On Arithmetic Progressions in the Generalized Thue-Morse Word

Olga G. Parshina(✉)

Novosibirsk State University, Pirogova Str. 2, 630090 Novosibirsk, Russia
parolja@gmail.com
http://www.nsu.ru/

Abstract. We consider the generalized Thue-Morse word on the alphabet $\Sigma = \{0, 1, 2\}$. The object of the research is arithmetic progressions in this word, which are sequences consisting of identical symbols. Let $A(d)$ be a function equal to the maximum length of an arithmetic progression with the difference d in the generalized Thue-Morse word. If n, d are positive integers and d is less than 3^n, then the upper bound for $A(d)$ is $3^n + 6$ and is attained with the difference $d = 3^n - 1$.

Keywords: Thue-Morse word · Arithmetic complexity · Arithmetic progression · Automatic word

1 Introduction

Arithmetic progressions in infinite words have been studied since the classical papers of Van der Waerden [1] and Szemerédi [2]. Results described in these papers say in particular, that we cannot constrain the length of a homogeneous arithmetic progression by a constant. For bounded differences and fixed word the question about the maximum possible length of an arithmetic progression depends on the structure of the word and the answer to it may be non-trivial. In some cases (e.g. Toeplitz words) the length of an arithmetic progression may be infinite. We are interested in ones having long progression with large differences only. Our research shows that such correlation between length of a progression and the value of its difference does not occur among automatic words.

The reported result refers to the Thue-Morse word, which was chosen because it is famous and the defining morphism is simple. More details on the Thue-Morse sequence and its applications can be found e.g. in [3]. It is known that the set of arithmetic progressions of the Thue-Morse word contains all binary words [4]. In [5] the results of [4] are generalized to symmetric D0L words including the generalized Thue-Morse word.

The original Thue-Morse word was studied earlier [6]. Here we considered its generalization on the alphabet $\Sigma_3 = \{0, 1, 2\}$, it was chosen to simplify the presentation, although the technique presented in the paper is easy to extend to Σ_q for an arbitrary prime q.

F. Manea and D. Nowotka (Eds.): WORDS 2015, LNCS 9304, pp. 191–196, 2015.
DOI: 10.1007/978-3-319-23660-5_16

We consider the maximal length of a homogeneous arithmetic progression of difference d. It is proven that the length grows quite rapidly *(see Theorem 1)* and reaches its maximums at points of specific kind. Since the Thue-Morse word over Σ_3 is easy to define by ternary representation of natural numbers, the proof of the theorem uses mostly this representation and arithmetic operations modulo 3. The proof has three stages: at first we reject the most part of the set of differences, which cannot provide the maximum of the length of a progression, then we present concrete values of the difference, of starting symbols and corresponding values of the length which are maxima of this function. The final part is exhaustion of the rest of initial symbols for this difference, there we give an upper bound of length for each type of starting number.

2 Preliminaries

Let $\Sigma_q = \{0, 1, ..., q-1\}$ be an alphabet. Consider a function $R_q : \mathbb{N} \to \Sigma_q$ which for every natural x gives its base-q expansion. The length of this word is denoted by $|R_q(x)|$. Also let $r_q(x)$ be the sum modulo q of the digits of x in its q-ary expansion. In other words $x = \sum\limits_{i=0}^{n-1} x_i \cdot q^i$, $R_q(x) = x_{n-1} \cdots x_1 x_0$ and $r_q(x) = \sum\limits_{i=0}^{n-1} x_i \bmod q$. The generalized Thue-Morse word is defined as: $w_{TM} = w_0 w_1 w_2 w_3 \cdots$, where $w_i = r_q(i) \in \Sigma$.

An arithmetic progression of length k with the starting number c and the difference d in arbitrary infinite word $v = v_0 v_1 v_2 v_3 \cdots$ is a sequence $v_c v_{c+d} v_{c+2d} \cdots v_{c+(k-1)d}$. We are interested in homogeneous progressions, i.e., in situations when $v_{c+id} = \alpha$ for each $i = 0, 1, ..., k-1$ and $\alpha \in \Sigma$.

Consider a function $A(c, d)$ which outputs the length of an arithmetic progression with starting symbol v_c and difference d for positive integers c and d. The function $A(d) = \max_c A(c, d)$ gives the length of the maximal arithmetic progression of difference d.

