On The Maximum Order Complexity Of Thue-Morse And Rudin-Shapiro Sequences Along Polynomial Values

Pierre Popoli

ON THE MAXIMUM ORDER COMPLEXITY OF THUE–MORSE AND RUDIN–SHAPIRO SEQUENCES ALONG POLYNOMIAL VALUES

PIERRE POPOLI

Abstract. Both the Thue–Morse and Rudin–Shapiro sequences are not suitable sequences for cryptography since their expansion complexity is small and their correlation measure of order 2 is large. These facts imply that these sequences are highly predictable despite the fact that they have a large maximum order complexity. Sun and Winterhof (2019) showed that the Thue–Morse sequence along squares keeps a large maximum order complexity. Since, by Christol's theorem, the expansion complexity of this rarefied sequence is no longer bounded, this provides a potentially better candidate for cryptographic applications. Similar results are known for the Rudin–Shapiro sequence and more general pattern sequences. In this paper we generalize these results to any polynomial subsequence (instead of squares) and thereby answer an open problem of Sun and Winterhof. We conclude this paper by some open problems.

1. Introduction

Pseudorandomness, i.e. the study of phenomena related to randomness for deterministic objects, has grown to a large and important subject in number theory and cryptography. In recent years, research focused in particular on automatic sequences (e.g. Thue–Morse sequence, Rudin–Shapiro sequence), i.e. sequences that are generated by a deterministic finite automaton. Such sequences are easy to generate and "regular" in some sense, but their behavior changes radically when the sequence is rarefied along a subsequence so that the rarefied sequence shows pseudorandom behavior. The aim of the present article is to study pseudorandomness in the context of such polynomially rarefied automatic sequences.

To begin with, we first introduce various measures for pseudorandomness and cite the results that are known in the context of classical automatic sequences.

Definition 1 (Maximum order complexity)**.** Let N be a positive integer with $N \geq 2$, and $\mathcal{S} = (s_n)_{n\geq 0}$ be a sequence over $\{0,1\}$ with $(s_0,\dots,s_{N-2}) \neq (a,\dots,a)$ for $a = 0$ or 1. The *Nth maximum order complexity* $M(\mathcal{S},N)$ is the smallest positive integer M such that there is a polynomial $f(x_1,\dots,x_M)$ with

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

If $s_i = a$ for $i = 0,\dots,N-2$, we define $M(\mathcal{S},N) = 0$ if $s_{N-1} = a$ and $M(\mathcal{S},N) = N-1$ else.

2010 *Mathematics Subject Classification.* 11A63, 11B85.

Key words and phrases. Automatic sequences, pseudorandomness, Thue–Morse sequence, Rudin–Shapiro sequence, polynomials.

A sequence with small maximum order complexity cannot be used in cryptography, since the sequence can be constructed from relatively short blocks of consecutive terms. However, a sequence with large maximum order complexity, is not automatically adapted in cryptography. It can still be very predictable as we will state later for the Thue–Morse sequence.

Diem [4] introduced the expansion complexity of a sequence as follows.

Definition 2 (Expansion complexity)**.** Let N be a positive integer, $\mathcal{S} = (s_n)_{n\geq 0}$ be a sequence over $\{0,1\}$ and $G(x)$ its generating function defined by

$$G(x) = \sum_{i\geq 0} s_i x^i.$$

The *Nth expansion complexity* $E(\mathcal{S}, N)$ is defined as the least total degree of a nonzero polynomial $h(x,y) \in \mathbb{F}_2[x,y]$ with

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

if $s_0, \ldots, s_{N-1}$ are not all equal to 0, and $E(\mathcal{S}, N) = 0$ otherwise.

Similarly to the maximum order complexity, a sequence with small Nth expansion complexity is predictable. By Christol's theorem (see [3]) automatic sequences over $\mathbb{F}_p$ are characterized by

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

This indicates that automatic sequences are not pseudorandom and may be considered as cryptographically weak.

Mauduit and Sárközy [9] introduced the correlation measure.

Definition 3 (Correlation measure of order 2)**.** Let N be a positive integer, $\mathcal{S} = (s_n)_{n\geq 0}$ be a sequence over $\{0,1\}$. The *N-th correlation measure of order 2* of $\mathcal{S}$ is

$$C_2(\mathcal{S}, N) = \max_{M,d_1,d_2} \left| \sum_{0\leq n\leq M} (-1)^{s_{n+d_1}+s_{n+d_2}} \right|,$$

where the maximum is taken over all M, d_1 and d_2 such that $0 \leq d_1 < d_2$ and $d_2 + M < N$.

