MATH CIRCLES 2016: PATTERS IN WORDS 2

BLAKE MADILL

ABSTRACT. This is lecture two of a three part lecture series on infinite sequences and patterns in words at the University of Waterloo. It will take place on March 2, 2016.

1. Recap from last time!

Welcome back! Let's recap what we did last time.

Definition 1. *Let Σ be a non-empty finite set (that is, just a finite collection of objects), which we call an* alphabet*. We call the elements of the alphabet* letters*. By an* infinite word *w on the alphabet Σ we mean an infinite sequence of elements from Σ:*

$$w = (a_1, a_2, a_2, \dots),$$

where each $a_i \in \Sigma$. We usually denote this sequence by $w = a_1a_2a_3a_4\cdots$, which gives it a more word-like appearance. By a finite word *on the alphabet Σ we mean a finite sequence*

$$w = (a_1, a_2, \dots, a_n),$$

where each $a_i \in \Sigma$. By a word *we mean either an infinite or finite word. We shall denote the "empty word" by ε. That is, this word is so finite that it is an empty sequence of elements.*

Definition 2. *Let w be a word. We say that u is a* subword *of w if there exists words x and y such that $w = xuy$. If $x = \varepsilon$ we call u a* prefix *of w. If $y = \varepsilon$ we call u a* suffix *of w.*

We answered the following questions:

Does there exist an infinite square-free word on the alphabet $\{0, 1\}$?

Does there exist an inifinite square-free word on the alphabet $\{0, 1, 2\}$?

Does there exist an infinite cube-free word on the alphabet $\{0, 1\}$?

Recall that a square is a finite word of the form uu, where u is a nonempty finite word. Similarly, a cube is a finite word of the form uuu, where u is a nonempty finite word.

Our main tool was the Thue-Morse word:

$$t = (t_n)_{n=0}^{\infty} = 0110100110010\cdots$$

Definition 3. *Let $n \in \mathbb{N} \cup \{0\} = \{0, 1, 2, 3, \dots\}$. Then n can be written as*

$$n = a_k2^k + a_{k-1}2^{k-1} + \cdots + a_22^2 + a_12 + a_0,$$

for some $k \in \mathbb{N} \cup \{0\}$, where $a_i \in \{0, 1\}$. We call the word $a_ka_{k-1}\cdots a_1a_0$ the base-2 expansion of n. We denote the base-2 expansion of $n \in \mathbb{N} \cup \{0\}$ by $[n]_2$.

By inspection we see that

$$t_n = \begin{cases} 0 & \text{if the number of 1's in the base-2 expansion is even} \\ 1 & \text{if the number of 1's in the base-2 expansion is odd} \end{cases}$$

for $1 \leq n \leq 10$. In fact, this is how we DEFINE the Thue-Morse word, in general.

Definition 4. *An* overlap *is a word of the form cxcxc, where x is a word (possiblly $x = \varepsilon$) and c is a single letter.*

Example 1. *In the English alphabet, the word "alfalfa" is an overlap with $x = lf$ and $c = a$. Since x may be empty, any cube of a letter is an overlap!*

A rather difficult result to prove is:

Theorem 1. *The Thue-Morse word is overlap-free.*

In particular, t is an example of an infinite word over the alphabet $\{0, 1\}$ which is cube-free!

We now construct an infinite square-free word on the alphabet $\{0, 1, 2\}$.

Theorem 2. *For $n \geq 1$, let c_n be the number 1's between the nth and (n+1)st occurrence of 0 in t. Set $c = (c_n)_{n=1}^{\infty}$. Then c is an infinite square-free word over the alphabet $\{0, 1, 2\}$.*

Proof. Notice that at first glimpse, there is no reason for us to even believe that c is over the alphabet $\{0, 1, 2\}$! That is, why can there only be at most two 1's between each occurrence of 0. Well, if 01^k0 is a subword of t for $k \geq 3$ then 1^3, which is an overlap, is a subword of t. We know that t is overlap-free! Therefore $k \leq 2$ and so c is indeed in the alphabet $\{0, 1, 2\}$.

Now, why must c be square-free? Well, assume that c is not square-free. We show that this cannot happen. Since c contains a square, there exists a nonempty word x such that xx is a subword of c. Say $x = x_1x_2 \cdots x_n$, where $x_i \in \{0, 1, 2\}$. Then

$$01^{x_1}01^{x_2}0 \cdots 01^{x_n}01^{x_1}01^{x_2}0 \cdots 01^{x_n}0,$$

is a subword of t. However, this is an overlap and so no square can exist in c. Therefore c is square-free. We did it! $\square$

2. Warm-up exercises

(1) Prove that any nonempty cube is an overlap. That is, let $w = x_1x_2\cdots x_n$ be a nonempty word over the alphabet Σ, where $x_i \in \Sigma$. Show that www contains an overlap.

