Discrete correlations of order 2 of generalized Golay-Shapiro sequences : a combinatorial approach

Irène Marcovici, Thomas Stoll, Pierre-Adrien Tahay

DISCRETE CORRELATIONS OF ORDER 2 OF GENERALIZED GOLAY–SHAPIRO SEQUENCES: A COMBINATORIAL APPROACH

Irène Marcovici
Université de Lorraine, CNRS, Inria, IECL, F-54000 Nancy, France
irene.marcovici@univ-lorraine.fr

Thomas Stoll
Université de Lorraine, CNRS, IECL, F-54000 Nancy, France
thomas.stoll@univ-lorraine.fr

Pierre-Adrien Tahay
Université de Lorraine, CNRS, IECL, F-54000 Nancy, France
pierre-adrien.tahay@univ-lorraine.fr

Abstract

We introduce a family of block-additive automatic sequences, that are obtained by allocating a weight to each couple of digits, and defining the nth term of the sequence as being the total weight of the integer n written in base k. Under an additional *difference condition* on the weight function, these sequences can be interpreted as generalized Golay–Shapiro sequences, and we prove that they have the same correlations of order 2 as sequences of symbols chosen uniformly and independently at random. The speed of convergence is very fast and is independent of the prime factor decomposition of k. This extends recent work of Tahay. The proof relies on direct observations about base-k representations of integers and combinatorial considerations. We also provide extensions of our results to higher-dimensional block-additive sequences.

1 Introduction

A k-automatic sequence on a finite set G is a sequence $u \in G^{\mathbb{N}}$ that can be computed by a deterministic finite automaton with output (DFAO) in the following way: the n-th term of the sequence is a function of the state reached by the automaton after reading the representation of the integer n in base k. Alternatively, a k-automatic sequence can also be defined as a sequence generated by a k-uniform morphism. We refer to the book of Allouche and Shallit [4] for a complete survey on automatic sequences.

Although automatic sequences are deterministic sequences having a very simple algorithmic description, some of them exhibit a complex behaviour. In this work, we are interested in exploring "how random" an automatic sequence

can look like. There are many different ways to measure the "random aspect" of a deterministic sequence. Here, we will study families of automatic sequences having the same discrete correlations of order 2 as sequences of symbols chosen uniformly and independently at random. We also provide explicit estimates for the speed of convergence.

The sequences we will consider are *block-additive sequences*. They are obtained by allocating a weight to each couple of digits, and defining the nth term of the sequence as being the total weight of the integer n written in base k. This weight is obtained by sliding the representation of the integer n in base k with a window of length 2 (or more generally, of length $L \geq 1$), and summing all the weights read. The name *block-additive* was already used in previous articles [7, 13]. With the terminology of Cateland [6], these sequences are *digital sequences*. In the special case where the weight matrix is a *difference matrix*, we will say that the automatic sequence obtained is a *generalized Golay–Shapiro sequence*, and prove that it has the same correlations of order 2 as a sequence of symbols chosen uniformly and independently at random.[1]

As we will comment on further in the article, our terminology of *generalized Golay–Shapiro sequences* is consistent with the definitions of [9, 15], and also intersects previous notions of generalized Golay–Shapiro sequences, such as the one of Queffélec [14] (see [9] for further references). For other generalizations of the Golay–Shapiro sequence that we will not investigate here, see Allouche and Liardet [2], Allouche and Shallit [3], and Mauduit and Rivat [12]. More specifically, we will not treat sequences such as those proposed by Mendès France that count $1w1$ in base 2 where w is an arbitrary finite non-empty binary string (these sequences are investigated in [2, Section 3]). Some alternative d-dimensional extensions of the Golay–Shapiro sequence are also proposed in [8] and [5].

As in the articles of Grant et al. [9] and Tahay [15], we study the correlations of order 2 of generalized Golay–Shapiro sequences, but rather than making use of exponential sums, we here only employ direct arguments relying on the base-k decomposition of the integers n and $n + r$, for a fixed r. This approach highlights the combinatorial role played by the *difference condition* defining a difference matrix, and allows to obtain more precise estimates on the correlations of order 2. Furthermore, in addition to studying the asymptotic proportion of integers n satisfying $u_n = u_{n+r}$, we provide results on the proportion of integers for which $(u_n, u_{n+r}) = (i, j)$, for any possible value of the couple $(i, j) \in G^2$. Precisely, we prove that the limit is equal to $1/|G|^2$ for all $(i, j) \in G^2$ and for any $r \in \mathbb{N} \setminus \{0\}$, as for an i.i.d. sequence of symbols uniformly drawn in G. After considering the one-dimensional case, we also mention extensions of our results to higher-dimensional block-additive sequences.

[1] *Golay–Shapiro sequences* are often referred to as *Rudin–Shapiro sequences*. In our paper, we prefer to give credit to the historical timeline of discovery as given by Allouche [1, Remark 1].

2 Definitions and presentation of the results

In all the article, we denote by $\mathbb{N}$ the set of non-negative integers, and for an integer $n \geq 2$, we use the notation $\mathbb{Z}_n = \mathbb{Z}/n\mathbb{Z}$.

2.1 Block-additive sequences of rank 2

For $k \in \mathbb{N} \setminus \{0\}$, we define $\Sigma_k = \{0, \ldots, k-1\}$, and we denote by $[n]_k$ the representation of the integer $n \in \mathbb{N}$ in base k. By definition, it is the unique sequence $x = (x_i)_{i\in\mathbb{N}} \in \Sigma_k^{\mathbb{N}}$ containing finitely many non-zero values, such that

$$n = \sum_{i\in\mathbb{N}} x_i k^i.$$

We will write

$$[n]_k = \quad x_0 \quad x_1 \quad x_2 \quad x_3 \quad \cdots .$$

We also introduce the notation $\ell_n = \min\{i \in \mathbb{N} : x_j = 0 \text{ for all } j > i\}$, and we define

$$\sigma_k(n) = \sum_{i\in\mathbb{N}} x_i = \sum_{i=0}^{\ell_n} x_i,$$

the k-ary sum-of-digits function.

Definition 1. Let $(G, +)$ be a finite abelian group, let $k \in \mathbb{N} \setminus \{0\}$, and let $f : \Sigma_k \times \Sigma_k \to G$ be a function satisfying $f(0, 0) = 0$. We say that the sequence $u = (u_n)_{n\in\mathbb{N}} \in G^{\mathbb{N}}$ is a *block-additive sequence (of rank 2) in base k* of *weight function (or matrix) f* if for any integer $n \in \mathbb{N}$, we have

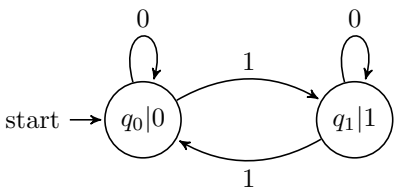
$$u_n = \sum_{i\in\mathbb{N}} f(x_i, x_{i+1}),$$

where $[n]_k = x$.

Example 1 (Prouhet–Thue–Morse sequence)**.** The Prouhet–Thue–Morse sequence (or Thue–Morse sequence) is given by

$$u_n \equiv \sigma_2(n) \pmod 2, \qquad n \in \mathbb{N}.$$

It is a block-additive sequence in base $k = 2$, with $G = \mathbb{Z}_2$, and weight function $f : \Sigma_2 \times \Sigma_2 \to G$ defined by $f(i, j) = i$, for any $(i, j) \in G^2$. The first terms are given by $u = (0, 1, 1, 0, 1, 0, 0, 1, 1, 0, 0, 1, 0, 1, 1, 0, \ldots)$. We represent below a DFAO computing this sequence.



