On the pseudorandom properties of filtered Legendre symbol sequences using three polynomials

Katalin Gyarmati and Károly Müllner

Abstract

This paper presents a further development of a well-known construction that relies on polynomials and the Legendre symbol. To address theoretical security concerns regarding the original method, which used a single polynomial, we introduce a new approach that combines three different polynomials to generate the sequence, thereby enhancing its security. We prove that the sequences produced by this new construction also exhibit strong pseudorandom properties with respect to the pseudorandom measures introduced by Mauduit and Sárközy.

1 Introduction

In 1997, Mauduit and Sárközy introduced the following quantitative measures to study the pseudorandomness of finite binary sequences.

2020 Mathematics Subject Classification: Primary: 11K45.

Keywords and phrases: polynomials, binary sequences.

Research supported by the Hungarian National Research Development and Innovation Fund KKP133819.

Definition 1. *For a binary sequence*

$$E_N = \{e_1, \dots, e_N\} \in \{-1, +1\}^N,$$

define the well-distribution measure of E_N *as*

$$W(E_N) = \max_{a,b,t} \left| \sum_{j=1}^{t} e_{a+jb} \right|,$$

where the maximum is taken over all a, b, t *such that* $a \in \mathbb{Z}$, $b, t \in \mathbb{N}$ *and* $1 \le a + b \le a + tb \le N$, *while the correlation measure of order* ℓ *of* E_N *is defined as*

$$C_\ell(E_N) = \max_{M,D} \left| \sum_{n=1}^{M} e_{n+d_1} \dots e_{n+d_\ell} \right|,$$

where the maximum is taken over all $D = (d_1, \dots, d_\ell)$ *and* M *such that* $0 \le d_1 < \dots < d_\ell < M + d_\ell \le N$.

These measures characterize the essential random properties of binary sequences in various applications (such as cryptography, Monte Carlo methods, and many others). According to papers by Cassaigne, Mauduit, Sárközy [2] and later Alon, Kohayakawa, and Mauduit, Moreira, and Rödl [1], the pseudorandomness of a sequence E_n is considered to be very strong if

$$W(E_N) \ll N^{1/2} (\log N)^c,$$
$$C_\ell(E_N) \ll N^{1/2} (\log N)^{c\ell}$$

hold at least for small ℓ's. In various applications, pseudorandom constructions are of great importance. Before Mauduit and Sárközy introduced the well-distribution and correlation measures, these constructions were tested a posteriori. This meant that after generation, computers were used to check if certain statistical properties were satisfied. According to the paper by Sárközy and Rivat [21], and later a paper of Mérai, Rivat and Sárközy [20] these a posteriori tests can be avoided if the well-distribution and correlation measures of the sequence are small. Thus, it has become important to construct sequences for which the W and C_k measures are provably small. In the

literature, there are numerous constructions with strong pseudorandomness properties (e.g., see [3], [4], [5], [10], [11], [12], [14], [15], [17], [18], [19], [22]). The following construction is perhaps the most natural and still the most widely used today among all constructions proposed.

Construction 1 (Hoffstein, Liemann)**.** *Let p be an odd prime and $f(x) \in \mathbb{F}_p[x]$ be a polynomial of degree k. Define $E_p = \{e_1, \ldots, e_p\} \in \{-1, +1\}^p$ by:*

$$e_n = \begin{cases} \left(\frac{f(n)}{p}\right) & \text{for } (f(n), p) = 1, \\ +1 & \text{for } p \mid f(n). \end{cases}$$

This construction was introduced by Hoffstein and Liemann, but nothing has been proven about the pseudorandom properties of the sequences. One year later, Goubin, Mauduit and Sárközy [7] proved the following:

Theorem A [Goubin, Mauduit, Sárközy] *Let p be an odd prime and $f(x) \in \mathbb{F}_p[x]$ be a polynomial of degree k, which is not of the form $cq(x)^2$, where $c \in \mathbb{F}_p$, $q(x) \in \mathbb{F}_p[x]$. Define $E_p = \{e_1, \ldots, e_p\} \in \{-1, +1\}^p$ by Construction 1. Then,*

$$W(E_p) \le 10kp^{1/2} \log p.$$

Assume that one of the following three conditions for ℓ, which is the order of the correlation, holds:
(i) $\ell = 2$;
(ii) $\ell < p$ and 2 is a primitive root modulo p;
(iii) $(4k)^\ell < p$.
Then,

$$C_\ell(E_p) \le 10k\ell p^{1/2} \log p.$$

Although Construction 1 has strong pseudorandom properties, a potential weakness exists. It is conceivable that a future algorithm could determine the polynomial f from just p^ε consecutive elements of the sequence E_p, especially if the polynomial's degree is under a certain bound. In this case, the entire sequence (e.g., used as the secret key) might be determined from a

few elements of the sequence. A solution to this problem could be to modify the sequence by using not one, but three different polynomials. For this, we introduce the following construction:

Construction 2. *Let p be an odd prime and $f(x), g(x), h(x) \in \mathbb{F}_p[x]$ be three polynomials of degree $\leq k$. Define $E_{f,g,h} = \{e_1, \ldots, e_p\} \in \{-1, +1\}^p$ by:*

$$e_n = \begin{cases} \left(\frac{f(n)}{p}\right), & \text{if } \left(\frac{h(n)}{p}\right) \in \{0, 1\} \text{ and } p \nmid f(n)g(n) \\ \left(\frac{g(n)}{p}\right), & \text{if } \left(\frac{h(n)}{p}\right) = -1 \text{ and } p \nmid f(n)g(n) \\ 1, & \text{if } p \mid f(n)g(n). \end{cases} \tag{1}$$

Note that if $f = g$, then this construction coincides with Construction 1.

