On the arithmetic autocorrelation of the Legendre sequence

Richard Hofer and Arne Winterhof

Johann Radon Institute for Computational and Applied Mathematics,
Austrian Academy of Sciences, Altenberger Str. 69, 4040 Linz, Austria
E-mail: {richard.hofer,arne.winterhof}@oeaw.ac.at

Abstract

The Legendre sequence possesses several desirable features of pseudorandomness in view of different applications such as a high linear complexity (profile) for cryptography and a small (aperiodic) autocorrelation for radar, gps, or sonar. Here we prove the first nontrivial bound on its arithmetic autocorrelation, another figure of merit introduced by Mandelbaum for error-correcting codes.

Keywords: arithmetic autocorrelation, Legendre sequence, pseudorandom sequences, pattern distribution

Mathematics Subject Classification: 94A55 (11T71, 94A05, 94A60)

1 Introduction

For a prime $p > 2$ let (ℓ_n) be the *Legendre sequence* defined by

$$\ell_n = \begin{cases} 1 & \text{if } \left(\frac{n}{p}\right) = 1, \\ 0 & \text{otherwise,} \end{cases} \qquad n \geq 0, \tag{1}$$

where $\left(\frac{\cdot}{\cdot}\right)$ is the Legendre symbol. Obviously, (ℓ_n) is p-periodic.

The Legendre sequence satisfies several desirable features of pseudorandomness. For example, Turyn [13] proved that it has a high linear complexity, see also [3] and [1, Chapter 9.3]. It also provides a high linear complexity profile, see [12, Theorem 9.2]. It is well known (see [10], [11]) that the (periodic) autocorrelation of the Legendre sequence is two-valued or three-valued,

$$\sum_{n=0}^{p-1}(-1)^{\ell_n+\ell_{n+t}} = \begin{cases} p & \text{if } t=0, \\ -1-\left(\frac{t}{p}\right)\left(1+(-1)^{\frac{p-1}{2}}\right) & \text{if } 1 \le t \le p-1, \end{cases}$$

depending on whether $p \equiv 3 \pmod 4$ or $p \equiv 1 \pmod 4$, and that the absolute value of the aperiodic autocorrelation

$$\sum_{n=0}^{M-1}(-1)^{\ell_n+\ell_{n+t}}$$

is of order of magnitude at most $p^{1/2}\log p$ for $1 \le t \le p-1$ and $1 \le M \le p-1$. Moreover, Mauduit and Sárközy [9] studied correlations of higher order. Ding [2] studied the pattern distribution of the Legendre sequence. More precisely, for $i_0, i_1, \ldots, i_{s-1} \in \{0,1\}$ and $0 < d_1 < d_2 < \ldots < d_{s-1} < p$, put

$$N = |\{0 \le n \le p-1 : \ell_n = i_0, \ell_{n+d_1} = i_1, \ldots, \ell_{n+d_{s-1}} = i_{s-1}\}|.$$

In [2, Proposition 2] Ding proved that

$$\left|N - \frac{p}{2^s}\right| \le \frac{p^{1/2}(2^{s-1}(s-3)+2)+2^{s-1}(s+1)-1}{2^s}. \tag{2}$$

In this article we study a different notion of autocorrelation, the arithmetic autocorrelation introduced by Mandelbaum [8]. Also, see the recent monograph by Goresky and Klapper [7] for more background and results on arithmetic correlations. Sequences with small arithmetic autocorrelation can be used to define codes over the integers (instead of finite fields) that can correct many errors. For an eventually T-periodic binary sequence (s_n) with preperiod T_0, that is $s_{n+T} = s_n$ for all $n \ge T_0$, the *imbalance* $Z(s_n)$ is defined by

$$Z(s_n) = N_0 - N_1,$$

where

$$N_i = |\{T_0 \le n \le T_0+T-1 : s_n = i\}|, \qquad i = 0, 1.$$

The *arithmetic autocorrelation function* $A(t)$ of a (purely) T-periodic binary sequence (a_n) is defined as follows. For $t \in \{1, 2, \ldots, T-1\}$ let (a_{n+t}) be the shift of (a_n) by lag t. Put

$$x_t = \sum_{n=0}^{T-1} a_{n+t}2^n \quad \text{and} \quad \alpha_t = \sum_{n=0}^{\infty} a_{n+t}2^n, \qquad 0 \le t < T. \tag{3}$$