Example 2 (Classical Golay–Shapiro sequence)**.** The (classical) Golay–Shapiro sequence (or Rudin–Shapiro sequence) on $G = \mathbb{Z}_2$ can be defined as the block-additive sequence in base $k = 2$ of weight function $f : \Sigma_2 \times \Sigma_2 \to G$ given by $f(i, j) = ij$, for any $(i, j) \in G^2$. In other words, u_n gives the parity count of the number of (possibly overlapping) occurrences of the block 11 in the binary expansion of n. The first terms are given by $u = (0, 0, 0, 1, 0, 0, 1, 0, 0, 0, 0, 1, 1, 1, 0, 1, \ldots)$.

The following proposition is straightforward. For the sake of completeness, we include the proof.

Proposition 1. *If a sequence is block-additive in base k, then it is a k-automatic sequence.*

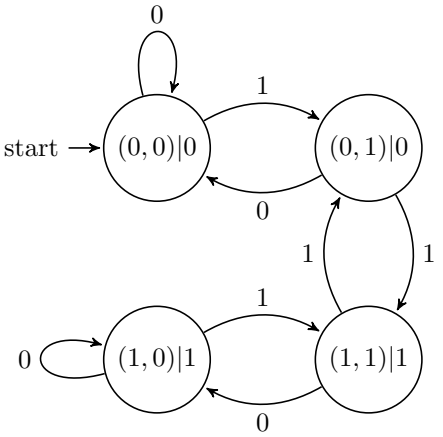
Proof. Let $Q = G \times \Sigma_k$, $q_0 = (0, 0)$, let $\delta : Q \times \Sigma_k \to Q$ be defined by

$$\delta((g, i), j) = (g + f(j, i), j),$$

and let $\tau : Q \to G$ be defined by $\tau(g, i) = g$. The DFAO $(Q, \Sigma_k, \delta, q_0, \tau)$ computes the block-additive sequence $u = (u_n)_{n\in\mathbb{N}}$ of weight function f, by reading the representation of the integer n in base k starting with the most significant digit, and using the output map τ. □

Remark 1. Alternatively, a block-additive sequence has the following morphic description. Again, let $Q = G \times \Sigma_k$ and $q_0 = (0, 0)$, and let $\phi : Q^* \to Q^*$ be the k-uniform morphism satisfying, for a state $s = (g, i) \in Q$, $\phi(s) = s_0 \cdots s_{k-1}$, with $s_j = (g + f(j, i), j)$. Consider the fixed point $\phi^\omega(q_0) \in Q^{\mathbb{N}}$. Then, the letter-to-letter projection of $\phi^\omega(q_0)$ by τ is the block-additive sequence of the function f.

Example 3. We represent below the DFAO given by the proof of Proposition 1 for the (classical) Golay–Shapiro sequence.



With the notations $q_0 = (0,0), q_1 = (0,1), q_2 = (1,0), q_3 = (1,1)$, the 2-uniform morphism described above is here given by

$$\phi(q_0) = q_0q_1,\ \phi(q_1) = q_0q_2,\ \phi(q_2) = q_3q_1,\ \phi(q_3) = q_3q_2,$$

with $\tau(q_0) = \tau(q_1) = 0, \tau(q_2) = \tau(q_3) = 1$.

2.2 Difference matrices and generalized Golay–Shapiro sequences

Definition 2. Let $(G,+)$ be a finite abelian group, and let $k \in \mathbb{N} \setminus \{0\}$. A *difference matrix* of size k is a matrix $D = (d(i,j))_{(i,j)\in\Sigma_k\times\Sigma_k} \in G^{\Sigma_k\times\Sigma_k}$ satisfying the following *difference condition*: for every $(i,j) \in \Sigma_k \times \Sigma_k$ with $i \neq j$, and every $g \in G$,

$$\operatorname{card}\Big\{h \in \Sigma_k : d(i,h) - d(j,h) = g\Big\} = \frac{k}{|G|}.$$

In other words, D is a difference matrix if for any $(i,j) \in \Sigma_k \times \Sigma_k$ with $i \neq j$, the set $\{d(i,h) - d(j,h) : h \in \Sigma_k\}$ contains every element of G equally often. Note that the difference condition requires the integer k to be a multiple of $|G|$. We introduce the notation $\pi = k/|G|$, and thus we have $\pi \in \mathbb{N} \setminus \{0\}$. We denote by $\mathcal{D}(G,k)$ the set of difference matrices of size k over the group G.

Definition 3. A block-additive sequence is a *generalized Golay–Shapiro sequence* if its weight function f is such that the matrix $(f(i,j))_{(i,j)\in\Sigma_k\times\Sigma_k} \in G^{\Sigma_k\times\Sigma_k}$ is a difference matrix.

Example 4.

1. The Thue–Morse sequence is *not* a generalized Golay–Shapiro sequence, since its weight function is given by the matrix $\begin{pmatrix} 0 & 0 \\ 1 & 1 \end{pmatrix}$, which does not belong to $\mathcal{D}(\mathbb{Z}_2, 2)$.
2. The classical Golay–Shapiro sequence is a generalized Golay–Shapiro sequence, since its weight function is given by the matrix $\begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}$, which belongs to $\mathcal{D}(\mathbb{Z}_2, 2)$.

Let us present different ways to construct difference matrices, and thus to define generalized Golay–Shapiro sequences.

Example 5. Let p be a prime number, and let $G = \mathbb{Z}_p$. Then, the matrix $D = (d(i,j))_{(i,j)\in\Sigma_p\times\Sigma_p}$ defined by $d(i,j) \equiv ij \pmod p$ is a difference matrix. The block-additive sequences thus obtained correspond to Queffélec's generalization of the Golay–Shapiro sequence [14, Section 4]. By definition, if $[n]_p = x$, we have $u_n \equiv \sum_{i\in\mathbb{N}} x_i x_{i+1} \pmod p$.

- As a particular case, for $p = 2$, the difference matrix is given by $\begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}$, and we recover the classical Golay–Shapiro sequence.

- For $p = 3$, the difference matrix is given by $\begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & 2 \\ 0 & 2 & 1 \end{pmatrix}$.

Example 6. For $k = 3$, another example of a difference matrix on $G = \mathbb{Z}_3$ is given by $\begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix}$. In the sequence obtained, the term u_n counts (modulo 3) the number of blocks of distinct digits in the base-3 decomposition of the integer n.

It can be seen that for an even integer $k \geq 4$, there exists *no* difference matrix of size k on $G = \mathbb{Z}_k$. Indeed, if k is even, we have $\sum_{i=0}^{k-1} i \equiv k/2 \pmod{k}$. But if $\sum_{h=0}^{k-1} (d(i_1, h) - d(i_2, h)) \equiv k/2 \pmod{k}$ and $\sum_{h=0}^{k-1} (d(i_2, h) - d(i_3, h)) \equiv k/2 \pmod{k}$, then $\sum_{h=0}^{k-1} (d(i_1, h) - d(i_3, h)) \equiv 0 \pmod{k}$, so that we obtain a contradiction.

However, the following theorem shows the existence of difference matrices at least for all powers of prime numbers. We include the proof for the sake of clearness.

