left(\frac{m}{\lambda^2 - m} \right)^k .$$

Again, since $\lambda \geq m - 1/2$, we have

$$\begin{aligned}
\lambda\left(\frac{m}{\lambda^2 - m}\right)^k &\leq \lambda\left(\frac{m}{(m-1/2)^2 - m}\right)^k \\
&= \lambda\left(\frac{m}{m^2 - 2m + 1/4}\right)^k \\
&\leq \lambda\left(\frac{m}{m^2 - 2m}\right)^k \\
&= \lambda/(m-2)^k. \qquad (1)
\end{aligned}$$

On the other hand,

$$\lambda = m \left(1 + \frac{1}{(m-2)^k} \right)^{-1},$$

whence

$$\lambda \left(1 + \frac{1}{(m-2)^k} \right) = m,$$

and so

$$\lambda/(m-2)^k = m - \lambda. \tag{2}$$

(1) and (2) establish

$$m - \lambda \geq \lambda \left(\frac{m}{\lambda^2 - m} \right)^k .$$

We conclude that $b_n \geq \lambda b_{n-1}$, which completes the proof.

Exponential growth of words avoiding patterns

Theorem (Bell and Goh 2007)

Let $k \geq 2$ and $m \geq 4$ be integers with $(k, m) \neq (2, 4)$. Let p be a pattern containing k distinct variables, each occurring at least twice in p. Then for $n \geq 0$, there are at least λ^n words of length n over an m-letter alphabet that avoid the pattern p, where

$$\lambda = \lambda(k, m) := m \left(1 + \frac{1}{(m-2)^k}\right)^{-1}.$$

Decidable properties

- Are there algorithms to decide if an infinite word
  - is aperiodic?
  - is recurrent?
  - avoids repetitions?
  - etc.
- Are there algorithms to compute its
  - complexity function?
  - recurrence function?
  - critical exponent?
  - etc.

Automatic sequences

- A sequence is k-automatic if it is generated by first iterating a k-uniform morphism and then renaming some of the symbols.

The Thue–Morse sequence

- the prototypical 2-automatic sequence:

$$0110100110010110\cdots$$

- generated by iterating the map

$$0 \to 01, \quad 1 \to 10$$

The characteristic sequence of the powers of 2

- Iterate the 2-uniform morphism

$$a \to ab, b \to bc, c \to cc$$

to get the infinite sequence

$$abbcbcccbcccccccbcccccccccccccccbcc \cdots .$$

- Now recode by $a, c \to 0$; $b \to 1$:

$$01101000100000001000000000000000100 \cdots .$$

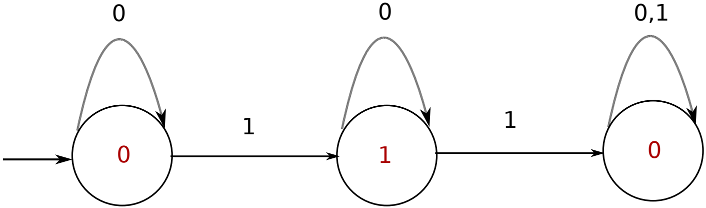
Determining periodicity

- Given a k-automatic sequence, can we tell if it is ultimately periodic?
- Honkala (1986) gave an algorithm.
- This result was often reproved: Muchnik (1991), Fagnot (1997), Allouche, R., and Shallit (2009).
- Leroux (2005) gave a polynomial time algorithm.

An automaton-based characterization

- The proof of Allouche et al. is perhaps the simplest.
- It is based on another characterization of automatic sequences:
- A sequence $\mathbf{a}$ is k-automatic if there exists a finite automaton with output that, when given the base-k representation of n as input, outputs the $(n+1)$-th term of $\mathbf{a}$.
- This is the original definition of an automatic sequence; the equivalence with the morphism-based definition is due to Cobham.

An automaton for the powers of 2



A logic-based characterization

- Another important characterization (Büchi–Bruyère):
- Let $V_k(x)$ denote the largest power of k that divides x.
- A sequence $\mathbf{a}$ is k-automatic if it is definable in the logical structure $\langle \mathbb{N}, +, V_k \rangle$.
- I.e., for each alphabet symbol b, there exists a first-order formula φ_b of $\langle \mathbb{N}, +, V_k \rangle$ such that

$$\mathbf{a}^{-1}(b) = \{n \in \mathbb{N} : \langle \mathbb{N}, +, V_k \rangle \models \varphi_b(n)\}.$$

Defining the powers of 2 using logic

- The characteristic sequence $\mathbf{a}$ of the powers of 2 has a simple definition in this formulation:

$$\begin{aligned}\mathbf{a}^{-1}(1) &= \{n \in \mathbb{N} : \langle \mathbb{N}, +, V_k \rangle \models (V_2(n) = n)\} \\ \mathbf{a}^{-1}(0) &= \{n \in \mathbb{N} : \langle \mathbb{N}, +, V_k \rangle \models \neg(V_2(n) = n)\}\end{aligned}$$

Decidability

Theorem (Bruyère 1985)

The first order theory of $\langle \mathbb{N}, +, V_k \rangle$ is decidable.

Applying these ideas

- We can now apply these ideas to obtain algorithms to determine periodicity, recurrence, etc.
- A sequence $\mathbf{a}$ is ultimately periodic if and only if there exist integers $p \geq 1$ and $n \geq 0$ such that $\mathbf{a}(i) = \mathbf{a}(i + p)$ for all $i \geq n$.
- Hence there exists a decision procedure for determining the periodicity of k-automatic sequences.

Critical exponent

- A word w with period p has an exponent $|w|/p$.
- The exponent of w is its largest exponent.
- The critical exponent of an infinite word is the supremum of the exponents of its finite factors.
- The Thue–Morse word has critical exponent 2.
- The Fibonacci word has critical exponent $2 + \varphi$.

An expression for the critical exponent

- Krieger showed that the critical exponent of the fixed point of a uniform morphism is either rational or infinite.
- For a sequence $\mathbf{a}$, let X be the set of all pairs (q, p) such that there exists a factor of $\mathbf{a}$ of length q with period p.
- If $\mathbf{a}$ is k-automatic, we can construct a finite automaton to accept $\{(q, p)_k : (q, p) \in X\}$.
- The critical exponent is $\sup\{q/p : (q, p) \in X\}$.

Calculating the critical exponent

Theorem (Shallit 2011)

Given a k-automatic sequence, its critical exponent is either rational or infinite and can be effectively computed.

The End