In this construction, if $p \nmid f(n)g(n)h(n)$, then the following formula can be proved:

$$e_n = \frac{1}{2}\left(1 + \left(\frac{h(n)}{p}\right)\right)\left(\frac{f(n)}{p}\right) + \frac{1}{2}\left(1 - \left(\frac{h(n)}{p}\right)\right)\left(\frac{g(n)}{p}\right). \tag{2}$$

From this formula, we will prove the following using multiplicative character techniques (e.g., Weil theorem):

Theorem 1. *Let p be an odd prime and $f(x), g(x), h(x) \in \mathbb{F}_p[x]$ be three polynomials of degrees between 1 and k, that have no multiple roots. Also assume that*

$$f(x) \nmid \prod_{t=1}^{p} g(x+t)h(x+t) \quad \text{and} \quad g(x) \nmid \prod_{t=1}^{p} h(x+t). \tag{3}$$

Define the sequence $E_{f,g,h} = \{e_1, \ldots, e_p\} \in \{-1, +1\}^p$ by Construction 2. Then,

$$W(E_{f,g,h}) \leq 10kp^{1/2} \log p. \tag{4}$$

Assume that one of the following three conditions for ℓ, which is the order of the correlation, holds:

(i) $\ell = 2$;

(ii) $\ell < p$ and 2 is a primitive root modulo p;

(iii) $(4k)^\ell < p$.

Then,

$$C_\ell(E_{f,g,h}) \leq 2^{\ell+3}\ell kp^{1/2} \log p. \tag{5}$$

For reasons of symmetry, the theorem also holds even if condition (3) is replaced by $g(x) \nmid \prod_{t=1}^{p} f(x+t)h(x+t)$ and $f(x) \nmid \prod_{t=1}^{p} h(x+t)$.

Security considerations. The construction $E_{f,g,h}$ was specifically designed to counter potential attacks aimed at reconstructing the underlying polynomial of a Legendre symbol sequence. In the original construction proposed by Goubin, Mauduit, and Sárközy [7], a single polynomial $f(n)$ is used, and the sequence elements are defined as $e_n = \left(\frac{f(n)}{p}\right)$. If an attacker can determine the values of e_n for a sufficient number of indices (roughly $\deg(f)$ values), they might be able to reconstruct the polynomial $f(x)$ using interpolation or other specialized algebraic algorithms.

In our proposed construction $E_{f,g,h}$, the attacker faces a significantly more complex task due to the following factors:

- **Triple Uncertainty:** The attacker does not know which of the three polynomials (f, g, or h) is responsible for a given element e_n. To even begin a reconstruction, one would first need to distinguish which indices n satisfy $\left(\frac{h(n)}{p}\right) = 1$ and which satisfy $\left(\frac{h(n)}{p}\right) = -1$.
- **Hidden Switching:** The "switching" polynomial $h(n)$ is itself hidden. Since e_n only reveals information about $f(n)$ or $g(n)$, the values of the Legendre symbol $\left(\frac{h(n)}{p}\right)$ are not directly observed. This adds an additional layer of protection, as the selector sequence is not public.
- **Combinatorial Explosion:** If an attacker attempts to guess the partition of N observed values into those coming from $f(n)$ and those from $g(n)$, they encounter a combinatorial explosion. For N observed elements, there are 2^N possible assignments. Without knowing the correct assignment, standard interpolation techniques for f and g cannot be applied effectively.

Thus, the interlacing of three polynomials ensures that even if a small number of values are leaked, the underlying algebraic structure remains computationally difficult to recover. This provide a significant security upgrade

over the single-polynomial case while maintaining the same level of pseudorandomness.

Thus, even though Construction 2 is slightly more complicated than Construction 1, it still provides strong bounds for the pseudorandom measures.

It is also clear that condition (3) cannot be completely dropped from Theorem 1. For instance, if $f = h$ and $g = nh$, where $n \in \mathbb{F}_p$ is a quadratic non-residue, then the elements of the sequence $E_{f,g,h}$ are all 1. The strength of Construction 2 will be further supported by numerical calculations, and we will compare the pseudorandom measures of some sequences in Constructions 1 and 2.

At first glance, checking condition (3) may be inconvenient. However, for small primes p, it does not require extensive computation and can be done using polynomial division. There are several ways to avoid this polynomial division, one of which is to use only irreducible polynomials of the form

$$x^r + a_{r-2}x^{r-2} + a_{r-3}x^{r-3} + \cdots + a_1 x + a_0$$

(so the coefficient of x^{r-1} is 0). Another possibility is to choose f, g, and h to all be products of second-degree irreducible polynomials. We will prove the following

Theorem 2. *Let p be prime, and $\mathcal{A}$, $\mathcal{B}$, $\mathcal{C}$ be sets containing only quadratic non-residues modulo p for which*

$$\mathcal{A} \not\subseteq \mathcal{B} \cup \mathcal{C} \quad \text{and} \quad \mathcal{B} \not\subseteq \mathcal{C}.$$

The polynomials f, g, and h are defined as follows.

$$f(x) = \prod_{n\in\mathcal{A}}(x^2 - n),\ g(x) = \prod_{n\in\mathcal{B}}(x^2 - n),\ h(x) = \prod_{n\in\mathcal{C}}(x^2 - n). \tag{6}$$

Define the sequence $E_{f,g,h} = \{e_1, \ldots, e_p\} \in \{-1, +1\}^p$ by Construction 2. Then,

$$W(E_{f,g,h}) \le 10kp^{1/2}\log p \tag{7}$$

$$C_\ell(E_{f,g,h}) \le 2^{\ell+3}\ell kp^{1/2}\log p. \tag{8}$$

Remark. The sequence $E_{f,g,h}$ given in Theorem 2 is symmetric, as f, g and h are even polynomials ($f(x) = f(-x),\ g(x) = g(-x),\ h(x) = h(-x)$). This implies that for every element of the sequence $e_n = e_{p-n}$ holds. Thus, for applications of the sequence given in Theorem 2, we suggest using at most the first $(p+1)/2$ elements.