For a random sequence, the correlation measure of order 2 is of order of $(N\log(N/2))^{1/2}$ (see [2]).

We introduce the symbolic complexity as an other measure of pseudorandomness.

Definition 4 (Symbolic complexity)**.** The *symbolic complexity*, or *subword complexity*, of a sequence $\mathcal{S}$ over $\{0,1\}$ is the function $p_{\mathcal{S}}$ defined for every positive integer k by

$$p_{\mathcal{S}}(k) = \mathrm{Card}\{(b_0, \ldots, b_{k-1}) \in \{0,1\}^k : \exists i, u(i) = b_0, \ldots, u(i+k-1) = b_{k-1}\}.$$

A sequence $\mathcal{S}$ over $\{0,1\}$ is *normal* if for every $k \geq 1$ and any $(b_0, \ldots, b_{k-1}) \in \{0,1\}^k$, we have

$$\lim_{N\to\infty} \frac{1}{N} \mathrm{Card}\{i < N :\ u(i) = b_0, \ldots, u(i+k-1) = b_{k-1}\} = \frac{1}{2^k}.$$

For a normal sequence, each block of length k appears and each block appears with the same frequency. A "good" pseudorandom sequence should have a large symbolic complexity.

We now look at these complexity measures for the Thue–Morse sequence. One possible definition of this emblematic sequence is as follows.

Definition 5 (Thue–Morse sequence)**.** For an integer $n \geq 0$, we write $n = \sum_{i\geq 0} \varepsilon_i 2^i$ with $\varepsilon_i \in \{0,1\}$ for all i and $(n)_2 = \cdots \varepsilon_1\varepsilon_0$. The binary sum-of-digits of n equals $s_1(n) = \sum_{i\geq 0} \varepsilon_i$. The *Thue–Morse sequence* $\mathcal{T} = (t(n))_{n\geq 0}$ is defined by $t(n) = s_1(n) \bmod 2$.

Note that the index in the sum-of-digits function relates to the length of the `1`-pattern that we consider (the length is 1 here; we will consider k consecutive `1`'s for *pattern sequences*). In what follows, we use Vinogradov's notation $f \ll g$ if there is a constant $c > 0$ such that $f \leq cg$.

Sun and Winterhof [13] showed that the Thue–Morse sequence has large maximum order complexity, $M(\mathcal{T}, N) \gg N$. Mauduit and Sárközy [8, Theorem 2] showed that the correlation measure of order 2 of the Thue–Morse sequence is large, $C_2(\mathcal{T}, N) \gg N$. On the other hand, it is well-known that $E(\mathcal{T}, N) \leq 5$ for all N since $h(x, y) = (x + 1)^3y^2 + (x + 1)^2y + x$ satisfies $h(x, G(x)) = 0$, where $G(x)$ is the generating function of $\mathcal{T}$. Also, its symbolic complexity is small, $p_{\mathcal{T}}(k) \ll k$ (see [1, Corollary 10.3.2] for a general result for all automatic sequences). The small upper bounds of the expansion complexity and of the symbolic complexity imply that the Thue–Morse sequence is far from being a pseudorandom sequence with respect to these measures.

Several of the mentioned results also hold true for more general pattern sequences (also called *Rudin–Shapiro sequences of degree k*, see [7]). For the sake of shortness, we refer to them as *pattern sequences*, as they were called by Sun and Winterhof [13].

Definition 6 (Pattern sequences)**.** Let $k \geq 1$. Denote by $P_k = \mathtt{1}\cdots\mathtt{1} \in \mathbb{F}_2^k$ the all `1`-pattern of length k and by $s_k(n)$ the number of occurrences of P_k in the binary digital representation of n. The *k-pattern sequence* $\mathcal{P}_k = (p_k(n))_{n\geq 0}$ (or, for short, *pattern sequence*) is defined by

$$p_k(n) = s_k(n) \bmod 2.$$

For $k = 1$ we get the Thue–Morse sequence $\mathcal{T} = \mathcal{P}_1$ and for $k = 2$ we get the Rudin–Shapiro (or Golay–Rudin–Shapiro) sequence $\mathcal{R} = \mathcal{P}_2 = (r(n))_n$. As the Thue–Morse sequence, the pattern sequence $\mathcal{P}_k$ is 2-automatic and has a large maximum order complexity (see [13, Theorem 2]). Its expansion complexity satisfies $E(\mathcal{P}_k, N) \leq 2^k + 3$ for $N \geq 1$ since $h(x, y) = (x+1)^{2^{k+1}+1}y^2+(x+1)^{2^k}y+x^{2^k-1}$ satisfies $h(x, G(x)) = 0$ with $G(x)$ the generating function of $\mathcal{P}_k$ (see [14]). Mérai and Winterhof [10, Corollary 4] showed that the correlation of order 2 for pattern sequences is still large. However, since pattern sequences are still automatic, their symbolic complexity is linear [1, Corollary 10.3.2]. Therefore, the pattern sequences are not pseudorandom with respect to each of the defined measures.

