Overlap-free symmetric D0L words

Anna Frid

Discrete Mathematics and Theoretical Computer Science **4**, 2001, 357–362

Overlap-Free Symmetric D0L words[†]

Anna E. Frid

Sobolev Institute of Mathematics, pr. Koptyuga, 4, 630090, Novosibirsk, Russia
`e-mail: frid@math.nsc.ru`

received Sep 27, 2000, revised Oct 1, 2001, accepted Oct 15, 2001.

A D0L word on an alphabet $\Sigma = \{0, 1, \ldots, q-1\}$ is called symmetric if it is a fixed point $w = \varphi(w)$ of a morphism $\varphi : \Sigma^* \to \Sigma^*$ defined by $\varphi(i) = \overline{t_1 + i}\,\overline{t_2 + i}\ldots\overline{t_m + i}$ for some word $t_1t_2\ldots t_m$ (equal to $\varphi(0)$) and every $i \in \Sigma$; here $\overline{a}$ means a mod q.

We prove a result conjectured by J. Shallit: if all the symbols in $\varphi(0)$ are distinct (i.e., if $t_i \neq t_j$ for $i \neq j$), then the symmetric D0L word w is overlap-free, i.e., contains no factor of the form $axaxa$ for any $x \in \Sigma^*$ and $a \in \Sigma$.

Keywords: overlap-free word, D0L word, symmetric morphism

1 Introduction

In his classical 1912 paper [15] (see also [3]), A. Thue gave the first example of an overlap-free infinite word, i. e., of a word which contains no subword of the form *axaxa* for any symbol *a* and word *x*. Thue's example is known now as the *Thue-Morse word*

$$w_{TM} = 0110100110010110100101100110100 1\ldots.$$

It was rediscovered several times, can be constructed in many alternative ways and occurs in various fields of mathematics (see the survey [1]).

The set of all overlap-free words was studied e. g. by E. D. Fife [8] who described all binary overlap-free infinite words and P. Séébold [13] who proved that the Thue-Morse word is essentially the only binary overlap-free word which is a fixed point of a morphism. Nowadays the theory of overlap-free words is a part of a more general theory of pattern avoidance [5].

J.-P. Allouche and J. Shallit [2] asked if the initial Thue's construction of an overlap-free word could be generalized and found a whole family of overlap-free infinite words built by a similar principle. This paper contains a further generalization of that result; its main theorem was conjectured by J. Shallit [14].

Let us give all the necessary definitions and state the main theorem. Consider a finite alphabet $\Sigma = \Sigma_q = \{0, 1, \ldots, q-1\}$. For an integer i, let $\bar{i}$ denote the residue of i modulo q. A morphism $\varphi : \Sigma_q^* \to \Sigma_q^*$ is called *symmetric* if for all $i \in \Sigma_q$ the equality holds

$$\varphi(i) = \overline{t_1 + i}\,\overline{t_2 + i}\ldots\overline{t_m + i},$$

[†]Supported in part by INTAS (grant 97-1001) and RFBR (grant 01-01-06018).

where $t_1t_2\dots t_m$ is an arbitrary word (equal to $\varphi(0)$). Clearly, if $t_1 = 0$, then φ has a fixed point, i. e., a (right) infinite word $w = w(\varphi)$ satisfying

$$w = \varphi(w).$$

Without loss of generality we assume that w starts with 0.

A symmetric morphism is *growing* if $|\varphi(0)| \geq 2$. We shall call a fixed point of a growing symmetric morphism a *symmetric D0L word*. For example, the Thue-Morse word w_{TM} is a fixed point of a symmetric morphism φ_{TM}:

$$\begin{cases} \varphi_{TM}(0) = 01, \\ \varphi_{TM}(1) = 10. \end{cases}$$

Symmetric D0L words include also other useful examples, such as the Dejean word [7], the Keränen word [11] and others (see Section 10.5 in [12], where in particular the term “symmetric” is introduced). Note that the class of symmetric D0L words is included in a wider class of uniform marked D0L words whose properties were studied e. g. in [10].

Note that an infinite word $w = w_1w_2\dots w_n\dots$, where $w_i \in \Sigma$, is the fixed point of the symmetric morphism φ if and only if

$$\forall k \geq 0\, \forall i \in \{1,\dots,m\}\;\; w_{km+i} = \overline{w_{k+1}+t_i}. \tag{1}$$

Indeed, this equality means that w_{km+i} is equal to the ith symbol of $\varphi(w_{k+1})$.