The analysis of Construction 2 prompted us to consider the generalizability of its underlying principle: generating a sequence by combining elements from multiple source sequences. This line of questioning led to the development of the following construction:

Construction 3. *Let $\mathcal{F} \subset \{-1,+1\}^N$ be a large family of binary sequences. By taking binary sequences $F_N = \{f_1, f_2, \ldots, f_N\}$, $G_N = \{g_1, g_2, \ldots, g_N\}$ and $H_N = \{h_1, h_2, \ldots, h_N\} \in \mathcal{F}$, we can define a new sequence $E_N = \{e_1, e_2, \ldots, e_N\} \in \{-1,+1\}^N$ with the following formula:*

$$e_n = \begin{cases} f_n, & \text{if } h_n = 1 \\ g_n, & \text{if } h_n = -1. \end{cases} \tag{9}$$

Similarly to Construction 2, there are special cases where the new sequence has weak pseudorandom properties. For example, if for every n, $f_n = h_n$ and $g_n = -h_n$, then all elements of our sequence are 1. To avoid such extreme cases, we must assume something about the large family $\mathcal{F}$ involved in the construction. For this, we will need the so-called cross-correlation measure. This family measure was introduced by Mauduit, Sárközy, and the first author of the present paper in [9].

Definition 2. *Let $N \in \mathbb{N}$, $\ell \in \mathbb{N}$, and for any ℓ binary sequences $E_N^{(1)}, \ldots, E_N^{(\ell)}$ with*

$$E_N^{(i)} = \left(e_1^{(i)}, \ldots, e_N^{(i)}\right) \in \{-1,+1\}^N \text{ (for } i = 1, 2, \ldots, \ell)$$

and any $M \in \mathbb{N}$ and ℓ-tuple $D = (d_1, \ldots, d_\ell)$ of non-negative integers with

$$0 \le d_1 \le \cdots \le d_\ell < M + d_\ell \le N, \tag{10}$$

write

$$V_\ell\left(E_N^{(1)}, \ldots, E_N^{(\ell)}, M, D\right) = \sum_{n=1}^{M} e_{n+d_1}^{(1)} \cdots e_{n+d_\ell}^{(\ell)} \tag{11}$$

Let

$$\tilde{C}_\ell\left(E_N^{(1)},\ldots,E_N^{(\ell)}\right)=\max_{M,D}\left|V_\ell\left(E_N^{(1)},\ldots,E_N^{(\ell)},M,D\right)\right| \tag{12}$$

where the maximum is taken over all $D=(d_1,\ldots,d_\ell)$ *and* $M\in\mathbb{N}$ *satisfying* (10) *with the additional restriction that if* $E_N^{(i)}=E_N^{(j)}$ *for some* $i\neq j$*, then we must not have* $d_i=d_j$*. Then the cross-correlation measure of order* ℓ *of the family* $\mathcal{F}$ *of binary sequences* $E_N\in\{-1,+1\}^N$ *is defined as*

$$\Phi_\ell(\mathcal{F})=\max\tilde{C}_\ell\left(E_N^{(1)},\ldots,E_N^{(\ell)}\right) \tag{13}$$

where the maximum is taken over all ℓ*-tuples of binary sequences* $\left(E_N^{(1)},\ldots,E_N^{(\ell)}\right)$ *with*

$$E_N^{(i)}\in\mathcal{F}\text{ for }i=1,\ldots,\ell.$$

Then we will prove the following:

Theorem 3. *Let* $\mathcal{F}\subset\{-1,+1\}^N$ *be a large family of binary sequences. For three distinct sequences* $F_N,G_N,H_N\in\mathcal{F}$ *define the sequence* $E_N\in\{-1,+1\}^N$ *by Construction 3. For this new sequence, we have*

$$C_\ell(E_N)\leq 2^\ell\max_{\ell\leq k\leq 2\ell}\Phi_k(\mathcal{F}).$$

This theorem is particularly useful when the sequences F_N,G_N,H_N are chosen from a large family $\mathcal{F}\subset\{-1,+1\}^N$ for which $\max_{1\leq k\leq 2\ell}\Phi_k(\mathcal{F})\ll N^{1/2+\varepsilon}$. Such families can be found, for example, in [6], [8], [9], [13], and [25].

2 Numerical Calculations

While Theorems 2 and 3 provide rigorous theoretical upper bounds for the pseudorandom measures, these bounds are derived using the Weil theorem and often include large constants. The primary objective of this section is to demonstrate that the actual pseudorandom measures of our construction are significantly smaller than the theoretical upper bounds derived from the Weil theorem. Regarding the family of sequences, we note that the construction $E_{f,g,h}$ allows for a large variety of sequences by choosing different triples of

polynomials. While the detailed analysis of the cross-correlation measure of such a family is a challenging problem and lies beyond the scope of the present paper, the structure of the construction suggests that sequences generated by different polynomials will remain nearly orthogonal.

In the following, we determine and compare the measures W and C_2 for a few specific triples of polynomials (f, g, h) and prime numbers p. The sequences E_f, E_g and E_h are defined according to Construction 1 using specifically given polynomials f, g, and h in place of f in (1). Meanwhile, $E_{f,g,h}$ is the sequence from Construction 2. Thus, here are the four tables of results:

Example 1. Let the three polynomials be defined by

$$f(x) = x^2 + 1,\ g(x) = x^2 + 3x + 1,\ h(x) = x^3 - 1.$$

Then,

p	$W(E_f)$	$W(E_g)$	$W(E_h)$	$W(E_{f,g,h})$	$C_2(E_f)$	$C_2(E_g)$	$C_2(E_h)$	$C_2(E_{f,g,h})$
2003	50	55	53	59	177	182	136	122
3001	51	108	129	60	174	194	138	183
4001	139	151	102	151	247	273	200	192
5003	67	86	90	92	348	264	190	269
6007	106	84	97	116	292	348	274	237

Example 2. Let the three polynomials defined by

$$f(x) = x^2 + x + 1,\ g(x) = x^3 - x + 1 \ \text{ and } \ h(x) = x^4 + x - 1.$$

Then,

p	$W(E_f)$	$W(E_g)$	$W(E_h)$	$W(E_{f,g,h})$	$C_2(E_f)$	$C_2(E_g)$	$C_2(E_h)$	$C_2(E_{f,g,h})$
2003	48	66	72	56	169	136	139	117
3001	75	133	51	184	203	147	152	172
4001	187	45	192	110	266	189	206	159
5003	72	101	81	79	274	209	220	195
6007	147	124	131	104	294	211	204	226