The behavior of these sequences regarding the defined pseudorandomness measures changes when these sequences are rarefied along specific subsequences. Sun and Winterhof [14] showed that the maximum order complexity of the Thue–Morse sequence and pattern sequences along squares remains large. Note that the largest possible order of magnitude of $M(\mathcal{S}, N)$ is N, while the expected value of $M(\mathcal{S}, N)$ is $\log N$ (see [14]).

Theorem 1 ([14], Theorem 1)**.** Let $\mathcal{T}' = (t(n^2))_n$ be the subsequence of the Thue–Morse sequence along squares. Then the Nth maximum order complexity of $\mathcal{T}'$ satisfies

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

Theorem 2 ([14], Theorem 2)**.** For $k \geq 2$ let $\mathcal{P}'_k = (p_k(n^2))_n$ be the subsequence of $\mathcal{P}_k$ along squares. Then the Nth maximum order complexity of $\mathcal{P}'_k$ satisfies

$$M(\mathcal{P}'_k, N) \geq \sqrt{\frac{N}{8}}, \qquad N \geq 2^{2k+2}.$$

Drmota, Mauduit and Rivat [6] showed that $\mathcal{T}'$ is a normal sequence, and Müllner [12] showed that $\mathcal{R}'$ and more general pattern sequences along squares are normal, too. These statements mean that $\mathcal{T}'$ and $\mathcal{R}'$ might be better candidates for cryptographic applications as the inherent weaknesses for automatic sequences disappear.

By [1, Theorem 6.10.1], $\mathcal{T}' = \mathcal{P}'_1$ is no longer automatic and so

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

By [12] and Christol's theorem we also have for $\mathcal{R}' = \mathcal{P}'_2$,

$$\sup_{N \geq 1} E(\mathcal{R}', N) = +\infty.$$

Passing from the subsequence of squares to more general polynomials seems to be a natural question. More specifically, Sun and Winterhof (Problem 4 in [14]) posed the problem to extend their results to this more general context.

In this paper we provide an answer to their problem.

Theorem 3. Let $d \geq 2$ and $P(X) \in \mathbb{Z}[X]$ be a monic polynomial of degree d with $P(\mathbb{N}) \subset \mathbb{N}$. Let $\mathcal{T}_P = (t(P(n)))_n$ be the subsequence of the Thue–Morse sequence along the polynomial subsequence $(P(n))_n$. Then $\mathcal{T}_P$ satisfies

$$M(\mathcal{T}_P, N) \gg N^{1/d},$$

where the implied constant only depends on P.

We recover the same bound for pattern sequences, too.

Theorem 4. Let $d \geq 2$ and $P(X) \in \mathbb{Z}[X]$ be a monic polynomial of degree d with $P(\mathbb{N}) \subset \mathbb{N}$. Let $\mathcal{P}_{k,P} = (p_k(P(n)))_n$ be the subsequence of the pattern sequence along the polynomial subsequence $(P(n))_n$. Then $\mathcal{P}_{k,P}$ satisfies

$$M(\mathcal{P}_{k,P}, N) \gg N^{1/d},$$

where the implied constant only depends on P and k.

It is possible to generalize our theorem for the case of integer-valued polynomials with rationals coefficients, we leave this rather straightforward extension to the interested reader (the proof runs along the same lines). We remark, however, that our construction crucially depends on the fact that the leading coefficient of the polynomial equals one.

The paper is structured as follows. In Section 2, we generalize Theorem 1 to any polynomial subsequence in place of the subsequence of squares and in Section 3 we establish the result for any pattern sequence. This answers a question posed by Sun

and Winterhof (Problem 4 in [14]). We finish the paper with a list of open problems in Section 4.

2. Thue–Morse sequence

The Thue–Morse sequence along arithmetic progressions is 2-automatic [1, Theorem 6.8.1]. By [1, Theorem 6.10.1], the Thue–Morse sequence along polynomial subsequences is non-automatic if and only if the polynomial is at least of degree 2. The problem raised by Sun and Winterhof [14] is to know whether there still holds a result such as Theorem 1 for general polynomial subsequences.