For every $m > 1$, let $\varphi_{m,q} : \Sigma_q^* \to \Sigma_q^*$ be the symmetric morphism defined by $\varphi_{m,q}(0) = 0\overline{1}\,\overline{2}\dots\overline{m-1}$. Note that $\varphi_{TM} = \varphi_{2,2}$. Let $w_{m,q}$ be the fixed point of $\varphi_{m,q}$ starting with 0; then the ith symbol of $w_{m,q}$ for each i can also be defined as $\overline{s_m(i)}$, where $s_m(i)$ is the sum of the digits in the base-m representation of i.

J.-P. Allouche and J. Shallit proved the following generalization of Thue’s result:

Theorem 1 ([2]) *The word $w_{m,q}$ is overlap-free if and only if $m \leq q$.*

J. Shallit conjectured also that symmetric D0L words of a much wider class are overlap-free. We turn this conjecture into

Theorem 2 *If $\varphi : \Sigma_q^* \to \Sigma_q^*$ is a growing symmetric morphism, and if all symbols occurring in $\varphi(0)$ are distinct, then the fixed point $w = w(\varphi)$ is overlap-free.*

The remaining part of the paper is devoted to the proof of this result.

2 Proof of Theorem 2

Let us start with introducing some more notions and citing a result by J. Berstel and L. Boasson [4] which we shall need later.

A *partial word* is a word on the alphabet $\Sigma \cup \{\diamond\}$, where the symbol $\diamond \notin \Sigma$ is called the *hole*‡. Each hole means an unknown symbol of Σ. A (partial) word $u = u_1\dots u_n$, where u_i are symbols, is called *(locally) p-periodic* if $u_i = u_{i+p}$ for all $i \in \{1,\dots,n-p\}$ such that $u_i \neq \diamond$ and $u_{i+p} \neq \diamond$.

The following result is a generalization of the classical Fine and Wilf’s theorem [9, 6]:

Theorem 3 ([4]) *Let u be a partial word of length n which is p-periodic and q-periodic. If u contains only one hole, and if $n \geq p+q$, then u is* $\gcd(p,q)$*-periodic.*

Now let us start the proof of Theorem 2 and first consider the easiest case:

‡ This definition slightly differs from the one given in [4].

Lemma 1 *If the symmetric morphism φ is defined by $\varphi(0) = 0\overline{c}\,\overline{2c}\ldots\overline{(m-1)c}$ for some integer $c > 0$, and if all the symbols of $\varphi(0)$ are distinct, then the fixed point w of φ is overlap-free.*

Proof. Let $S \subset \Sigma$ be the set of symbols occurring in w and q' be its cardinality. Denote $\Sigma' = \{0, \ldots, q'-1\}$ and define $h : (\Sigma')^* \to S^*$ as the symbol-to symbol morphism transforming each symbol $i \in \Sigma'$ to $h(i) = \overline{ci}$. Since the cardinalities of S and Σ' coincide, and since each symbol of S can be represented as $\overline{ci}$ for some i, h is a one-to-one mapping. But it can be easily checked that $\varphi h = h\varphi_{m,q'}$. Since $w_{m,q'} = \varphi_{m,q'}(w_{m,q'})$, we have $h(w_{m,q'}) = h(\varphi_{m,q'}(w_{m,q'})) = \varphi(h(w_{m,q'}))$, so $h(w_{m,q'})$ is the fixed point of φ; it starts with 0 since $h(0) = 0$. We see that $h(w_{m,q'}) = w$, that is, w is obtained from $w_{m,q'}$ by renaming symbols. It is overlap-free due to Theorem 1. □

A *block* is an image of symbol under a morphism. Let $S(m)$ denote the class of all symmetric morphisms on Σ of block length m with all the symbols in a block distinct. We assume also that the image of 0 always starts with 0, so that all the morphisms of $S(m)$ admit fixed points. Clearly, the class $S(m)$ is non-empty only if $m \le q$.