Example 3. Let the three polynomials defined by

$$f(x) = x^4 - 1,\ g(x) = x^6 - 4x^3 + 3, \quad \text{and} \quad h(x) = x^3 - 6x^2 + 15x - 14.$$

Then,

p	$W(E_f)$	$W(E_g)$	$W(E_h)$	$W(E_{f,g,h})$	$C_2(E_f)$	$C_2(E_g)$	$C_2(E_h)$	$C_2(E_{f,g,h})$
2003	51	47	53	54	178	121	164	105
3001	186	80	132	146	254	181	171	190
4001	212	98	54	170	217	210	171	200
5003	62	69	84	70	281	165	183	244
6007	88	149	130	74	293	192	173	231

Example 4. Let the three polynomials defined by

$$f(x) = x^2 - 1,\ g(x) = x^3 + x^2 + 1 \quad \text{and} \quad h(x) = x^4 + x^3 + 1.$$

Then,

p	$W(E_f)$	$W(E_g)$	$W(E_h)$	$W(E_{f,g,h})$	$C_2(E_f)$	$C_2(E_g)$	$C_2(E_h)$	$C_2(E_{f,g,h})$
2003	35	64	55	62	201	126	125	132
3001	120	62	78	94	206	178	157	166
4001	198	185	118	109	236	187	225	213
5003	74	77	79	83	276	223	233	202
6007	126	83	142	115	348	226	242	268

We found that in every case we studied, we have

$$W(E_{f,g,h}) \leq 2\max\{W(E_f), W(E_g), W(E_h)\}$$

and

$$C_2(E_{f,g,h}) \leq 2\max\{C_2(E_f), C_2(E_g), C_2(E_h)\}.$$

The computational results demonstrate that, apart from a few exceptional cases, the sequences $E_{f,g,h}$ have pseudorandom properties almost as

good as those of the sequences from the original Construction 1. It is conjectured that these constructions are cryptographically secure, as recovering the full sequence from partial knowledge would necessitate computationally expensive algorithms.

3 Proofs

Proof of Theorem 1. Consider a triple of numbers a, b, t for which

$$W(E_{f,g,h}) = \left| \sum_{j=1}^{t} e_{a+jb} \right|. \tag{14}$$

Let $\mathcal{L}$ be the following set:

$$\mathcal{L} = \{n : \ p \mid f(n)g(n)h(n)\}.$$

Since the degrees of the polynomials f, g, and h are $\leq k$, we have $|\mathcal{L}| \leq 3k$. Furthermore, by (2), (14) and the triangle inequality,

$$\begin{aligned} W(E_{f,g,h}) \leq & \left| \sum_{j=1}^{t} \tfrac{1}{2} \left(1 + \left(\tfrac{h(a+jb)}{p}\right)\right) \left(\tfrac{f(a+jb)}{p}\right) + \tfrac{1}{2} \left(1 - \left(\tfrac{h(a+jb)}{p}\right)\right) \left(\tfrac{g(a+jb)}{p}\right) \right| \\ & + 2 \sum_{a+jb \in \mathcal{L}} 1 \\ & \leq \frac{1}{2} \left| \sum_{j=1}^{t} \left(\frac{f(a+jb)}{p}\right) \right| + \frac{1}{2} \left| \sum_{j=1}^{t} \left(\frac{g(a+jb)}{p}\right) \right| \\ & + \frac{1}{2} \left| \sum_{j=1}^{t} \left(\frac{f(a+jb)h(a+jb)}{p}\right) \right| + \frac{1}{2} \left| \sum_{j=1}^{t} \left(\frac{g(a+jb)h(a+jb)}{p}\right) \right| + 6k \end{aligned} \tag{15}$$

The theorem of Weil [23] on character sums and polynomials can be extended to incomplete sums using the Vinogradov method. In this extended theorem, Winterhof [24] optimized the value of the constant factor, proving the following:

Lemma 1. *Suppose that p is a prime, χ is a non-principal character modulo p of order d, $f \in \mathbb{F}_p[x]$ has s distinct roots in $\overline{\mathbb{F}}_P$, and it is not a constant multiple of the d-th power of a polynomial over $\mathbb{F}_p$. Let y be a real number with $0 < y \leq p$. Then for any $x \in \mathbb{R}$.*

$$\left| \sum_{x<n\leq x+y} \chi(f(n)) \right| < sp^{1/2}(1+\log p).$$

Since none of the polynomials f, g and h have multiple roots, none of them are of the form cq^2 where $c \in \mathbb{F}_p$ and $q \in \mathbb{F}_p[x]$. We also know that the polynomials fh or gh could only be of the form cq^2 (where $c \in \mathbb{F}_p$ and $q \in \mathbb{F}_p[x]$) if f is a constant multiple of h (for fh) or if g is a constant multiple of h (for gh). However, this contradicts (3). Thus, by applying Lemma 1, we get that

$$\left| \sum_{j=1}^{t} \left(\frac{f(a+jb)}{p} \right) \right|, \left| \sum_{j=1}^{t} \left(\frac{g(a+jb)}{p} \right) \right| < kp^{1/2}(1+\log p)$$

$$\left| \sum_{j=1}^{t} \left(\frac{f(a+jb)h(a+jb)}{p} \right) \right|, \left| \sum_{j=1}^{t} \left(\frac{g(a+jb)h(a+jb)}{p} \right) \right| < 2kp^{1/2}(1+\log p). \tag{16}$$

By (14), (15) and (16)

$$W(E_{f,g,h}) < 3kp^{1/2}(1+\log p) + 6k < 10kp^{1/2}\log p,$$

from which (4) follows.

