



ELSEVIER

Contents lists available at ScienceDirect

Journal of Number Theory

www.elsevier.com/locate/jnt



General Section

Algorithmic classification of noncorrelated binary pattern sequences



Jakub Konieczny*

Camille Jordan Institute, Claude Bernard University Lyon 1, 43 Boulevard du 11 novembre 1918, 69622 Villeurbanne Cedex, France

Faculty of Mathematics and Computer Science, Jagiellonian University in Kraków, Łojasiewicza 6, 30-348 Kraków, Poland

ARTICLE INFO

Article history:

Received 30 June 2020

Received in revised form 7 October 2020

Accepted 12 October 2020

Available online 25 January 2021

Communicated by F. Pellarin

MSC:

primary 47B15

secondary 11B50

Keywords:

Pattern sequence

Automatic sequence

Regular sequence

Correlation

Spectral measure

Thue-Morse sequence

Rudin-Shapiro sequence

ABSTRACT

The main subject of this paper are binary pattern sequences, that is, sequences of the form $(-1)^{\#(n,A)}$ where A is a set of strings of 0s and 1s, and $\#(n,A)$ denotes the total number of times patterns from A appear in the binary expansion of n . A sequence is said to be noncorrelated if the corresponding spectral measure is equal to the Lebesgue measure.

We show that it is possible to algorithmically verify if a given binary pattern sequence is noncorrelated. As an application, we compute that there are exactly 2272 noncorrelated binary pattern sequences of length ≤ 4 . If we restrict our attention to patterns that do not end with 0, we put forward a sufficient condition for a pattern sequence to be noncorrelated. We conjecture that this condition is also necessary, and verify this conjecture for lengths ≤ 5 .

© 2020 Elsevier Inc. All rights reserved.

* Correspondence to: Camille Jordan Institute, Claude Bernard University Lyon 1, 43 Boulevard du 11 novembre 1918, 69622 Villeurbanne Cedex, France.

E-mail address: jakub.konieczny@gmail.com.

1. Introduction

Uniformity properties of sequences defined in terms of digital expansions have long been studied. Consider, for instance, the Thue–Morse sequence $t(n) = (-1)^{s_2(n)}$, where $s_2(n)$ denotes the sum of binary digits of n , discussed at length by Allouche and Shallit in the survey paper [AS99]. It was shown by Gelfond [Gel68] (see also [MS98]) that $t(n)$ is equidistributed in arithmetic progressions:

$$\lim_{N \rightarrow \infty} |\{0 \leq n < N \mid t(An + B) = +1\}| / N = 1/2 \tag{1}$$

for all $A \in \mathbb{N}$ and $B \in \mathbb{N}_0$, and the rate of convergence can be made explicit. Analogous results hold also for other bases, with mild additional assumptions to account for the fact that $s_k(n) \equiv n \pmod{k - 1}$. Mauduit and Sárközy [MS98] also observed that the Thue–Morse sequence admits large self-correlations. Here, the (self-)correlation coefficients of a sequence $a: \mathbb{N} \rightarrow \mathbb{C}$ are defined by

$$\gamma_a(m) := \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{n=0}^{N-1} a(n)\overline{a(n+m)}, \tag{2}$$

and a simple computation shows that $\gamma_t(1) = -1/3 \neq 0$ (see Section 3 for details). By the same token,

$$\gamma_t(2^\ell) = -1/3 \text{ for all } \ell \in \mathbb{N}_0, \tag{3}$$

meaning in particular that $\gamma(m) \not\rightarrow 0$ as $m \rightarrow \infty$. On the other hand, the coefficients $\gamma_t(m)$ tend to be rather small; in particular

$$\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{m=0}^{N-1} \gamma_t(m)^2 = 0, \tag{4}$$

which follows e.g. from results in [Coq76]. The spectral measure μ_a on \mathbb{R}/\mathbb{Z} associated to a sequence $a: \mathbb{N} \rightarrow \mathbb{C}$ is characterised by the identity $\int_{\mathbb{R}/\mathbb{Z}} \exp(2\pi imt) d\mu_a(t) = \gamma_a(m)$, and (4) is equivalent to absolute continuity of μ_t .

Many other notions of uniformity have been investigated for the Thue–Morse sequence. In an influential paper, Mauduit and Rivat showed that $t(n)$ and its analogues in different bases are equidistributed along the primes [MR10]. Drmota, Mauduit and Rivat [DMR19] showed that $t(n^2)$ is a normal sequence, meaning that each finite sequence of ± 1 s appears with the expected frequency. Spiegelhofer [Spi18] proved that $t(n)$ has level of distribution 1, which is a far-reaching quantitative generalisation of (1) and can be used to show equidistribution along Piatetski–Shapiro sequences $[n^c]$, $1 < c < 2$ (see [FM96] for analogous, but somewhat weaker, results in different bases). It was also shown by the author [Kon19] that $t(n)$ has small Gowers norms, meaning that it is uniform from the point of view of higher order Fourier analysis.

Another oft-studied sequence carries the name of Rudin–Shapiro and is given by $r(n) = (-1)^{\#(11,n)}$, where $\#(11,n)$ denotes the number of times the pattern 11 appears in the binary expansion of n , allowing overlaps. Similarly to the Thue–Morse sequence, the Rudin–Shapiro sequence is equidistributed in arithmetic progressions and along the primes [MR15], and has small Gowers norms [Kon19]. However, in contrast to (3), the Rudin–Shapiro sequence is *noncorrelated*, by which we mean that $\gamma_r(m) = 0$ for all $m \geq 1$ or, equivalently, that the spectral measure μ_r is the Lebesgue measure. Intuitively, noncorrelated sequences are free of any sort of periodic behaviour.

The Thue–Morse and the Rudin–Shapiro sequences are special cases of what we call binary pattern sequences. In general, a binary pattern sequence takes the form

$$a(n) = a_A(n) = (-1)^{\#(A,n)},$$

where A is a finite set of patterns over the alphabet $\{0, 1\}$ and $\#(A, n)$ denotes the total number of appearances of patterns from A in the binary expansion of n (see Section 2 for details). Pattern sequences were studied in a more general context by Morton and Mourant [MM89, Mor90], Coquet, Kamae and Mendès France [CKMF77], and Boyd, Cook and Morton [BCM89]. Generalised Rudin–Shapiro sequences and their correlation coefficients were studied by Allouche and Liardet [AL91]. Finally, Zheng, Peng and Kamae [ZPK18] studied correlation coefficients of binary pattern sequences, and obtained a complete classification of noncorrelated sequences corresponding to sets of patterns of length ≤ 3 . Examples of sets A that give rise to noncorrelated sequences include:

- $\{11\}$ (then $a_A(n) = r(n)$ is the Rudin–Shapiro sequence);
- $\{11, 1\}$ (then $a_A(n) = r(n)t(n)$) and $\{10, 1\}$ (then $a_A(n) = (-1)^n r(n)$);
- $\{101, 111\}$, or more generally $\{101, 111\} \cup B$ for a set $B \subseteq \{0, 1\}^2$.

In this paper, we extend the result of [ZPK18] to patterns of length ≤ 4 and put the findings in a wider context provided by the theory of automatic and regular sequences. Many of the ideas we use have their analogues and prototypes in [ZPK18]; throughout the paper we give references to the relevant results therein.

Unfortunately, there does not appear to be a simple criterion that determines if a given pattern sequence is noncorrelated (except for the partial information suggested by Conjecture 1.2 below). Due to practical limitations we only state a counting result here, as opposed to a complete list.¹

Theorem A. *There are precisely 2272 noncorrelated binary pattern sequences corresponding to patterns of length ≤ 4 .*

¹ The list, together with the code which can be used to produce it, is available from the author.