As we mentioned before, here we consider the alphabet Σ_3 and the generalized Thue-Morse word:

$$w_{TM} = 012120201120201012201012120 \cdots$$

Since the word is cube-free, $A(1) = 2$. We can see that $A(0, 5) = 2$, $A(2, 2) = A(20, 2) = 3$ and we state that $A(2) = 3$.

3 Main Result

Here we formulate the main theorem, the remainder of the paper is devoted to its proof.

Theorem 1. *For all numbers $n \geq 1$ the following holds:*

$$\max_{d<3^n} A(d) = \begin{cases} 3^n + 6, & n \equiv 0 \, mod \, 3, \\ 3^n, & otherwise. \end{cases}$$

As we will see later, the function $A(d)$ reaches its maxima with differences of the form $3^n - 1$ for natural n. Let us prove that if $d \neq 3^n - 1$, the value of $A(d)$ will be not more than 3^n for fixed n.

3.1 Case of $d \neq 3^n - 1$

We need to prove that $A(d) \leq 3^n$.

At first we note that subsequences of the w_{TM} which are composed of letters with indexes having the same remainder of the division by three are equivalent to the word, so we do not need to consider differences which are divisible by three.

Every number may be represented in such a way: $c = y \cdot 3^n + x$, where x, y are arbitrary positive integers, $x < 3^n$. We will call x a suffix of c.

Consider the set $\mathbf{X} = \{0, 1, 2, ..., 3^n - 1\}$, $|\mathbf{X}| = 3^n$. Each difference d and suffix x belong to $\mathbf{X}$ and d is prime to $|\mathbf{X}|$. So the $\mathbf{X}$ is an additive cyclic group and d is a generator of $\mathbf{X}$, thus for every $x \in \mathbf{X}$ the set $\{x + i \cdot d\}_{i=0}^{3^n-1}$ is precisely $\mathbf{X}$. We will prove the statement if for each $d \neq 3^n - 1$ we find an element $x \in \mathbf{X}$ with following properties:

(a) $x + d < 3^n$;
(b) $r_3(x + d) \neq r_3(x)$.

Indeed, consider the starting number of the form c. Because of (a), $c + d = y \cdot 3^n + (x + d)$. Hence, $r_3(c) = r_3(y) + r_3(x) \bmod 3$, $r_3(c + d) = r_3(y) + r_3(x + d) \bmod 3$. Because of (b), $r_3(c + d) \neq r_3(c)$, and the homogeneity of a progression will hold by at most 3^n steps.

If $r_3(d) \neq 0$, then a suitable x is a zero. In another case we use the inequation $d \neq 3^n - 1$ which means that $R_3(d) = d_{n-1} \cdots d_1 d_0$ has at least one letter $d_j, j \in \{0, 1, ..., n - 1\}$: $d_j \neq 2$. There are two possibilities:

1. $\exists j : d_j \neq 2, d_{j-1} = 2$, in this case $x = 3^{j-1}$.
2. $\forall j \ (d_j \neq 2 \Rightarrow \forall k < j \ d_k \neq 2)$.

If $j > 1$, then, as soon as d is not divisible by three, $d_1 d_0$ may be equal to 01 or 11. In the first case we choose $x = 2$, in the second $x = 1$. But we can not find x satisfying (a) and (b) then $j = 1$, i.e., then $R_3(d) = \underbrace{2 \cdots 2}_{n-1} 1$. So we find three suffixes x with properties (i) $x + d$, $x + 2d > 3^n$ and (ii) $r_3(x)$ is equal to the suffix of $x + d$ and $x + 2d$. Let us explain the necessity of these properties.

Consider a number c with such a suffix x, $r_3(c) = r_3(y) + r_3(x) \bmod 3$. The condition (i) gives to us equalities $r_3(c + d) = r_3(y + 1) + r_3(x + d) \bmod 3$ and $r_3(c + 2d) = r_3(y + 2) + r_3(x + 2d) \bmod 3$. With (ii) we need $r_3(y) = r_3(y + 1) = r_3(y + 2)$ for saving the homogeneity of a progression, but there is no number with such property. So we know that by at most three steps the value of r_3 will change. Suitable suffixes are $6, 15, 24$, and all described above guarantee that the arithmetic progression will be not longer than 3^n.