Let's move on to the proof of (5). This inequality trivially holds if $\ell k \geq \dfrac{p^{1/2}}{13\log p}$. Thus, for the rest of the proof, we assume that

$$\ell k < \frac{p^{1/2}}{13\log p}. \tag{17}$$

Consider numbers M and $0 \leq d_1 < d_2 < \cdots < d_\ell < M + d_\ell \leq p$ for which

$$C_\ell(E_{f,g,h}) = \left| \sum_{n=1}^{M} e_{n+d_1} \ldots e_{n+d_\ell} \right|. \tag{18}$$

Let $\mathcal{H}$ be the following set

$$\mathcal{H} = \{n : \exists d_i \text{ such that } p | f(n+d_i)g(n+d_i)h(n+d_i)\}$$

Then, $|\mathcal{H}| \leq 3k\ell$. By (2) and (18) we get

$$\begin{aligned} C_\ell(E_{f,g,h}) = &\left| \frac{1}{2^\ell} \sum_{n=1}^{M} \right. \\ &\left. \prod_{j=1}^{\ell} \left(\frac{f(n+d_j) + g(n+d_j) + f(n+d_j)h(n+d_j) - g(n+d_j)h(n+d_j)}{p} \right) \right| \\ &+ 2 \sum_{n \in \mathcal{H}} 1 \\ \leq &\left| \frac{1}{2^2} \sum_{n=1}^{M} \right. \\ &\left. \prod_{j=1}^{\ell} \left(\frac{f(n+d_j) + g(n+d_j) + f(n+d_j)h(n+d_j) - g(n+d_j)h(n+d_j)}{p} \right) \right| \\ &+ 6k\ell. \end{aligned} \tag{19}$$

Expanding the product, we get a sum of 4^ℓ Legendre symbol sums of the form:

$$\sum_{n=1}^{M} \left(\frac{f(n+\widetilde{d}_1)\dots f(n+\widetilde{d}_i)g(n+d'_1)\dots g(n+d'_v)h(n+d''_1)\dots h(n+d''_z)}{p} \right),$$

where

$$i, v, z \leq \ell, \ i+v+z \leq 2\ell$$

and $\widetilde{d}_r \neq \widetilde{d}_s, d'_r \neq d'_s, d''_r \neq d''_s$. Let G be the following set of polynomials:

$$\begin{aligned} G = \{F : \ F = &f(x+\widetilde{d}_1)\dots f(x+\widetilde{d}_i)g(x+d'_1)\dots g(x+d'_v)h(x+d''_1)\dots h(x+d''_z), \\ &\text{where } i, v, z \leq \ell, \ i+v+z \leq 2\ell \text{ and } \widetilde{d}_r \neq \widetilde{d}_s, d'_r \neq d'_s, d''_r \neq d''_s\}. \end{aligned}$$

Then by (19)

$$C_\ell(E_{f,g,h}) \leq \frac{4^\ell}{2^\ell} \max_{F \in G} \left| \sum_{n=1}^{M} \left(\frac{F(n)}{p} \right) \right| + 6k\ell. \tag{20}$$

In order to apply Lemma 1, we will use the following lemma.

Lemma 2. *Suppose that the conditions of Theorem 1 hold. Then every polynomial $F \in G$ is not of the form cq^2 with $c \in \mathbb{F}_p$ and $q \in \mathbb{F}_p[x]$.*

Proof of Lemma 2. Consider a polynomial $F \in G$ of the form

$$F = f(x+\widetilde{d}_1)\dots f(x+\widetilde{d}_i)g(x+d'_1)\dots g(x+d'_v)h(x+d''_1)\dots h(x+d''_z), \tag{21}$$

where $i, v, z \le \ell,\ i+v+z \le 2\ell$ and $\widetilde{d}_r \ne \widetilde{d}_s, d'_r \ne d'_s, d''_r \ne d''_s$. We will distinguish three cases.

Case I: $i \ge 1$, i.e. in (21) the polynomial F contains a factor $f(x+\widetilde{d}_r)$.

In this case, since $f(x) \nmid \prod_{i=1}^{n} g(x+t)h(x+t)$, $f(x)$ has an irreducible factor that does not divide $\prod_{i=1}^{n} g(x+t)h(x+t)$. Let this irreducible factor be $f_0(x)$. Thus,

$$f_0(x) \nmid \prod_{t=1}^{p} g(x+t)h(x+t) \tag{22}$$

Let us introduce the following equivalence relation: Two irreducible polynomials $\phi(x), \psi(x) \in \mathbb{F}_p(x)$ are equivalent if there exists a $\tau \in \mathbb{F}_p$ such that $\phi(x) = \psi(x+\tau)$. By (22), the polynomials $g(x+d'_i)$ and $h(x+d''_i)$ have no irreducible factor that is equivalent to $f_0(x)$. Let $\overline{f}(x)$ denote the product of the irreducible factors of $f(x)$ that are equivalent to $f_0(x)$.

Then the product of the irreducible factors of $f(x+\widetilde{d}_i)$ that are equivalent to $f_0(x)$ is $\overline{f}(x+\widetilde{d}_i)$. Thus, the product of the irreducible factors of $f(x+\widetilde{d}_1)\dots f(x+\widetilde{d}_i)$ that are equivalent to $f_0(x)$ is $\overline{f}(x+\widetilde{d}_1)\overline{f}(x+\widetilde{d}_2)\dots\overline{f}(x+\widetilde{d}_i)$. Suppose that, contrary to Lemma 2, the polynomial $F(x) = f(x+\widetilde{d}_1)\dots f(x+\widetilde{d}_i)g(x+d'_1)\dots g(x+d'_v)h(x+d''_1)\dots h(x+d''_z)$ is of the form cq^2. This implies that $\overline{f}(x+\widetilde{d}_1)\overline{f}(x+\widetilde{d}_2)\dots\overline{f}(x+\widetilde{d}_i)$ is also of that form. Then, define the sequence $\widetilde{E}_p = \{\widetilde{e}_1, \dots, \widetilde{e}_p\}$ by the formula

$$\widetilde{e}_n \stackrel{\text{def}}{=} \begin{cases} \left(\dfrac{\overline{f}(n)}{p}\right) & \text{if } (n,p) = 1 \\ 1 & \text{if } (n,p) > 1 \end{cases} \tag{23}$$

then

$$\begin{aligned} C_i(\widetilde{E}_p) &\geq \left| \sum_{n=1}^{p} \widetilde{e}_{n+\widetilde{d}_1} \cdots \widetilde{e}_{n+\widetilde{d}_i} \right| \\ &\geq \left| \sum_{n=1}^{p} \left(\frac{\overline{f}(n+\widetilde{d}_1) \cdot \overline{f}(n+\widetilde{d}_i)}{p} \right) \right| - 2ik \\ &\geq p - 3ik \geq p - 3\ell k \end{aligned}$$

