The Expansion Complexity of Ultimately Periodic Sequences over Finite Fields

Yi Zhang

Department of Foundational Mathematics
Xi'an Jiaotong-Liverpool University, Suzhou, China

Joint work with Zhimin Sun, Xiangyong Zeng, Chunlei Li, and Lin Yi

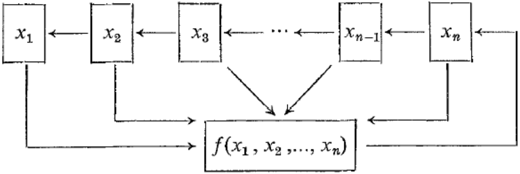
Background

Let $\mathbb{F}_q$ be a finite field of q elements. Pseudo-random sequences are widely used in cryptography. E.g., key stream generator.

- Question 1: How to generate pseudo-random sequences?
- Question 2: How to measure randomness of a given cryptographic sequence?

For Question 1, periodic pseudo-random sequences over $\mathbb{F}_q$ can be generated by *feedback shift registers* (FSRs).

Background



feedback shift register (FSR) with n stages

- The feedback function $f \in \mathbb{F}_q[x_1, x_2, \ldots, x_n]$.
- If f is linear, call it a *linear feedback shift register* (LFSR). The corresponding sequence is a C-finite sequence over $\mathbb{F}_q$.
- If f is nonlinear, call it a *nonlinear feedback shift register* (NFSR).

Background

Question 2: How to measure randomness of a given cryptographic sequence?

- Linear complexity: the length of the shortest LFSRs that generate the sequence.
  - The Berlekamp-Massey (BM) algorithm: an efficient algorithm for computing the shortest LFSR of a given sequence.
  - The Berlekamp-Massey-Sakata (BMS) algorithm: a generalization of the BM algorithm to the multivariate case using the idea of Gröbner bases.
- Nonlinear complexity: the length of the shortest NFSRs that generate the sequence.
  - Blumer's algorithm: an efficient algorithm for computing the shortest NFSR of a given sequence in linear time and memory.
  - Nonlinear complexity $\leq$ linear complexity.

Motivation

Based on the function expansion into expansion series, Xing and Lam (1999) constructed sequences with optimal linear complexity, which are close to half of their lengths.

Diem (2012) showed that this type of sequences can be computed from a shorter subsequence and then introduced

- Expansion complexity: the least total degree of annihilating polynomials for the generating function of a given sequence.
    - Mérai *et al*. (2017) studied the lower and upper bound for expansion complexity of ultimately period sequences and aperiodic sequences.
    - Gómez-Pérez *et al*. (2018) gave an upper bound for expansion complexity of any sequence and introduced the notion of *irreducible expansion complexity*.
    - Gómez-Pérez and Mérai (2020) studied the expansion complexity of any sequence using Gröbner bases.

Motivation

Our contribution:

- An explicit formula of the irreducible expansion complexity of ultimately periodic sequences.
- A tighter upper bound for the Nth expansion complexity of arbitrary sequence with given nonlinear complexity.

Preliminaries

Let $\mathcal{S} = (s_i)$ be a sequence over $\mathbb{F}_q$.

- The *Nth linear complexity* $L_N(\mathcal{S})$ is the minimal order of the linear recurrence satisfies

$$s_{i+N} + \sum_{j=0}^{L_N-1} c_j s_{i+j} = 0, \text{ for } 0 \le i \le N - L_N - 1$$

  with $c_j \in \mathbb{F}_q$. If $s_0 = \cdots = s_{N-1} = \alpha$, then $L_N(\mathcal{S}) := 0$ when $\alpha = 0$ and $L_N(\mathcal{S}) := 1$ when $\alpha \neq 0$.

- The *linear complexity* of $\mathcal{S}$ is given by

$$L(\mathcal{S}) = \sup_{N \ge 1} L_N(\mathcal{S}).$$

Preliminaries

Let $\mathcal{S} = (s_i)$ be a sequence over $\mathbb{F}_q$.

- The *Nth nonlinear complexity* $C_N(\mathcal{S})$ is the length of the shortest NFSRs that generate $\{s_i\}_{i=0}^{N-1}$.
  If $s_0 = \cdots = s_{N-2} = \alpha$, then $C_N(\mathcal{S}) := 0$ when $s_{N-1} = \alpha$ and $C_N(\mathcal{S}) := N-1$ when $s_{N-1} \neq \alpha$.
- The *nonlinear complexity* of $\mathcal{S}$ is given by

$$C(\mathcal{S}) = \sup_{N \geq 1} C_N(\mathcal{S}).$$

- $C_N(\mathcal{S}) \leq L_N(\mathcal{S})$ for $N = 1, 2, \ldots$.