As a key step towards obtaining the above result, we reduce the task of verifying whether a given binary pattern sequence is noncorrelated to a finite computation, which can then be automated. The time complexity of the resulting algorithm is polynomial in 2^ℓ , where ℓ denotes the length of patterns under consideration. Since it takes approximately 2^ℓ bits to specify a binary pattern sequence, this is optimal up to improvements in the exponent.

Theorem B. *There exists an algorithm which, given a finite set of patterns $A \subseteq \{0, 1\}^\ell$, performs $2^{O(\ell)}$ operations and decides if the corresponding pattern sequence a_A is noncorrelated.*

While we keep the exposition fairly self-contained, we also wish to emphasize that the above problem can be seen as a part of a larger theory. We note that binary pattern sequences are 2-automatic (see Section 2 for the relevant definitions). A crucial component of our reasoning is Theorem 3.5, which asserts that the correlation sequences coming from automatic sequences are regular. While this result will not come as a surprise to the experts in the field, to the best of our knowledge it does not appear in print elsewhere. Its importance stems from the fact that a regular sequence admits a simple recursive description, and hence many properties are easily verified for such a sequence. In our particular application, we reduce the task of determining if a pattern sequence is noncorrelated to the ostensibly simpler task of determining if a 2-regular sequence is identically zero.

The problem of classifying noncorrelated pattern sequences becomes more tractable if we impose additional assumptions on the set of patterns under consideration. Let us call a binary pattern sequence $a: \mathbb{N}_0 \rightarrow \{+1, -1\}$ *dilation-invariant* if $a(2n) = a(n)$ for all $n \in \mathbb{N}_0$, or equivalently, if $a = a_A$ for a set A that contains only patterns that begin and end with 1 (see Section 2.4 for details). In the dilation-invariant case, we have a conjectural classification, which we are able to confirm in one direction in full generality, and in the opposite direction for patterns of length ≤ 5 .

Theorem C. *Let A be a set of patterns over the alphabet $\{0, 1\}$, all of which begin and end with 1. Let ℓ be the length of the longest word in A and let $a = a_A$ be the corresponding binary pattern sequence. If $\ell \geq 2$ and $1\{0, 1\}^{\ell-2}1 \subseteq A$ then a is noncorrelated. Conversely, if $2 \leq \ell \leq 5$ and a is noncorrelated then $1\{0, 1\}^{\ell-2}1 \subseteq A$.*

Conjecture 1.1. *Let A , ℓ and a be as in Theorem C. If a is noncorrelated then $\ell \geq 2$ and $1\{0, 1\}^{\ell-2}1 \subseteq A$.*

If $A = 1\{0, 1\}^{\ell-2}1$ then the fact that a_A is noncorrelated follows from [ZPK18]. More generally, Theorem 1.3 in [ZPK18] (see also [AL91]) provides a classification of all noncorrelated binary pattern sequences a_A for sets of patterns of the form $A = w_1\{0, 1\}^{\ell_1}w_2\{0, 1\}^{\ell_2} \dots w_s\{0, 1\}^{\ell_s}w_{s+1}$ where $s \in \mathbb{N}$, w_i are words over the alphabet $\{0, 1\}$ and $\ell_1, \ell_2, \dots, \ell_s \in \mathbb{N}_0$. Conjecture 1.1 is consistent with said classification.

Returning to the general case, we notice that each binary pattern sequence $a: \mathbb{N}_0 \rightarrow \{+1, -1\}$ can be written as the product of a periodic sequence h and a dilation-invariant pattern sequence b (Lemma 2.9). The correlation coefficients of a and b are closely related (see also Remark 5.6), and in all cases that we were able to check (i.e., $\ell \leq 4$), if a is noncorrelated then so is b . This motivates us to put forward the following conjecture.

Conjecture 1.2. *Let $a: \mathbb{N}_0 \rightarrow \{+1, -1\}$ be a noncorrelated binary pattern sequence. Then a is the product of a periodic sequence and an dilation-invariant noncorrelated binary pattern sequence*

Above we restricted our attention to base 2 for the sake of brevity. In the remaining part of the paper, we work in arbitrary base $k \geq 2$. In particular, the natural base- k variant of Theorem B holds true. The same applies to the first part of Theorem C, except that it is less clear what the base- k variant should be and the resulting statement is vacuous for many values of k (see Proposition 5.3). When it comes to computations, we only consider base 2 since for larger bases the number of distinct pattern sequences becomes so large that merely listing them all is already infeasible even for modest pattern lengths.

Acknowledgements. While writing this paper, the author was supported by the ERC grant ErgComNum 682150 at the Hebrew University of Jerusalem. During the review process, the author was working within the framework of the LABEX MILYON (ANR-10-LABX-0070) of Université de Lyon, within the program “Investissements d’Avenir” (ANR-11-IDEX-0007) operated by the French National Research Agency (ANR). The author also acknowledges support from the Foundation for Polish Science (FNP).

The author wishes to express his gratitude to Boris Adamczewski, Jakub Byszewski, Aihua Fan and Tamar Ziegler for helpful conversations and to the anonymous Referee for the careful reading of this paper and constrictive suggestions.

2. Background and definitions

Convention: *Throughout the paper, k denotes the base and is considered to be fixed. In particular, all constructions and constants are allowed to depend on k unless explicitly stated otherwise.*

2.1. Pattern sequences

We let $\Sigma_k = \{0, 1, \dots, k - 1\}$ denote the set of digits in base k . For a set X , we let X^* denote the monoid consisting of words over the alphabet X , equipped with the operation of concatenation and neutral element ϵ , the empty word. For $v \in X^*$, $|v|$ denotes the length of v . For $n \in \mathbb{N}_0$, $(n)_k \in \Sigma_k^*$ denotes the expansion of n in base k (without leading zeros). Conversely, for $v \in \Sigma_k^*$, $[v]_k \in \mathbb{N}_0$ denotes the integer encoded by v .

Let X be a set. We say that a word $v \in X^*$ appears in another word $w \in X^*$, or that v is a factor of w , if there exist $x, y \in X^*$ such that $w = xvy$. We call v a prefix (resp. suffix) of w if we may take $x = \epsilon$ (resp. $y = \epsilon$). We further define $\#(v, w)$ to be the number of times v appears in w , that is, the number distinct of pairs $(x, y) \in X^* \times X^*$ such that $w = xvy$. We note that this definition allows for overlaps, so for instance $\#(010, 01010) = 2$. More generally, for a finite set $A \subseteq X^*$, we define $\#(A, w) = \sum_{v \in A} \#(v, w)$.

Accordingly, for $n \in \mathbb{N}_0$ and $v \in \Sigma_k^* \setminus \{0\}^*$, $\#(n, v)$ denotes the number of times that v appears in the base- k expansion of n padded with sufficiently many leading zeros, that is, $\#(v, n) = (v, 0^{|v|-1}(n)_k)$. The inclusion of the leading zeros in the expansion of n ensures better behaviour of the map $n \mapsto \#(v, n)$; in particular, for each $n, m \in \mathbb{N}_0$ and sufficiently large $\alpha \in \mathbb{N}_0$ we have $\#(v, k^\alpha m + n) = \#(v, k^\alpha m) + \#(v, n)$. The assumption that v is not a string of zeros ensures that $\#(v, n)$ is well-defined, in the sense that for fixed n , the sequence $(v, 0^\alpha(n)_k)$ ($\alpha \in \mathbb{N}_0$) is eventually constant.