Note that with respect to the 2-norm of $\mathbb{Q}$, that is

$$|x|_2 = 2^{-k} \quad \text{if } x = 2^k \frac{u}{v} \in \mathbb{Q} \setminus \{0\} \text{ with odd } u \text{ and } v,$$

the geometric series $\sum_{n=0}^{\infty} x^n$ converges for any even integer x to

$$\sum_{n=0}^{\infty} x^n = -\frac{1}{x-1}, \qquad |x|_2 < 1.$$

In particular we have $\sum_{n=0}^{\infty} 2^n = -1$, or more general

$$\sum_{n=0}^{\infty} 2^{nk} = -\frac{1}{2^k - 1}, \qquad k = 1, 2, \ldots$$

and therefore we get

$$\alpha_t = \sum_{n=0}^{T-1} a_{n+t} 2^n \sum_{m=0}^{\infty} 2^{mT} = -\frac{x_t}{2^T - 1}, \qquad 0 \le t < T.$$

We write

$$\alpha_0 - \alpha_t = \sum_{n=0}^{\infty} s_{n,t} 2^n \tag{4}$$

with unique $s_{n,t} \in \{0, 1\}$.

If $x_0 \ge x_t$, note that $(s_{n,t})$ is (purely) periodic with period T since

$$\sum_{n=0}^{\infty} s_{n,t} 2^n = (x_0 - x_t) \sum_{n=0}^{\infty} 2^{nT}.$$

If $x_0 < x_t$, note that

$$0 < \sum_{n=0}^{T-1} s_{n,t} 2^n = 2^T + \sum_{n=0}^{T-1} (a_n - a_{n+t}) 2^n = 2^T + x_0 - x_t < 2^T,$$

and thus $(s_{n,t})$ is eventually periodic with period T from T on (see also Goresky and Klapper [4, Proposition 2]) since

$$\sum_{n=T}^{\infty} s_{n,t} 2^{n-T} = -1 + \sum_{n=0}^{\infty} (a_n - a_{n+t}) 2^n = (2^T - 1 + x_0 - x_t) \sum_{n=0}^{\infty} 2^{nT}. \tag{5}$$

In both cases we define

$$A(t) = Z(s_{n,t}), \qquad 1 \le t \le T - 1.$$

For the Legendre sequence (ℓ_n) we will prove that

$$|A(t)| \le 4p^{3/4} (\log_2 p)^{1/2}, \qquad 1 \le t \le p - 1,$$

which is the main result of this article. For very small $\min\{t, p-t\}$ we improve this bound.

We start with a preliminary result on the symmetry of the arithmetic autocorrelation and add its proof for the convenience of the reader.

2 Symmetry of the arithmetic autocorrelation

Proposition 1. *The arithmetic autocorrelation function of a periodic binary sequence* (a_n) *of least period* T *satisfies*

$$A(t) = -A(T-t) \quad \textit{for } 1 \le t \le T-1.$$

Proof. For $0 \le t < T$, let x_t and α_t be defined by (3). If $x_0 > x_t$, then we have

$$-2^{T+t} < \sum_{n=0}^{T+t-1} (a_n - a_{n+T-t})2^n = \sum_{n=0}^{t-1} (a_n - a_{n+T-t})2^n - 2^t(x_0 - x_t) < 0.$$