However, $\overline{f}$ is a polynomial that satisfies the conditions of Theorem A, and thus

$$\begin{aligned} C_i(\widetilde{E}_p) &\leq 10ikp^{1/2} \log p \\ &\leq 10\ell k p^{1/2} \log p. \end{aligned}$$

Thus,

$$\begin{aligned} p - 3\ell k &\leq 10\ell k p^{1/2} \log p \\ p &\leq 10\ell k p^{1/2} \log p + 3\ell k < 13\ell k p^{1/2} \log p \\ \frac{p^{1/2}}{13 \log p} &< \ell k, \end{aligned}$$

which contradicts (17). Thus, in Case I we proved that the polynomial F is not of the form cq^2, where $c \in \mathbb{F}_p$ and $q \in \mathbb{F}_p[x]$.

Case II: The polynomial F does not contain a factor of the form $f(x+d_r)$ and $v \geq 1$, so F is of the form $g(x+d_1') \dots g(x+d_v')h(x+d_1'') \dots h(x+d_z'')$.

In this case, $g(x)$ has an irreducible factor that is not a divisor of $\prod_{t=1}^{p} h(x+t)$. Let's call this irreducible factor $g_0(x)$. Let $\overline{g}(x)$ denote the product of the irreducible factors of $g(x)$ that are equivalent to $g_0(x)$. Then the product of the irreducible factors of $g(x+d_i')$ that are equivalent to $g_0(x)$ is $\overline{g}(x+d_i')$. Thus, the product of the irreducible factors of $g(x+d_1') \dots g(x+d_i')$ that are equivalent to $g_0(x)$ is $\overline{g}(x+d_1')\overline{g}(x+d_2') \dots \overline{g}(x+d_i')$.

Similarly to Case I, if the polynomial $F(x) = g(x+d_1') \dots g(x+d_i')h(x+d_1'') \cdots h(x+d_z'')$ is of the form $cq^2(x)$, then the product $\overline{g}(x+d_1')\overline{g}(x+$

$d_2') \dots \overline{g}(x + d_i')$ must also be of that form. We define $\widetilde{E}_p = \{\widetilde{e}_1, \dots, \widetilde{e}_p\}$ using the formula (23) (but with $\overline{g}$ instead of $\overline{f}$). This would imply, similarly to Case I, that $p - 3\ell k \leq C_i(\widetilde{E}_p) < 10\ell k p^{1/2}(\log p)$, which is a contradiction. Thus, in this case as well, the polynomial F cannot be of the form $cq^2(x)$, where $c \in \mathbb{F}_p$ and $q \in \mathbb{F}_p[X]$.

Case III: The polynomial F contains only factors of the form $h(x + d_r)$, so F is of the form $h(x + d_1'') \dots h(x + d_z'')$.

We define $\widetilde{E}_p = \{\widetilde{e}_1, \dots, \widetilde{e}_p\}$ using the formula (23) (but with h instead of $\overline{f}$), If $h(x + d_1'') \dots h(x + d_z'')$ is of the form cq^2 (with $c \in \mathbb{F}_p$, $q \in \mathbb{F}_p[x]$) then $p - 3\ell k \leq C_z\left(\widetilde{E}_p\right)$. Using Theorem A we immediately get that $C_z(\widetilde{E}_p) \leq 10zkp^{1/2} \log p$. Thus, $p - 3\ell k < 10\ell k p^{1/2} \log p$, which is a contradiction. Thus, we have proven in every case that $F(x)$ is indeed not a polynomial of the form cq^2 with $c \in \mathbb{F}_p$, $q \in \mathbb{F}_p[x]$. This completes the proof of Lemma 2.

By using Lemma 1, we get

$$\left| \sum_{n=1}^{M} \left(\frac{F(n)}{p} \right) \right| \leq 2\ell k \sqrt{p}(1 + \log p).$$

Thus, by (20)

$$\begin{aligned} C_\ell(E_f, g, h)) &\leq 2^{\ell+1} \ell k p^{1/2}(1 + \log p) + 6k\ell \\ &< 2^{\ell+3} \ell k p^{1/2} \log p, \end{aligned}$$

which completes the proof of the theorem.

Proof of Theorem 2. We want to apply Theorem 1 to the polynomials given in (6). First, let's check if the conditions of (3) in Theorem 1 are satisfied. Since $\mathcal{A} \not\subseteq \mathcal{B} \cup \mathcal{C}$, there exists an $n_1 \in \mathcal{A}$ such that $n_1 \notin \mathcal{B} \cup \mathcal{C}$. Then, we have

$$x^2 - n_1 \mid f(x) \quad \text{but} \quad x^2 - n_1 \nmid \prod_{t=1}^{T} g(x + t)h(x + t), \tag{24}$$

since the polynomial $x^2 - n_1$ is irreducible and the irreducible factors of the polynomial $g(x+t)h(x+t)$ are of the form $(x+t)^2 - n$, where $n \in \mathcal{B} \cup \mathcal{C}$. Clearly, none of the irreducible polynomials $(x+t)^2 - n = x^2 - 2tx + t^2 - n$ are identical to the irreducible polynomial $x^2 - n_1$. Thus, condition (3) indeed holds. Using Theorem 1, the result (7) follows immediately.