The following trivial identity will be essential for our general proof. Let a, b be positive integers and $0 \le b < 2^r$, then we have

$$s_1(a2^r + b) = s_1(a) + s_1(b). \tag{1}$$

If we have such a and b we say that the sum is *non-interfering*. The proof of our main result is based both on non-interfering sums and on carry propagation.

Lemma 5. Let $d \ge 2$ and $P(X) \in \mathbb{Z}[X]$ with $P(X) = X^d + \alpha_{d-1}X^{d-1} + \cdots + \alpha_1 X + \alpha_0$ such that $P(\mathbb{N}) \subset \mathbb{N}$ and all $\alpha_i \ge 0$. Put $\alpha_{\max} = \max(\alpha_i)$. Then there exists a positive integer $l_0(P)$ such that for all $l > l_0(P)$ the following two properties hold:

(i) For all $1 \le n < \frac{1}{2(2\alpha_{\max})^{1/d}}\, 2^l$ and for all $r \ge 1$,

$$t(P(n + 2^{dl})) = t(P(n + 2^{dl+r})).$$

(ii) There are nonnegative integers y and r depending only on P, such that

$$t(P(1 + y2^l + 2^{dl})) \ne t(P(1 + y2^l + 2^{dl+r})).$$

Proof. Set $\alpha_d = 1$. For the first part we write

$$\begin{aligned} P(n + 2^{dl}) &= \sum_{0 \le j \le d} \alpha_j (n + 2^{dl})^j, \\ &= \sum_{0 \le i \le d} \left(\sum_{i \le j \le d} \binom{j}{i} \alpha_j n^{j-i} \right) 2^{idl}. \end{aligned} \tag{2}$$

Set $\beta_i = \sum_{i \le j \le d} \binom{j}{i} \alpha_j n^{j-i}$. We note that for all $0 \le i \le d$,

$$\beta_i \le \alpha_{\max}\, n^{d-i} \sum_{i \le j \le d} \binom{j}{i} \le \alpha_{\max}\, n^d \binom{d+1}{i+1} \le \alpha_{\max}\, n^d 2^{d+1}.$$

There exists a positive integer $l_0(P)$ such that for all $l > l_0(P)$ and all integers n with

$$1 \le n < \frac{1}{2(2\alpha_{\max})^{1/d}}\, 2^l, \tag{3}$$

we have $\beta_i < 2^{dl}$. Thus, for $l > l_0(P)$ and all n with (3) the sum (2) is non-interfering, thus we get for all $r \ge 1$,

$$t(P(n + 2^{dl})) = \sum_{0 \le i \le d} t(\beta_i) = t(P(n + 2^{dl+r})),$$

which shows property (i). As for the second part, we write

$$P(1+y2^l+2^{dl}) = \sum_{0\le i\le d}\left(\sum_{i\le j\le d}\binom{j}{i}\alpha_j(1+y2^l)^{j-i}\right)2^{idl}. \tag{4}$$

We regroup terms by powers of 2 and check for possible interferences. The general term is $2^{idl+l(j-i)}$, for $0\le i\le j\le d$, whereas the coefficients depend only on P and y. We represent the general terms in the following table:

i \ j	0	1	2	$\cdots$	d
0	2^0	2^l	2^{2l}		2^{dl}
1		2^{dl}	2^{dl+l}		$2^{dl+(d-1)l}$
2			2^{2dl}		$2^{2dl+(d-2)l}$
$\vdots$				$\ddots$	
d					$2^{d\cdot dl}$

The only possible interference, for $l > l_1(P)$, is between $(i,j)=(0,d)$ and $(i,j)=(1,1)$ that both correspond to the general term 2^{dl}. Each coefficient in front of $2^{idl+(j-i)l}$ does not depend on l since y will be chosen later to depend only on P and the gap between any two distinct general terms in (4) is at least 2^l. The interfering term is the term in front of 2^{dl} in the expansion of

$$\begin{aligned}&(\alpha_0+\alpha_1(1+y2^l)+\cdots+(1+y2^l)^d)\,2^{0\cdot dl}\\&+(\alpha_1+2\alpha_2(1+y2^l)+\cdots+d(1+y2^l)^{d-1})\,2^{1\cdot dl},\end{aligned}$$

since all the other general terms are at least of size 2^{2dl}. Therefore, the interfering term is $y^d+\sum_{1\le i\le d} i\alpha_i$.