We will call a set $A \subseteq \Sigma_k^*$ *admissible* if A is finite and $A \cap \{0\}^* = \emptyset$, so that we may define $\#(A, n) = \sum_{v \in A} \#(v, n)$. For any admissible set A , the corresponding *pattern sequence* is defined by (cf. [ZPK18, Definition 1.1])

$$a_A(n) = (-1)^{\#(A, n)}. \tag{5}$$

If additionally $|u| \leq \ell$ for all $u \in A$ then we say that a is a pattern sequence of length $\leq \ell$, or equivalently we define the *length* of a as the least possible value of $\max_{u \in A} |u|$ among all representations of a in the form (5), where $A \subseteq \Sigma_k^*$ is an admissible set. Note that one pattern sequence can have multiple representations of the aforementioned form.

For two sets A, B , we let $A \oplus B$ denote the symmetric difference $(A \setminus B) \cup (B \setminus A)$.

Lemma 2.1. *The class of pattern sequences $\mathbb{N}_0 \rightarrow \{+1, -1\}$ is closed under multiplication.*

Proof. It is enough to note that for any admissible sets $A, B \subseteq \Sigma_k^*$ we have

$$a_A \cdot a_B = a_{A \oplus B}. \quad \square$$

It will usually be convenient to impose further restrictions on the set of patterns A . Depending on the context we require that either A has not leading zeros (in the sense that 0 is not a prefix of v for any $v \in A$) or that A has constant length (in the sense that there is some $\ell \in \mathbb{N}$ such that $|v| = \ell$ for all $v \in A$).

Lemma 2.2. *Let $\ell \in \mathbb{N}_0$ and let $a: \mathbb{N}_0 \rightarrow \{+1, -1\}$ be a pattern sequence of length $\leq \ell$. Then there exist admissible sets $B, C \subseteq \Sigma_k^*$ such that B has no leading zeros, C has constant length ℓ , and $a = a_B = a_C$. Moreover, B and C are uniquely determined by a .*

Proof. Pick any admissible set $A \subseteq \Sigma_k^*$ with $a = a_A$. Note that for each $v \in \Sigma_k^*$ and each $n \in \mathbb{N}_0$ we have

$$\#(v, n) = \sum_{i=0}^{k-1} \#(iv, n) \tag{6}$$

To construct B , begin with A and as long as A contains at least one word starting with 0 , say $0v$, replace A with $A \oplus \{iv \mid i \in \Sigma_k\} \oplus \{v\}$. Because of (6), this operation does not change the sequence a_A . Since each iteration decreases the total number of leading zeros in the patterns in A , after a finite number of steps this procedure must terminate and the resulting set of patterns has no leading zeros.

To construct C , likewise, begin with A and as long as A contains at least one word v with length $|v| < \ell$, pick the shortest such word v and replace A with $A \oplus \{iv \mid i \in \Sigma_k\} \oplus \{v\}$. Like before, this operation does not change the sequence a_A . Each iteration either decreases the number of words in A with least possible length, or increases the length of the shortest word in A . At the same time, no words of length larger than ℓ are introduced. Hence, after a finite number of steps this procedure must terminate and the resulting set of patterns has constant length equal to ℓ .

It remains to show uniqueness. Using Lemma 2.1, we may assume that $A = \emptyset$. For the sake of contradiction, suppose that one of B and C is non-empty. Consider first the case when $B \neq \emptyset$ and let v be the shortest word in B . Then $1 = a_B([v]_k) = -1$, since $0^\ell v$ contains exactly one pattern from B , namely v . Hence, we have reached a contradiction. Consider next the case when $C \neq \emptyset$ and choose the word $0^m v \in C$ where m is largest possible. Then we again reach the contradiction: $1 = a_C([v]_k) = -1$. \square

Remark 2.3. We focus our attention on ± 1 -valued sequences for two basic reasons. The first one is practical: The noncorrelation phenomenon that we are interested in relies on occurrence of certain arithmetic coincidences, which become less likely as the number of possible values increases; accordingly, the computational part of the problem becomes increasingly resource-intensive as sequences under consideration become more complicated. The second reason is conceptual: For a ± 1 -valued sequence a with mean $M_a = 0$, noncorrelation is tantamount to equidistribution of the pairs $a(n), a(n + m)$. More precisely, for each $m \in \mathbb{N}$, if we additionally assume that the limits mentioned above exist then $\gamma_a(m) = 0$ if and only if

$$\lim_{N \rightarrow \infty} \frac{|\{n < N \mid a(n) = i, a(n + m) = i'\}|}{N} = \frac{1}{4} \quad \text{for all } i, i' \in \{+1, -1\}.$$

The analogous characterisation is false without the assumption that a is allowed to take more than 2 values.

2.2. Automatic sequences

In this section we briefly discuss the basics of the theory of automatic sequences; for extensive background see [AS03a]. For $i \in \Sigma_k$, we define the operators Λ_i acting on sequences $\mathbb{N}_0 \rightarrow \mathbb{C}$ by

$$\Lambda_i a(n) = a(kn + i). \tag{7}$$

The *k*-kernel of a sequence $a: \mathbb{N}_0 \rightarrow \mathbb{C}$ consists of all sequences $\mathbb{N}_0 \rightarrow \mathbb{C}$ that can be obtained from a by repeated application of Λ_i 's, that is,

$$\mathcal{N}_k(a) = \{n \mapsto a(k^\alpha n + r) \mid \alpha, r \in \mathbb{N}_0, 0 \leq r < k^\alpha\} \subseteq \mathbb{C}^{\mathbb{N}_0}. \tag{8}$$

It will also be convenient to introduce the shift operator S acting on sequences $\mathbb{N}_0 \rightarrow \mathbb{C}$ by $Sa(n) = a(n + 1)$. For future reference, we record how the introduced operators interact.

Lemma 2.4. *For each $0 \leq i < k - 1$ we have $\Lambda_i S = \Lambda_{i+1}$. Moreover, $\Lambda_{k-1} S = S \Lambda_1$.*

Proof. Direct computation. \square

A sequence a is *k*-automatic (or just automatic, if k is clear from the context) if and only if $\mathcal{N}_k(f)$ is finite. Many equivalent definitions of automaticity are possible, and we briefly mention some of them to provide context. Details and terminology can be found in [AS03a]. As the name suggests, a sequence is *k*-automatic if and only if it is computed by a deterministic finite *k*-automaton with output. Any fixed point of a *k*-uniform morphism is *k*-automatic, and conversely any *k*-automatic sequence can be obtained as a letter-to-letter coding of a fixed point of a *k*-uniform morphism. When k is a prime and a is a sequence taking values in a finite field \mathbb{F} of characteristic k , yet another criterion due to Christol shows that a is automatic if and only if the associated formal power series is algebraic over \mathbb{F} .

It is a well-known fact that the class of *k*-automatic complex-valued sequences is closed under addition, multiplication, conjugation and restriction to subsequences, that is, if $a, b: \mathbb{N}_0 \rightarrow \mathbb{C}$ are *k*-automatic, then so are $n \mapsto a(n) + b(n)$, $n \mapsto a(n) \cdot b(n)$, $n \mapsto \bar{a}(n)$ and $n \mapsto a(An + B)$ for any $A \in \mathbb{N}$, $B \in \mathbb{N}_0$. More generally, if $a, b: \mathbb{N}_0 \rightarrow \mathbb{C}$ are *k*-automatic and $h: \mathbb{C}^2 \rightarrow \mathbb{C}$ is arbitrary, then the sequence $n \mapsto h(a(n), b(n))$ is *k*-automatic.