To prove (8), we will follow the notation and proof of Theorem 1. Let G be the following set of polynomials:

$$G = \{F : \ F = f(x+\widetilde{d}_1)\dots f(x+\widetilde{d}_i)g(x+d'_1)\dots g(x+d'_v)h(x+d''_1)\dots h(x+d''_z),$$
$$\text{where } i, v, z \le \ell, \ i+v+z \le 2\ell \text{ and } \widetilde{d}_r \ne \widetilde{d}_s, d'_r \ne d'_s, d''_r \ne d''_s\}.$$

Then by (19),

$$C_\ell(E_{f,g,h}) \le \frac{4^\ell}{2^\ell} \max_{F \in G} \left| \sum_{n=1}^{M} \left(\frac{F(n)}{P} \right) \right| + 6k\ell. \tag{25}$$

In order to apply Lemma 1, we must show that for any $F \in G$, the polynomial F is not of the form cq^2, where $c \in \mathbb{F}_p$ and $q \in \mathbb{F}_p[x]$. To do this, let's consider a polynomial $F \in G$ of the form

$$F = f(x+\widetilde{d}_1)\dots f(x+\widetilde{d}_i)g(x+d'_1)\dots g(x+d'_v)h(x+d''_1)\dots h(x+d''_z), \tag{26}$$

where $i, v, z \le \ell, \ i+v+z \le 2\ell$ and $\widetilde{d}_r \ne \widetilde{d}_s, d'_r \ne d'_s, d''_r \ne d''_s$. Next we distinguish three cases.

Case I: $i \ge 1$, i.e. in (21) the polynomial F contains a factor $f(x+\widetilde{d}_r)$.

Case II: The polynomial F does not contain factor $f(x+\widetilde{d}_r)$ and $v \ge 1$, so F is of the form $g(x+d'_1)\dots g(x+d'_v)h(x+d''_1)\dots h(x+d''_z)$.

Case III: The polynomial F contains only factors of the form $h(x+d_r)$, so F is of the form $h(x+d''_1)\dots h(x+d''_z)$.

First, we prove that in Case I F is not of the form cq^2, where $c \in \mathbb{F}_p$ and $q \in \mathbb{F}_p[X]$. Let $f_0(x) = x^2 - n_1$, where $n_1 \in \mathcal{A}$ but $n_1 \notin \mathcal{B} \cup \mathcal{C}$. By

(24), $g(x+d_1')\dots g(x+d_v')h(x+d_1'')\dots h(x+d_z'')$ has no irreducible factor equivalent to $f_0(x)$. It's also clear that $f_0(x)$ is not equivalent to any other irreducible factor of $f(x)$. Then the irreducible factor of $f(x+\widetilde{d}_r)$ that is equivalent to $f_0(x)$ is $f_0(x+\widetilde{d}_r)$. Thus, the product of the irreducible factors of $f(x+\widetilde{d}_1)\dots f(x+\widetilde{d}_i)$ that are equivalent to $f_0(x)$ is

$$\begin{aligned} &f_0(x+\widetilde{d}_1)f_0(x+\widetilde{d}_2)\dots f_0(x+\widetilde{d}_i) \\ &\qquad = \left((x+\widetilde{d}_1)^2-n_1\right)\left((x+\widetilde{d}_2)^2-n_1\right)\cdots\left((x+\widetilde{d}_i)^2-n_1\right). \end{aligned} \tag{27}$$

Thus, if the polynomial $F(x) = f(x+\widetilde{d}_1)\dots f(x+\widetilde{d}_i)g(x+d_1')\dots g(x+d_v')h(x+d_1'')\dots h(x+d_z'')$ is of the form $cq^2(x)$, then the polynomial in (27) must also be of that form. But this contradicts the fact that the irreducible polynomials $(x+\widetilde{d}_1)^2-n_1$, $(x+\widetilde{d}_2)^2-n_1,\dots,(x+\widetilde{d}_i)^2-n_1$ are distinct. This completes the proof for Case I, showing that F is not of the form cq^2, where $c\in\mathbb{F}_p$ and $q\in\mathbb{F}_p[X]$.

Cases II and III can be handled similarly to Case I. Thus, in all the three cases, we get that $F\in G$ is not of the form cq^2 with $c\in\mathbb{F}_p$, $q\in\mathbb{F}_p[x]$. Thus, we can apply Lemma 1, from which

$$\left|\sum_{n=1}^{M}\left(\frac{F(n)}{p}\right)\right| \le 2\ell k p^{1/2}(1+\log p)$$

Thus, by (25)

$$\begin{aligned} C_\ell(E_{f,g,h}) &\le 2^{\ell+1}\ell k p^{1/2}(1+\log p)+6k\ell \\ &< 2^{\ell+3}\ell k p^{1/2}\log p, \end{aligned}$$

which completes the proof of the theorem.

Proof of Theorem 3. Based on formula (9), the elements of the sequence $E_N=\{e_1,\dots,e_N\}$ satisfy

$$\begin{aligned} e_n &= \frac{1}{2}\left((1+h_n)f_n+(1-h_n)g_n\right) \\ &= \frac{1}{2}\left(f_n+g_n+f_nh_n-g_nh_n\right). \end{aligned} \tag{28}$$

Consider numbers M and $0 \le d_1 < d_2 < \cdots < d_\ell < M + d_\ell \le p$ for which

$$C_\ell(E_{f,g,h}) = \left| \sum_{n=1}^{M} e_{n+d_1} \dots e_{n+d_\ell} \right| \tag{29}$$

By (28) and (29)

$$C_\ell(E_N) = \frac{1}{2^\ell} \left| \sum_{n=1}^{M} \prod_{j=1}^{\ell} \left(f_{n+d_j} + g_{n+d_j} + f_{n+d_j} h_{n+d_j} - g_{n+d_j} h_{n+d_j} \right) \right|. \tag{30}$$

Expanding the product yields a sum of 4^ℓ terms, each of the form:

$$\sum_{n=1}^{M} f_{n+\widetilde{d}_1} \cdots f_{n+\widetilde{d}_i} g_{n+d'_i} \cdots g_{n+d'_v} h_{n+d''_1} \cdots h_{n+d''_\ell},$$

where

$$i, v, z \le \ell, \ i + v + z \le 2\ell$$

and $\widetilde{d}_r \ne \widetilde{d}_s, d'_r \ne d'_s, d''_r \ne d''_s$. By the definition of the cross-correlation measure, each of these sums has an absolute value $\le \max_{\ell \le k \le 2\ell} \Phi_k(\mathcal{F})$. Thus, by (30) we get

$$\begin{aligned} C_\ell(E_N) &\le \frac{4^\ell}{2^\ell} \max_{\ell \le k \le 2\ell} \Phi_k(\mathcal{F}) \\ &= 2^\ell \max_{\ell \le k \le 2\ell} \Phi_k(\mathcal{F}), \end{aligned}$$

which completes the proof.

