

On the maximum order complexity of subsequences of the Thue-Morse and Rudin-Shapiro sequence along squares

Zhimin Sun¹ and Arne Winterhof²

¹ Faculty of Mathematics and Statistics,
Hubei Key Laboratory of Applied Mathematics,
Hubei University, Wuhan, 430062, China
e-mail: zmsun@hubu.edu.cn

² Johann Radon Institute for Computational and Applied Mathematics,
Austrian Academy of Sciences,
Altenberger Straße 69, A-4040 Linz, Austria
e-mail: arne.winterhof@oeaw.ac.at

Abstract

Automatic sequences such as the Thue-Morse sequence and the Rudin-Shapiro sequence are highly predictable and thus not suitable in cryptography. In particular, they have small expansion complexity. However, they still have a large maximum order complexity.

Certain subsequences of automatic sequences are not automatic anymore and may be attractive candidates for applications in cryptography. In this paper we show that subsequences along the squares of certain pattern sequences including the Thue-Morse sequence and the Rudin-Shapiro sequence have also large maximum order complexity but do not suffer a small expansion complexity anymore.

Keywords. Thue-Morse sequence, Rudin-Shapiro sequence, automatic sequence, maximum order complexity, measures of pseudorandomness

MSC 2010. 11B85, 11K45

1 Introduction

For a positive integer N and a sequence $\mathcal{S} = (s_i)_{i=0}^\infty$ over the finite field \mathbb{F}_2 of two elements with $(s_0, \dots, s_{N-2}) \neq (a, \dots, a)$, $a \in \{0, 1\}$, the N th maximum order complexity $M(\mathcal{S}, N)$ (or N th nonlinear complexity) is the smallest positive integer M such that there is a polynomial $f(x_1, \dots, x_M) \in \mathbb{F}_2[x_1, \dots, x_M]$ with

$$s_{i+M} = f(s_i, s_{i+1}, \dots, s_{i+M-1}), \quad 0 \leq i \leq N - M - 1,$$

see [8, 9]. If $s_i = a$ for $i = 0, \dots, N - 2$, we define $M(\mathcal{S}, N) = 0$ if $s_{N-1} = a$ and $M(\mathcal{S}, N) = N - 1$ if $s_{N-1} \neq a$. A sequence with small N th maximum order complexity (for sufficiently large N) is predictable and thus unsuitable in cryptography. However, there are predictable sequences with large N th maximum order complexity and further quality measures for cryptographic sequences have to be studied.

Diem [4] introduced the expansion complexity of the sequence \mathcal{S} as follows. We define the *generating function* $G(x)$ of \mathcal{S} by

$$G(x) = \sum_{i=0}^{\infty} s_i x^i,$$

viewed as a formal power series over \mathbb{F}_2 . (Note the change by the factor x compared to the definition in [4].) For a positive integer N , the N th expansion complexity $E(\mathcal{S}, N)$ is 0 if $s_0 = \dots = s_{N-1} = 0$ and otherwise the least total degree of any nonzero polynomial $h(x, y) \in \mathbb{F}_2[x, y]$ with

$$h(x, G(x)) \equiv 0 \pmod{x^N}.$$

A sequence with small N th expansion complexity is predictable.

Automatic sequences such as the Thue-Morse sequence and the Rudin-Shapiro sequence have a large N th maximum order complexity of order of magnitude N , see [14]. However, by Christol's theorem [3] they are characterized by

$$\sup_{N \geq 1} E(\mathcal{S}, N) < \infty,$$

see also [1, Theorem 12.2.5].

For example, the *Thue-Morse sequence* $\mathcal{T} = (t_i)_{i=0}^\infty$ over \mathbb{F}_2 is defined by

$$t_i = \begin{cases} t_{i/2} & \text{if } i \text{ is even,} \\ t_{(i-1)/2} + 1 & \text{if } i \text{ is odd,} \end{cases} \quad i = 1, 2, \dots$$

with initial value $t_0 = 0$. An explicit formula for $M(\mathcal{T}, N)$ is given in [14, Theorem 1]. In particular, it satisfies

$$M(\mathcal{T}, N) \geq \frac{N}{5} + 1, \quad N \geq 4.$$

However, taking

$$h(x, y) = (x + 1)^3 y^2 + (x + 1)^2 y + x,$$

its generating function $G(x)$ satisfies $h(x, G(x)) = 0$ and thus

$$E(\mathcal{T}, N) \leq 5, \quad N = 1, 2, \dots$$

Hence, despite of a large N th maximum order complexity, the Thue-Morse sequence is highly predictable. Other indicators for its predictability are a linear subword complexity, see [1, Exercise 10.11.10] or [2, 10], and a large correlation measure of order 2 [11].