For a sequence $a: \mathbb{N}_0 \rightarrow \mathbb{C}$, we define the mean and the logarithmic mean:

$$M_a := \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{n=0}^{N-1} a(n), \quad M_a^{\log} := \lim_{N \rightarrow \infty} \frac{1}{\log N} \sum_{n=0}^{N-1} \frac{1}{n+1} a(n). \tag{9}$$

We note that M_a are not guaranteed to exist, even when the sequence a is automatic. (Consider, for instance, the sequence defined by $a(0) = 0$ and $a(n) = (-1)^\alpha$ if $k^\alpha \leq n < k^{\alpha+1}$, $\alpha \in \mathbb{N}_0$.) On the other hand, we have the following positive result for logarithmic means.

Theorem 2.5 ([AS03a, Thm. 8.4.8]). *Let $a: \mathbb{N}_0 \rightarrow \mathbb{C}$ be a *k*-automatic sequence. Then M_a^{\log} exists.*

We also record the fact that that if $a: \mathbb{N}_0 \rightarrow \mathbb{C}$ is a bounded sequence and M_a exists then M_a^{\log} also exists and $M_a^{\log} = M_a$, see e.g. [AS03a, Prop. 8.4.4 (a)].

Pattern sequences are, unsurprisingly, automatic. In fact, we have the following characterisation of pattern sequences in terms of their k -kernels (cf. [ZPK18, Lemma 2.2]).

Lemma 2.6. *Let $a: \mathbb{N}_0 \rightarrow \{+1, -1\}$ be a sequence with $a(0) = +1$ and $\ell \in \mathbb{N}$. Then the following conditions are equivalent:*

- (i) *There exists a set $A \subseteq \Sigma_k^\ell \setminus \{0^\ell\}$ with $a = a_A$.*
- (ii) *For each $b \in \mathcal{N}_k(a)$, the sequence b/a has period $k^{\ell-1}$.*

Proof. (i) \Rightarrow (ii): Let $i \in \Sigma_k$ and $n \in \mathbb{N}_0$. Then each factor of $(n)_k$ is also a factor of $(kn + i)_k = (n)_k i$ and conversely each factor of $(kn + i)_k$ that is not a suffix is a factor of $(n)_k$. More precisely, for each $v \in \Sigma_k^* \setminus \{0\}^*$ we have

$$\#(v, kn + i) = \begin{cases} \#(v, n) + 1 & \text{if } v \text{ is a suffix of } 0^{|v|-1}(kn + i)_k, \\ \#(v, n) & \text{otherwise.} \end{cases}$$

Consequently, $\Lambda_i a(n)/a(n) = -1$ if the suffix of $0^{\ell-1}(kn + i)_k$ of length ℓ belongs to A and $\Lambda_i a(n)/a(n) = +1$ otherwise. It follows that $h_i := \Lambda_i a/a$ is $k^{\ell-1}$ -periodic. Since for each $i \in \Sigma_k$, the operator Λ_i maps $k^{\ell-1}$ -periodic sequences to $k^{\ell-2}$ -periodic sequences (or constant sequences, if $\ell = 1$), it follows that all sequences in the k -kernel of a take the form $a \cdot h$ where h is $k^{\ell-1}$ -periodic.

(ii) \Rightarrow (i): For each $i \in \Sigma_k$, let $h_i := \Lambda_i a/a$. Note that $h_0(0) = a(0)/a(0) = +1$ and that the sequences h_i take values in $\{+1, -1\}$. Conversely, given any k -tuple of $k^{\ell-1}$ -periodic $\{+1, -1\}$ -valued sequences h'_i ($i \in \Sigma_k$) with $h'_0(0) = +1$, we can inductively construct a sequence $a: \mathbb{N}_0 \rightarrow \{+1, -1\}$ with $a(0) = +1$ and $h_i = h'_i$ for all $i \in \Sigma_k$. Hence, the number of sequences that satisfy (ii) is

$$\left(2^{k^{\ell-1}}\right)^{k-1} \cdot 2^{k^{\ell-1}-1} = 2^{k^\ell-1}.$$

On the other hand, the number of subsets of $\Sigma_k^\ell \setminus \{0^\ell\}$ is also equal to $2^{k^\ell-1}$, and by the previously proven implication and Lemma 2.2, each of these choices gives rise to a different sequence satisfying (ii). It follows that each sequence satisfying (ii) has a representation as in (i). \square

2.3. Regular sequences

The class of k -regular sequences was introduced by Allouche and Shallit [AS92,AS03b] as a natural generalization of the class of k -automatic sequences.

Let R be a ring contained in \mathbb{C} . A sequence $f: \mathbb{N}_0 \rightarrow \mathbb{C}$ is (R, k) -regular if $\mathcal{N}_k(f)$ is contained in a finitely generated R -module. Note that if $R' \subseteq \mathbb{C}$ is another ring and

$R \subseteq R'$ then any (R, k) -regular sequence is also (R', k) -regular. In our context, the choice of the ring R does not play a major role: For the sake of brevity, we set $R = \mathbb{Q}$ throughout the paper and omit R from the notation. (Strictly speaking we could have worked with $R = \mathbb{Z}[1/k]$, making some results marginally stronger.) The fact that the ring under consideration is in fact a field leads to a slightly more succinct definition of regularity: A sequence $f: \mathbb{N}_0 \rightarrow \mathbb{C}$ is k -regular if and only if its k -kernel spans a finite dimensional vector space over \mathbb{Q} : $\dim \text{span}_{\mathbb{Q}} \mathcal{N}_k(f) < \infty$.

The class of k -regular sequences enjoys closure properties analogous to k -automatic sequences: If $f, g: \mathbb{N}_0 \rightarrow \mathbb{C}$ are k -regular, then so are $n \mapsto f(n) + g(n)$, $n \mapsto f(n) \cdot g(n)$, $n \mapsto \overline{f}(n)$, $n \mapsto zf(n)$ ($z \in \mathbb{C}$) and $n \mapsto f(An + B)$ ($A \in \mathbb{N}$, $B \in \mathbb{N}_0$). In particular, k -regular sequences $\mathbb{N}_0 \rightarrow \mathbb{C}$ form an involutive algebra over \mathbb{C} (with addition and multiplication defined pointwise).

We will need a method to verify if a given regular sequence is identically zero. The following lemma provides a simple criterion.

Lemma 2.7. *Let $f: \mathbb{N}_0 \rightarrow \mathbb{C}$ be k -regular and non-zero. Then there exists $g \in \mathcal{N}_k(f)$ with $g(0) \neq 0$.*

Proof. For the sake of contradiction, suppose that $g(0) = 0$ for all $g \in \mathcal{N}_k(f)$. We show by induction on α that $g(n) = 0$ for all $g \in \mathcal{N}_k(f)$ and $0 \leq n < k^\alpha$. If $\alpha = 0$ then $n = 0$, so there is nothing to prove. If $\alpha > 0$ and $n < k^\alpha$ then $n = kn' + i$ where $i \in \Sigma_k$ and $n' < k^{\alpha-1}$. Hence, $g(n) = \Lambda_i g(n') = 0$ by the inductive assumption. \square

2.4. Invariant sequences

We will say that a sequence $a: \mathbb{N}_0 \rightarrow \mathbb{C}$ is *dilation-invariant* if $a(kn) = a(n)$ for all $n \in \mathbb{N}_0$. The dilation-invariant pattern sequences admit a simple description. Following the convention in Section 2.1, we will say that a set $A \subseteq \Sigma_k^*$ has no trailing zeros if 0 is not a suffix of any $v \in A$.

Lemma 2.8. *Let $a: \mathbb{N}_0 \rightarrow \{+1, -1\}$ be a pattern sequence. Then a is dilation-invariant if and only if there exists a set $A \subseteq \Sigma_k^*$ that has no leading and no trailing zeros and such that $a = a_A$.*

Proof. If $A \subseteq \Sigma_k^*$ has no trailing zeros then $\#(v, n) = \#(v, kn)$ for all $v \in A$ and $n \in \mathbb{N}_0$, so a_A is dilation-invariant.