Preliminaries

Let $G(x) = \sum_{i=0}^{\infty} s_i x^i$ be the generating function of $\mathcal{S}$.

- The *Nth expansion complexity* $E_N(\mathcal{S}) := 0$ if $s_i = 0$ for $i = 0, \dots, N-1$ and otherwise the least total degree of $h(x,y) \in \mathbb{F}_q[x,y]$ with $h(x, G(x)) \equiv 0 \pmod{x^N}$.
- The *expansion complexity* of $\mathcal{S}$ is $E(\mathcal{S}) = \sup_{N \geq 1} E_N(\mathcal{S})$.
- If $h(x,y)$ is irreducible, call it the *Nth irreducible expansion complexity* $E_N^*(\mathcal{S})$ and the irreducible expansion complexity $E^*(\mathcal{S})$, respectively.
- $E_N(\mathcal{S}) \leq E_N^*(\mathcal{S}) \leq E_{N+1}^*(\mathcal{S})$ for $N = 1, 2, \dots$.

Ultimately Periodic Sequences

A sequence $\mathcal{S} = (s_i)$ over $\mathbb{F}_q$ is called *ultimately or eventually periodic* if there exist $n > 0$ and $u \geq 0$ s.t.

$$s_{i+n} = s_i \quad \text{for} \quad i \geq u. \tag{1}$$

It is said to have parameters (n, u) if n is the least period satisfying (1). When $u = 0$, call $\mathcal{S}$ a *(purely) periodic* sequence.

Example 1.Let $\mathcal{S} = (s_i)$ be the Legendre sequence of a prime period p defined by

$$s_i = \begin{cases} [1 + (\frac{i}{p})]/2, & \text{if } i \not\equiv 0 \pmod{p}; \\ 0, & \text{otherwise,} \end{cases}$$

where $(\frac{i}{p})$ is the Legendre symbol.

Ultimately Periodic Sequences

For an ultimately periodic sequence $\mathcal{S} = (s_i)$ with parameters (n, u), denote

$$s_u(x) = \sum_{i=0}^{u-1} s_i x^i \quad \text{and} \quad s_n(x) = \sum_{i=0}^{n-1} s_{i+u} x^i.$$

Then the generating function of $\mathcal{S}$ is

$$\begin{aligned} G(x) &= \sum_{i=0}^{u-1} s_i x^i + x^u \sum_{i=0}^{n-1} s_{i+u} x^i (1 + x^n + \cdots) \\ &= s_u(x) + x^u \cdot \frac{s_n(x)}{1 - x^n} \\ &= \frac{f_0(x)}{f_1(x)}, \end{aligned} \tag{2}$$

where $\gcd(f_0, f_1) = 1$. We can show that $\deg(f_1) = L(\mathcal{S}) - u$.

Ultimately Periodic Sequences

Theorem 1 (Sun-Zeng-Li-Z.-Yi 2021) Let $\mathcal{S} = (s_i)$ be an ultimately periodic sequence with parameters (n, u) and linear complexity $L(\mathcal{S})$. Set $\ell = L(\mathcal{S}) - u + \max\{1, k - n\}$, where $k \in \mathbb{Z}_{\geq 0}$ can be derived from $\{s_i\}_{i=0}^{n+u-1}$. Then the Nth irreducible expansion complexity

$$E_N^*(\mathcal{S}) = \ell \quad \text{for} \quad N > \ell(\ell - 1).$$

Moreover, the irreducible expansion complexity $E^*(\mathcal{S}) = \ell$.

Ingredients for the proof:

- Let $G(x) = f_0(x)/f_1(x)$ with $f_0(x), f_1(x)$ given in (2). Set $h(x, y) = f_1(x)y - f_0(x)$. Show that $\deg(h) = \ell$.
- Use psedo-division algorithm over $\mathbb{F}_q[x, y]$ to show that $E_N^*(\mathcal{S}) = \ell$ when N is large enough and $E^*(\mathcal{S}) = \ell$.

Ultimately Periodic Sequences

- Theorem 1 implies that $E_N(\mathcal{S}) = E_N^*(\mathcal{S})$ for $N > \ell(\ell - 1)$.