Subsequences of automatic sequences may be not automatic anymore and can look much more random. For example, the subsequence of the Thue-Morse sequence along squares $\mathcal{T}' = (t_{i^2})_{i=0}^{\infty}$ is not automatic by [1, Theorem 6.10.1], that is,

$$\sup_{N \geq 1} E(\mathcal{T}', N) = \infty,$$

it has the largest possible subword complexity [12] and is even normal [5]. In Section 2 we prove a lower bound on $M(\mathcal{T}', N)$ of order of magnitude $N^{1/2}$, which indicates that \mathcal{T}' is rather unpredictable.

More generally, for a positive integer k we study subsequences along the squares of the *pattern sequences* $\mathcal{P}_k = (p_n)_{n=0}^{\infty}$ over \mathbb{F}_2 defined by

$$p_n \equiv s_k(n) \pmod{2},$$

where $P_k = 11 \dots 1 \in \mathbb{F}_2^k$ is the all 1 pattern of length k and $s_k(n)$ is the number of occurrences of P_k in the binary representation of n . For $k = 1$ we get the Thue-Morse sequence and for $k = 2$ the *Rudin-Shapiro sequence*. In Section 3 for $k \geq 2$ we prove a lower bound on the maximum order complexity of \mathcal{P}'_k of order of magnitude $N^{1/2}$. Note that the proof is slightly different than for $k = 1$.

We finish this paper with a list of open problems in Section 4.

2 The Thue-Morse sequence along squares

Theorem 1. *Let $\mathcal{T}' = (t_{i^2})_{i=0}^{\infty}$ be the subsequence of the Thue-Morse sequence along squares. Then the N th maximum order complexity of \mathcal{T}' satisfies*

$$M(\mathcal{T}', N) \geq \sqrt{\frac{2N}{5}}, \quad N \geq 21.$$

Proof. Let $\ell \geq 2$ be the integer defined by

$$5 \cdot 2^\ell < N \leq 5 \cdot 2^{\ell+1} \tag{1}$$

and note that the Thue-Morse sequence satisfies

$$t_n \equiv s_1(n) \pmod{2}, \quad n = 0, 1, \dots,$$

where $s_1(n)$ denotes the number of $n_i = 1$ in the binary expansion of n , that is,

$$n = \sum_{i=0}^{\infty} n_i 2^i \quad \text{with} \quad n_i \in \{0, 1\}.$$

(Note that only finitely many n_i are nonzero.)

For $i = 0, 1, \dots, \left\lfloor \sqrt{2^{\ell+2}} - 1 \right\rfloor$ we have (since $\ell \geq 2$)

$$t_{(i+2^{\ell+1})^2} \equiv s_1(i^2 + i2^{\ell+2} + 2^{2\ell+2}) \equiv s_1(i^2) + s_1(i) + 1 \pmod{2}$$

and

$$t_{(i+2^{\ell+2})^2} \equiv s_1(i^2 + i2^{\ell+3} + 2^{2\ell+4}) \equiv s_1(i^2) + s_1(i) + 1 \pmod{2}$$

and thus

$$t_{(i+2^{\ell+1})^2} = t_{(i+2^{\ell+2})^2}, \quad i = 0, 1, \dots, \left\lfloor \sqrt{2^{\ell+2}} - 1 \right\rfloor. \tag{2}$$

Moreover, we have

$$s_1((2^\ell + 2^{\ell+1})^2) = s_1(2^{2\ell} + 2^{2\ell+3}) = 2$$

but

$$s_1((2^\ell + 2^{\ell+2})^2) = s_1(2^{2\ell} + 2^{2\ell+3} + 2^{2\ell+4}) = 3.$$

