Maximum order complexity for some automatic and morphic sequences along polynomial values

Pierre Popoli

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

OWNS, March 15, 2022







Summary



Automatic and morphic sequences

2 Complexities







The Thue–Morse, or Prouhet–Thue–Morse, sequence is given by the following automaton





The Thue–Morse, or Prouhet–Thue–Morse, sequence is given by the following automaton





The Thue–Morse, or Prouhet–Thue–Morse, sequence is given by the following automaton





The Thue–Morse, or Prouhet–Thue–Morse, sequence is given by the following automaton





The Thue–Morse, or Prouhet–Thue–Morse, sequence is given by the following automaton



Input: $(13)_2 = 1101$, we read left to right.



Then t(13) = 1.

 \mathcal{T} can also be generated by the following morphism $f : \begin{cases} 0 \mapsto 01 \\ 1 \mapsto 10. \end{cases}$ This morphism is *uniform*: images have same length. We apply recursively the corresponding morphism

 $egin{aligned} 0 \ f(0) &= 01 \ f(01) &= 01 \ 10 \ f(0110) &= 0110 \ 1001 \ dots \ f^{\omega}(0) &= 0110100110010110 \dots \end{aligned}$

Thus t(n) is the *n*-th value of $f^{\omega}(0)$.

An *automatic* sequence is a sequence that can be generated by an automaton or equivalently as a projection of a fixed point of a uniform morphism.

Another example: Golay–Rudin–Shapiro sequence \mathcal{R} .



For $(13)_2 = 1101$, we have r(13) = 1. $\mathcal{R} = (r(n))_n = 000100100011101...$ Or with the morphism $f : \begin{cases} a \mapsto ab, \\ b \mapsto ac, \\ c \mapsto db, \\ d \mapsto dc. \end{cases}$ and $\pi : \begin{cases} a, b \mapsto 0, \\ c, d \mapsto 1. \\ d \mapsto dc. \end{cases}$ Thus $\begin{cases} f^{\omega}(a) = abacabdbabacdcac... \\ \pi(f^{\omega}(a)) = 000100100011101.... \end{cases}$

Generalization

Both the Thue–Morse and the Golay–Rudin–Shapiro sequence come from a larger family of automatic sequences.

Let w be a word in base 2, $e_w(n)$ counts the number of occurences of w in the expansion of n in base 2.

Example: w = 101, $(21)_2 = 10101$, $e_w(n) = 2$.

Then $\mathcal{S} = (s_n)_n = (e_{q,\omega}(n) \pmod{q})_n$ is an automatic sequence.

Particular word: $w_k = \underbrace{1 \cdots 1}_{k}$

- For k = 1 we have the Thue–Morse sequence, t(n) counts the number of 1 in $(n)_2 \mod 2$.
- For k = 2 we have the Golay-Rudin-Shapiro sequence, r(n) counts the number of 11 in $(n)_2 \mod 2$.

For a general *k*, these sequences are called *pattern sequences*.

Morphic sequences

Morphic sequences are generated by morphisms that are not necessarily uniform.

Example 1: Fibonacci word
$$f: \begin{cases} 0 \mapsto 01 \\ 1 \mapsto 0 \end{cases}$$
, thus we have $f^{\omega}(0) = 0100101001001 \cdots$

We have $|f^n(0)| = F_n$ the *n*-th Fibonacci number and this sequence is not automatic since the frequencies of letters are irrational.