Conversely, suppose that a is dilation-invariant and let $A \subseteq \Sigma_k^*$ be a set of patterns without leading zeros such that $a = a_A$, which exists by Lemma 2.2. Suppose for the sake of contradiction that A contains a pattern ending with 0, say $u0 \in A$ for some $u \in \Sigma_k^*$, and let u be as short as possible. Since a is dilation-invariant, we have

$$a([u0]_k) = a([u]_k). \tag{10}$$

On the other hand, each $v \in A$ either ends in a non-zero digit (in which case $\#(v, u0) = \#(v, u)$), or ends in 0 and is not a factor of $u0$ (in which case $\#(v, u0) = \#(v, u) = 0$), or is equal to $u0$ (in which case $\#(v, u0) = 1$ and $\#(v, u) = 0$). As a consequence,

$$a([u0]_k) = (-1)^{\#(A, u0)} = (-1)^{\#(A, u)+1} = -a([u]_k), \tag{11}$$

which contradicts (10) and finishes the argument. \square

We also record the fact that every pattern sequence is the product of a dilation-invariant sequence and a periodic sequence. As we will see (cf. Remark 5.6) the introduction of the multiplicative factor affects the correlation coefficients in a relatively simple way, which motivates our focus on dilation-invariant sequences.

Lemma 2.9. *Let $\ell \in \mathbb{N}_0$ and let $a: \mathbb{N}_0 \rightarrow \{+1, -1\}$ be a pattern sequence of length $\leq \ell$. Then there exists a unique dilation-invariant pattern sequence of length $\leq \ell$ such that a/b is $k^{\ell-1}$ -periodic.*

Proof. By Lemma 2.2, we may assume that $a = a_A$ for a set $A \subseteq \Sigma_k^*$ without leading zeros. Reasoning along similar lines as in the proof of Lemma 2.2, we note that for any word v , we have

$$\sum_{i=0}^{k-1} \#(vi, n) - \#(v, n) = \begin{cases} 1 & \text{if } v \text{ is a suffix of } n, \\ 0 & \text{otherwise.} \end{cases}$$

In particular, letting $D(v) := \{vi \mid i \in \Sigma_k\} \cup \{v\}$, we see that the sequence $a_{D(v)}$ is $k^{|v|}$ -periodic. We construct a sequence of sets $A := A_0, A_1, \dots, A_t =: B$, where $A_{j+1} = A_j \oplus D(v)$ if A_j contains the word $v0$ for some $v \in \Sigma_k^*$ and t is the first index such that no word in A_t ends with 0. This construction is guaranteed to terminate because each step decreases the total length of words in A_j that end with 0. Letting $b := a_B$ we observe that a/b is the product of $k^{\ell-1}$ -periodic sequences and hence $k^{\ell-1}$ -periodic. \square

3. Correlation coefficients

In this section we study correlation coefficients of k -automatic sequences and show that they are k -regular (Corollary 3.5). This allows us to reduce the task of verifying if a given k -automatic sequence is noncorrelated to checking if a given k -regular sequence is identically zero on \mathbb{N} , which can be accomplished with the help of Lemma 2.7.

3.1. Definitions

For two sequences $a, b: \mathbb{N}_0 \rightarrow \mathbb{C}$, we define the correlation coefficients:

$$\gamma_{a,b}(m) := \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{n=0}^{N-1} a(n)\bar{b}(n+m), \tag{12}$$

if the limit exists (otherwise, $\gamma_{a,b}(m)$ is considered undefined). We are often interested in the case where $a = b$, when we write γ_a in place of $\gamma_{a,a}$. Unfortunately, the limit defining $\gamma_{a,b}(m)$ is not guaranteed to converge even if a and b are automatic. This motivates us to consider the logarithmic correlation coefficients, defined by

$$\gamma_{a,b}^{\log}(m) := \lim_{N \rightarrow \infty} \frac{1}{\log N} \sum_{n=0}^{N-1} \frac{1}{n+1} a(n)\bar{b}(n+m). \tag{13}$$

If a and b are automatic and $m \in \mathbb{N}_0$, then the sequence $n \mapsto a(n)\bar{b}(n+m)$ is also automatic. Since Theorem 2.5 guarantees existence of logarithmic means of automatic sequences, we have the following fact.

Corollary 3.1. *Let $a, b: \mathbb{N}_0 \rightarrow \mathbb{C}$ be k -automatic sequences. Then the coefficients $\gamma_{a,b}^{\log}(m)$ are well-defined for all $m \in \mathbb{N}_0$. Moreover, if the coefficient $\gamma_{a,b}(m)$ is well-defined for some $m \in \mathbb{N}_0$ then $\gamma_{a,b}(m) = \gamma_{a,b}^{\log}(m)$.*

3.2. Recurrence

Our next goal is to obtain a recursive description of the correlation coefficients discussed above. Recall that for a k -automatic sequence a , the kernel $\mathcal{N}_k(a)$ is finite and closed under the operators Λ_i defined in (7) for all $i \in \Sigma_k$.

Lemma 3.2. *Let \mathcal{N} be a finite set of sequences $\mathbb{N}_0 \rightarrow \mathbb{C}$, closed under the operators Λ_i for all $i \in \Sigma_k$. Then for all $a, b \in \mathcal{N}$ it holds that*

$$\gamma_{a,b}^{\log}(m) = \frac{1}{k} \sum_{i=0}^{k-1} \gamma_{a'_i, b'_i}^{\log}(m'_i), \tag{14}$$

where a'_i, b'_i and m'_i are given by

$$a'_i = \Lambda_i a, \quad b'_i = \Lambda_{m+i \bmod k} b, \quad m'_i = \left\lfloor \frac{m+i}{k} \right\rfloor. \tag{15}$$

Proof. Rescaling if necessary, we assume that all sequences in \mathcal{N} are 1-bounded (that is, $|a(n)| \leq 1$ for all $a \in \mathcal{N}$ and $n \in \mathbb{N}_0$). For each $N > 0$, splitting $[N]$ into residue classes modulo k we obtain

$$\begin{aligned} \sum_{n=0}^{N-1} \frac{a(n)\bar{b}(n+m)}{n+1} &= \sum_{i=0}^{k-1} \sum_{n=0}^{\lfloor N/k \rfloor - 1} \frac{a(kn+i)\bar{b}(kn+i+m)}{kn+i+1} + O\left(\frac{1}{N}\right). \\ &= \frac{1}{k} \sum_{i=0}^{k-1} \sum_{n=0}^{\lfloor N/k \rfloor - 1} \frac{a'_i(n)\bar{b}'_i(n+m'_i)}{n+1} + O(1), \end{aligned}$$

where a'_i, b'_i and m'_i are given by (15), and we use the estimate $1/(kn+i+1) = 1/k(n+1) + O(1/(n+1)^2)$ together with the fact that $\sum_{n=0}^{\infty} 1/(n+1)^2$ is summable. Dividing by $\log N$ and recalling that $1/\log(N/k) = 1/\log N + O(1/\log^2 N)$ we obtain

$$\frac{1}{\log N} \sum_{n=0}^{N-1} \frac{a(n)\bar{b}(n+m)}{n+1} = \frac{1}{k} \sum_{i=0}^{k-1} \frac{1}{\log \lfloor N/k \rfloor} \sum_{n=0}^{\lfloor N/k \rfloor - 1} \frac{a'_i(n)\bar{b}'_i(n+m'_i)}{n+1} + O\left(\frac{1}{\log N}\right).$$

Letting $N \rightarrow \infty$, we obtain (14). \square

While the coefficients $\gamma_{a,b}^{\log}(m)$ are better-behaved in general, our original motivation concerns the coefficients $\gamma_{a,b}(m)$ (where additionally $a = b$). Fortunately, existence of the latter is easy to ensure under mild additional assumptions.