Hence,

$$t_{(2^\ell + 2^{\ell+1})^2} \neq t_{(2^\ell + 2^{\ell+2})^2}. \tag{3}$$

Now assume

$$M = M(\mathcal{T}', N) \leq \left\lfloor \sqrt{2^{\ell+2} - 1} \right\rfloor + 1,$$

that is, there is a polynomial $f(x_1, \dots, x_M)$ in M variables with

$$t_{(j+M)^2} = f(t_{j^2}, \dots, t_{(j+M-1)^2}), \quad j = 0, 1, \dots, N - M - 1. \quad (4)$$

Note that for $0 \leq k \leq N - M - 1$ the values of $t_{(k+M)^2}, t_{(k+M+1)^2}, \dots, t_{(N-1)^2}$ are uniquely determined by the values of $t_{k^2}, \dots, t_{(k+M-1)^2}$ and by applying successively the recurrence (4) for $j = k, \dots, N - M - 1$. In particular, if

$$(t_{k_1^2}, \dots, t_{(k_1+M-1)^2}) = (t_{k_2^2}, \dots, t_{(k_2+M-1)^2})$$

for some k_1 and k_2 with $0 \leq k_1 < k_2 \leq N - M - 1$, we get also

$$(t_{(k_1+M)^2}, \dots, t_{(k_1+N-k_2-1)^2}) = (t_{(k_2+M)^2}, \dots, t_{(N-1)^2}).$$

Taking $k_1 = 2^{\ell+1}$ and $k_2 = 2^{\ell+2}$ we get from (2):

$$(t_{(2^{\ell+1}+M)^2}, \dots, t_{(N-2^{\ell+1}-1)^2}) = (t_{(2^{\ell+2}+M)^2}, \dots, t_{(N-1)^2}).$$

Since $N - 1 \geq 2^\ell + 2^{\ell+2}$ (by the lower bound in (1)) and $M \leq 2^\ell$ this includes

$$t_{(2^\ell+2^{\ell+1})^2} = t_{(2^\ell+2^{\ell+2})^2}$$

which contradicts (3) and we get (using the upper bound in (1))

$$M(\mathcal{T}', N) \geq \left\lfloor \sqrt{2^{\ell+2} - 1} \right\rfloor + 2 \geq \sqrt{\frac{2N}{5}},$$

which completes the proof. \square

Remarks.

1. Since the N th linear complexity is lower bounded by the N th maximum order complexity, this result shows that an attack via the Berlekamp-Massey algorithm fails for sufficiently large N .
2. Our experimental results support the conjecture that Theorem 1 is (up to the constant) best possible, that is, $M(\mathcal{T}', N)$ is of order of magnitude \sqrt{N} .

3. Our lower bound is strong enough to guarantee that \mathcal{T}' is not vulnerable under any known algorithm that calculates a shortest recurrence relation. This is even true if we consider the simpler problem of finding a shortest linear recurrence, see Remark 1. However, it does not guarantee that there is no other efficient way to attack our sequence although we are not aware of any such possible attack. Hence it is still important to study further features of this sequence such as its expansion complexity or its correlation measure of order k . For further discussions about predictability and measures of pseudorandomness we refer to the surveys [7, 13].
4. Further experiments indicate that also the N th expansion complexity of \mathcal{T}' is quite large, that is, we believe that its order of magnitude is close to the best possible order $N^{1/2}$. (We have $E(\mathcal{S}, N) \leq \sqrt{2N}$ for any sequence \mathcal{S} by [6, Theorem 1].)

3 Pattern sequences along squares for $k \geq 2$

Theorem 2. *For $k \geq 2$ let $\mathcal{P}'_k = (p_{i^2})_{i=0}^\infty$ be the subsequence of the pattern sequence \mathcal{P}_k along the squares. Then the N th maximum order complexity of \mathcal{P}'_k satisfies*

$$M(\mathcal{P}'_k, N) \geq \left(\frac{N}{8}\right)^{1/2}, \quad N \geq 2^{2k+2}.$$