On the other hand, by a similar calculation, we have

$$P(1+y2^l+2^{dl+r}) = \sum_{0\le i\le d}\left(\sum_{i\le j\le d}\binom{j}{i}\alpha_j(1+y2^l)^{j-i}\right)2^{i(dl+r)}.$$

We regroup terms by powers of 2 as before. The general term is $2^{i(dl+r)+(j-i)l}$ and we have the following table:

i \ j	0	1	2	$\cdots$	d
0	2^0	2^l	2^{2l}		2^{dl}
1		2^{dl+r}	$2^{dl+l}2^r$		$2^{dl+(d-1)l}2^r$
2			$2^{2dl}2^{2r}$		$2^{2dl+(d-2)l}2^{2r}$
$\vdots$				$\ddots$	
d					$2^{d\cdot dl}2^{dr}$

Once again the only interference possible is for $(i,j) = (0,d)$ and $(i,j) = (1,1)$ for $l > l_2(P)$. Here, the interfering term is $y^d + 2^r \sum_{1 \le i \le d} i\alpha_i$. All the coefficients in the second table are identical to the ones given in the first table up to a multiplicative factor (a power of 2). Since $t(2^\mu n) = t(n)$ for all $\mu \ge 0$, the contributions coming from the non-interfering terms are the same as in the former case. Put $z = \sum_{1 \le i \le d} i\alpha_i > 0$; we note that z is a positive integer that only depends on P and that $z \ge 2$. Summing up, for $l > l_3(P)$, we have

$$\begin{aligned} &t(P(1 + y2^l + 2^{dl})) + t(P(1 + y2^l + 2^{dl+r})) \\ &\qquad \equiv t(y^d + z) + t(y^d + 2^r z) \pmod 2. \end{aligned} \tag{5}$$

Our final aim is to guarantee the existence of r and y, only depending on P, such that right hand side of (5) equals 1 (mod 2). Let $\lambda \ge 1$ be the unique integer with $2^\lambda \le z < 2^{\lambda+1}$. Thus the most significant bit of z is at position 2^λ. Let $y = 2^\lambda$, thus $z < 2^{\lambda d}$, then we have

$$t(y^d + z) = t(2^{\lambda d} + z) \equiv 1 + t(z) \pmod 2.$$

Let $r = \lambda d - \lambda \ge 1$, we have

$$t(y^d + 2^r z) = t(2^{\lambda d} + 2^{\lambda d - \lambda} z) = t(2^\lambda + z) = t(z)$$

since the most significant bit of z is at position 2^λ. We therefore conclude

$$t(P(1 + y2^l + 2^{dl})) + t(P(1 + y2^l + 2^{dl+r})) \equiv 1 + 2\,t(z) \equiv 1 \pmod 2$$

and we get property (ii). □

We have now all we need to prove Theorem 3.

Proof of Theorem 3. We first note that we can suppose $\alpha_i \ge 0$ for all $0 \le i < d-1$ and $\alpha_d = 1$. Indeed, for positive integers n, a we have $P(n+a) = \sum_{0 \le i \le d} \beta_i n^i$ with $\beta_i = \sum_{i \le j \le d} \binom{j}{i} \alpha_j a^{j-i}$ and $\alpha_d = 1$, such that for sufficiently large a,

$$\beta_i = a^{d-i} \left(\binom{d}{i} \alpha_d + \sum_{i \le j < d} \binom{j}{i} \alpha_j a^{j-d} \right) \gg_P a^{d-i},$$

and therefore we have $\beta_i \ge 0$ for all $i \ge 0$. This translation by the positive integer a that only depends on P does not affect the measures of complexity that we study since we will suppose N sufficiently large. Hence, without loss of generality, we can assume that all the coefficients of P are positive integers.

Let $\alpha_{\max}, z, \lambda, y, r$ be such as in the proof of the second part of Lemma 5. Note that all these quantities only depend on P and not on l. Let $N > N_0(P)$ be large enough and $M(\mathcal{T}_P, N) = M$. Let $l \ge 2$ be the integer defined by

$$1 + y2^l + 2^{dl+r} < N \le 1 + y2^{l+1} + 2^{d(l+1)+r}.$$

We follow the argument in the proof of Sun and Winterhof [14, Theorem 1]. Assume that

$$M < \frac{1}{2(2\alpha_{\max})^{1/d}}\, 2^l,$$

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

$$t(P(j+M)) = f(t(P(j)),\dots,t(P(j+M-1))), \quad j=0,1\dots,N-M-1. \tag{6}$$

Note that for $0\le k\le N-M-1$ the values of $t(P(k+M)),\dots,t(P(N-1))$ are uniquely determined by the values of $t(P(k)),\dots,t(P(k+M-1))$ by applying (6) successively for $j=k,\dots,N-M-1$. In particular, if

$$(t(P(k_1)),\dots,t(P(k_1+M-1))) = (t(P(k_2)),\dots,t(P(k_2+M-1))) \tag{7}$$

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

$$(t(P(k_1+M)),\dots,t(P(k_1+N-k_2-1))) = \\ (t(P(k_2+M)),\dots,t(P(N-1))).$$

Take $k_1=2^{ld}$ and $k_2=2^{ld+r}$. By the first part of Lemma 5, (k_1,k_2) satisfies (7). Then we have

$$(t(P(2^{ld}+M)),\dots,t(P(N+2^{ld}(1-2^r)-1))) = \\ (t(P(2^{ld+r}+M)),\dots,t(P(N-1))).$$