Example 2: The characteristic sequence of squares is a morphic sequence with

$$f: \begin{cases} a \mapsto abcc, \\ b \mapsto bcc, \\ c \mapsto c. \end{cases} \text{ et } \pi: \begin{cases} a, b \mapsto 1, \\ c \mapsto 0. \end{cases}$$

$$\frac{n \quad 0 \ 1 \ 2 \ 3 \ 4 \ 5 \ 6 \ 7 \ 8 \ 9 \quad 10 \quad 11 \quad 12 \quad 13 \quad 14 \quad 15 \quad 16 \\ \pi(f^{\omega}(a)) \quad 1 \ 1 \ 0 \ 0 \ 1 \ 0 \ 0 \quad 0 \quad 1 \quad 0 \quad 0 \quad 0 \quad 0 \quad 1 \\ \text{This sequence is not automatic either.} \end{cases}$$

Remark

Notice that it is not trivial in general to determine whether a morphic sequence is automatic or not. An example is the sequence generated by

 $f: 0 \mapsto 12, 1 \mapsto 102, 2 \mapsto 0$ $f^{\omega}(1) = 102120102012...$

which is automatic, Berstel (1978) with the image reduction modulo 3 of

$$g: 0 \mapsto 12, 1 \mapsto 13, 2 \mapsto 20, 3 \mapsto 21$$

 $g^{\omega}(1) = 13212013012...$

Summary

1 Automatic and morphic sequences

2 Complexities

3 Pseudorandom sequences



Subword complexity

Let $w = a_0 a_1 a_2 \dots$ be an infinite word on an alphabet Σ . A finite word $u = b_0 b_1 \dots b_{k-1} \in \Sigma^k$ is a *subword* of w if there exists i such that

$$a_i = b_0, \ a_{i+1} = b_1, \ \ldots, \ a_{i+k-1} = b_{k-1}.$$

Subword complexity

Let S be a sequence Σ . For $k \ge 0$, we define p_S by

 $p_{\mathcal{S}}(k) = \#\{u \in \Sigma^k : u \text{ is a subword of } w\}.$

Obviously $p_{\mathcal{S}}(k) \leq \operatorname{Card}(\Sigma)^k$ for all k.

Examples:

- S = 01010101010..., we have $p_S(k) = 2$ for all k.
- If $p_{\mathcal{S}}(k) = k + 1$, \mathcal{S} is said to be *sturmian*.

A sequence is said to be *normal* if for every word $b_0 \dots b_{k-1} \in \Sigma^k$:

$$\lim_{N \to +\infty} \frac{1}{N} \# \{ i < N, s_i = b_0, \dots, s_{i+k-1} = b_{k-1} \} = \frac{1}{\operatorname{Card}(\Sigma)^k},$$

i.e. every word appears in the sequence and each word of a fixed length appears with the same frequency.

Borel (1909) showed that almost every sequence is normal but relatively few constructions are known.

Example: The Champernowne sequence (1933)

 $\mathcal{S}=0$ 1 10 11 100 \ldots

is a normal sequence on $\{0, 1\}$.

It is conjectured that π is a normal number but still not known.

Linear complexity

Let p be prime, $S = (s_n)_n$ be a sequence over \mathbb{F}_p and $N \ge 1$.

Linear complexity at rank N

L(S, N) is the smallest integer L such that

$$s_{i+L}=c_0s_i+\cdots+c_{L-1}s_{i+L-1},$$

with $c_j \in \mathbb{F}_p$ and $0 \le i \le N - L - 1$. It represents the shortest linear recurrence that we need to get the first N terms.

This complexity corresponds to the length of the shortest Linear Feedback Shift Register (LFSR) that generates the sequence. This complexity is then used as an indactor of the unpredictability of the sequence.

Maximum order complexity

Maximum order complexity at rank N

 $M(\mathcal{S}, N)$ is the smallest integer M such that

$$s_{i+M}=f(s_i,\ldots,s_{i+M-1}),$$

with $f(X_1,\ldots,X_M) \in \mathbb{F}_p[X_1,\ldots,X_M]$ and $0 \le i \le N-M-1$.

Same definition as before but not restricted to linear recurrence. We are not interested in the degree of the polynomial, only the number of variables.

We have, trivially,

$$M(\mathcal{S}, N) \leq L(\mathcal{S}, N) \leq N.$$

The maximum order complexity of a sequence can be used to determine the pseudorandomness of a sequence but it is not sufficient.