- The calculation of $E_N^*(\mathcal{S})$ can be converted to that of $L(\mathcal{S})$, which can be computed by the BM algorithm, and the comparison of some terms of $\mathcal{S}$.

- Constructing an ultimately periodic sequence $\mathcal{S} = (s_i)_{i=0}^{\infty}$ with large irreducible expansion complexity can be done by:
  - Choose an n-periodic sequence $(s_i)_{i=u}^{\infty}$ with large linear complexity.
  - Choose a pre-periodic sequence $(s_0, \ldots, s_{u-1})$ with $s_{u-1} \neq s_{u+n-1}$.

Ultimately Periodic Sequences

Example 1 (continued). Let $\mathcal{S} = (s_i)_{i=0}^{\infty}$ be the Legendre sequence of a prime period p defined by

$$s_i = \begin{cases} [1 + (\frac{i}{p})]/2, & \text{if } i \not\equiv 0 \pmod{p}; \\ 0, & \text{otherwise}, \end{cases}$$

where $(\frac{i}{p})$ is the Legendre symbol. By Theorem 1, we have

- if $p \equiv 1 \pmod{8}$, then $E_N^*(\mathcal{S}) = \frac{p+1}{2}$ for $N > \frac{(p+1)(p-1)}{4}$;
- if $p \equiv 3 \pmod{8}$, then $E_N^*(\mathcal{S}) = p + 1$ for $N > p(p+1)$;
- if $p \equiv 5 \pmod{8}$, then $E_N^*(\mathcal{S}) = p$ for $N > p(p-1)$;
- if $p \equiv 7 \pmod{8}$, then $E_N^*(\mathcal{S}) = \frac{p+3}{2}$ for $N > \frac{(p+3)(p+1)}{4}$.

Upper Bound

Let $\mathcal{S}$ be a sequence over $\mathbb{F}_q$.

Gómez-Pérez *et al*. (2018) show that the Nth expansion complexity $E_N(\mathcal{S}) \leq \sqrt{2N}$.

However, if the first N terms of $\mathcal{S}$ is $(0, 0, \ldots, 0, 1)$ and $N \geq 3$, then $C_N(\mathcal{S}) = N - 1$ but $E_N(\mathcal{S}) = 2$, which is far less than $\sqrt{2N}$.

Question 3: What is the relation between $C_N(\mathcal{S})$ and $E_N(\mathcal{S})$?

Upper Bound

Theorem 2 (Sun-Zeng-Li-Z.-Yi 2021). Let $\mathcal{S} = (s_i)$ be a sequence over $\mathbb{F}_q$. If the Nth nonlinear complexity $C_N(\mathcal{S}) = N - k$ with $1 \leq k < \sqrt{2N} - 2$, then Nth expansion complexity

$$E_N(\mathcal{S}) \leq k + 2.$$

Idea for the proof: Induction and construction theorem (Yi-Zeng-Sun 2021) of sequence $\mathcal{S}$ with $C_N(S) \geq N/2$.

- Upper bound of $E_N(\mathcal{S})$ only depends on the number of distinct subsequences of length $N - k$ appearing in $\{s_i\}_{i=0}^{N-1}$.
- Non-randomness property of $\mathcal{S}$ with large $C_N(\mathcal{S})$ may be detected by computing its $E_N(\mathcal{S})$.

Upper Bound

The upper bound in Theorem 2 is tight.

Example 2. Let $\mathcal{S}$ be a binary sequence. Its first N elements satisfy

$$s_0 = 0, s_1 = 1, s_i = s_{i+2}, 0 \leq i \leq N-4, s_{N-1} \neq s_{N-3}.$$

Then $C_N(\mathcal{S}) = N-2$ by Jansen's Ph.D. thesis. By detailed calculation, we find $E_N(\mathcal{S}) = 2$, which is exactly the upper bound in Theorem 2.

C. J. A. Jansen, *Investigations on Nonlinear Stream Cipher Systems: Construction and Evaluation Methods*, Ph.D. dissertation, Technical University of Delft, Delft, 1989.

Conclusion

- An explicit formula of the (irreducible) expansion complexity of ultimately periodic sequences over finite fields.

- A tighter upper bound of the Nth expansion complexity for arbitrary sequences with given Nth nonlinear complexity.

Conclusion

- An explicit formula of the (irreducible) expansion complexity of ultimately periodic sequences over finite fields.

- A tighter upper bound of the Nth expansion complexity for arbitrary sequences with given Nth nonlinear complexity.

Thanks!