References

[1] N. Alon, Y. Kohayakawa, C. Mauduit, C. G. Moreira, and V. Rödl, *Measures of pseudorandomness for finite sequences: typical values*, Proc. Lond. Math. Soc. 95(3) (2007), 778–812.

[2] J. Cassaigne, C. Mauduit, and A. Sárközy *On finite pseudorandom binary sequences VII: The measures of pseudorandomness*, Acta Arith. 103 (2) (2001), 97–118.

[3] Z.-X. Chen, *Elliptic curve analogue of Legendre sequences*, Monatsh. Math. 154 (2008), 1–10.

[4] Z. Chen, S. Li, and G. Xiao, *Construction of pseudorandom binary sequences from elliptic curves by using the discrete logarithms*, in: Sequences and their applications - SETA 2006, LNCS 4086, Springer, 2006; pp. 285–294.

[5] Z. Chen, A. Ostafe, and A. Winterhof, *Structure of pseudorandom numbers derived from Fermat quotients*, Lecture Notes in Comput. Sci., 6087, Springer, Berlin, 2010, Arithmetic of finite fields, 73–85.

[6] K. Doğan, M. Şahin, and O. Yayla, *Families of sequences with good family complexity and cross-correlation measure*, AIMS Mathematics 10(1) 2025, 38–55.

[7] L. Goubin, C. Mauduit and A. Sárközy, *Construction of large families of pseudorandom binary sequences*, J. Number Theory 106 (2004), 56-69.

[8] K. Gyarmati, *On the Cross-Combined Measure of Families of Binary Lattices and Sequences*, In: Pomykala, J; Pieprzyk, J; Kaczorowski, J (szerk.) Number-Theoretic Methods in Cryptology: First International Conference, NuTMiC 2017, Warsaw, Poland, Lecture Notes in Computer Science 10737 (2018), 217–238.

[9] K. Gyarmati, C. Mauduit, and A. Sárközy, *The cross-correlation measure for families of binary sequences*, Applications of Algebra and Number Theory (Lectures on the occasion of Harald Niederreiter's 70th Birthday), eds.: G. Larcher et al., Cambridge University Press, Camebridge, (2014), 126–143.

[10] K. Gyarmati, A. Pethő, and A. Sárközy, *On linear recursion and pseudorandomness*, Acta Arith. 118(4) (2005), 359–374.

[11] H. Liu, *New pseudorandom sequences constructed using multiplicative inverses*, Acta Arith. 125 (2006), 11–19.

[12] H. Liu, *A family of pseudorandom binary sequences constructed by the multiplicative inverse*, Acta Arith. 130 (2007), 167–180.

[13] H. Liu and X Liu, *Binary sequence family with both small cross-correlation and large family complexity*, Finite Fields Appl. 97, Article 102440 (2024).

[14] S. Louboutin, J. Rivat and A. Sárközy, *On a problem of D. H. Lehmer*, Proc. Amer. Math. Soc. 135 (2007), 969–975.

[15] C. Mauduit, J. Rivat and A. Sárközy, *Construction of pseudorandom binary sequences using additive characters*, Monatsh. Math. 141 (2004), 197–208.

[16] C. Mauduit and A. Sárközy, *On finite pseudorandom binary sequences, I. Measures of pseudorandomness, the Legendre symbol*, Acta Arith. 82 (4) (1997), 365–377.

[17] L. Mérai, *A construction of pseudorandom binary sequences using both additive and multiplicative characters*, Acta Arith. 139 (2009), 241–252.

[18] L. Mérai, *A construction of pseudorandom binary sequences using rational functions*, Unif. Distrib. Theory 4 (2009), 35–49.

[19] L. Mérai, *Construction of large families of pseudorandom binary sequences*, Ramanujan J. 18 (2009), 341–349.

[20] L. Mérai, J. Rivat, and A. Sárközy, *The measures of pseudorandomness and the NIST tests*, Kaczorowski, Jerzy (ed.) et al., Number-theoretic methods in cryptology. First international conference, NuTMiC 2017, Warsaw, Poland, Cham: Springer. Lect. Notes Comput. Sci. 10737, (2018), 197–216.

[21] J. Rivat and A. Sárközy, *On pseudorandom sequences and their application*, Ahlswede, Rudolf (ed.) et al., General theory of information transfer and combinatorics. Berlin: Springer, Lecture Notes in Computer Science 4123, (2006) 343–361.

[22] J. Rivat and A. Sárközy, *Modular constructions of pseudorandom binary sequences with composite moduli*, Period. Math. Hungar. 51 (2005), 75–107.

[23] A. Weil, *Sur les courbes algébriques et les variétés qui s'en déduisent*, Act. Sci. et Ind. 1041, Publ. Inst. Math. Univ. Strasbourg 7 (1945). Paris: Hermann & Cie. iv, 85 p. (1948).

[24] A. Winterhof, *Some estimates for character sums and applications*, Des. Codes Cryptography 22, No. 2, 123–131 (2001).

[25] A. Winterhof and O. Yayla, *Family complexity and cross-correlation measure for families of binary sequences*, Ramanujan J. 39(3), 639–645 (2016).

Katalin Gyarmati
Eötvös Loránd University, Institute of Mathematics,
H-1117 Budapest Pázmány Péter sétány 1/C, Hungary
Email: katalin.gyarmati@gmail.com

Károly Müllner
Eötvös Loránd University, Institute of Mathematics,
H-1117 Budapest Pázmány Péter sétány 1/C, Hungary
Email: mullni@student.elte.hu