Example 1/2

Let S = 01011... We have M(S, 2) = 1 since the two first letters are not identical.

In order that M(S,3) = 1, we have to find a polynomial f(x) such that

$$\begin{cases} s_2 = f(s_1), \\ s_1 = f(s_0). \end{cases} \implies \begin{cases} 0 = f(1), \\ 1 = f(0). \end{cases}$$

Thus the polynomial f(x) = -x + 1 is convenient.

In order that $M(\mathcal{S},4) = 1$, we have to find a polynomial f(x) such that

$$\begin{cases} s_3 = f(s_2), \\ s_2 = f(s_1), \\ s_1 = f(s_0). \end{cases} \quad \implies \begin{cases} 1 = f(0), \\ 0 = f(1), \\ 1 = f(0). \end{cases}$$

Thus the same polynomial as before is convenient.

4

Example 2/2

 $\mathcal{S} = \texttt{01011}\ldots$

In order that M(S,5) = 1, we have to find a polynomial f(x) such that

$$\left\{\begin{array}{ll} 1=f(1),1=f(0)\\ 0=f(1),1=f(0) \end{array}\right \implies \text{not possible}.$$

We have to increase the number of variables.

In order that M(S,5) = 2, we have to find a polynomial f(x,y) such that

$$\begin{cases} s_4 = f(s_3, s_2) \\ s_3 = f(s_2, s_1) \\ s_2 = f(s_1, s_0) \end{cases} \implies \begin{cases} 1 = f(1, 0) \\ 1 = f(0, 1) \\ 0 = f(1, 0) \end{cases} \implies \text{again not possible.} \end{cases}$$

In order that M(S,5) = 3, we have to find a polynomial f(x, y, z) such that

$$\begin{cases} s_4 = f(s_3, s_2, s_1) \\ s_3 = f(s_2, s_1, s_0) \end{cases} \implies \begin{cases} 1 = f(1, 0, 1) \\ 1 = f(0, 1, 0) \end{cases}$$

Thus f(x, y, z) = x + y is convenient.

Special factor and maximum order complexity

A finite word u is to be a *special factor* of a word w if there exists at least two different letters α and β such that $u\alpha$ and $u\beta$ are subwords of w.

Theorem: Jansen (1989)

Let $S = (s_n)_n$ be a sequence on Σ . Let k be the length of the largest special factor of the word $s_0 s_1 \dots s_{N-1}$. Then M(S, N) = k + 1.

For example for w = 01011 we notice that 01 is the longest special factor of w. Therefore M(w, 5) = 2 + 1 = 3.

Expansion complexity

Let G(x) be the generating series of S : $G(x) = \sum_{n \ge 0} s_n x^n$.

Expansion complexity at rank N

 $E(\mathcal{S}, N)$ is the least total degree of $h(x, y) \in \mathbb{F}_p[X, Y]$ such that

 $h(x, G(x)) \equiv 0 \mod x^N$.

Christol's theorem(1979)

$$\begin{split} \mathcal{S} \text{ is } p\text{-automatic} \Leftrightarrow & \text{The generating series of } \mathcal{S} \text{ is algebraic over } \mathbb{F}_p. \\ \Leftrightarrow & \textit{E}(S,\textit{N}) < +\infty \text{ pour tout } \textit{N} \geq 1. \end{split}$$

We have the following inequalities

 $E(\mathcal{S}, N) \leq L(\mathcal{S}, N) + 1,$ $M(\mathcal{S}, N) \leq L(\mathcal{S}, N).$

Natural question: M(S, N) weaker than E(S, N)? No, take the Thue–Morse sequence. We need both of theses complexities.

Summary

1 Automatic and morphic sequences

2 Complexities







A sequence S is said *pseudo-random* if S has similar complexities as a truly random sequence and can be easily generated.