Hence,

$$\alpha_0 - \alpha_{T-t} = \underbrace{2^{T+t} + \sum_{n=0}^{T+t-1} (a_n - a_{n+T-t})2^n}_{< 2^{T+t}} + 2^t \sum_{n=T}^{\infty} (1 - s_{n,t})2^n = \sum_{n=0}^{\infty} s_{n,T-t}2^n$$

with $(s_{n,k})$ defined by (4). Both $(s_{n,t})$ and $(s_{n,T-t})$ are (eventually) periodic with period T from T on and the number of ones in a period of $(s_{n,t})$ equals the number of zeros in a period of $(s_{n,T-t})$. Hence,

$$A(T-t) = Z(s_{n,T-t}) = -Z(s_{n,t}) = -A(t), \qquad t = 1, \ldots, T-1.$$

If $x_0 < x_t$, then we have

$$2^{T+t} > \sum_{n=0}^{T+t-1} (a_n - a_{n+T-t})2^n = \sum_{n=0}^{t-1} (a_n - a_{n+T-t})2^n - 2^t(x_0 - x_t) > 0$$

and thus

$$\alpha_0 - \alpha_{T-t} = \sum_{n=0}^{T+t-1} (a_n - a_{n+T-t})2^n + 2^t \sum_{n=T}^{\infty} (1 - s_{n,t})2^n = \sum_{n=0}^{\infty} s_{n,T-t}2^n$$

by (5) and the result follows as in the first case. □

3 A bound on the arithmetic autocorrelation of the Legendre sequence

For $t = 1$ the arithmetic autocorrelation of the Legendre sequence (ℓ_n) is easy to determine. Then

$$x_0 - x_1 = x_0/2 = x_1 = \sum_{n=0}^{p-1} \ell_{n+1}2^n,$$

$N_0 = N_1 + 1 = (p+1)/2$ and thus

$$A(1) = 1 = -A(p-1). \tag{6}$$

Now we deal with any $1 \le t \le p-1$.

Theorem 2. *The arithmetic autocorrelation function of the p-periodic sequence (ℓ_n) defined by (1) satisfies*

$$|A(t)| \le \begin{cases} 4p^{3/4}(\log_2 p)^{1/2} & \text{if } r > m, \\ 2^r(4\log_2 p + 2(m^2 - r^2))p^{1/2} & \text{if } r \le m, \end{cases}$$

where $m = \lfloor 1/4 \log_2 p - 1/2 \log_2 \log_2 p \rfloor$ and $r = \min\{t, p-t\}$ for $1 \le t \le p-1$.

Proof. By (6) and Proposition 1 we may assume $2 \le t \le (p-1)/2$. In the following we derive a lower bound on the number N_1 of ones in a period of the p-periodic sequence $(s_{n,t})$ defined by (4).

If $p \le 4p^{3/4}(\log_2 p)^{1/2}$ or $p \le 2^t(4\log_2 p + 2(m^2 - t^2))p^{1/2}$, respectively, then the result follows immediately since the trivial bound $|A(t)| \le p$ always holds. Thus it is enough to prove the inequality for $p^{1/4} > 4(\log_2 p)^{1/2}$ or $p^{1/2} > 2^t(4\log_2 p + 2(m^2 - t^2))$, respectively.

Note that $1 \le m \le 1/4 \log_2 p$. Take $a \in \{0,1\}$. For some k and m with $0 \le k < m$ and $p \le n < 2p$ assume

$$\begin{aligned} &(\ell_{n-k-1}, \ell_{n-k-1+t}) = (a, 1-a), \\ &\ell_{n-k+j} = \ell_{n-k+j+t}, \quad j = 0, \dots, k-1, \\ &(\ell_n, \ell_{n+t}) \in \{0,1\}^2. \end{aligned} \tag{7}$$

We consider only patterns of length $4 \le s = 2k+4 \le 1/2 \log_2 p + 2$ and therefore we can further estimate (2) by $sp^{1/2}/2$, that is

$$\left| N - \frac{p}{2^s} \right| \le \frac{p^{1/2}(2^{s-1}(s-3)+2) + 2^{s-1}(s+1) - 1}{2^s} \le \frac{s}{2} p^{1/2} \tag{8}$$

since $p^{1/4} > 4(\log_2 p)^{1/2}$ or $p^{1/2} > 2^t(4\log_2 p + 2(m^2 - t^2))$, respectively.