Our goal is to prove that, for any fixed m, all the fixed points of morphisms of $S(m)$ are overlap-free. Suppose the opposite and consider the minimal counter-example, i. e., a morphism $\varphi \in S(m)$ and its fixed point w containing an overlap $v = axaxa$ of minimal length (so that overlaps occurring in other fixed points of morphisms of $S(m)$ are not shorter). Here $a \in \Sigma$ and $x \in \Sigma^*$; we denote the length $|ax|$ by l, and thus have $|v| = 2l+1$. Let us fix an occurrence of v to w and its position with respect to blocks of φ: we consider v as a word obtained from $\varphi(s)$, where s is a factor of w, by erasing $\alpha - 1$ symbols from the left and $m - \beta$ symbols from the right, where $1 \le \alpha, \beta \le m$. So, v starts with the symbol numbered α of a block and ends with the symbol numbered β.

Claim 1 *The inequality $l \ge m$ holds.*

Proof. Suppose that $l < m$. The 1st, $(l+1)$th, and $(2l+1)$th symbols of v are equal and thus must lie in three different blocks. So, v contains a complete block. But this block must be l-periodic since v is l-periodic; hence it must contain two equal symbols since $l < m$. A contradiction. □

Claim 2 *The block length m does not divide l.*

Proof. Suppose the opposite: let $l = mk$. Then the length of the "inverse image" s of v is equal to $2k+1$. Since v is an overlap, its $(mi+1)$th symbol is equal to the $(m(i+k)+1)$th one for any $i \in \{0, \ldots, k\}$; they are symbols numbered α of respectively the $(i+1)$th and the $(i+k+1)$th blocks of $\varphi(s)$. Since the morphism φ is symmetric, each block is uniquely determined by its αth symbol, so $(i+1)$th and $(i+k+1)$th symbols of s are equal. Thus, s is an overlap in w shorter than v, a contradiction. □

For every word $u = u_1u_2\ldots u_{n+1} \in \Sigma^{n+1}$, where $u_1, \ldots u_{n+1} \in \Sigma$, let us define the word $r(u) \in \Sigma^n$ as obtained from u by subtraction of consecutive symbols:

$$r(u) = \overline{u_2 - u_1}\,\overline{u_3 - u_2}\ldots\overline{u_{n+1} - u_n}.$$

Clearly, u can be reconstructed from its first symbol u_1 and the word $r(u) = r_1 \ldots r_n$, where $r_1, \ldots, r_n \in \Sigma$:

$$u = u_1\overline{u_1 + r_1}\,\overline{u_1 + r_1 + r_2}\ldots\overline{u_1 + r_1 + \ldots + r_n}. \tag{2}$$

Let us consider the word $r(v) = r(axaxa)$. Its length is equal to $2l$, and it is l-periodic as well as v. Since φ is symmetric, the word $r(\varphi(i))$ does not depend on the symbol $i \in \Sigma$; we denote $r(\varphi(i)) = b = b_1 \ldots b_{m-1}$, where $b_1,\dots,b_{m-1}\in\Sigma$. Since v starts with the symbol number α of a block and ends with the symbol number β, we have

$$r(v)=b_\alpha\dots b_{m-1}c_1bc_2b\dots bc_nb_1\dots b_{\beta-1},$$

where $|s|=n+1$ and $c_1\dots c_n$ are symbols of Σ depending on pairs of consecutive blocks in $\varphi(s)$; if $\alpha=m$, then $r(v)$ just starts with c_1, and if $\beta=1$, $r(v)$ just ends with c_n. Let n' be the last number such that $c_{n'}$ is situated in the first occurrence of $r(axa)$ in $r(v)$. Since $r(v)$ is l-periodic, for all $i\in\{1,\dots,n'\}$ the symbol c_i is equal to the symbol of $r(v)$ situated at distance l from it. Due to Claim 2, $l\not\equiv 0 \pmod m$, and thus all these symbols are equal to $b_{l'}$, where $l\equiv l' \pmod m$. So, the word $r(axa)$ (equal to the prefix of length l of $r(v)$) is m-periodic:

$$r(axa)=b_\alpha\dots b_{m-1}(b_{l'}b)^{n'-1}b_{l'}b_1\dots b_{\gamma-1},$$

where $\gamma-\alpha\equiv l \pmod m$, $\gamma\in\{1,\dots,m\}$.