Theorem 1. *[10, Theorem 6.6] For any prime number p and any integers $m, n \in \mathbb{N} \setminus \{0\}$ such that $m \leq n$, there exists a finite abelian group G of order p^m such that the set $\mathcal{D}(G, p^n)$ is non-empty.*

Proof. Let $H = \mathbb{F}_{p^m}$, and $G = \mathbb{F}_{p^n}$ be the finite fields with respectively p^m and p^n elements. We can represent the elements of G by polynomials of the form $\beta_0 + \beta_1 x + \cdots + \beta_{n-1} x^{n-1}$, with $\beta_0, \ldots, \beta_{n-1} \in \mathbb{Z}_p$. The group $(H, +)$ can be seen as the subgroup of $(G, +)$ made of the polynomials of degrees smaller or equal to m. Let $\varphi : G \to H$ be the function which maps the element $\beta_0 + \beta_1 x + \cdots + \beta_{n-1} x^{n-1}$ to the element $\beta_0 + \beta_1 x + \cdots + \beta_{m-1} x^{m-1}$, and for two polynomials $(\alpha(x), \beta(x)) \in G^2$, let $d(\alpha(x), \beta(x)) = \varphi(\alpha(x) \cdot \beta(x))$, where $\cdot$ denotes the multiplication in the field G. Then, one can check that the matrix $D = (d(i, j))_{(i,j) \in \Sigma_{p^n} \times \Sigma_{p^n}}$ (we identify Σ_{p^n} with G, using any bijection) is a difference matrix on $(G, +) \cong ((\mathbb{Z}_p)^n, +)$. □

Note that there exist difference matrices which do not belong to the families described in the proof of Theorem 1 (see [11, p.127] and [10, Table 6.37]). For example, the matrix

$$\begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 2 & 2 \\ 0 & 1 & 0 & 2 & 1 & 2 \\ 0 & 1 & 2 & 0 & 2 & 1 \\ 0 & 2 & 1 & 2 & 0 & 1 \\ 0 & 2 & 2 & 1 & 1 & 0 \end{pmatrix}$$

is an element of $\mathcal{D}(\mathbb{Z}_3, 6)$ that is not covered by Theorem 1.

The enumeration and the classification of difference matrices is a complex task. We refer to [10, 11] for an indepth study of these questions and various examples of difference matrices.

2.3 Main results

We can now state our main results, in the one-dimensional case. We use the notation $\log_k(N)$ for the logarithm of N to base k.

Theorem 2. *If u is a generalized Golay–Shapiro sequence, then for any $r \in \mathbb{N} \setminus \{0\}$, $g \in G$, and $N \in \mathbb{N}$,*

$$\Big| \frac{1}{N} \ \mathrm{card} \Big\{ n \in [\![0, N-1]\!] : u_{n+r} - u_n = g \Big\} - \frac{1}{|G|} \Big| \ \leq \ r\, k\, \frac{1 + \log_k(N)}{N}.$$

The limit $1/|G|$ is thus the same as for an i.i.d. sequence of symbols uniformly distributed in G. But the convergence is here much faster than in the random case, since the error term is of order $\log(N)/N$, while for i.i.d. sequences, the central limit theorem tells us that it is of order $1/\sqrt{N}$.

Remark 2. For k prime or a prime power, the bound in Theorem 2 is the same as the one obtained by Tahay [15, Theorem 4]. This is natural since the underlying objects (generalizations of the Golay–Shapiro sequence) are the same. However, our generalization of the Golay–Shapiro sequence to other composed k is different from Tahay [15]: it is directly based on one single difference matrix of size k, while Tahay's construction uses the prime factor decomposition of k and, as a side effect, the error term in his result is $N^{-1/d}$ where d denotes the number of different primes appearing in the prime factor decomposition of k [15, Theorem 5]. The size of our error term for our generalized objects is $\log(N)/N$, as $N \to \infty$, which is much smaller for fixed r and is independent of the arithmetic structure of k.

Theorem 3. *If u is a generalized Golay–Shapiro sequence, then for any $r \in \mathbb{N} \setminus \{0\}$, and any $(i, j) \in G^2$,*

$$\lim_{N \to \infty} \frac{1}{N} \ \mathrm{card} \Big\{ n \in [\![0, N-1]\!] : (u_n, u_{n+r}) = (i, j) \Big\} = \frac{1}{|G|^2}.$$

Remark 3. Tahay obtained several results on the mean value of the discrete correlation coefficients along the integers. The *discrete correlation coefficient* equals 0 if two symbols are identical, and 1 otherwise [15, Definition 1]. Theorem 3 gives a local result that is uniform in the values of the two symbols.

3 Discrete correlations of order 2 of generalized Golay–Shapiro sequences

The aim of this section is to prove Theorem 2 and Theorem 3. Namely, we prove that generalized Golay–Shapiro sequences have the same discrete correlations of order 2 as i.i.d. sequences of symbols, and give a tight estimate of the speed of convergence.

3.1 Frequencies of letters in generalized Golay–Shapiro sequences

In this section, we present some first general results on generalized Golay–Shapiro sequences, that we will need afterwards.

Lemma 1. *A generalized Golay–Shapiro sequence is a primitive morphic sequence.*

Proof. As in the proof of Proposition 1, let $Q = G \times \Sigma_k$, and let M be the matrix indexed by Q and with values in $\{0, 1\}$, defined by $M((g, i), (g', i')) = 1$ if and only if there exists $j \in \Sigma_k$ such that $(g', i') = (g + f(j, i), j)$. Equivalently, $M((g, i), (g', i'))$ is given by the number of $j \in \Sigma_k$ such that $(g', i') = (g + f(j, i), j)$. This matrix thus describes the allowed transitions in the DFAO given in the proof of Proposition 1, or equivalently, the incidence matrix of the k-uniform morphism defined in Remark 1. We prove that all the entries of $M^{2|G|+3}$ are positive (the bound might be not optimal), where we recall that the entries of M^n correspond to the number of paths (i.e. consecutive transitions in the DFAO) of length n from one state to another.

Let $s_1 = (g_1, i_1)$ and $s_2 = (g_2, i_2)$ be two elements of Q. By the difference condition, there exists at least one $h \in G$ such that $f(1, h) - f(0, h) = g_2 - g_1 - f(0, 1) - f(0, i_1) - f(i_2, 0)$. From the state i_1, let us read in the DFAO the sequence $(0, h, 0, h, 0, h, \ldots, 0, h, 0, h, 1, 0, i_2)$, made of $|G|$ times the pattern $(0, h)$, followed by the pattern $(1, 0, i_2)$. Then, the new state will be s_2, since

$$\begin{aligned}
&f(0, i_1) + f(h, 0) + f(0, h) + f(h, 0) + \cdots + f(0, h) + f(h, 0) + f(1, h) + f(0, 1) + f(i_2, 0) \\
&= |G| f(h, 0) + (|G| - 1) f(0, h) + f(1, h) + f(0, 1) + f(0, i_1) + f(i_2, 0) \\
&= f(1, h) - f(0, h) + f(0, 1) + f(0, s_1) + f(s_2, 0) = g_2 - g_1.
\end{aligned}$$

The conclusion follows. □

Proposition 2. *If u is a generalized Golay–Shapiro sequence, then any pattern has a frequency in the sequence u. Furthermore, the frequency of each element of G (corresponding to patterns of length 1) is equal to $1/|G|$.*

Proof. The existence of the frequencies for all patterns follows from the fact that the sequence $\phi^\omega(q_0) \in Q^{\mathbb{N}}$ is a primitive morphic sequence, where ϕ is the morphism given in Remark 1. Furthermore, each element of Q has exactly k preimages, since to state $s = (g, j) \in Q$, one can arrive from the state $(g - f(j, i), i)$, for any $i \in G$ (by reading j). The vector of frequencies being the unique eigenvector to the eigenvalue k, all the elements of Q have the same frequency in $\phi^\omega(q_0)$, and consequently, each element of G has the same frequency in the image of $\phi^\omega(q_0)$ by τ. □

3.2 Fibre of an integer

We now introduce the notion of *fibre of an integer*, that will be useful in our context to study correlations of order 2 of generalized Golay–Shapiro sequences.

