

Combinatorics on Words:

Applications to Number Theory and Ramsey Theory

Narad Rampersad

Department of Mathematics and Statistics
University of Winnipeg

9 May 2008

Words

- A **word** is a sequence of symbols from an **alphabet**.
- Words may be finite, like the word `finite` or infinite, like the word

abaabaaabaaaab...

- In 1906, Norwegian mathematician Axel Thue constructed an infinite word over a three letter alphabet that contained no repetitions.
- Here is how to construct such a word using the symbols 0, 1, and 2.
- First, define a substitution rule (or **morphism**)

$$0 \rightarrow 1, \quad 1 \rightarrow 20, \quad 2 \rightarrow 210.$$

- This means for every 0 substitute a 1, for every 1 substitute 20, and for every 2 substitute 210.

Avoiding Repetitions in Words

- Now, starting with 0, repeatedly apply this rule, obtaining the words

$$0 \Rightarrow 1 \Rightarrow 20 \Rightarrow 2101 \Rightarrow 21020120 \Rightarrow \\ 2102012101202101 \Rightarrow \dots$$

- In the limit we obtain the infinite word

$$21020121012021020120210121020120\dots$$

- This word contains no repetitions (to prove this takes a little more work).

Avoiding Repetitions in Words

- This result has been independently rediscovered numerous times in the last century, for instance by Arshon in 1937 and Morse and Hedlund in 1944.
- The existence of such a word was used in 1968 by Novikov and Adian to solve a longstanding open problem in group theory posed by Burnside in 1902.
- The repetitions discussed so far have been words of the form xx .
- We may write this algebraically as x^2 and call it a **square**.
- Analogously, we may define other types of repetitions, such of those of the form $xxx = x^3$, called a **cube**, and so on.

Avoiding Cubes

- In 1912, Axel Thue showed that the word

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

obtained by repeatedly applying the morphism $0 \rightarrow 01, 1 \rightarrow 10$ contains no cubes.

- This result was rediscovered by Morse in 1921, and so this word is usually referred to as the **Thue–Morse** word.
- In fact, The Thue–Morse contains no **overlap**: i.e., no occurrence of xxa , where x is a word and a is the first letter of x .
- In 1929, chess grandmaster Max Euwe used this word to show that under a certain (plausible) set of rules it was possible to play an infinite game of chess.

The Thue–Morse Word

- The Thue–Morse word can be defined in several other ways.
- Let t_i denote the i -th symbol of \mathbf{t} .
- Then t_i is the number of 1's mod 2 in the binary expansion of i .
- We thus have the recursive definition $t_{2i} = t_i$ and $t_{2i+1} = 1 - t_i$.

The Thue–Morse Word in Number Theory

- Let β be an irrational real number between 1 and 2. Let us write the expansion of 1 in base- β :

$$1 = \sum_{n=1}^{\infty} a_n \beta^{-n},$$

where $a_n \in \{0, 1\}$.

- β -expansions are not always unique!
- Let $\varphi = (1 + \sqrt{5})/2$ be the golden ratio.
- Then $1 = \varphi^{-1} + \varphi^{-2}$ and $1 = \sum_{n \geq 2} \varphi^{-n}$, so .11 and .01111... are both φ -expansions of 1.

The Thue–Morse Word in Number Theory

- Curiously, there are certain β for which 1 has a unique β -expansion.
- Moreover, there is a *least* β between 1 and 2 with this property.
- This β is the unique solution (≈ 1.78723) to

$$1 = \sum_{n=1}^{\infty} t_n \beta^{-n},$$

where $t_0 t_1 t_2 \dots$ is the Thue–Morse word!

- This β is called the Komornik–Loreti constant, and was discovered in 1998.
- It was proved to be transcendental by Allouche and Cosnard, using a result of Mahler.

The Thue–Morse Word in Number Theory

- Consider the power series $F(z) = \sum_{n \geq 0} t_n z^{-n}$, where $t_0 t_1 t_2 \dots$ is the Thue–Morse word.
- Using standard results from analysis, one can show that F is a transcendental function.
- Using the identities $t_{2i} = t_i$ and $t_{2i+1} = 1 - t_i$, one can easily show (exercise!) that F satisfies the functional equation $F(z) = (1 - z)F(z^2)$.
- In 1929, Mahler showed that a transcendental function satisfying a functional equation of this sort takes transcendental values at every non-zero algebraic point in its disc of convergence.

The Thue–Morse Word in Number Theory

- We deduce the transcendence of the Komornik–Loreti constant as follows.
- If β is the Komornik–Loreti constant, then $F(\beta) = 1$.
- If β were algebraic, then $F(\beta) = 1$ would be transcendental by Mahler's argument.
- This contradiction implies that β is transcendental.
- Consider now the real number $F(2) = \sum_{n \geq 0} t_n 2^{-n}$, the so-called Thue–Morse constant, whose binary expansion

0.1101001100101101001011001101001...

is given by the Thue–Morse word.

