Notes for CS 462, January 16 2012

Today we proved Theorem 2.5.1 using a proof different from that in the text. Here it is:

We define $\overline{0} = 1$ and $\overline{1} = 0$.

Lemma 1. *Let μ be the morphism defined by $\mu(0) = 01$ and $\mu(1) = 10$. Then $\mu(t_n) = t_{2n}t_{2n+1}$.*

Proof. By induction on n.

The base case is $n = 0$. Then we have $\mu(t_0) = \mu(0) = 01 = t_0 t_1$.

For the induction step, assume the result is true for $i < n$; we prove it for $i = n$.

Now the binary expansion of $2n$ is the same as that for n, except with an extra 0 on the end. The binary expansion of $2n + 1$ is that same as that for n, except with an extra 1 on the end. Hence $t_{2n} = t_n$ and $t_{2n+1} = \overline{t_n}$.

It follows that $\mu(t_n) = t_n \, \overline{t_n} = t_{2n}t_{2n+1}$. □

Note that this lemma implies that

- If $t_r = a$ and r is even, then $t_{r+1} = \overline{a}$;
- If $t_r = a$ and r is odd, then $t_{r-1} = \overline{a}$.

Theorem 2. *The Thue-Morse word $\mathbf{t}$ is overlap-free.*

Proof. We assume that $\mathbf{t}$ has an overlap $axaxa$ beginning at position k, with $|ax| = n$. Thus it looks like

$$\mathbf{t} = t_0 t_1 \cdots t_{k-1} \overbrace{t_k}^{a} \; \overbrace{t_{k+1} \cdots t_{k+n-1}}^{x} \; \overbrace{t_{k+n}}^{a} \; \overbrace{t_{k+n+1} \cdots t_{k+2n-1}}^{x} \; \overbrace{t_{k+2n}}^{a} \cdots . \quad (1)$$

To get a contradiction we assume that (a) this overlap is smallest among all overlaps in $\mathbf{t}$ and (b) among all overlaps of this size, it appears earliest in $\mathbf{t}$. In other words, we assume that n is as small as possible and k as small as possible for this n.

There are a number of cases to consider:

Case 1: k even, n even.

Since $k + 2n$ is even, we get $t_{k+2n+1} = \overline{t_{k+2n}} = \overline{a}$.

Letting $u = \mathbf{t}[\frac{k}{2} + 1..\frac{k}{2} + \frac{n}{2} - 1]$ and $v = \mathbf{t}[\frac{k}{2} + \frac{n}{2}..\frac{k}{2} + n - 1]$, by the lemma we see $\mu(auava) = axaxa\overline{a} = \mathbf{t}[k..k + 2n + 1]$. Furthermore, since $\mu(au) = ax = \mu(av)$, we must have $u = v$. So $auaua = \mathbf{t}[\frac{k}{2}..\frac{k}{2} + n]$ is an overlap with $|au| = n/2 < n$, a contradiction.

Case 2: k odd, n even.

Since k is odd, we have $t_{k-1} = \overline{t_k} = \overline{a}$. Similarly, since $k + n$ and $k + 2n$ are both odd, we get $t_{k+n-1} = \overline{t_{k+n}} = \overline{a}$ and $t_{k+2n-1} = \overline{t_{k+2n}} = \overline{a}$. So Eq. (1) above can be rewritten as

$$\mathbf{t} = t_0 t_1 \cdots \overbrace{t_{k-1}}^{\overline{a}} \; \overbrace{t_k}^{a} \; \overbrace{t_{k+1} \cdots t_{k+n-2}}^{y} \; \overbrace{t_{k+n-1}}^{\overline{a}} \; \overbrace{t_{k+n}}^{a} \; \overbrace{t_{k+n+1} \cdots t_{k+2n-2}}^{y} \overbrace{t_{k+2n-1}}^{\overline{a}} \; \overbrace{t_{k+2n}}^{a} \cdots , \quad (2)$$

where $x = y\overline{a}$.

Now $\mathbf{t}[k-1..k+2n-1] = \overline{a}ay\overline{a}ay\overline{a}$ is an overlap of the same length as before, but occurring one place earlier than before, that is, $k-1 < k$. This is a contradiction.

Case 3: k even, n odd.

There are two subcases here: $n = 1$ and $n > 1$. If $n = 1$ then the overlap is just aaa. But from the lemma we know that $\mathbf{t} \in (01+10)^\omega$, so we cannot have three consecutive identical symbols in $\mathbf{t}$.

The second subcase is $n > 1$. The idea here is to "ping-pong" back and forth between the two copies of x, learning more and more symbols of x.

We start with the first copy.

Since $k+n-1$ is even, the lemma gives us $t_{k+n-1} = \overline{t_{k+n}} = \overline{a}$. So the last symbol of x must be $\overline{a}$, which tell us that in the second copy of x we have $t_{k+2n-1} = \overline{a}$. Now $k+2n-1$ is odd, so the lemma gives us $t_{k+2n-2} = \overline{t_{k+2n-1}} = a$. Back in the first copy of x, this gives us $t_{k+n-2} = a$.

Now $k+n-2$ is odd, so the lemma gives us $t_{k+n-3} = \overline{t_{k+n-2}} = \overline{a}$. In the second copy of x, we get $t_{k+2n-3} = \overline{a}$, too.

We continue ping-ponging back and forth, learning more and more symbols of x. You can see that the situation above continues, giving

$$x = \cdots a\,\overline{a}\,a\,\overline{a}\,a\,\overline{a}.$$

But since $n = |ax|$ is odd, this means that ax must begin and end with the same symbol. But ax begins with a and ends with $\overline{a}$, a contradiction.

Case 4: k odd, n odd.

This is just like the previous case. In the second subcase, we "ping-pong" back and forth in the same manner, except that we start with the second copy.

Thus, all cases give us a contradiction, so there is no overlap in $\mathbf{t}$.

□