Let $r \in \mathbb{N} \setminus \{0\}$ be a fixed integer. For $n \in \mathbb{N}$, let us introduce the representations of n and $n+r$ in base k as follows

$$[n]_k = x,$$
$$[n+r]_k = y.$$

We define the integer

$$c_n = \min\{i \in \mathbb{N} : x_j = y_j \text{ for all } j > i\}.$$

Note that c_n depends on r, but that for the sake of shortness, we do not mention this dependence in the notation. The integer c_n measures how far the carry propagates when adding r to n. By definition, $x_{c_n} \neq y_{c_n}$ and for every $j > c_n$, we have $x_j = y_j$. We illustrate the definition of c_n below:

$$\begin{array}{rccccccc} [n]_k = & x_0 & x_1 & \cdots & x_{c_n} & x_{c_n+1} & x_{c_n+2} & \cdots \\ [n+r]_k = & y_0 & y_1 & \cdots & y_{c_n} & x_{c_n+1} & x_{c_n+2} & \cdots \end{array} \tag{1}$$

We define the *fibre* of n as the set

$$\begin{aligned} \mathcal{F}_r(n) &= \{m \in \mathbb{N} : x' = [m]_k \text{ satisfies } x'_i = x_i \text{ for every } i \in \mathbb{N} \setminus \{c_n + 1\}\} \\ &= \{n + (\alpha - x_{c_n+1})\, k^{c_n+1} : \alpha \in \Sigma_k\}. \end{aligned}$$

Thus, we have

$$\begin{array}{rcccccccc} \mathcal{F}_r(n) = \{ & x_0 & x_1 & \cdots & x_{c_n} & 0 & x_{c_n+2} & x_{c_n+3} & \cdots, \\ & x_0 & x_1 & \cdots & x_{c_n} & 1 & x_{c_n+2} & x_{c_n+3} & \cdots, \\ & x_0 & x_1 & \cdots & x_{c_n} & 2 & x_{c_n+2} & x_{c_n+3} & \cdots, \\ & & & & & \vdots & & & \\ & x_0 & x_1 & \cdots & x_{c_n} & k-1 & x_{c_n+2} & x_{c_n+3} & \cdots\}. \end{array}$$

Note that if $m \in \mathcal{F}_r(n)$, then $c_m = c_n$, so that

$$m \in \mathcal{F}_r(n) \iff n \in \mathcal{F}_r(m).$$

Furthermore, let $m \in \mathcal{F}_r(n)$, and let $x' = [m]_k$, $y' = [m+r]_k$. Then by definition, $y'_{c_n+1} = x'_{c_n+1}$, and for every $i \in \mathbb{N} \setminus \{c_n + 1\}$, we have $y'_i = y_i$, as represented below:

$$\begin{array}{rccccccc} [m]_k = x' = & x_0 & x_1 & \cdots & x_{c_n} & x'_{c_n+1} & x_{c_n+2} & \cdots \\ [m+r]_k = y' = & y_0 & y_1 & \cdots & y_{c_n} & x'_{c_n+1} & x_{c_n+2} & \cdots \end{array} \tag{2}$$

Let u be a block-additive sequence in base k of weight f, and recall the notation $\pi = k/|G|$. For $n \in \mathbb{N}$, we also introduce the notation $\Delta_r(n) = u_{n+r} - u_n$.

Proposition 3. *If u is a generalized Golay–Shapiro sequence, then for any $n \in \mathbb{N}$, and any $g \in G$, we have*

$$\operatorname{card}\{m \in \mathcal{F}_r(n) : \Delta_r(m) = g\} = \pi.$$

Proof. By definition of a block-additive sequence, with the notations of Equation (1), we have

$$\begin{aligned}\Delta_r(n) &= \sum_{i\in\mathbb{N}} f(y_i, y_{i+1}) - \sum_{i\in\mathbb{N}} f(x_i, x_{i+1}) \\ &= \sum_{i=0}^{c_n} \Big(f(y_i, y_{i+1}) - f(x_i, x_{i+1})\Big).\end{aligned}$$

If $m \in \mathcal{F}_r(n)$, with the notations of Equation (2), we have

$$\Delta_r(m) = \sum_{i=0}^{c_n} (f(y'_i, y'_{i+1}) - f(x'_i, x'_{i+1})),$$

so that

$$\begin{aligned}\Delta_r(m) - \Delta_r(n) &= \Big(f(y'_{c_n}, y'_{c_n+1}) - f(x'_{c_n}, x'_{c_n+1})\Big) - \Big(f(y_{c_n}, y_{c_n+1}) - f(x_{c_n}, x_{c_n+1})\Big) \\ &= \Big(f(y_{c_n}, x'_{c_n+1}) - f(x_{c_n}, x'_{c_n+1})\Big) - \Big(f(y_{c_n}, x_{c_n+1}) - f(x_{c_n}, x_{c_n+1})\Big).\end{aligned}$$

It follows that for all $g \in G$,

$$\operatorname{card}\{m \in \mathcal{F}_r(n) : \Delta_r(m) - \Delta_r(n) = g\} = \operatorname{card}\Big\{\alpha \in \Sigma_k : f(y_{c_n}, \alpha) - f(x_{c_n}, \alpha) - A_n = g\Big\},$$

with $A_n = f(y_{c_n}, x_{c_n+1}) - f(x_{c_n}, x_{c_n+1})$.

Consequently, if u is a generalized Golay–Shapiro sequence, then for any $n \in \mathbb{N}$, and any $g \in G$, we have

$$\operatorname{card}\{m \in \mathcal{F}_r(n) : \Delta_r(m) - \Delta_r(n) = g\} = \pi,$$

and Proposition 3 follows. □

3.3 Proof of Theorem 2

Using the notion of fibre developed above, we obtain the following proposition, from which Theorem 2 directly follows, since $\sum_{g\in G} \operatorname{card}\Big\{n \in [\![0, N-1]\!] : \Delta_r(n) = g\Big\} = N$.

Proposition 4. *If u is a generalized Golay–Shapiro sequence, then for any $g \in G$,*

$$\begin{aligned}\operatorname{card}\Big\{n \in [\![0, N-1]\!] : \Delta_r(n) = g\Big\} &\geq \frac{\pi N}{k} - \pi\, r\, k - \pi\, r\, \sigma_k(N) \\ &\geq \frac{N}{|G|} - \pi\, r\, k(1 + \log_k(N)).\end{aligned}$$

Proof. Let $N \in \mathbb{N} \setminus \{0\}$, and let $a = [N]_k$. We determine the conditions under which an integer $n \in [\![0, N-1]\!]$ satisfies $\mathcal{F}_r(n) \subset [\![0, N-1]\!]$. Recall the notation $\ell_N = \min\{i \in \mathbb{N} : a_i = 0 \text{ for all } j > i\}$. We can thus write

$$[N]_k = a_0\ a_1\ \cdots\ a_{\ell_N-1}\ a_{\ell_N}\ 0\ 0\ \cdots$$

- If $n = a'_{\ell_N}\, k^{\ell_N} + \alpha\, k^{\ell_N-1} + \gamma$, for some $\alpha \leq k-1$, $a'_{\ell_N} < a_{\ell_N}$, and $\gamma < k^{\ell_N-1} - r$, then $c_n \leq \ell_N - 2$, so that $\mathcal{F}_r(n) \subset [\![0, N-1]\!]$.

$$[n]_k = \underbrace{x_0\ x_1\ \cdots\ x_{\ell_N-2}}_{\gamma < k^{\ell_N-1}-r}\ \alpha\ \underbrace{a'_{\ell_N}}_{<a_{\ell_N}}\ 0\ 0\ \cdots$$
$$[n+r]_k = x'_0\ x'_1\ \cdots\ x'_{\ell_N-2}\ \alpha\ a'_{\ell_N}\ 0\ 0\ \cdots$$

- If $n = a_{\ell_N}\, k^{\ell_N} + a'_{\ell_N-1}\, k^{\ell_N-1} + \alpha\, k^{\ell_N-2} + \gamma$, for some $\alpha \leq k-1$, $a'_{\ell_N-1} < a_{\ell_N-1}$, and $\gamma < k^{\ell_N-2} - r$, then $c_n \leq \ell_N - 3$, so that $\mathcal{F}_r(n) \subset [\![0, N-1]\!]$.