Since $N-1\ge 1+y2^l+2^{dl+r}$ and $M\le 1+y2^l$, this includes

$$t(P(1+y2^l+2^{dl})) = t(P(1+y2^l+2^{dl+r})),$$

which contradicts the second part of Lemma 5 and we get

$$M \ge \frac{1}{2(2\alpha_{\max})^{1/d}}\,2^l \gg_P N^{1/d}. \tag{8}$$

This finishes the proof of Theorem 3. □

3. Pattern sequences

The identity (1) is not true in general for s_k instead of s_1. For example, we have $s_2(4+2)=s_2(\mathtt{110})=1$ and $s_2(4)+s_2(2)=s_2(\mathtt{100})+s_2(\mathtt{10})=0$. However, a very similar identity holds true when we add a $\mathtt{0}$-bit in (1) between the expansions of $a2^r$ and b : Let a,b be positive integers and $0\le b<2^r$, then we have for all $k\ge 2$,

$$s_k(a2^{r+1}+b) = s_k(a)+s_k(b). \tag{9}$$

Lemma 6. Let $d\ge 2$ and $P(X)\in\mathbb{Z}[X]$ with $P(X)=X^d+\alpha_{d-1}X^{d-1}+\cdots+\alpha_1X+\alpha_0$ such that $P(\mathbb{N})\subset\mathbb{N}$ and all $\alpha_i\ge 0$. Put $\alpha_{\max}=\max(\alpha_i)$. Then there exists a positive integer $l_0(P,k)$ such that for all $l>l_0(P,k)$ the following two properties hold:

(i) For all $1\le n<\frac{1}{4(2\alpha_{\max})^{1/d}}\,2^l$ and for all $s\ge 1$,

$$p_k(P(n+2^{dl})) = p_k(P(n+2^{dl+s})).$$

(ii) There are nonnegative integers $y=y(P,k)$ and $s=s(P,k)$ such that

$$p_k(P(1+y2^l+2^{dl})) \ne p_k(P(1+y2^l+2^{dl+s})).$$

Proof. We can directly proceed as in the proof of Lemma 5 with (1) replaced by (9). The "digital gap" in (9) translates into a change by a factor 2. We use the same notation as in Lemma 5. For second part we get that for $l > l_0(P)$,

$$\begin{aligned} p_k(P(1+y2^l+2^{dl})) + p_k(P(1+y2^l+2^{dl+s})) \\ \equiv p_k(y^d+z) + p_k(y^d+2^s z) \pmod 2. \end{aligned}$$

Let $f_a(x) = ax^3 + ax^2 - x + a$ be a polynomial where a is a suitable positive integer that we will chose later. We write

$$f_a(x)^d = \sum_{0 \le i \le 3d} \mu_i x^i$$

with $\mu_i \in \mathbb{Z}$. For $a > a_0(d)$, we have $\mu_i > 0$ for $i \neq 1$ and $\mu_1 = -da^{d-1} < 0$. It follows that for $u > u_0(d)$ we have

$$(f_a(2^u)^d)_2 = \eta_1 \mathtt{0} \ldots \mathtt{0} \eta_2 \mathtt{0} \ldots \mathtt{0} \eta_{t-1} \mathtt{1} \ldots \mathtt{1} \eta_t \mathtt{0} \ldots \mathtt{0} \eta_{t+1} \tag{10}$$