First we assume $m+1 \le t \le (p-1)/2$. From (8) we know that (for fixed a) the number of patterns

$$\begin{pmatrix} \ell_{n-k-1} & \ell_{n-k} & \dots & \ell_{n-1} & \ell_n \\ \ell_{n-k-1+t} & \ell_{n-k+t} & \dots & \ell_{n-1+t} & \ell_{n+t} \end{pmatrix} \tag{9}$$

satisfying the assumptions (7) in

$$\begin{matrix} \ell_{p-k-1} & \ell_{p-k} & \dots & \ell_{p-1} & \ell_p & \dots & \ell_{2p-2} & \ell_{2p-1} \\ \ell_{t+p-k-1} & \ell_{t+p-k} & \dots & \ell_{t+p-1} & \ell_{t+p} & \dots & \ell_{t+2p-2} & \ell_{t+2p-1} \end{matrix} \tag{10}$$

is at least $p/2^{2k+4} - (k+2)p^{1/2}$. We have to distinguish between two cases.

If $a = 1$, then $(\ell_{n-k-1}, \ell_{n-k-1+t}) = (1,0)$. The subtraction of 0 from 1 gives no carry, no matter if there was a carry in the previous step. Hence

$$s_{n,t} = \begin{cases} 1 & \text{if } \ell_n \neq \ell_{n+t}, \\ 0 & \text{if } \ell_n = \ell_{n+t}. \end{cases}$$

Since there are 2^{k+1} possible choices for the pattern (9) we count at least $p/2^{k+3} - (k+2)2^{k+1}p^{1/2}$ different $p \leq n < 2p$ with $s_{n,t} = 1$.

If $a = 0$, then $(\ell_{n-k-1}, \ell_{n-k-1+t}) = (0,1)$. The subtraction of 1 from 0 gives a carry, no matter if there was a carry in the previous step. Hence

$$s_{n,t} = \begin{cases} 1 & \text{if } \ell_n = \ell_{n+t}, \\ 0 & \text{if } \ell_n \neq \ell_{n+t}. \end{cases}$$

Just as before there are 2^{k+1} possible choices for the pattern (9) and so we get at least $p/2^{k+3} - (k+2)2^{k+1}p^{1/2}$ additional n with $s_{n,t} = 1$.

Thus in total we have at least $p/2^{k+2} - (k+2)2^{k+2}p^{1/2}$ different $p \leq n < 2p$ with $\ell_{n-k-1} \neq \ell_{n-k-1+t}$, $(\ell_{n-k+j}, \ell_{n-k+j+t}) \in \{(0,0),(1,1)\}$ for $j = 0, \ldots, k-1$ and $s_{n,t} = 1$.

Summing up all the contributions we get the formula

$$N_1 \geq \frac{1}{4}\left(\sum_{k=0}^{m-1} 2^{-k}\right) p - 2\left(\sum_{k=0}^{m-1} 2^{k+1}(k+2)\right) p^{1/2}.$$

The first sum on the right hand side of the inequality is a geometric series, hence we have

$$\frac{1}{4}\sum_{k=0}^{m-1} \frac{1}{2^k} = \frac{1}{2} - 2^{-m-1}.$$

The second sum can be estimated by

$$\sum_{k=0}^{m-1} 2^{k+1}(k+2) = m2^{m+1} \leq 2^{m-1}\log_2 p$$

where we used $m \leq 1/4 \log_2 p$. Thus by the definition of m we get

$$\begin{aligned} N_1 &\geq \frac{1}{2}p - 2^{-m-1}p - 2^m p^{1/2}\log_2 p \\ &\geq \frac{p}{2} - p^{3/4}(\log_2 p)^{1/2} - p^{3/4}(\log_2 p)^{1/2} = \frac{p}{2} - 2p^{3/4}(\log_2 p)^{1/2}. \end{aligned}$$