$$[n]_k = \underbrace{x_0\ x_1\ \cdots\ x_{\ell_N-3}}_{\gamma < k^{\ell_N-2}-r}\ \alpha\ \underbrace{a'_{\ell_N-1}}_{<a_{\ell_N-1}}\ a_{\ell_N}\ 0\ 0\ \cdots$$
$$[n+r]_k = x'_0\ x'_1\ \cdots\ x'_{\ell_N-3}\ \alpha\ a'_{\ell_N-1}\ a_{\ell_N}\ 0\ 0\ \cdots$$

- If $n = a_{\ell_N}\, k^{\ell_N} + a_{\ell_N-1}\, k^{\ell_N-1} + a'_{\ell_N-2}\, k^{\ell_N-2} + \alpha\, k^{\ell_N-3} + \gamma$, for some $\alpha \leq k-1$, $a'_{\ell_N-2} < a_{\ell_N-2}$, and $\gamma < k^{\ell_N-3} - r$, then $c_n \leq \ell_N - 4$, so that $\mathcal{F}_r(n) \subset [\![0, N]\!]$.

$$[n]_k = \underbrace{x_0\ x_1\ \cdots\ x_{\ell_N-4}}_{\gamma < k^{\ell_N-3}-r}\ \alpha\ \underbrace{a'_{\ell_N-2}}_{<a_{\ell_N-2}}\ a_{\ell_N-1}\ a_{\ell_N}\ 0\ 0\ \cdots$$
$$[n+r]_k = x'_0\ x'_1\ \cdots\ x'_{\ell_N-4}\ \alpha\ a'_{\ell_N-2}\ a_{\ell_N-1}\ a_{\ell_N}\ 0\ 0\ \cdots$$

- And so on, the last condition that will be of interest for us being that if $n = a_{\ell_N}\, k^{\ell_N} + a_{\ell_N-1}\, k^{\ell_N-1} + \ldots + a_{\ell_r+3}\, k^{\ell_r+3} + a'_{\ell_r+2}\, k^{\ell_r+2} + \alpha\, k^{\ell_r+1} + \gamma$, for some $\alpha \leq k-1$, $a'_{\ell_r+2} < a_{\ell_r+2}$, and $\gamma < k^{\ell_r+1} - r$, then $c_n \leq \ell_r$, so that $\mathcal{F}_r(n) \subset [\![0, N-1]\!]$.

The number of different integers $n \in [\![0, N-1]\!]$ satisfying $\mathcal{F}_r(n) \subset [\![0, N-1]\!]$ that we have exhibited above is equal to

$$\begin{aligned}
& a_{\ell_N} k(k^{\ell_N-1} - r) + a_{\ell_N-1} k(k^{\ell_N-2} - r) + a_{\ell_N-2} k(k^{\ell_N-3} - r) + \ldots + a_{\ell_r+2} k(k^{\ell_r+1} - r) \\
= \ & N - (a_{\ell_r+1} k^{\ell_r+1} + a_{\ell_r} k^{\ell_r} + \ldots + a_1 k + a_0) - r\, k\, (a_{\ell_N} + a_{\ell_N-1} + a_{\ell_N-2} + \ldots + a_{\ell_r+2}) \\
> \ & N - r\, k^2 - r\, k\, \sigma_k(N).
\end{aligned}$$

For the last inequality, observe that $a_{\ell_r+1} k^{\ell_r+1} + a_{\ell_r} k^{\ell_r} + \ldots + a_1 k + a_0 < k^{\ell_r+2} \leq r\, k^2$. Proposition 4 then directly follows from Proposition 3. □

3.4 Correlation matrix

In order to prove Theorem 3, we first introduce the notion of correlation matrix, and formulate the previous results using this terminology.

Let $u \in G^{\mathbb{N}}$ be a fixed sequence. For $r \in \mathbb{N} \setminus \{0\}, (i,j) \in G^2$ and $n \in \mathbb{N}$, we define

$$\delta_{i,j}^r(n) = \begin{cases} 1 & \text{if } (u_n, u_{n+r}) = (i,j), \\ 0 & \text{otherwise;} \end{cases}$$

and

$$C_{i,j}^r(N) = \frac{1}{N} \sum_{n=0}^{N-1} \delta_{i,j}^r(n).$$

As a consequence of Proposition 2, if u is a generalized Golay–Shapiro sequence, then for any $r \in \mathbb{N} \setminus \{0\}$ and $(i,j) \in G^2$, the sequence $C_{i,j}^r(N)$ converges when N goes to infinity, so that we can also introduce

$$C_{i,j}^r = \lim_{N \to \infty} C_{i,j}^r(N).$$

We define the *correlation matrix* as the matrix $C^r = (C_{i,j}^r)_{(i,j) \in G^2}$ of size $|G| \times |G|$. By Proposition 2, for any $i \in G$, the asymptotic frequency of the symbol i is equal to

$$\sum_{j \in G} C_{i,j}^r = \frac{1}{|G|}.$$

As a consequence of Proposition 4, we obtain the following results.

Corollary 1. *If u is a generalized Golay–Shapiro sequence, then for any $(i,j) \in G^2$,*

$$\sum_{\ell \in G} C_{i-\ell, j-\ell}^r(N) \geq \frac{1}{|G|} - \pi\, r\, k \frac{1 + \log_k(N)}{N}.$$

Corollary 2. *If u is a generalized Golay–Shapiro sequence, then for any $(i,j) \in G^2$,*

$$\sum_{\ell \in G} C_{i-\ell, j-\ell}^r = \frac{1}{|G|}.$$

Proof. It is a consequence from Corollary 1 and the observation that

$$\sum_{(i,j) \in G^2} C_{i,j}^r = 1.$$

□

Note that this result refines the estimates of Tahay concerning the discrete correlation coefficient (cf. Remark 3) that detects whether two symbols differ or not. In our language, he proved that

$$\sum_{i \in G} C_{i,i}^r = \frac{1}{|G|}.$$

3.5 Proof of Theorem 3

With the notations above, Theorem 3 is equivalent to the next proposition, that we now prove. Note that this result is stronger than Corollary 2 as it gives the values of the individual terms in the sum.

Proposition 5. *If u is a generalized Golay–Shapiro sequence, then for any $(i,j) \in G^2$,*

$$C_{i,j}^r = \frac{1}{|G|^2}.$$

Proof. Let us fix some $\alpha \in \Sigma_k$ and consider the integers $n \in [\![0, k^{2N+1}-1]\!]$ that are such that the base-k decomposition $x = [n]_k$ of n satisfies $x_{N+1} = \alpha$. In other words, $n = m_1\, k^{N+1} + \alpha\, k^N + m_2$, for some integers $m_1, m_2 \in [\![0, k^N - 1]\!]$. Assuming furthermore that $m_2 < k^N - r$, we will have $c_n < N$, so that

$$(u_n, u_{n+r}) = (u_{km_1+\alpha}, u_{km_1+\alpha}) + (u_{\alpha k^N+m_2}, u_{\alpha k^N+m_2+r}),$$

by definition of a block-additive sequence.

The proof will be based on the following idea: when taking independently at random some integers m_1, m_2 uniformly distributed in $[\![0, k^N-1]\!]$, the distribution of $u_{km_1+\alpha}$ converges to the uniform distribution on G when N goes to infinity, while for the second term $(u_{\alpha k^N+m_2}, u_{\alpha k^N+m_2+r})$, the distribution is asymptotically given by the values $C_{i,j}$ of the correlation matrix. Now, we have $(u_n, u_{n+r}) = (i,j)$ if $u_{km_1+\alpha} = \ell$ for some ℓ and $(u_{\alpha k^N+m_2}, u_{\alpha k^N+m_2+r}) = (i-\ell, j-\ell)$. Using the independence of m_1 and m_2, we thus obtain

$$C_{i,j}^r = \sum_{\alpha\in\Sigma_k} \frac{1}{k} \sum_{\ell\in G} \frac{1}{|G|} C_{i-\ell,j-\ell} = \sum_{\alpha\in\Sigma_k} \frac{1}{k}\frac{1}{|G|}\frac{1}{|G|} = \frac{1}{|G|^2},$$

since we already know by Corollary 2 that for any $(i,j) \in G^2, \sum_{\ell\in G} C_{i-\ell,j-\ell} = \frac{1}{|G|}$.