- Again, Mahler's method shows that $F(2)$ is transcendental.

A Combinatorial Criterion for Transcendence

- Look at the squares in the Thue–Morse word:

01101001100101101001011001101001 . . .

- We have:
 - ▶ 11 at position 1;
 - ▶ 1010 at position 2;
 - ▶ 10011001 at position 4;
 - ▶ 1001011010010110 at position 8; etc.
- In general, there are larger and larger squares occurring not too far from the beginning of the sequence.
- This property turns out to be a sufficient condition for transcendence.

A Combinatorial Criterion for Transcendence

- Truncate the binary expansion of the Thue–Morse constant after the square 11, and append $11111111 \dots$ to get the rational number $0.1111111111 \dots$.
- Now truncate after the square 1010, and append $10101010 \dots$ to get the rational number $0.1101010101010 \dots$.
- Truncate after the square 10011001, and append $10011001 \dots$ to get the rational number $0.1101001100110011001 \dots$, and so on.
- This gives a sequence of very good rational approximations to the Thue–Morse constant.
- Recall that if an irrational number is approximated too well by rationals, it cannot be algebraic.
- It is possible to use standard (but deep!) results in Diophantine approximation to show that these rational numbers approximate the Thue–Morse constant too well.

A Combinatorial Criterion for Transcendence

- The following criterion for transcendence, due to Adamczewski, Bugeaud, and Luca, formalizes the above observations.
- Let $b > 1$ be an integer and let $\mathbf{a} = a_0 a_1 \cdots$ be an infinite word over $\{0, 1, \dots, b-1\}$.
- Let $w > 1$ be a real number. Suppose there exist two sequences of finite words $(U_n)_{n \geq 1}, (V_n)_{n \geq 1}$ such that:
 - (i) For any $n \geq 1$, the word $U_n V_n^w$ is a prefix of the word \mathbf{a} ;
 - (ii) The sequence $(|U_n|/|V_n|)_{n \geq 1}$ is bounded from above;
 - (iii) The sequence $(|V_n|)_{n \geq 1}$ is increasing.
- Then the real number

$$\sum_{n \geq 0} \frac{a_n}{b^n}$$

is either rational or transcendental.

The Thue–Morse Word in Number Theory

- Recall that the Thue–Morse word is overlap-free: it has no occurrence of xxa , where a is the first symbol of x .
- Is it possible that all real numbers whose binary expansions are overlap-free are transcendental?
- Stated another way, must the binary expansion of any algebraic number contain infinitely many occurrences of overlaps?
- Binary overlap-free words in general have much the same structure as the Thue–Morse word.
- They thus also satisfy the Adamczewski–Bugeaud–Luca criterion for transcendence.

Applications to Transcendental Number Theory

Theorem (Adamczewski and R. 2007)

The binary expansion of an algebraic number contains infinitely many occurrences of overlaps (even $7/3$ -powers).

Theorem (Adamczewski and R. 2007)

The ternary expansion of an algebraic number contains either infinitely many occurrences of squares or infinitely many occurrences of one of the blocks 010 or 02120.

Words as Colourings

- An infinite word over a finite alphabet A is a map \mathbf{w} from \mathbb{N} to A .
- A infinite word can be thus be viewed as a “colouring” of the natural numbers using the set A of “colours”.
- An infinite word avoiding repetitions is thus a **non-repetitive colouring** of the natural numbers.
- We may also consider non-repetitive colourings of other mathematical structures such as:
 - ▶ the real line;
 - ▶ the d -dimensional integer lattice;
 - ▶ graphs, etc.

Nonrepetitive Colourings of the Real Line

Theorem (Rote, see Grytczuk and Śliwa 2003)

There exists a colouring f of \mathbb{R} , $f : \mathbb{R} \rightarrow \{0, 1\}$, 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 $f(x + t) \neq f(y + t)$.

- 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.

Nonrepetitive Colourings of the Real Line

- 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 $q_2 \neq q_1$.
- 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.
- But the classical Lindemann–Weierstrass theorem asserts the linear independence of algebraic powers of e over the algebraic numbers.
- This contradiction proves the theorem.

van der Waerden's Theorem

- van der Waerden's theorem asserts that if the natural numbers are partitioned into finitely many sets, then one set contains arbitrarily large arithmetic progressions.
- Rephrased as a theorem regarding words, it asserts that 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.
- We thus cannot avoid repetitions in arithmetic progressions.
- Suppose we only try to avoid repetitions in certain types of arithmetic progressions: e.g. arithmetic progressions of odd difference.

Repetitions in Arithmetic Progressions

Theorem (Carpi 1988)

For every integer $n \geq 2$, there exists an infinite word over a finite alphabet that contains no squares in any arithmetic progression except those whose difference is a multiple of n .

Folding a Piece of Paper

- Take an ordinary 8.5×11 piece of paper and fold it in half.
- Now unfold the paper and record the pattern of hills and valleys created, writing 0 for a hill and 1 for a valley.