Lemma 3.3. *Let \mathcal{N} be a finite set of sequences $\mathbb{N}_0 \rightarrow \mathbb{C}$, closed under the operators Λ_i for all $i \in \Sigma_k$. Suppose that $\gamma_{a,b}(0)$ exists for each $a, b \in \mathcal{N}$. Then also $\gamma_{a,b}(m)$ exists for all $a, b \in \mathcal{N}$ and $m \in \mathbb{N}_0$ and, using the notation from (15), satisfy*

$$\gamma_{a,b}(m) = \frac{1}{k} \sum_{i=0}^{k-1} \gamma_{a'_i, b'_i}(m'_i), \tag{16}$$

Proof. Rescaling if necessary, we may assume that all sequences in \mathcal{N} are 1-bounded. Generalizing the definition of $\gamma_{a,b}(m)$ slightly, for $x \geq 1$ let us put

$$\gamma_{a,b}(m; x) = \frac{1}{\lfloor x \rfloor} \sum_{n=0}^{\lfloor x \rfloor - 1} a(n)\bar{b}(n+m). \tag{17}$$

Then, following the same reasoning as in Lemma 3.2, we find the recursive relation

$$\gamma_{a,b}(m; x) = \frac{1}{k} \sum_{i=0}^{k-1} \gamma_{a'_i, b'_i}(m'_i; x/k) + O(1/x), \tag{18}$$

where a'_i, b'_i and m'_i are given by (15).

In particular, for $m = 1$ we obtain

$$\gamma_{a,b}(1; x) = \frac{1}{k} \sum_{i=0}^{k-2} \gamma_{a'_i, b'_i}(0; x/k) + \frac{1}{k} \gamma_{a'_{k-1}, b'_{k-1}}(1; x/k) + O(1/x). \tag{19}$$

Iterating (19) t times, we conclude that there exist weights $w_{a',b'}^{(t)} \geq 0$ ($a', b' \in \mathcal{N}$) with $\sum_{a',b' \in \mathcal{N}} w_{a',b'}^{(t)} = 1 - 1/k^t$ and sequences $a^{(t)}, b^{(t)} \in \mathcal{N}$ such that

$$\gamma_{a,b}(1; x) = \sum_{a',b' \in \mathcal{N}} w_{a',b'}^{(t)} \gamma_{a',b'}(0; x/k^t) + \frac{1}{k^t} \gamma_{a^{(t)},b^{(t)}}(1; x/k^t) + O(k^t/x). \tag{20}$$

Since $\gamma_{a',b'}(0; y) \rightarrow \gamma_{a',b'}(0)$ as $y \rightarrow \infty$ for each $a', b' \in \mathcal{N}$, letting $x \rightarrow \infty$ in (20) we conclude that there exists a number $\gamma_{a,b}^{(t)}(1) = \sum_{a',b' \in \mathcal{N}} w_{a',b'}^{(t)} \gamma_{a',b'}(0)$ such that

$$\limsup_{x \rightarrow \infty} \left| \gamma_{a,b}(1; x) - \gamma_{a,b}^{(t)}(1) \right| = O(1/k^t). \tag{21}$$

It follows that the sequence $\gamma_{a,b}^{(t)}(1)$ ($t \in \mathbb{N}$) is Cauchy, and $\gamma_{a,b}(1)$ is well-defined:

$$\gamma_{a,b}(1) = \lim_{x \rightarrow \infty} \gamma_{a,b}(1; x) = \lim_{t \rightarrow \infty} \gamma_{a,b}^{(t)}(1). \tag{22}$$

We are now ready to prove by induction on m that the coefficients $\gamma_{a,b}(m)$ are well-defined for all $m \in \mathbb{N}_0$ and $a, b \in \mathcal{N}$. The case $m = 0$ is included in the assumptions, and we have dealt with $m = 1$ above. Suppose now that $m \geq 2$. For each $i \in \Sigma_k$, since $\lfloor x/k \rfloor \leq x/k < x$ for all $x > 0$, we have

$$m'_i = \left\lfloor \frac{m+i}{k} \right\rfloor \leq \left\lfloor \frac{m+k-1}{k} \right\rfloor = \left\lfloor \frac{m-1}{k} \right\rfloor + 1 < (m-1) + 1 = m. \tag{23}$$

Hence, existence of $\gamma_{a,b}(m)$ follows from (19) and the inductive assumption. Finally, to obtain (16) it remains to pass to the limit $x \rightarrow \infty$ in (18) (or use Lemma 3.2 combined with the remark after Theorem 2.5). \square

3.3. Regularity

We are now ready to show that the logarithmic correlation sequences coming from k -automatic sequences are k -regular. In fact, bearing in mind applications in Section 4 we record a slightly more precise statement. Recall that for a sequence a , the sequence Sa is given by $Sa(n) = a(n+1)$. Similar ideas can be seen in [AS03b, Thm. 6].

Proposition 3.4. *Let \mathcal{N} be a finite set of sequences $\mathbb{N}_0 \rightarrow \mathbb{C}$, closed under the operators Λ_i for all $i \in \Sigma_k$. Let $\mathcal{M} = \left\{ S^e \gamma_{a,b}^{\log} \mid a, b \in \mathcal{N}, e \in \{0, 1\} \right\}$. Then $\text{span}_{\mathbb{Q}} \mathcal{M}$ is closed under the operators Λ_i for all $i \in \Sigma_k$.*

Proof. Pick any $g = S^e \gamma_{a,b}^{\log} \in \mathcal{M}$ ($a, b \in \mathcal{N}, e \in \{0, 1\}$) and $j \in \Sigma_k$. It follows from Lemma 3.2 that

$$\Lambda_j g(n) = \gamma_{a,b}^{\log}(kn + j + e) = \frac{1}{k} \sum_{i=0}^{k-1} \gamma_{a'_i, b'_i}^{\log}(n + e'_i) \tag{24}$$

where for each $i \in \Sigma_k$, $a'_i, b'_i \in \mathcal{N}$ and

$$e'_i = \left\lfloor \frac{j+i+e}{k} \right\rfloor \leq \left\lfloor \frac{(k-1) + (k-1) + 1}{k} \right\rfloor = 1.$$

It remains to note that each of the functions of n appearing under the sum on the right hand side of (24) belongs to \mathcal{M} . \square

Theorem 3.5. *If $a: \mathbb{N}_0 \rightarrow \mathbb{C}$ is k -automatic then the sequence γ_a^{\log} is k -regular and $\dim \text{span}_{\mathbb{Q}} \mathcal{N}_k(\gamma_{a,a}^{\log}) \leq 2 |\mathcal{N}_k(a)|^2$.*

4. Verifying noncorrelation

We now discuss the practical details of how one can check if a given pattern sequence is noncorrelated. We begin by setting up the notation and adapting the general results from previous sections to the situation at hand; this is done in subsections 4.1 and 4.2. Then, in subsections 4.3 and 4.4 we discuss how the relevant computations can be performed. Finally, in subsection 4.5 we discuss the complexity of the resulting algorithm, which finishes the proof of Theorem B. Implementation of this algorithm allows us to verify Theorem A by direct computation.

4.1. Setup

Throughout this section, $A \subseteq \Sigma_k^\ell$ denotes an admissible set and $a: \mathbb{N}_0 \rightarrow \{+1, -1\}$ denotes the corresponding pattern sequence:

$$a(n) = a_A(n) = (-1)^{\#(A,n)} \quad (n \in \mathbb{N}_0).$$

We also introduce the sequence $f: \mathbb{N}_0 \rightarrow \mathbb{R}$ given by

$$f := \mathbf{1}_{\mathbb{N}} \cdot \gamma_a.$$

Our task amounts to verifying that f is well-defined (i.e., that the limits defining $\gamma_a(m)$ exist for all $m \in \mathbb{N}$) and determining whether it is identically zero. The existence question is easily accounted for (cf. [ZPK18, Section 3]).