Expected order for a truly random binary sequence

- Linear complexity: $L(S, N) \simeq N/2$.
- Maximum order complexity: $M(S, N) \simeq \log N$.
- Subword complexity: $p_{\mathcal{S}}(N) \simeq 2^N$.
- Expansion complexity: $E(S, N) \simeq \sqrt{N}$.

Classical Thue-Morse

Vinogradov's notation: $f \ll g$ means $|f| \leq C|g|$, for some $C \geq 0$ and for N large enough.

Measures for the Thue–Morse sequence

For $N \ge 4$, we have

$$\begin{split} & \mathcal{M}(\mathcal{T}, N) \gg N: \text{ Sun-Winterhof (2019)}, \\ & p_{\mathcal{T}}(N) \ll N: \text{ Brlek, de Luca-Varrichio (1989)} \\ & \text{ automatic, Allouche-Shallit (2003)}, \\ & E(\mathcal{T}, N) \leq 5. \end{split}$$

These measures are very similar for pattern sequences.

What happens along squares ?

Let us denote $\mathcal{T}_2 = (t(n^2))_{n \ge 0}$.

0 1 1 0 1 0 0 1 1 0 0 1 0 1 1 0 1 0 0 1...

Theorem: Drmota, Mauduit and Rivat (2019)

 \mathcal{T}_2 is a normal sequence.

 $\mathcal{T}_2 \text{ is no longer automatic } \implies E(\mathcal{T}_2, N) \rightarrow +\infty.$

Theorem: Sun and Winterhof (2019)

 $M(\mathcal{T}_2, N) \gg N^{1/2}.$

Thus the Thue–Morse sequence along squares is a better candidate for a pseudorandom sequence.

Again, same phenomenon appears for *pattern sequences*.

Generalization to other polynomial subsequences

Let $P \in \mathbb{Z}[X]$, $P(\mathbb{N}) \subset \mathbb{N}$ of degree $d \geq 2$. Let us denote $\mathcal{T}_P = (t(P(n)))_n$. Then

- The subword complexity \mathcal{T}_P is exponential: $p_{\mathcal{T}_P}(N) \ge c^N$ with $c = 2^{1/2^{d-2}}$, Moshe (2007).
- \mathcal{T}_P is not automatic $E(\mathcal{T}_P, N) \to +\infty$.

Theorem: P. (2020)

Let $P \in \mathbb{Z}[X]$, $P(\mathbb{N}) \subset \mathbb{N}$ of degree d monic. Let $\mathcal{T}_P = (t(P(n)))_n$ and $\mathcal{P}_{k,P} = (p_k(P(n)))_n$, then we have for $N \ge N_0(k, P)$,

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

Sum of digits function properties

Since $t_{2n} = t_n$ and $t_{2n+1} = 1 - t_n$, we have $t(n) \equiv s_2(n) \pmod{2}$, where $s_2(n)$ is the sum of digits function in base 2.

$$(n)_2 = \varepsilon_r \dots \varepsilon_0 \implies \begin{cases} (2n)_2 &= \varepsilon_r \dots \varepsilon_0 0\\ (2n+1)_2 &= \varepsilon_r \dots \varepsilon_0 1. \end{cases}$$

For $a, b \ge 0$, $b < 2^r$ we have $s_2(a2^r + b) = s_2(a) + s_2(b)$.

$$(a)_2 \quad 0 \cdots 0 = a2^r \\ + \qquad (b)_2 = b \\ (a)_2 \quad (b)_2 = a2^r + b.$$

In this case the sum is said to be *non-interfering* since digits of a and b do not interact.