Let us consider the prefix of $r(v)$ of length $m+l$. It exists due to Claim 1 and is equal to

$$r(axa)b_\gamma\dots b_{m-1}c_{n'+1}b_1\dots b_{\gamma-1}.$$

Subsituting the unknown symbol $c_{n'+1}$ by a hole $\diamond$, we obtain a partial word

$$b_\alpha\dots b_{m-1}(b_{l'}b)^{n'}\diamond b_1\dots b_{\gamma-1},$$

which is l-periodic as well as $r(v)$. But at the same time, it is m-periodic; thus, due to Theorem 3 it is p-periodic, where $p=\gcd(l,m)$. Consequently, $b=r(\varphi(0))$ is also p-periodic: $b=(b_1\dots b_p)^{m'-1}b_1\dots b_{p-1}$, where $m'=m/p$. Let us return to $\varphi(0)$ and denote $g_1=0$, $g_k=\overline{b_1+b_2+\dots+b_{k-1}}$ for $k\in\{2,\dots,p\}$, and $c=\overline{b_1+b_2+\dots+b_p}$; due to (2), we see that $\varphi(0)$ is of the form

$$\varphi(0)=g_1\dots g_p\,\overline{g_1+c}\dots\overline{g_p+c}\dots\overline{g_1+(m'-1)c}\dots\overline{g_p+(m'-1)c}. \tag{3}$$

Here $g_1=0$ since φ has a fixed point, and $m'=m/p$. The words of the form $\overline{g_1+ic}\dots\overline{g_p+ic}$, where $i\in\{0,\dots,m'-1\}$, will be called *subblocks*. Note that for all $k\in\{1,\dots,p\}$, a subblock is uniquely determined by its kth symbol, and that w consists of consecutive subblocks.

Let w_i denote the ith symbol of the fixed point w of φ, i. e., let $w=w_1\dots w_n\dots$, where $w_i\in\Sigma$. Consider the arithmetical subsequence

$$w'=w_1w_{p+1}w_{2p+1}\dots w_{np+1}\dots.$$

Claim 3 *The word w' is the fixed point of a morphism $\varphi'\in S(m)$.*

Proof. Let us define the symmetric morphism φ' by

$$\varphi'(0)=g_1\overline{g_1+c}\dots\overline{g_1+(m'-1)c}\,g_2\overline{g_2+c}\dots\overline{g_2+(m'-1)c}\dots g_p\overline{g_p+c}\dots\overline{g_p+(m'-1)c}.$$

Since $\varphi'(0)$ is obtained from $\varphi(0)$ by permuting symbols, and all the symbols of $\varphi(0)$ are distinct, so are the symbols of $\varphi'(0)$. Since $g_1=0$, and φ' is symmetric by definition, $\varphi'\in S(m)$. So we must prove only that w' is its fixed point, i. e., that

$$\forall k\geq 0\ \forall i\in\{1,\dots,m\}\ \ w'_{km+i} \text{ is equal to the } i\text{th symbol of } \varphi'(w'_{k+1}), \tag{4}$$

where w'_k is the kth symbol of $w'=w'_1w'_2\dots w'_n\dots$.

Clearly, each $i \in \{1, \dots, m\}$ can be uniquely represented as $i = jm' + \delta$, where $j \in \{0, \dots, p-1\}$ and $\delta \in \{1, \dots, m'\}$. Since by definition of w' for all ν we have $w'_\nu = w_{p(\nu-1)+1}$, for any $k \geq 0$

$$w'_{km+i} = w'_{km+jm'+\delta} = w_{p(km+jm'+\delta-1)+1} = w_{(pk+j)m+p(\delta-1)+1}.$$

By Equality (1), $w_{(pk+j)m+p(\delta-1)+1}$ is equal to the $(p(\delta-1)+1)$th symbol of $\varphi(w_{pk+j+1})$, that is, to $\overline{(\delta-1)c + w_{pk+j+1}}$ (recall that $g_1 = 0$). In its turn, w_{pk+j+1} is the $(j+1)$th symbol of the subblock starting with $w_{pk+1} = w'_{k+1}$. It is equal to $\overline{w'_{k+1} + g_{j+1}}$, and thus, $w'_{km+i} = \overline{w'_{k+1} + (\delta-1)c + g_{j+1}}$. By the definition of φ', it is equal to the symbol numbered $jm' + \delta = i$ of $\varphi'(w'_{k+1})$. We have proved (4) and Claim 3. □