with $t = 3d$ and some $\eta_i = \eta_i(a,d)$. For $i \in \{1, \ldots, t-2, t+1\}$, we can choose η_i the binary expansion of μ_i, η_{t-1} such that its first bit is $\mathtt{1}$ and its last bit is $\mathtt{0}$ and η_t such that its first bit is $\mathtt{0}$ and its last bit is $\mathtt{1}$. Note that the decomposition in (10) is not unique in general. For $a = 2^\lambda$ with $\lambda > \lambda_0(z)$ we have $p_k((f_a(2^u)^d + z) = p_k((f_a(2^u)^d) + p_k(z)$. Furthermore, the length of the $\mathtt{1}$-block depends on u and the transition from u to $(u+1)$ adds exactly one $\mathtt{1}$-bit to that $\mathtt{1}$-block. Let $y = f_a(2^u)$ for $a = 2^\lambda > \max(a_0, 2^{\lambda_0})$ and $u > u_0(a,d)$. We write in the following $y^d = \omega_1 \mathtt{0} \ldots \mathtt{0} \omega_2 \mathtt{0} \mathtt{1}^{(\alpha)} \mathtt{0} \omega_3$ where $\mathtt{1}^{(\alpha)}$ is the block of $\mathtt{1}$-bits in (10) between η_{t-1} and η_t. For $u > u_1(a,d,z,k)$ sufficiently large, we can ensure that $\alpha > \max(\lceil \log_2(z) \rceil, k)$. We write $z = \mathtt{1}^{(i)} \mathtt{0} \omega'$ or $z = \mathtt{1}^{(i)}$ for some $i \ge 1$ and digital block ω' which can be possibly empty. We choose s in a way that the first i $\mathtt{1}$-bits interfere with the last i $\mathtt{1}$-bits of the inner $\mathtt{1}$-bits block of y^d. Note that this is always possible since α is larger than the length of z. When $z = \mathtt{1}^{(i)} \mathtt{0} \omega'$ we get

			$\overbrace{\mathtt{1}\ldots\mathtt{1}}^{\alpha-i}$	$\overbrace{\mathtt{1}\ldots\mathtt{11}}^{i}$		
	$\omega_1\mathtt{0}\ldots\mathtt{0}$	w_2	$\mathtt{1}\ldots\mathtt{1}$	$\mathtt{1}\ldots\mathtt{11}$	$\mathtt{0}\omega_3$	$= y^d$
$+$				$\mathtt{1}\ldots\mathtt{11}$	$\mathtt{0}\omega'$	$= 2^s z$
	$\omega_1\mathtt{0}\ldots\mathtt{0}$	(w_2+1)	$\mathtt{0}\ldots\mathtt{0}$	$\mathtt{1}\ldots\mathtt{10}$	$\mathtt{0}(\omega_3+\omega')$	$= y^d + 2^s z$

When $z = \mathtt{1}^{(i)}$ we get

			$\overbrace{\mathtt{1}\ldots\mathtt{1}}^{\alpha-i}$	$\overbrace{\mathtt{1}\ldots\mathtt{11}}^{i}$		
	$\omega_1\mathtt{0}\ldots\mathtt{0}$	w_2	$\mathtt{1}\ldots\mathtt{1}$	$\mathtt{1}\ldots\mathtt{11}$	$\mathtt{0}\omega_3$	$= y^d$
$+$				$\mathtt{1}\ldots\mathtt{11}$	$\mathtt{0}\cdots\mathtt{0}$	$= 2^s z$
	$\omega_1\mathtt{0}\ldots\mathtt{0}$	(w_2+1)	$\mathtt{0}\ldots\mathtt{0}$	$\mathtt{1}\ldots\mathtt{10}$	$\mathtt{0}\omega_3$	$= y^d + 2^s z$

In both cases, for $u > u_2(a,d,z,k)$ sufficiently large, $p_k(y^d + 2^s z)$ is constant with respect to u since no more k-pattern is created or canceled. In fact, only the length of the $\mathtt{0}$-blocks and the one of the $\mathtt{1}$-block change with u in (10) but the sum $y^d + 2^s z$ reduces the $\mathtt{1}$-block of length α to a $\mathtt{1}$-block of length $(i-1)$ which does not depend on u anymore. Furthermore, the number of k-patterns in $\omega_1, \omega_2, \omega_3$ and ω' are independent from u. Hence $p_k(y^d + 2^s z)$ is constant for all $u > u_2(a,d,z,k)$. What matters now is the number of k-patterns we cancel in y^d by adding $2^s z$. Since $\alpha > k$, the transition

from u to $(u+1)$ implies that if we add $2^s z$ to y^d, we cancel one more k-pattern in y^d. Thus we can choose $u > u_2(a,d,z,k)$ in such a way that the parity of the number of k-patterns that are canceled in y^d by adding $2^s z$ changes. This leads to a solution of $p_k(y^d + 2^s z) \equiv p_k(y^d) + p_k(z) + 1 \pmod 2$ and the lemma is proved.