(2) Find t_{19} and $t_{2^{97}+2}$.

3. Subword Complexity

To understand an infinite word, it is quite often sufficient to understand its finite subwords. For example, to see that the Thue-Morse word contains a square you can just consider all the length-2 subwords of t: $\{01, 10, 11, 00\}$. Namely, 11 is a square in t. Let us start with a basic question about subwords:

How many subwords of length n does a given infinite word have?

We make the following definition.

Definition 5. *Let w be an infinite word over the alphabet Σ. For $n \in \mathbb{N} \cup \{0\}$ we denote the number of subwords of w of length n by $\rho_w(n)$. We call the ρ_w the subword complexity function of w.*

Example 2. *From above, we see that $\rho_t(2) = |\{01, 10, 11, 00\}| = 4$. However, $\rho_t(3) = |\{001, 010, 011, 100, 101, 110\}| = 6$ since 111 and 000 are (cubes) not subwords of t.*

Example 3. *For any infinite word w, $\rho_w(0) = 1$ since the only subword of length 0 is ε.*

Exercise 1. *Let $q = 1001000100000001000000000000000100\cdots$. Find $\rho_q(n)$ for $n = 0, 1, 2, \ldots, 5$.*

How big can $\rho_w(n)$ possibly be? Well, of course this depends on w and n! But we do get the following bound.

Theorem 3. *Let w be an infinite word on the alphabet Σ. Then $\rho_w(n) \leq |\Sigma|^n$ for every $n \in \mathbb{N}$.*

Exercise 2. *Let w be an infinite word on the alphabet Σ. If $\rho_w(2) = 2$ then what does w look like?*

Theorem 4. *Let w be an infinite word on Σ. Then $\rho_w(n) \leq \rho_w(n+1) \leq |\Sigma|\rho_w(n)$.*

Exercise 3. *Show that the above bounds can be achieved. That is, find a word u such that $\rho_u(n) = \rho_u(n+1)$ for every $n \in \mathbb{N}$ and find a word v over $\{0,1\}$ such that $\rho_v(n) = 2^n$ for every $n \in \mathbb{N}$.*

In general, it is quite difficult to determine the subword complexity of an infinite word. Our goal now is to develop computational tools to compute complexities. We shall split these computations into "baby steps", where we start with relatively manageable infinite words.

4. Periodic Words

In this section we discuss the subword complexity of periodic words. Recall that we usually says a mathematical object is periodic if some sort of repetition continues to happen forever. For example, you likely know that the function $\sin(x)$ is periodic with period 2π. We thus form the following natural definition.

Definition 6. *Let z be a nonempty finite word. We denote by z^ω, the infinite word $z^\omega = zzzzz\cdots$. We call such a word* periodic. *We say that an infinite word w is* ultimately periodic *if $w = uz^\omega$, where u is a possibly empty finite word and z is a non-empty finite word. Clearly, any periodic word is ultimately periodic.*

Example 4. *If*

$$w_1 = 011001100110011001100110\cdots$$

and

$$w_2 = 110110011001100110011\cdots$$

then we see that $w_1 = (0110)^\omega$ is periodic and $w_2 = 11(0110)^\omega$ is ultimately periodic.

Exercise 4. *Let w_1 and w_2 be as above. Compute $\rho_{w_1}(n)$ and $\rho_{w_2}(n)$ for every $n \in \mathbb{N}$.*

Definition 7. *Let w be a finite word. We denote the length of w (i.e. the number of letters in w) by $\ell(w)$.*

What is really going on in the above exercise is the following theorem.

Theorem 5. *Let $w = uz^\omega$ be an ultimately periodic word. Assume z cannot be written as a power of a shorter word. Moreover, assume that u is either empty or ends with a letter different from the last letter of z. (Any ultimately periodic word can be written in this way). Then $\rho_w(n) = \ell(uz)$, for all $n \geq \ell(uz)$. That is, the sequence $(\rho_w(n))_{n=1}^\infty$ is eventually constant.*

Exercise 5. *Using the above theorem, compute $\rho_w(n)$ where*

$$w = 12345345345345\cdots$$

is the infinite word over the alphabet $\{1, 2, 3, 4, 5\}$.