Of course, there are many suffixes satisfy (a),(b) or (i),(ii), but it is not necessary to consider all of them.

3.2 Case of $d = 3^n - 1$

Here we use the notation $R_3(x) = X = x_{n-1} \cdots x_1 x_0$, $R_3(y) = Y = y_{n-1} \cdots y_1 y_0$.

Lemma 1. *Let $d = 3^n - 1$, $c = z \cdot 3^{2n} + y \cdot 3^n + x$, where $x + y = 3^n - 1$, z is a non-negative integer, then*

$$\max_z A(c,d) = \begin{cases} 3^n + 3 - y, & n \equiv 0 \, mod \, 3, \\ 3^n - y, & otherwise. \end{cases}$$

Proof. Because of the value of d we can regard the action $c + d$ as two simultaneous actions: $x - 1$ and $y + 1$. Thus while $y \leq 3^n - 1$ the sum of digits in YX (the concatenation of ternary notation of y and x) equals $2n$. This property provides us with the arithmetic progression of length $3^n - y$.

Let $y = 3^n - 1$, hence $x = 0$. If we add the difference d to such a number, the sum of digits in result's ternary notation will be equal to $4n$. To save the required property of members of the progression we need $2n \equiv 4n$ mod 3, i.e., $n \equiv 0$ mod 3. After next addition of d, z increases to $z + 1$, y becomes 0 and $x = 3^n - 2$. We may define z arbitrary for holding the homogeneity of the progression (for example if $r_3(z) = 1$ we need $r_3(z + 1) = 2$, in this case z may be equal to 1). Now $r_3(y) + r_3(x) = 2n + 1$ mod 3. Let us add the difference once more: $Y = \underbrace{0 \cdots 0}_{n-1} 1$, $X = \underbrace{2 \cdots 2}_{n-1} 0$ and homogeneity holds. Next addition of d changes the value of $r_3(y) + r_3(x)$ because $Y = \underbrace{0 \cdots 0}_{n-1} 2$ and $X = \underbrace{2 \cdots 2}_{n-2} 12$.

So if $3|n$, the length of an arithmetic progression equals $3^n + 3 - y$ and equals $3^n - y$ otherwise.

Lemma 2. *Let $n \equiv 0$ mod 3, $d = 3^n - 1$, $c = z \cdot 3^{2n} + y \cdot 3^n + x$, $y = 3^n - 2$, $x = 2$, z is arbitrary non-negative, then* $\max_z A(c,d) = 3^n + 6$.

Proof. Let us add to the c the difference d three times and look at the result:

1. $y = 3^n - 1$, $x = 1$;
2. $z \to z + 1$, $y = 0$, $x = 0$;
3. $y = 0$, $x = 3^n - 1$.

Sums of digits in ternary notation of described numbers are the same for a suitable z and coincide with the similar sum of initial c. After these steps we get into conditions of *Lemma* 1 with $y = 0$ which provide us with an arithmetical progression of length $3^n + 3$. Now we subtract d from the initial c to make sure that $r_3(c - d) \neq r_3(c)$ and we cannot get an arithmetical progression longer. Indeed, $c - d = z \cdot 3^{2n} + (3^n - 3) \cdot 3^n + 3$ and the sum of digits in its ternary notation is $2n - 1$, while in c it is $2n + 1$.

So we find starting numbers for the difference $d = 3^n - 1$ which provide us with arithmetical progressions of the length mentioned in the Theorem.

Now let us prove that we can not get an arithmetical progression with difference $d = 3^n - 1$ longer than in the statement of the Theorem.

Here we represent a starting number of progression c like this: $c = y \cdot 3^n + x$, $x < 3^n$.

The case of initial number c with $x_j + y_j = 2$, $j = 0, 1, ..., n-1$ is described in *Lemma* 1. In another case there is at least one index j: $x_j + y_j \neq 2$. We choose j which is the minimal. There are six possibilities of values (y_j, x_j): $(0,0), (0,1), (1,0), (1,2), (2,1), (2,2)$.

For c of each type we find numbers k and h: $r_3(c + k \cdot d) \neq r_3(c + h \cdot d)$.