Thus, for $\ell \geq 0$, we have for all $n < 2^{\ell}$

$$s_2(n+2^{\ell}) = s_2(n+2^{\ell+1}).$$

Building two same blocks

Let $P = x^d + \alpha_{d-1}x^{d-1} + \cdots + \alpha_0$ and $P \in \mathbb{N}[X]$, we have for all $\ell, r > 0$ and $0 \le n < c_P 2^{\ell}$, for some $c_P > 0$,

$$t(P(n+2^{d\ell})) = t(P(n+2^{d\ell+r})).$$

Sketch of proof:

$$P(n+2^{d\ell}) = \sum_{0 \le j \le d} \alpha_j (n+2^{d\ell})^j = \sum_{0 \le i \le d} \underbrace{\left(\sum_{i \le j \le d} \binom{j}{i} \alpha_j n^{j-i}\right)}_{\beta_i} 2^{idl}.$$
Then $P(n+2^{d\ell}) = \beta_0 + \beta_1 2^{d\ell} + \dots + \beta_d 2^{d^2\ell}$ and each $\beta_i < 2^{d\ell}$ by hypothesis on n . Thus we have

$$t(P(n+2^{d\ell})) \equiv \sum_{0 \leq i \leq d} t(\beta_i) \pmod{2}.$$

By the same proof we have $t(P(n+2^{d\ell+r})) \equiv \sum_{0 \le i \le d} t(\beta_i) \pmod{2}$.

Looking for two different successors 1/2.

Hardest part in general. We look for $n > c_P 2^\ell$ such that

$$t(P(n+2^{d\ell})) \equiv t(P(n+2^{d\ell+r})) + 1 \pmod{2}.$$

Sketch of proof: We try to find n of the form $n = 1 + y2^{\ell}$ such that

$$t(P(1+y2^{\ell}+2^{d\ell})) \equiv t(P(1+y2^{\ell}+2^{d\ell+r}))+1 \pmod{2}.$$

By similar calculations, it remains to find (y, r) such that

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

with z = P'(1).

Looking for two different successors 2/2.

Let $y = 2^s$, s large enough such that

$$t(2^{sd} + z) = t(2^{sd}) + t(z) = 1 + t(z).$$

Thus we need $t(2^{sd} + 2^r z) = t(z)$ by a suitable choice of r. Let us write z = 1z' and choose r such that

$$\begin{array}{ccccccc} 1 & 0 \cdots 0 & 0 \cdots 0 & = 2^{sd} \\ + & 1 & z' & 0 \cdots 0 & = 2^r z \\ \hline 1 & 0 & z' & 0 \cdots 0 & = 2^{sd} + 2^r z \end{array}$$

Thus $t(y^d + 2^r z) = 1 + t(z') = t(z)$.

We have 2 blocks of length $c_1 2^\ell$ among the first $c_2 2^{d\ell}$ values \Longrightarrow

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

Summary

1 Automatic and morphic sequences

2 Complexities







Numeration system based on the Fibonnaci sequence: $F_n = 0, 1, 1, 2, 3, 5, 8, 13, 21, \ldots$

Zeckendorf base

 $\mathcal{F} = (F_n)_n$ Fibonacci sequence, $F_0 = 0$ et $F_1 = 1$. Each integer *n* can be represented uniquely by

$$n = \sum_{i\geq 0} \varepsilon_i(n) F_{i+2},$$
 with $\varepsilon_{i+1}(n) \varepsilon_i(n) = 0.$

 $\varepsilon_{i+1}(n)\varepsilon_i(n) = 0$ hypothesis is to ensure the unicity.

Binary	Zeckendorf			
0	0	8	1000	10000
1	1	9	1001	10001
10	10	10	1010	10010
11	100	11	1011	10100
100	101	12	1100	10101
101	1000	13	1101	10000
110	1001	14	1110	10001
111	1010	15	1111	10010
	Binary 0 1 10 11 100 101 110 111	Binary Zeckendorf 0 0 1 1 10 10 11 100 100 101 101 1000 110 1001 110 1001	BinaryZeckendorf008119101010111001110010112101100013110100114111101015	BinaryZeckendorf00810001191001101010101011100111011100101121100101100013110111010011411101111010151111