Claim 4 *The word w' contains an overlap of length $2l'+1$, where $l' = l/p$.*

Proof. Let our occurrence of the overlap v to w start with the kth symbol of a subblock, i. e., let $\alpha \equiv k \pmod p$, where $k \in \{1, \dots, p\}$. It means that $v = w_{jp+k} w_{jp+k+1} \dots w_{(j+2l')p+k}$ for some $j \geq 0$; since v is an overlap, $w_{(j+\nu)p+1} = w_{(j+\nu+l')p+1}$ for all $\nu \in \{1, \dots, l'\}$. But we have also $w_{jp+k} = w_{(j+l')p+k}$, and since a subblock is uniquely determined by its kth symbol, $w_{jp+1} = w_{(j+l')p+1}$. So, the word $w_{jp+1} w_{(j+1)p+1} \dots w_{(j+2l')p+1}$ is l'-periodic, and it is the needed overlap in w'. □

As it follows from Claims 3 and 4, we have found a fixed point of a morphism of $S(m)$ containing an overlap of length $l' = l/p$. But if $p > 1$, this contradicts to the minimality of our counter-example. On the other hand, if $p = 1$, then it follows from (3) that

$$\varphi(0) = 0\overline{c}\,\overline{2c} \dots \overline{(m-1)c}.$$

But a fixed point of such a morphism cannot be a counter-example according to Lemma 1. A contradiction. Theorem 2 is proved. □

Acknowledgements

I thank J. Shallit for introducing me to the problem, J.-P. Allouche for useful comments, and the referee for correcting the proof of Claim 1.

References

[1] J.-P. Allouche and J. Shallit. The ubiquitous Prouhet-Thue-Morse sequence. In C. Ding, T. Helleseth and H. Niederreiter, eds., *Sequences and Their Applications, Proceedings of SETA'98*, Springer-Verlag, 1999, 1–16.

[2] J.-P. Allouche and J. Shallit. Sums of digits, overlaps, and palindromes. *Discr. Math. Theoret. Comput. Sci.* **4** (2000), 1–10.

[3] J. Berstel. Axel Thue's work on repetitions in words. In P. Leroux and C. Reutenauer, eds., *Séries Formelles et Combinatoire Algébrique*, no. 11 in Publications du LACIM, Université du Québec à Montréal, 1992, 65–80.

[4] J. Berstel and L. Boasson. Partial words and a theorem of Fine and Wilf. *Theoret. Comput. Sci.* **218** (1999), no. 1, 135–141.

[5] J. Cassaigne. Unavoidable patterns. In M. Lothaire, *Algebraic Combinatorics on Words*, Cambridge University Press, to appear.

[6] C. Choffrut and J. Karhumäki. Combinatorics on words. In G. Rozenberg and A. Salomaa, eds., *Handbook of Formal Languages*, v. 1, chapter 6. Springer-Verlag, 1997.

[7] F. Dejean. Sur un théorème de Thue. *J. Combin. Theory. Ser. A.* **13** (1972), 25–36.

[8] E. D. Fife. Binary sequences which contain no *BBb*. *Trans. Amer. Math. Soc.* **261** (1980), 115–136.

[9] N. J. Fine and H. S. Wilf. Uniqueness theorem for periodic functions. *Proc. Amer. Math. Soc.* **16** (1965), 109–114.

[10] A. Frid. Applying a uniform marked morphism to a word. *Discr. Math. Theoret. Comput. Sci.* **3** (1999), 125–140.

[11] V. Keränen. Abelian squares are avoidable on 4 letters. In *Automata, Languages and Programming, Proceedings of ICALP'92* Lecture Notes in Comput. Sci.; V. 700, Berlin: Springer, 1992, 41–52.

[12] G. Lallement. *Semigroups and Combinatorial Applications*, Wiley, 1979.

[13] P. Séébold. Sequences generated by infinitely iterated morphisms. *Discrete Appl. Math.* **11** (1985), 255–264.

[14] J. Shallit, private communication.

[15] A. Thue. Über die gegenseitige Lage gleicher Teile gewisser Zeichenreihen. *Norske vid. Selsk. Skr. Mat. Nat. Kl.* **1** (1912), 1–67. Reprinted in *Selected Mathematical Papers of Axel Thue*, T. Nagell, ed., Universitetsforlaget, Oslo, 1977, 413–478.