Analogously N_0 can be bounded below by

$$N_0 \geq \frac{p}{2} - 2p^{3/4}(\log_2 p)^{1/2}$$

and therefore since $N_0 + N_1 = p$

$$|A(t)| = |N_0 - N_1| = |p - 2N_1| = |p - 2N_0| \le 4p^{3/4}(\log_2 p)^{1/2}.$$

Now we assume $2 \le t \le m$, that means some indices in (9) coincide and so we have to deal with shorter patterns. From (8) we know that (for fixed a) the number of patterns (9) satisfying the assumptions (7) in (10) is at least

$$\begin{aligned} &p/2^{2k+4} - (k+2)p^{1/2}, && k \le t-2,\\ &p/2^{k+t+2} - \frac{k+t+2}{2}p^{1/2}, && k \ge t-1. \end{aligned}$$

Similarly as before if $a = 1$, then

$$s_{n,t} = \begin{cases} 1 & \text{if } \ell_n \ne \ell_{n+t},\\ 0 & \text{if } \ell_n = \ell_{n+t}, \end{cases}$$

and if $a = 0$, then

$$s_{n,t} = \begin{cases} 1 & \text{if } \ell_n = \ell_{n+t},\\ 0 & \text{if } \ell_n \ne \ell_{n+t}. \end{cases}$$

For each case we have 2^{k+1} possible choices for the pattern (9) if $k \le t-2$ and 2^{t-1} possible choices if $k \ge t-1$ and thus in total we count at least

$$\begin{aligned} &p/2^{k+2} - (k+2)2^{k+2}p^{1/2}, && k \le t-2,\\ &p/2^{k+2} - (k+t+2)2^{t-1}p^{1/2}, && k \ge t-1, \end{aligned}$$

different $p \le n < 2p$ with $\ell_{n-k-1} \ne \ell_{n-k-1+t}$, $(\ell_{n-k+j}, \ell_{n-k+j+t}) \in \{(0,0),(1,1)\}$ for $j = 0, \dots, k-1$ and $s_{n,t} = 1$.

Put $m' = 2m - t + 1$. Summing up all the contributions we get

$$\begin{aligned} N_1 &\ge \frac{1}{4}\left(\sum_{k=0}^{m'-1} 2^{-k}\right) p - 2\left(\sum_{k=0}^{t-2} 2^{k+1}(k+2) + 2^{t-2}\sum_{k=t-1}^{m'-1}(k+t+2)\right) p^{1/2}\\ &= \frac{p}{2} - 2^{-m'-1}p - 2(2^t(t-1) + 2^{t-3}((m')^2 + (2t+3)m' + 2 + t - 3t^2))p^{1/2}\\ &= \frac{p}{2} - 2^{-m'-1}p - 2^{t-2}((m'+t-1)^2 + 5m' - 4t^2 + 11t - 7)p^{1/2}\\ &\ge \frac{p}{2} - 2^{-2m+t-2}p - 2^{t-2}(4m^2 - 4t^2 + 16m)p^{1/2}\\ &\ge \frac{p}{2} - 2^{-2m+t-2}p - 2^{t-1}(2(m^2 - t^2) + 2\log_2 p)p^{1/2} \end{aligned}$$

where we used $m \le 1/4 \log_2 p$. Thus by the definition of m

$$N_1 \ge \frac{p}{2} - 2^{t-1}(4\log_2 p + 2(m^2 - t^2))p^{1/2}.$$

Analogously N_0 can be bounded below by

$$N_0 \geq \frac{p}{2} - 2^{t-1}(4\log_2 p + 2(m^2 - t^2))p^{1/2}$$

and therefore

$$|A(t)| = |N_0 - N_1| \leq 2^t(4\log_2 p + 2(m^2 - t^2))p^{1/2}.$$

Thus the result follows. □

4 Final remarks

For fixed $1 \leq t < T$, Goresky and Klapper [5, 6] proved that the expected arithmetic autocorrelation, averaged over all binary sequences of period T, is

$$\frac{T}{2^{T-\gcd(t,T)}}.$$

Sequences with *ideal arithmetic autocorrelation* equal to zero for all nontrivial shifts t are known, see [4]. However, the maximum absolute value of the (classical) autocorrelation of these so-called *ℓ-sequences* equals the period since the second half of a period is the bit-wise complement of the first half [4, Proposition 1]. Hence, these sequences are far away from looking random. In contrast to these sequences, the Legendre sequence of (almost) perfect (classical) autocorrelation still guarantees a rather small arithmetic autocorrelation with respect to its period p if p is sufficiently large.