Proof. Let ℓ be the positive integer defined by

$$2^{2k+\ell+1} \leq N < 2^{2k+\ell+2}. \quad (5)$$

For $0 \leq i \leq \left\lfloor \sqrt{2^{\ell+2k-1}} - 1 \right\rfloor$ we have

$$s_k((i + 2^{\ell+2k-1})^2) = s_k(i^2 + i2^{\ell+2k} + 2^{2\ell+4k-2}) = s_k(i^2) + s_k(i)$$

as well as

$$s_k((i + 2^{\ell+2k})^2) = s_k(i^2) + s_k(i).$$

Thus

$$p_{(i+2^{\ell+2k-1})^2} = p_{(i+2^{\ell+2k})^2}, \quad i = 0, 1, \dots, \left\lfloor \sqrt{2^{\ell+2k-1}} - 1 \right\rfloor. \quad (6)$$

For $k = 2$ we have

$$s_2((2^{\ell+2} + 2^{\ell+3})^2) = s_2(1 + 2^3) = 0$$

but

$$s_2((2^{\ell+2} + 2^{\ell+4})^2) = s_2(1 + 2^3 + 2^4) = 1$$

and thus

$$p_{(2^{\ell+2}+2^{\ell+3})^2} \neq p_{(2^{\ell+2}+2^{\ell+4})^2}. \quad (7)$$

For even $k > 2$ we have

$$s_k(((2^k - 1)2^\ell + 2^{2k-1+\ell})^2) = s_k(1 + 2^{3k} - 2^{k+1} + 2^{4k-2}) = k \equiv 0 \pmod{2}$$

but

$$s_k(((2^k - 1)2^\ell + 2^{2k+\ell})^2) = s_k(1 + (2^{2k} - 2^{k+1}) + (2^{3k+1} - 2^{2k+1}) + 2^{4k}) = 1$$

and thus

$$p_{((2^k-1)2^\ell+2^{2k-1+\ell})^2} \neq p_{((2^k-1)2^\ell+2^{2k+\ell})^2}. \quad (8)$$

For $k = 3$ we have

$$s_3((7 \cdot 2^{\ell+3} + 2^{\ell+5})^2) = s_3(1 + 2^7 - 2^3) = 2 \equiv 0 \pmod{2}$$

but

$$s_3((7 \cdot 2^{\ell+3} + 2^{\ell+6})^2) = s_3(1 + 2^8 - 2^5) = 1$$

and thus

$$p_{(7 \cdot 2^{\ell+3}+2^{\ell+5})^2} \neq p_{(7 \cdot 2^{\ell+3}+2^{\ell+6})^2}. \quad (9)$$

For odd $k > 3$ we have

$$s_k(((2^{k-1} - 1)2^{\ell+2} + 2^{2k-1+\ell})^2) = s_k(1 + (2^{3k-3} - 2^k) + 2^{4k-6}) = k - 2 \equiv 1 \pmod{2}$$

but

$$\begin{aligned} & s_k(((2^{k-1} - 1)2^{\ell+2} + 2^{2k+\ell})^2) \\ &= s_k(1 + (2^{2k-2} - 2^k) + (2^{3k-2} - 2^{2k-1}) + 2^{4k-4}) \\ &= 0 \end{aligned}$$

and thus

$$p_{((2^{k-1}-1)2^{\ell+2}+2^{2k-1+\ell})^2} \neq p_{((2^{k-1}-1)2^{\ell+2}+2^{2k+\ell})^2}. \quad (10)$$

Now the result follows the same way as Theorem 1 as follows.

Assume $M = M(\mathcal{P}'_k, N) \leq \left\lfloor \sqrt{2^{\ell+2k-1}-1} \right\rfloor + 1$ and thus there is a recurrence of order M which successively continues (6) to get

$$p_{(i+2^{\ell+2k-1})^2} = p_{(i+2^{\ell+2k})^2}, \quad i = 0, 1, \dots, N - 2^{\ell+2k} - 1.$$

Choosing

$$i = \begin{cases} 2^{\ell+2}, & k = 2, \\ (2^k - 1)2^\ell, & k > 2 \text{ and } k \text{ even}, \\ 7 \cdot 2^{\ell+3}, & k = 3, \\ ((2^{k-1} - 1)2^{\ell+2}), & k > 3 \text{ and } k \text{ odd}, \end{cases}$$

we get a contradiction to (7), (8), (9) or (10), respectively. Hence,

$$M \geq \left\lfloor \sqrt{2^{\ell+2k-1}-1} \right\rfloor + 2 \geq \left(\frac{N}{8} \right)^{1/2}$$

by (5). □

4 Open problems

Problem 1. [5, Conjecture 1] *Show that the subsequences of the Thue-Morse sequence (pattern sequence) along any polynomial of degree $d \geq 2$ are normal.*

This problem may be out of reach and we state some weaker problems. It is known that the subword complexity is maximal if $d = 2$ [12]. For $d \geq 3$ a lower bound on the subword complexity is given in [12], as well.