0

- Now fold the paper twice, unfold, and record the pattern of hills and valleys.

0 0 1

- Now fold three times, unfold, and record the pattern.

0 0 1 0 0 1 1

- Now fold infinitely (!) many times. After unfolding, you get the following infinite sequence, called the **paperfolding sequence**.

0 0 1 0 0 1 1 0 0 0 1 1 0 1 1 ...

A Recursive Definition

- A **paperfolding word** $f = f_0 f_1 f_2 \cdots$ over the alphabet $\{0, 1\}$ satisfies the following recursive definition: there exists $a \in \{0, 1\}$ such that

$$f_{4n} = a, \quad n \geq 0$$

$$f_{4n+2} = \bar{a}, \quad n \geq 0$$

$(f_{2n+1})_{n \geq 0}$ is a paperfolding word.

- The **ordinary paperfolding word**

0010011000110110...

is the paperfolding word uniquely characterized by $f_{2^m-1} = 0$ for all $m \geq 0$.

Repetitions in the Paperfolding Words

- Allouche and Bousquet-Mélou (1994) showed that for any paperfolding word \mathbf{f} , if ww is a non-empty subword of \mathbf{f} , then $|w| \in \{1, 3, 5\}$.
- Let us now see how to establish the following specific case of Carpi's result: There exists an infinite word over a four letter alphabet that avoids squares in all arithmetic progressions of odd difference.
- Let $\mathbf{f} = f_0 f_1 f_2 \dots$ be any paperfolding word over $\{1, 4\}$. Define $\mathbf{v} = v_0 v_1 v_2 \dots$ by

$$\begin{aligned}v_{4n} &= 2 \\v_{4n+2} &= 3 \\v_{2n+1} &= f_{2n+1},\end{aligned}$$

for all $n \geq 0$.

Repetitions in the Paperfolding Words

- For example, if

$$\mathbf{f} = 1141144111441441 \dots$$

is the ordinary paperfolding word over $\{1, 4\}$, then

$$\mathbf{v} = 2131243121342431 \dots$$

Theorem (Kao, R., Shallit, and Silva 2008)

Let \mathbf{v} be any word obtained from a paperfolding word \mathbf{f} by the construction described above. Then the word \mathbf{v} contains no squares in any arithmetic progression of odd difference.

Repetitions in the Paperfolding Words

- Recall that any square ww in \mathbf{f} has $|w| \in \{1, 3, 5\}$.
- Recoding the even indexed positions of \mathbf{f} by mapping $1 \rightarrow 2$ and $4 \rightarrow 3$ destroys all these squares with odd periods.
- Thus \mathbf{v} is squarefree.
- The recursive, self-similar nature of \mathbf{v} ensures that when we look at subsequences of \mathbf{v} in odd difference arithmetic progressions we get nothing that wasn't already in \mathbf{v} to begin with.
- So \mathbf{v} avoids squares in any arithmetic progression of odd difference.

Higher Dimensions

- A **2-dimensional word** is a map \mathbf{w} from \mathbb{N}^2 to A , where we write $w_{m,n}$ for $\mathbf{w}(m, n)$.
- 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 $t \geq 0$

$$x_t = w_{i_1+j_1t, i_2+j_2t}.$$

Higher Dimensions

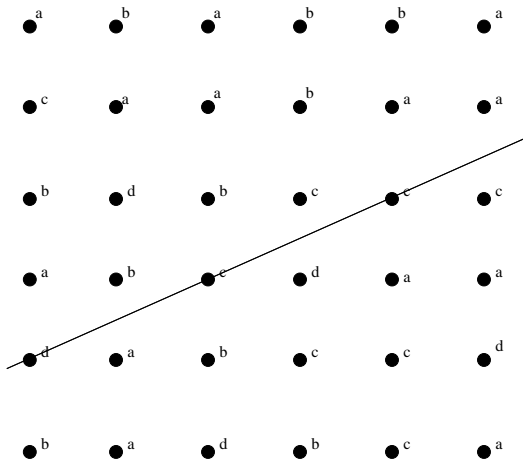


Figure: Here $\mathbf{x} = dcc \dots$ is a line.

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.

- Let $\mathbf{u} = u_0 u_1 u_2 \dots$ and $\mathbf{v} = v_0 v_1 v_2 \dots$ be any infinite words over the alphabet $A = \{1, 2, 3, 4\}$ that avoid squares in all arithmetic progressions of odd difference.
- We define \mathbf{w} over the alphabet $A \times A$ by

$$w_{m,n} = (u_m, v_n).$$

Proof of Carpi's 2D construction

- 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 the word $(u_{i_1+j_1t})_{t \geq 0}$ is an arithmetic subsequence of odd difference of \mathbf{u} and hence is squarefree.
- The line \mathbf{x} is therefore also squarefree.

Thank you!