We need two more parameters: l and m which are defined from these notations: $Y = y_{s-1} \cdots y_{j+l+1} 2 \cdots 2 y_j \cdots y_0$, $X = x_{n-1} \cdots x_{j+m+1} 0 \cdots 0 x_j \cdots x_0$. Of course m and l may be equal to zero, $y_{j+l+1} \neq 2$ and $x_{j+m+1} \neq 0$.

We will act such a way: we add $3^{j+1} \cdot d$ to the c, and the block of twos in Y transforms to the block of zeros, the block of zeros in X transforms to the block of twos. In cases $(y_j, x_j) \in \{(0,0), (1,0), (2,1), (2,2)\}$ after next $3^j \cdot d$ addition we get a number with the sum of digits different from the previous one. So suitable values of k and h are 3^{j+1} and $4 \cdot 3^j$. In cases $(y_j, x_j) \in \{(0,1), (1,2)\}$ to change the sum, we need to add $3^j \cdot d$ once more, and suitable k and h are $4 \cdot 3^j$ and $5 \cdot 3^j$. Let us consider an example for $(y_j, x_j) = (0,1)$.
Here $n = 7$, $m = l = 3$, $j = 1$, $d = 2186$, $R_3(d) = 2222222$.

number	$\mathbf{R_3}$	$\mathbf{r_3}$
$c = 2640685$	11222011100011	1
	+ 222222200	
$c + 3^{j+1} \cdot d$	12000011022211	1
	+ 22222220	
$c + 4 \cdot 3^j \cdot d$	12000111022201	1
	+ 22222220	
$c + 5 \cdot 3^j \cdot d$	12000211022121	0

But these values of k, h satisfy the Theorem if and only $j < n-1$, the case $j = n-1$ needs a special consideration.

We will act the following way: we add the $x \cdot d$ to the c and nullify x by that, then add d necessary number of times. So the worst case is then $X = x_{n-1} 2 \cdots 2$, $Y = y_{n-1} 0 \cdots 0$. Here is the table of values k, h.

(y_{n-1}, x_{n-1})	(0,0), (1,0)		(0,1), (2,1)		(2,2), (1,2)	
parameters	k	h	k	h	k	h
$3 \nmid n$	$3^{n-1} - 1$	3^{n-1}	$2 \cdot 3^{n-1} - 1$	$2 \cdot 3^{n-1}$	$3^n - 1$	3^n
$3 \mid n$	3^{n-1}	$3^{n-1} + 1$	$2 \cdot 3^{n-1}$	$2 \cdot 3^{n-1} + 1$	3^n	$3^n + 1$

One can see that these values satisfy the Theorem.

We have considered all the possible cases and thus completed the proof.

4 Conclusion

This result helps to better understand the structure of the well-known Thue-Morse word and its generalization. The result and the technique of the proof may be generalized on the Thue-Morse word over an arbitrary alphabet of prime cardinality q, the upper bound for the length of a progression in this case is $q^n + 2 \cdot q$.

References

1. Van der Waerden, B.L.: Beweis einer Baudetschen Vermutung. Nieuw Arch. Wisk. **15**, 212–216 (1927)
2. Szemerédi, E.: On sets of integers containing no k elements in arithmetic progression. Collection of articles in memory of Juriĭ Vladimirović Linnik. Acta Arith. **27**, 199–245 (1975)
3. Allouche, J.-P., Shallit, J.: The ubiquitous Prouhet-Thue-Morse sequence. In: Ding, C., Helleseth, T., Niederreiter, H. (eds.) Proceedings of SETA 1998, pp. 1–16. Springer, New York (1999)
4. Avgustinovich, S. V., Fon-der-Flaass, D. G., Frid, A. E.: Arithmetical complexity of infinite words. In: Proceedings of Words, Languages and Combinatorics III (2000), pp. 51–62 . World Scientific, Singapore (2003)
5. Frid, A.E.: Arithmetical complexity of symmetric D0L words. Theoret. Comput. Sci. **306**(1–3), 535–542 (2003)
6. Паршина, О.Г.: Однородные арифметические прогрессии в слове Туэ-Морса. Новосибирск: Материалы 53-й Международной научной студенческой конференции МНСК-2015, Математика, С. 218. (2015)