Problem 2. [12, Open Question 4] *Show that the subword complexity of the subsequence of the Thue-Morse sequence along any polynomial of degree $d \geq 3$ is maximal.*

Problem 3. [5, above Conjecture 1] *Determine the frequency of 0 and 1 in the subsequence of the Thue-Morse sequence along any polynomial of degree $d \geq 3$.*

Problem 4. *Extend Theorems 1 and 2 to any polynomial of degree $d \geq 2$.*

Problem 5. *Prove upper bounds on the correlation measure of order k for subsequences of the Thue-Morse sequence along squares (polynomials).*

Problem 6. *Prove lower bounds on the expansion complexity of the Thue-Morse sequence along squares (polynomials).*

Acknowledgments

We thank the referees for their careful reading and their valuable remarks. The first author is supported by China Scholarship Council and the National Natural Science Foundation of China Grant 61472120. The second author is supported by the Austrian Science Fund FWF Project P 30405-N32.

References

- [1] J.-P. Allouche, J. Shallit, Automatic sequences. Theory, applications, generalizations. Cambridge University Press, Cambridge, 2003.
- [2] S. Brlek, Enumeration of factors in the Thue-Morse word. First Montreal Conference on Combinatorics and Computer Science, 1987. Discrete Appl. Math. 24 (1989), 83–96.
- [3] G. Christol, Ensembles presque periodiques k -reconnaissables. Theoret. Comput. Sci. 9 (1979), no. 1, 141–145.
- [4] C. Diem, On the use of expansion series for stream ciphers. LMS J. Comput. Math. 15 (2012), 326–340.
- [5] M. Drmota, C. Mauduit, J. Rivat, Normality along squares. J. Eur. Math. Soc. (JEMS), to appear.
- [6] D. Gómez-Pérez, L. Mérai, H. Niederreiter, On the expansion complexity of sequences over finite fields, IEEE Trans. Inform. Theory 64 (2018), 4228–4232.

- [7] R. Hofer, L. Mérai, A. Winterhof, Measures of pseudorandomness: arithmetic autocorrelation and correlation measure. Number theory-Diophantine problems, uniform distribution and applications, 303–312, Springer, Cham, 2017.
- [8] C. J. A. Jansen, Investigations on nonlinear streamcipher systems: construction and evaluation methods, Ph.D. dissertation, Technical University of Delft, Delft, 1989.
- [9] C. J. A. Jansen, The maximum order complexity of sequence ensembles. D.W. Davies (Ed.): Advances in Cryptology - EUROCRYPT '91, Lect. Notes Comput. Sci. 547, pp. 153–159, Springer-Verlag, Berlin Heidelberg, 1991.
- [10] A. de Luca, S. Varricchio, Some combinatorial properties of the Thue-Morse sequence and a problem in semigroups. Theoret. Comput. Sci. 63 (1989), 333–348.
- [11] C. Mauduit, A. Sárközy, On finite pseudorandom binary sequences. II. The Champernowne, Rudin-Shapiro, and Thue-Morse sequences, a further construction. J. Number Theory 73 (1998), 256–276.
- [12] Y. Moshe, On the subword complexity of Thue-Morse polynomial extractions. Theoret. Comput. Sci. 389 (2007), 318–329.
- [13] A. Topuzoğlu, A. Winterhof, Pseudorandom sequences. Topics in geometry, coding theory and cryptography, 135–166, Algebr. Appl., 6, Springer, Dordrecht, 2007.
- [14] Z. Sun, A. Winterhof, On the maximum order complexity of the Thue-Morse and Rudin-Shapiro sequence, Preprint.