More formally, let us introduce the following notations, for any $i, j, \ell \in G$,

$$A_\ell^\alpha(N) = \mathrm{card}\{m \in [\![0, k^N-1]\!] : u_{km+\alpha} = \ell\}$$
$$B_{i,j}^{r,\alpha}(N) = \mathrm{card}\{m \in [\![0, k^N-r-1]\!] : \delta_{i,j}^r(\alpha\, k^N + m) = 1\}.$$

We claim that for any $\alpha \in \Sigma_k$,

$$\lim_{N\to\infty} \frac{A_\ell^\alpha(N)}{k^N} = \frac{1}{|G|}, \quad \text{and} \quad \lim_{N\to\infty} \sum_{\ell\in G} \frac{B_{i-\ell,j-\ell}^{r,\alpha}(N)}{k^N} = \frac{1}{|G|}.$$

For the first limit, we use the same tools as for Proposition 2. Let ϕ be the primitive morphism given in Remark 1, so that the sequence u is the image of $\phi^\omega(q_0)$ by τ. One can see that the sequence $(u_{kn+\alpha})_{n\in\mathbb{N}}$ is the image of $\phi^\omega(q_0)$ by the function $\tau' : Q \to G$ defined by $\tau'(g,i) = g + f(i,\alpha)$. As we have already seen in the proof of Proposition 2, all the elements of Q have the same frequency in $\phi^\omega(q_0)$. Consequently, each element of G has the same frequency in the image

of $\phi^\omega(q_0)$ by τ'. Indeed, for any $g' \in G$ and $i \in \Sigma_k$, there exists exactly one $g \in G$ such that $\tau'(g,i) = g'$, so that the cardinal of $\tau'^{-1}(\{g'\})$ does not depend on the choice of g'.

For the second limit, observe that

$$\begin{aligned}
&\lim_{N\to\infty} \sum_{\ell\in G} \frac{B^{r,\alpha}_{i-\ell,j-\ell}(N)}{k^N} \\
&= \lim_{N\to\infty} \sum_{\ell\in G} \frac{(\alpha k^N + k^N - r)\cdot C^r_{i-\ell,j-\ell}(\alpha k^N + k^N - r) - \alpha k^N \cdot C^r_{i-\ell,j-\ell}(\alpha k^N)}{k^N} \\
&= \sum_{\ell\in G} (\alpha+1)\cdot C^r_{i-\ell,j-\ell} - \alpha \cdot C^r_{i-\ell,j-\ell} \\
&= \frac{1}{|G|}.
\end{aligned}$$

Now, for any $(i,j) \in G^2$, we have

$$\sum_{n=0}^{k^{2N+1}-1} \delta^r_{i,j}(n) \geq \sum_{\alpha\in\Sigma_k} \sum_{\ell\in G} A^N_\ell(\alpha)\ B^N_{i-\ell,j-\ell}(\alpha).$$

It follows that

$$C^r_{i,j}(k^{2N+1}) \geq \frac{1}{k} \sum_{\alpha\in\Sigma_k} \sum_{\ell\in G} \frac{A^N_\ell(\alpha)}{k^N}\ \frac{B^N_{i-\ell,j-\ell}(\alpha)}{k^N}.$$

When N goes to infinity, we know that the limit of the left term exists and is equal to $C^r_{i,j}$. We thus obtain

$$C^r_{i,j} \geq \frac{1}{|G|^2}.$$

Since $\sum_{(i,j)\in G^2} C^r_{i,j} = 1$, this ends the proof. □

4 Higher dimensional generalized Golay–Shapiro sequences

We propose the following natural extension of Def. 1 and 3 in dimension d. For greater readability, we represent the elements of Σ_k^d as column vectors.

Definition 4. Let $(G,+)$ be a finite abelian group, and let $k \in \mathbb{N}\setminus\{0\}$. We say that the sequence $u = (u_{n_1,\dots,n_d})_{(n_1,\dots,n_d)\in\mathbb{N}^d} \in G^{\mathbb{N}^d}$ is a *d-dimensional block-additive sequence in base k* if there exists a map $f : \Sigma_k^d \times \Sigma_k^d \to G$ satisfying

$f\left(\begin{pmatrix}0\\ \vdots\\ 0\end{pmatrix},\begin{pmatrix}0\\ \vdots\\ 0\end{pmatrix}\right)=0$, such that for any integer $n\in\mathbb{N}^d$, we have

$$u_{n_1,\dots,n_d}=\sum_{i\in\mathbb{N}} f\left(\begin{pmatrix}x_i^1\\ \vdots\\ x_i^d\end{pmatrix},\begin{pmatrix}x_{i+1}^1\\ \vdots\\ x_{i+1}^d\end{pmatrix}\right)=\sum_{i\in\mathbb{N}} f(x_i,x_{i+1}),$$

where $x=(x_i)_{i\in\mathbb{N}}=\begin{pmatrix}x^1\\ \vdots\\ x^d\end{pmatrix}=\begin{pmatrix}(x_i^1)_{i\in\mathbb{N}}\\ \vdots\\ (x_i^d)_{i\in\mathbb{N}}\end{pmatrix}=\begin{pmatrix}[n_1]_k\\ \vdots\\ [n_d]_k\end{pmatrix}$.

We say furthermore that the sequence u is a *generalized d-dimensional Golay–Shapiro sequence* if the function f satisfies, for every $(i,j)\in\Sigma_k^d\times\Sigma_k^d$ with $i\neq j$, and for every $g\in G$,

$$\operatorname{card}\Big\{h\in\Sigma_k^d : f(i,h)-f(j,h)=g\Big\}=\frac{k^d}{|G|}.$$

Equivalently, this amounts to saying that the matrix $(f(i,j))_{(i,j)\in\Sigma_k^d\times\Sigma_k^d}$ is a difference matrix.

The definition above can be extended to d-dimensional sequences for which the base k_i may depend on the component $i\in\{1,\dots,d\}$ (in the difference condition, k^d is then replaced by the product $k_1\dots k_d$). However, for simplicity, we restrict ourselves to the case of a unique base k.

Let $r\in\mathbb{N}^d\setminus\{(0,\dots,0)\}$. For $n=(n_1,\dots,n_d)\in\mathbb{N}^d$, we introduce the representations of n and $n+r$ in base k as follows

$$[n]_k=x=\begin{pmatrix}x^1\\ \vdots\\ x^d\end{pmatrix},\qquad [n+r]_k=y=\begin{pmatrix}y^1\\ \vdots\\ y^d\end{pmatrix},$$

and we define the integer

$$c_n=\min\{i\in\mathbb{N} : x_j=y_j \text{ for all } j>i\},$$

which measures how far the carry propagates when adding r to n.

We define again the *fibre* of n as the set

$$\mathcal{F}_r(n)=\{m\in\mathbb{N} : x'=[m]_k \text{ satisfies } x'_i=x_i \text{ for every } i\in\mathbb{N}\setminus\{c_n+1\}\},$$

and use the notation $\Delta_r(n)=u_{n+r}-u_n$.

Since the d-dimensional sequence has d components that are all 1-dimensional and independent, the previous arguments can be repeated verbatim.

Proposition 6. *If u is a generalized d-dimensional Golay–Shapiro sequence, then for any $n\in\mathbb{N}$, and any $g\in G$, we have*

$$\operatorname{card}\{m\in\mathcal{F}_r(n) : \Delta_r(m)=g\}=\pi.$$

We also extend the notations δ^r and C^r to d-dimensional sequences. Precisely, for $N = (N_1, \dots, N_d)$, we define

$$C^r_{i,j}(N) = \frac{1}{N_1 \cdots N_d} \sum_{\{n \in \mathbb{N}^d \,:\, n < N\}} \delta^r_{i,j}(n),$$

where the notation $n < N$ means that for all $i \in \{1, \dots, d\}, n_i < N_i$. We also introduce

$$C^r_{i,j} = \lim_{N_1, \dots, N_d \to \infty} C^r_{i,j}(N).$$

Following the previous lines, one can show as in the one-dimensional case that if u is a generalized d-dimensional Golay–Shapiro sequence, then for any $(i, j) \in G^2$,

$$\sum_{\ell \in G} C^r_{i-\ell, j-\ell} = \frac{1}{|G|},$$

which also allows to obtain the following extension of Proposition 5.