□

Proof of Theorem 4. The proof follows the lines of the proof of Theorem 3 where we replace Lemma 5 by Lemma 6. We note that the size of y only occurs in the final step of the proof, namely, inequality (8). Thus we have

$$M \geq \frac{1}{4(2\alpha_{\max})^{1/d}}\, 2^l \gg_{P,k} N^{1/d},$$

which proves Theorem 4. □

4. Open problems

We can define a larger family of automatic sequences on $\{0,\dots,q-1\}$ for $q \geq 2$ by the following procedure. Let $\omega \in \{0,\dots,q-1\}^k \setminus \{0\dots0\}$ be a pattern of length k and $e_\omega(n)$ be the number of occurrences of ω in the q-ary representation of n. We define the sequence $(\rho(n))_n$ by

$$\rho(n) \equiv e_\omega(n) \pmod m, \tag{11}$$

where $m \geq 2$ is a fixed integer. From our reasoning in the proofs, one can extend our result for $\omega = (q-1)\dots(q-1)$ and any $m \geq 2$. Indeed in the proof of Lemma 6, we use a 1-bit carry propagation along a large block of 1-bits. The same argument works to cancel bits of digits $(q-1)$ in base-q expansions.

Problem 1. Extend Theorem 4 to general pattern sequences as defined in (11). It is essential to our proof that the pattern must appear at least once in y^d. For prime m and d such $m \nmid d$, Hensel's lifting lemma can be useful to make sure that it appears. It seems that other ideas are needed in the case $m \mid d$.

Problem 2 ([6], Conjecture 1)**.** Show that the subsequences of the Thue–Morse sequence along any polynomial of degree $d \geq 3$ are normal. A lower bound on the subword complexity was established by Moshe [11, Corollary 3] and [5] gives a partial answer for generalized Thue–Morse sequence with large q.

Problem 3 ([14])**.** Prove lower bounds on the expansion complexity of the Thue–Morse sequence along polynomials of degree $d \geq 2$. This question seems to be hard even for small degree polynomials.

Acknowledgements. This work was supported partly by the french PIA project " Lorraine Université d'Excellence ", reference ANR-15-IDEX-04-LUE. The author expresses his gratitude to D. Jamet and T. Stoll for the supervision of this work, and A. Winterhof for the right path to the proof.

References

1. J.-P. Allouche and J. Shallit, *Automatic Sequences: Theory, Applications, Generalizations*, Cambridge University Press, Cambridge, 2003.

2. N. Alon, Y. Kohayakawa, C. Mauduit, C. G. Moreira, and V. Rödl, *Measures of pseudorandomness for finite sequences: typical values*, Proc. Lond. Math. Soc. (3) **95** (2007), no. 3, 778–812. MR 2368283
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, and J. Rivat, *The sum-of-digits function of polynomial sequences*, J. Lond. Math. Soc. (2) **84** (2011), no. 1, 81–102.
6. ———, *Normality along squares*, J. Eur. Math. Soc. (JEMS) **21** (2019), no. 2, 507–548.
7. C. Mauduit and J. Rivat, *Rudin–Shapiro sequences along squares*, Transactions of the American Mathematical Society **370** (2018), no. 11, 7899–7921.
8. C. Mauduit and A. Sárközy, *On finite pseudorandom binary sequences. II. The Champernowne, Rudin-Shapiro, and Thue-Morse sequences, a further construction*, J. Number Theory **73** (1998), no. 2, 256–276.
9. Christian Mauduit and András Sárközy, *On finite pseudorandom binary sequences. I. Measure of pseudorandomness, the Legendre symbol*, Acta Arith. **82** (1997), no. 4, 365–377. MR 1483689
10. L. Mérai and A. Winterhof, *On the pseudorandomness of automatic sequences*, Cryptogr. Commun. **10** (2018), no. 6, 1013–1022.
11. Y. Moshe, *On the subword complexity of Thue-Morse polynomial extractions*, Theoret. Comput. Sci. **389** (2007), no. 1-2, 318–329.
12. Clemens Müllner, *The Rudin-Shapiro sequence and similar sequences are normal along squares*, Canad. J. Math. **70** (2018), no. 5, 1096–1129. MR 3831916
13. Z. Sun and A. Winterhof, *On the maximum order complexity of subsequences of the Thue-Morse and Rudin-Shapiro sequence*, Unif. Distrib. Theory **14** (2019), no. 2, 33–42.
14. ———, *On the maximum order complexity of subsequences of the Thue-Morse and Rudin-Shapiro sequence along squares*, Int. J. Comput. Math. Comput. Syst. Theory **4** (2019), no. 1, 30–36.

Institut Élie Cartan de Lorraine, Université de Lorraine, Vandœuvre-lès-Nancy, France