Lemma 4.1. *For each $b, c \in \mathcal{N}_k(a)$ and $m \in \mathbb{N}_0$, the coefficient $\gamma_{b,c}(m)$ exists.*

Proof. By Lemma 2.6, all sequences in $\mathcal{N}_k(a)$ are products of a and $k^{\ell-1}$ -periodic sequences. Hence, there is a $k^{\ell-1}$ -periodic sequence h such that $b(n)c(n) = a(n)^2 h(n) = h(n)$ for all $n \in \mathbb{N}_0$, and consequently

$$\gamma_{b,c}(0) = \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{n=0}^{N-1} h(n) = \frac{1}{k^{\ell-1}} \sum_{n=0}^{k^{\ell-1}-1} h(n)$$

exists. Existence of $\gamma_{b,c}(m)$ for $m \in \mathbb{N}$ now follows from Lemma 3.3. \square

Recall that $f = \mathbf{1}_{\mathbb{N}} \cdot \gamma_a^{\log}$ is k -regular by Theorem 3.5. In principle, in order to decide if f is identically zero, it is now enough to follow the arguments in Section 3 to describe the structure of the k -kernel of f and then apply Lemma 2.7. In practice, we essentially follow this route, but we also take advantage of the fact that f is a k -regular sequence of a rather specific form.

4.2. Recursive relations

As a first step towards describing the recursive relations that define f , we introduce a set that spans $\mathcal{N}_k(f)$, in analogy to Proposition 3.4. It will be convenient to introduce the restricted averages

$$\gamma^{(r)}(m) := \lim_{N \rightarrow \infty} \frac{k^\ell}{\log N} \sum_{n=0}^{N-1} \frac{1}{n+1} a(n)a(n+m) \mathbf{1}_{k^\ell \mathbb{N}_0+r}(n). \tag{25}$$

Note that these averages are well-defined thanks to Theorem 2.5. Additionally, it follows from Lemma 2.6 and Lemma 4.1 that the logarithmic averages can be replaced with unweighted averages:

$$\gamma^{(r)}(m) = \lim_{N \rightarrow \infty} \frac{k^\ell}{N} \sum_{n=0}^{N-1} a(n)a(n+m) \mathbf{1}_{k^\ell \mathbb{N}_0+r}(n). \tag{26}$$

As a direct consequence of the relevant definitions, we have

$$f = \frac{\mathbf{1}_{\mathbb{N}}}{k^\ell} \cdot \sum_{r=0}^{k^\ell-1} \gamma^{(r)} = \frac{1}{k^\ell} \sum_{q=1}^{k^\ell} \sum_{r=0}^{k^\ell-1} \mathbf{1}_{k^\ell \mathbb{N}_0+q} \cdot \gamma^{(r)}. \tag{27}$$

Proposition 4.2. *Each sequence in $\mathcal{N}_k(f)$ is a linear combination of the sequences $\mathbf{1}_{k^\ell \mathbb{N}_0+q} \cdot S^{e\gamma^{(r)}}$, where $e \in \{0, 1\}$, $0 \leq r < k^\ell$ and $0 \leq q \leq k^\ell$. In particular, $\dim \text{span}_{\mathbb{Q}} \mathcal{N}_k(f) \leq 2k^\ell(k^\ell + 1)$.*

The proof of the above proposition will follow directly once we describe the behaviour of the base sequences $\mathbf{1}_{k^\ell \mathbb{N}_0+q} \cdot S^{e\gamma^{(r)}}$ under the operators Λ_i ($i \in \Sigma_k$). To simplify this description, it will be convenient to introduce the auxiliary sequence $h: \mathbb{N}_0 \rightarrow \{+1, -1\}$, given by

$$h(n) := a(n)/a(\lfloor n/k \rfloor). \tag{28}$$

The following basic fact is analogous to [ZPK18, Lemma 2.1].

Lemma 4.3. *The sequence h given by (28) is k^ℓ -periodic.*

Proof. Follows immediately from Lemma 2.6. \square

Lemma 4.4. *Let $e \in \{0, 1\}$, $i \in \Sigma_k$, $0 \leq q \leq k^\ell$ and $0 \leq r < k^\ell$. If $i \not\equiv q \pmod k$ then $\Lambda_i \mathbf{1}_{k^\ell \mathbb{N}_0 + q} = 0$. If $i \equiv q \pmod k$ then*

$$\Lambda_i \left(\mathbf{1}_{k^\ell \mathbb{N}_0 + q} \cdot S^e \gamma^{(r)} \right) = \frac{h(r)h(r+q+e)}{k} \sum_{q'} \sum_{r'} \mathbf{1}_{k^\ell \mathbb{N}_0 + q'} \cdot S^{e'} \gamma^{(r')}, \tag{29}$$

where the value of e and the ranges of the summations are given by

$$e' := \left\lfloor \frac{i + e + (r \bmod k)}{k} \right\rfloor, \quad q' \in k^{\ell-1} \Sigma_k + \lfloor q/k \rfloor, \quad r' \in k^{\ell-1} \Sigma_k + \lfloor r/k \rfloor.$$

Proof. The case $e = 0$ follows by a standard adaptation of the proof of Lemma 3.2. Then, the case $e = 1$ is derived using Lemma 2.4. \square

4.3. Small shifts

Bearing in mind that we hope to apply Lemma 2.7, we need to be able to compute the values $S^e \gamma^{(r)}(0) = \gamma^{(r)}(e)$ for $e \in \{0, 1\}$ and $0 \leq r \leq k^\ell$. This can, in principle, be accomplished by straightforward adaptations of the arguments in Lemma 3.3 and Lemma 4.1. Here, we discuss the practical details of how the computations are performed. Recall that $\gamma^{(r)}(0) = 1$, so we only need to compute $\gamma^{(r)}(1)$.

For $0 \leq r < k^\ell$, let $\nu = \nu(r)$ denote the first position where a digit distinct from $k - 1$ appears in the base- k expansion $(r)_k$; if $r = k^\alpha - 1$ for some $\alpha \geq 0$ then $\nu = \alpha$. We consider r in nondecreasing order with respect to $\nu(r)$. We have three ranges to consider: $r = 0$, $1 \leq r < \ell$ and $r = \ell$.

If $\nu(r) = 0$ then it follows from Lemma 4.4 that

$$\gamma^{(r)}(1) = \frac{h(r)h(r+1)}{k} \sum_{r'} \gamma^{(r')}(0) = h(r)h(r+1); \tag{30}$$

here and elsewhere, the summation over r' runs through $r' \in k^{\ell-1} \Sigma_k + \lfloor r/k \rfloor$. Since we can readily compute $h(r)$ and $h(r+1)$, we can compute $\gamma^{(r)}(1)$.

If $1 \leq \nu(r) < \ell$ then another application of Lemma 4.4 yields

$$\gamma^{(r)}(1) = \frac{h(r)h(r+1)}{k} \sum_{r'} \gamma^{(r')}(1). \tag{31}$$

For all r' appearing in the above sum we have $\nu(r') = \nu(r) - 1$, and hence $\gamma^{(r')}(1)$ has been previously computed. Hence, again, we can directly compute $\gamma^{(r)}(1)$.