Proposition 7. *If u is a generalized d-dimensional Golay–Shapiro sequence, then for any $(i, j) \in G^2$,*

$$C^r_{i,j} = \frac{1}{|G|^2}.$$

Example 7. In Figures 1 and 2, we present four different examples of generalized Golay–Shapiro sequences, for $d = 2$, $k = 2$, $G = \mathbb{Z}_2$. For each example, the values of the function $f : \Sigma_2^2 \to \mathbb{Z}_2$ is given by a matrix, with the elements of Σ_2^2 sorted in the lexicographic order. On the first line of the matrix, one can thus read successively

$$f\left(\binom{0}{0}, \binom{0}{0}\right), f\left(\binom{0}{0}, \binom{0}{1}\right), f\left(\binom{0}{0}, \binom{1}{0}\right), f\left(\binom{0}{0}, \binom{1}{1}\right),$$

and then on the second line

$$f\left(\binom{0}{1}, \binom{0}{0}\right), f\left(\binom{0}{1}, \binom{0}{1}\right), \dots$$

and so on. On the pictures, the cell $(n_1, n_2) \in \mathbb{N}^2$ is colored in blue if $u_{n_1,n_2} = 1$ and in white if $u_{n_1,n_2} = 0$. The corner corresponding to the value $u_{0,0}$ is the bottom-left corner.

Let us present in more detail the first example. For $i, j \in \Sigma_2^2$, the weight function satisfies $f(i, j) = 0$ if $i = j$, and $f(i, j) = 1$ otherwise. As an example, we compute below $u_{436,48}$.

$$\begin{array}{rcccccccccccccc}
[436]_2 & = & 0 & 0 & 1 & 0 & 1 & 1 & 0 & 1 & 1 & 0 & 0 & \cdots & & \\
[48]_2 & = & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & \cdots & & \\
\hline
u_{436,48} & \equiv & 0 & + & 1 & + & 1 & + & 0 & + & 1 & + & 0 & + & \cdots & \equiv 0 \pmod 2
\end{array}$$

The following table gives the first values of u_{n_1,n_2}, for $(n_1, n_2) \in [\![0, 2^3 - 1]\!]^2$.

$n_2 \diagdown n_1$	0	1	2	3	4	5	6	7
7	1	0	1	0	0	1	0	1
6	0	0	1	1	1	1	0	0
5	1	1	1	1	1	1	1	1
4	0	1	1	0	0	1	1	0
3	1	0	0	1	0	1	1	0
2	0	0	0	0	1	1	1	1
1	1	1	0	0	1	1	0	0
0	0	1	0	1	0	1	0	1

These values are also contained in the bottom-left 8×8-squares of the two pictures that are on the first part of Figure 1.

Concerning the second example (second part of Figure 1), it can be seen that the weight function satisfies

$$f\left(\binom{i_1}{i_2}, \binom{j_1}{j_2}\right) \equiv i_1 j_1 + i_2 j_2 \pmod 2.$$

As a consequence, the sequence obtained can also be computed by $u_{m,n} = v_m + v_n$, where v is the classical one-dimensional Golay–Shapiro sequence.

5 Extensions and open questions

5.1 Block-additive sequences of rank larger than 2

Until now, we have only considered block-additive functions of rank 2. More generally, we can consider the notion of block-additive functions of rank L, for an integer $L \geq 1$, in the sense of Cateland [6].

Definition 5. Let $(G, +)$ be a finite abelian group, let $k \in \mathbb{N} \setminus \{0\}$, and for an integer $L \geq 1$, let $f : \Sigma_k^L \to G$ be a function satisfying $f(0, 0, \ldots, 0) = 0$. We say that the sequence $u = (u_n)_{n\in\mathbb{N}} \in G^{\mathbb{N}}$ is a *block-additive sequence (of rank L) in base k* of *weight function f* if for any integer $n \in \mathbb{N}$, we have

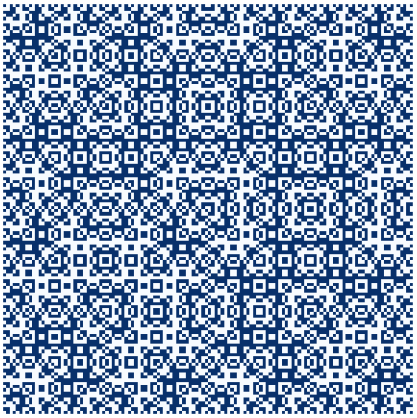
$$u_n = \sum_{i\in\mathbb{N}} f(x_i, x_{i+1}, \ldots, x_{i+L-1}),$$

where $[n]_k = x$.

Let $(G, +)$ be a finite abelian group, and let $k \in \mathbb{N} \setminus \{0\}$. For an integer $L \geq 2$, we say that the function $d : \Sigma_k^L \to G$ satisfies the *difference condition (of rank L)* if for every $(i, j) \in \Sigma_k \times \Sigma_k$ with $i \neq j$, for every $(x_2, \ldots, x_{L-1}) \in \Sigma_k^{L-2}$, and for every $g \in G$,

$$\operatorname{card}\{h \in \Sigma_k : d(i, x_2, \ldots, x_{L-1}, h) - d(j, x_2, \ldots, x_{L-1}, h) = g\} = \frac{k}{|G|}.$$

The difference condition is a sufficient condition for obtaining the same results as in Section 3, for block-additive sequences of rank L with $L \geq 2$.

Matrix $\begin{pmatrix} 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 \\ 1 & 1 & 1 & 0 \end{pmatrix}$	
Terms in $[\![0, 2^5 - 1]\!]^2$	Terms in $[\![0, 2^7 - 1]\!]^2$
	

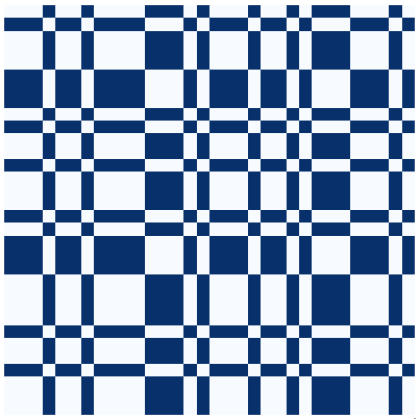
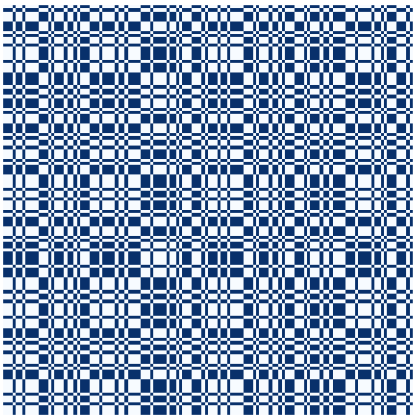
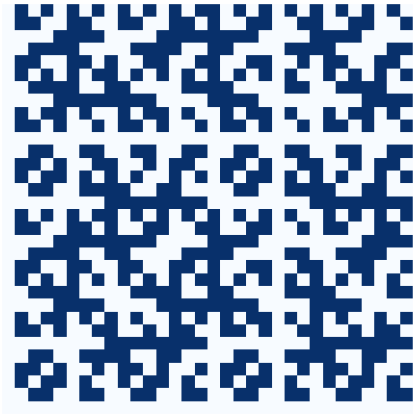
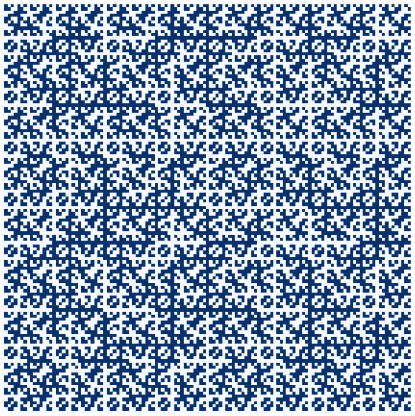
Matrix $\begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 \\ 0 & 1 & 1 & 0 \end{pmatrix}$	
Terms in $[\![0, 2^5 - 1]\!]^2$	Terms in $[\![0, 2^7 - 1]\!]^2$
	