The following table of maximum absolute values of the arithmetic autocorrelation of the Legendre sequence of period p for all primes $p < 150$ may lead to the conjecture that it is bounded by $p^{1/2}\ln p$ which we actually checked for all primes $p < 1000$:

p	3	5	7	11	13	17	19	23	29	31	37	41
$\max\limits_{1\leq t<p} \lvert A(t)\rvert$	1	3	3	5	7	7	9	9	7	13	15	15
$\lfloor p^{1/2}\ln p\rfloor$	1	3	5	7	9	11	12	15	18	19	21	23

p	43	47	53	59	61	67	71	73	79	83	89	97
$\max\limits_{1\leq t<p} \lvert A(t)\rvert$	17	15	13	17	15	17	17	13	23	21	21	27
$\lfloor p^{1/2}\ln p\rfloor$	24	26	28	31	32	34	35	36	38	40	42	45

p	101	103	107	109	113	127	131	137	139	149
$\max\limits_{1\leq t<p} \lvert A(t)\rvert$	21	23	23	21	25	35	29	27	27	27
$\lfloor p^{1/2}\ln p\rfloor$	46	47	48	48	50	54	55	57	58	61

5 Conclusion

We showed that the Legendre sequence of period p has a maximal (absolute value of the) arithmetic autocorrelation of order of magnitude at most $p^{3/4}\log_2 p$. Besides many previously known nice properties including a very small (classical) autocorrelation, this is another desirable feature of pseudorandomness (for sufficiently large p).

Acknowledgement

The authors are partially supported by the Austrian Science Fund FWF Project 5511-N26 which is part of the Special Research Program ”Quasi-Monte Carlo Methods: Theory and Applications”.

References

[1] T. W. Cusick, C. Ding, A. Renvall, *Stream ciphers and number theory*, revised ed., North-Holland Mathematical Library 66, Elsevier Science B. V., Amsterdam, 2004.

[2] C. Ding, *Pattern distributions of Legendre sequences*, IEEE Trans. Inform. Theory 44, 1693–1698, 1998.

[3] C. Ding, T. Helleseth, W. Shan, *On the linear complexity of Legendre sequences*, IEEE Trans. Inform. Theory 44, 1276–1278, 1998.

[4] M. Goresky, A. Klapper, *Arithmetic crosscorrelations of feedback with carry shift register sequences*, IEEE Trans. Inform. Theory 43, 1342–1345, 1997.

[5] M. Goresky, A. Klapper, *Some results on the arithmetic correlation of sequences (extended abstract)*, Sequences and their applications–SETA 2008, 71–80, Lecture Notes in Comput. Sci., 5203, Springer, Berlin, 2008.

[6] M. Goresky, A. Klapper, *Statistical properties of the arithmetic correlation of sequences*, Internat. J. Found. Comput. Sci. 22, no. 6, 1297–1315, 2011.

[7] M. Goresky, A. Klapper, *Algebraic shift register sequences*, Cambridge University Press, Cambridge, 2012.

[8] D. Mandelbaum, *Arithmetic codes with large distance*, IEEE Trans. Inform. Theory 13, 237–242, 1967.

[9] C. Mauduit, A. Sárközy, *On finite pseudorandom binary sequences. I. Measure of pseudorandomness, the Legendre symbol*, Acta Arith. 82, 365–377, 1997.

[10] R. E. A. C. Paley, *On orthogonal matrices*, J. Math. Physics 12, 311–320, 1933.

[11] O. Perron, *Bemerkungen über die Verteilung der quadratischen Reste*, Math. Z. 56, 122–130, 1952.

[12] I. Shparlinski, *Cryptographic applications of analytic number theory. Complexity lower bounds and pseudorandomness*, Progress in Computer Science and Applied Logic 22, Birkhäuser Verlag, Basel, 2003.

[13] R. J. Turyn, *The linear generation of Legendre sequence*, J. Soc. Indust. Appl. Math. 12, 115–116, 1964.