Such as for the Thue–Morse sequence, we define the sum of digits function

$$s_Z(n) = \sum_{i\geq 0} \varepsilon_i(n).$$

 $\mathcal{S}_Z = (s_Z(n) \pmod{2})_{n \geq 0}$ is a morphic sequence with

$$f: \left\{ \begin{array}{ll} a\mapsto ab, b\mapsto c\\ c\mapsto cd, d\mapsto a \end{array} \right. \quad \text{and} \quad \pi: \left\{ \begin{array}{ll} a\mapsto 0, b\mapsto 1\\ c\mapsto 1, d\mapsto 0. \end{array} \right.$$

 $S_Z = \pi \circ f(a) = 011101001000110001011...$ S_Z is not an automatic sequence, Drmota-Müllner-Spiegelhofer (2018).

Carry propagation

			1	0	0	1	0 =	= 10		
•	Transversality:	+					1 =	- 1		
			1	0	1	0	0 =	= 11.		
	due to $F_{n+2} = F_{n+1}$	$+F_n$.								
						1	0	0	0	= 5
•	Right carry propagation:	ion		+		1	0	0	1	= 6
		.1011.			1	0	0	1	1	
					1	0	1	0	0	= 11.
		_								

due to $2F_n = F_{n+1} + F_{n-2}$.

Results

Let $\varphi = \frac{1+\sqrt{5}}{2}$ denote the golden ratio.

Theorem 1: Jamet, P., Stoll (2021)

There exists $N_0 > 0$ such that for all $N > N_0$ we have

$$M(\mathcal{S}_Z, N) \geq rac{1}{arphi + arphi^3}N + 1.$$

Theorem 2: Jamet, P., Stoll (2021)

Let $P \in \mathbb{Z}[X]$, $P(\mathbb{N}) \subset \mathbb{N}$ monic of degree d. $S_{Z,P} = (s_Z(P(n) \pmod{2}))_n$, then we have for $N \ge N_0(P)$,

 $M(\mathcal{S}_{Z,P}, N) \gg N^{1/2d}.$

The factor 2d instead of d comes from the carry propagation to the right.

Summary

1 Automatic and morphic sequences

2 Complexities

3 Pseudorandom sequences

4 Zeckendorf base



DAWG

We can find the special factors of a word by building its DAWG (Direct Acyclic Word Graph).

Let $w = a_1 \dots a_n \in \Sigma^n$. For a subword y of w, we define the set $E_w(y) = \{i : y = a_{i-|y|+1} \dots a_i\}$, i.e. the set of ending positions of y. Two subwords y and z are said *suffix-equivalents* if $E_w(y) = E_w(z)$. The DAWG of a word is the smallest graph that recognizes every subword of a word (Blumer et al.). The edges are subwords and the vertices are letters. Two subwords are in the same edge if they are suffix-equivalents.











w = 01011.Sub(w) = {0, 1, 01, 10, 11, 010, 101, 011, 0101, 1011, 01011}.



Thus the longest factor special is the deepest node with at least two outgoing arrows, here this is 01.



Figure: Maximum order complexity at rank N of the Thue–Morse sequence along squares.



Figure: Maximum order complexity at rank N of the Thue–Morse sequence along cubes.



Figure: Maximum order complexity at rank N of Zeckendorf along $\ensuremath{\mathbb{N}}$

$$\tfrac{1}{1+\varphi^2}=0,27639\ldots$$



Figure: Maximum order complexity at rank N of Zeckendorf along squares.



Figure: Maximum order complexity at rank N of Zeckendorf along cubes.

Conjectures

• The maximum order complexity of the Thue–Morse sequence along polynomial degree satisfies

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

• The maximum order complexity of the Zeckendorf sequence along polynomial degree satisfies

$$M(\mathcal{S}_{Z,P}, N) \asymp N^{1/2d}.$$

Thank you for your attention !

Contact:

Pierre Popoli pierre.popoli@univ-lorraine.fr Institut Élie Cartan de Lorraine Université de Lorraine