Figure 1: Examples of generalized 2-dimensional Golay–Shapiro sequences in base 2

Matrix $\begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & 1 & 1 & 0 \end{pmatrix}$	
Terms in $[\![0, 2^5 - 1]\!]^2$	Terms in $[\![0, 2^7 - 1]\!]^2$
	

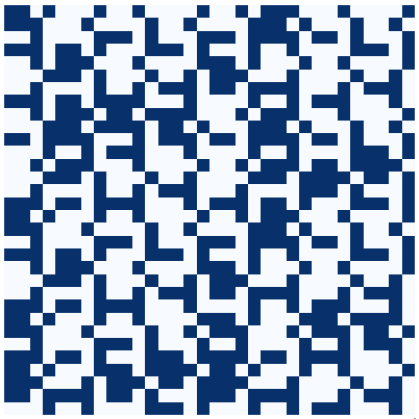
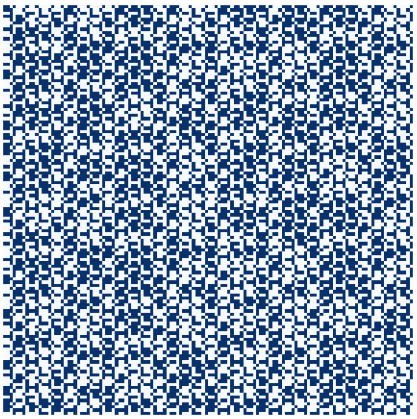
Matrix $\begin{pmatrix} 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 \end{pmatrix}$	
Terms in $[\![0, 2^5 - 1]\!]^2$	Terms in $[\![0, 2^7 - 1]\!]^2$
	

Figure 2: Examples of generalized 2-dimensional Golay–Shapiro sequences in base 2

Example 8. Let us set $k = 2$, $G = \mathbb{Z}_2$, and let $f : \Sigma_k^3 \to G$ be defined by

$$f(x, y, z) = \begin{cases} 0 \text{ if } x = y = z, \\ 1 \text{ otherwise.} \end{cases}$$

This function satisfies the difference condition. Consequently, the block-additive sequence $u = (u_n)_{n\in\mathbb{N}}$ of weight function f, which is such that u_n counts (modulo 2) the number of blocks different from 000 and 111 in the binary representation of n, has the same correlations of order 2 as a binary sequence chosen uniformly at random.

Open question 1. How can we generate functions satisfying the difference condition of rank L? Could there be a weaker condition on the weight function for which the block-additive sequences obtained would have the same correlations of order 2?

5.2 Can an automatic sequence look even more random?

Another possible direction of research consists in trying to construct block-additive sequences for which not only the correlations of order 2, but also correlations of higher order would be the same as for uniform random sequences. Precisely, for integers $0 < r_1 < \cdots < r_{\ell-1}$, and for a choice $(i_0, \ldots, i_{\ell-1}) \in G^\ell$, we introduce

$$\delta^r_{i_0,\ldots,i_{\ell-1}}(n) = \begin{cases} 1 & \text{if } (u_n, u_{n+r_1}, \ldots, u_{n+r_{\ell-1}}) = (i_0, \ldots, i_{\ell-1}), \\ 0 & \text{otherwise,} \end{cases}$$

and we look at the asymptotic behaviour of $\frac{1}{N}\sum_{n=0}^{N-1} \delta^r_{i_0,\ldots,i_{\ell-1}}(n)$, when N goes to infinity. We say that a sequence has the same correlations of order ℓ as a uniform random sequence if for any choice of $0 < r_1 < \cdots < r_{\ell-1}$, and for any $(i_0, \ldots, i_{\ell-1}) \in G^\ell$,

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} \delta^r_{i_0,\ldots,i_{\ell-1}}(n) = \frac{1}{|G|^\ell}.$$

Open question 2. For a given $\ell \geq 3$, is it possible to construct a block-additive sequence having the same correlations of order ℓ as a uniform random sequence?

Note that it is not possible to construct an automatic sequence such that *for every* $\ell \geq 1$, the correlations of order ℓ would be the same as for a uniform random sequence. Indeed, this would in particular imply the sequence to be normal, while the complexity of an automatic sequence is at most linear.

Acknowledgement. This work was supported partly by the French PIA project "Lorraine Université d'Excellence", reference ANR-15-IDEX-04-LUE, and by the projects ANR-18-CE40-0018 (EST) and ANR-20-CE91-0006 (ArithRand).

References

[1] J.-P. Allouche, On a Golay-Shapiro-like sequence, *Unif. Distrib. Theory* **11(2)** (2016), 205-210.

[2] J.-P. Allouche and P. Liardet, Generalized Rudin-Shapiro sequences, *Acta Arith.* **60(1)** (1991), 1-27.

[3] J.-P. Allouche and J. Shallit, Complexité des suites de Rudin-Shapiro généralisées, *J. Théor. Nombres Bordeaux* **5(2)** (1993), 283-302.

[4] J.-P. Allouche and J. Shallit, *Automatic Sequences*, Theory, Applications, Generalizations, Cambridge University Press, Cambridge, 2003.

[5] A. Barbé and F. von Haeseler, Correlation and spectral properties of higher-dimensional paperfolding and Rudin-Shapiro sequences, *J. Phys. A* **38(12)** (2005), 2599-2622.

[6] E. Cateland, *Digital sequences and k-regular sequences*, PhD thesis, Université Sciences et Technologies - Bordeaux I, 1992.

[7] M. Drmota, P. J. Grabner, and P. Liardet, Block additive functions on the Gaussian integers, *Acta Arith.* **135(4)** (2008), 299-332.

[8] N. Priebe Frank, Substitution sequences in $\mathbb{Z}^d$ with a non-simple Lebesgue component in the spectrum, *Ergodic Theory Dynam. Systems* **23(2)** (2003), 519-532.

[9] E. Grant, J. Shallit, and T. Stoll, Bounds for the discrete correlation of infinite sequences on k symbols and generalized Rudin-Shapiro sequences, *Acta Arith.* **140(4)** (2009), 345-368.

[10] A. S. Hedayat, N. J. A. Sloane, and J. Stufken, *Orthogonal Arrays*, Springer Series in Statistics, Springer-Verlag, New York, 1999.

[11] P. H. J. Lampio, *Classification of difference matrices and complex Hadamard matrices*, PhD thesis, Aalto University, 2015.

[12] C. Mauduit and J. Rivat, Prime numbers along Rudin-Shapiro sequences, *J. Eur. Math. Soc. (JEMS)* **17(10)** (2015), 2595-2642.

[13] C. Müllner, The Rudin-Shapiro sequence and similar sequences are normal along squares, *Canad. J. Math.* **70(5)** (2018), 1096-1129.

[14] M. Queffélec, Une nouvelle propriété des suites de Rudin-Shapiro, *Ann. Inst. Fourier (Grenoble)* **37(2)** (1987), 115-138.

[15] P.-A. Tahay, Discrete correlation of order 2 of generalized Rudin-Shapiro sequences on alphabets of arbitrary size, *Unif. Distrib. Theory* **15(1)** (2020), 1-26.