Finally, if $\nu = \ell$ (meaning that $r = k^\ell - 1$) then (31) continues to hold, and we have $\nu(r') = \ell - 1$ for all summands on the right-hand-side except for the one corresponding to $r' = r$. Hence, we can compute $\gamma^{(r)}$ as

$$\gamma_{k^\ell-1}(1) = \frac{1}{kh(r)h(r+1)-1} \sum_{i=0}^{k-2} \gamma_{k^\ell-2(ki+1)-1}(1). \tag{32}$$

4.4. *Basis construction*

Recall that our general strategy calls for a construction of a spanning set of $\text{span}_{\mathbb{Q}} \mathcal{N}_k(f)$. For technical reasons, it appears to be slightly more convenient and efficient to instead work with the potentially larger space

$$\mathcal{M} := \text{span}_{\mathbb{Q}} \{ \mathbf{1}_{k^\ell \mathbb{N}_0+q} \cdot g \mid g \in \mathcal{N}_k(f), 0 \leq q \leq k^\ell \}.$$

It remains true that $f = 0$ if and only if $\mathcal{M} = \{0\}$, and that \mathcal{M} is closed under Λ_i for all $i \in \Sigma_k$. Additionally, \mathcal{M} admits a decomposition

$$\mathcal{M} = \bigoplus_{q=0}^{k^\ell} \mathcal{M}_q, \quad \mathcal{M}_q := \eta_q \cdot \text{span}_{\mathbb{Q}} \mathcal{N}_k(f),$$

where the sequences η_q are given by

$$\eta_0 = \mathbf{1}_{\{0\}}, \quad \eta_q = \mathbf{1}_{k^\ell \mathbb{N}_0+q} \text{ for } 1 \leq q \leq k^\ell.$$

By Lemma 2.7, to show that $\mathcal{M} = \{0\}$ it suffices to verify that $g(0) = 0$ for each $g \in \mathcal{M}$, which is trivially satisfied for $g \in \mathcal{M}_q$ for all $1 \leq q \leq k^\ell$.

We proceed to construct a list of sequences $f_1, f_2, \dots \in \mathcal{M}$ which spans \mathcal{M} . Additionally, we ensure that for each $t \geq 1$, the sequence f_t belongs to \mathcal{M}_{q_t} for some $0 \leq q_t \leq k^\ell$ and we keep track the value of q_t . By Proposition 4.2, each f_t has a decomposition

$$f_t = \sum_{r=0}^{k^\ell-1} \sum_{e=0}^1 w_{r,e}^{(t)} \eta_t S^e \gamma^{(r)}, \tag{33}$$

for some coefficients $w_{r,e}^{(t)}$, which we also keep track of. While we cannot ensure that f_1, f_2, \dots are linearly independent (in fact, we are primarily interested in the case when $f_1 = f_2 = \dots = 0$), we will ensure that for each $1 \leq q \leq k^\ell$, the (multi-)set of coefficient vectors $\{w^{(t)} \mid q_t = q\} \subseteq \mathbb{R}^{2k^\ell}$ is linearly independent.

We start by setting for $1 \leq t \leq k^\ell$,

$$f_t = \mathbf{1}_{k^\ell \mathbb{N}_0+t} \cdot \sum_{r=0}^{k^\ell} \gamma^{(r)}, \quad q_t = t, \tag{34}$$

and accordingly $w_{r,e}^{(t)} = \mathbf{1}_{\{0\}}(e)$ ($0 \leq r < k^\ell, e \in \{0, 1\}$).

Suppose next that at a certain stage we have constructed f_1, f_2, \dots, f_u and that for all $1 \leq t \leq u$ we have ensured that $\Lambda_i f_t \in \text{span}_{\mathbb{Q}} \{f_1, f_2, \dots, f_u\}$ for all $i \in \Sigma_k$.

(Initially, $v = k^\ell$ and $u = 0$.) If $u = v$ then $\text{span}_{\mathbb{Q}}\{f_1, f_2, \dots, f_v\}$ is a subset of \mathcal{M} that is closed under Λ_i ($i \in \Sigma_k$) and under multiplication by $\mathbf{1}_{k^\ell\mathbb{N}_0+q}$ ($0 \leq q \leq k^\ell$), hence $\text{span}_{\mathbb{Q}}\{f_1, f_2, \dots, f_v\} = \mathcal{M}$ and the construction is complete.

Let us next consider the case when $u < v$. Put $q = q_{u+1}$, $g = f_{u+1}$ and $w = w^{(u+1)}$. Recall that the only value of i for which $\Lambda_i g$ could be non-zero is $i = q \pmod k$. If $q = 0$ then $g = g(0)\mathbf{1}_{\{0\}}$. Hence, either $g(0) \neq 0$, in which case a is not noncorrelated and we are done; or $g(0) = 0$, in which case $g = 0$ and so $\Lambda_i g = 0$ as well. Suppose now that $1 \leq q \leq k^\ell$. Applying Lemma 4.4, we obtain a representation of $\Lambda_i g$ in the form

$$\Lambda_i g = \sum_{q'} \sum_{r'} \sum_{e'} w'_{q',r',e'} \mathbf{1}_{k^\ell\mathbb{N}_0+q'} \cdot S^{e'} \gamma^{(r')}, \tag{35}$$

where the ranges of summation are given by $0 \leq q' \leq k^\ell$, $0 \leq r' < k^\ell$ and $0 \leq e' \leq 1$, and the coefficients w' are given by explicit formulae coming from (29). Bearing in mind that $\mathbf{1}_{k^\ell\mathbb{N}_0} = \mathbf{1}_{k^\ell\mathbb{N}} + \mathbf{1}_{\{0\}}$, we find the decomposition

$$\Lambda_i g = \sum_{q'} g'_{q'}, \quad g'_{q'} = \sum_{r'} \sum_{e'} w''_{q',r',e'} \eta_{q'} S^{e'} \gamma^{(r')}, \tag{36}$$

where the coefficients $w''_{q',r',e'}$ are given by:

$$w''_{q',r',e'} = w'_{q',r',e'} \text{ if } q' \neq k^\ell, \quad w''_{k^\ell,r',e'} = w'_{k^\ell,r',e'} + w'_{0,r',e'}. \tag{37}$$

For each q' , we append $g'_{q'}$ to the list f_1, f_2, \dots, f_v if (and only if)

$$(w''_{q',r',e'})_{r',e'} \notin \text{span}_{\mathbb{Q}} \left\{ \left(w_{r,e}^{(t)} \right)_{r,e} \mid 1 \leq t \leq v, q_t = q' \right\}. \tag{38}$$

If (38) holds then we also record $g'_{q'} \in \mathcal{M}_{q'}$ (that is, we append q' to the list q_1, q_2, \dots, q_v) and that the decomposition of $g'_{q'}$ as the sum of basis sequences is given by (36) (what is, we append $w''_{q'}$ to the list $w^{(1)}, w^{(2)}, \dots, w^{(v)}$). Each time a new sequence is added, v increases by 1 and after all q' have been processed, u increases by 1.

The linear independence condition (38) ensures that for each $1 \leq q \leq k^\ell$, there are at most $2k^\ell$ values of t with $q_t = q$, and hence the construction needs to terminate after a bounded number of steps. As the result, we either find, for some $t \geq 1$, a sequence $f_t \in \mathcal{M}$ with $f_t(0) \neq 0$ (in which case a is not noncorrelated) or we construct a finite list of sequences $f_1, f_2, \dots, f_N \in \mathcal{M}$ that spans \mathcal{M} and satisfies $f_t(0) = 0$ for all $1 \leq t \leq N$ (in which case a is noncorrelated). In either case, we are able to determine whether a is noncorrelated.

4.5. Complexity

We now provide quantitative estimates for the amount of computational power needed to verify if the pattern sequence a is noncorrelated using the method described above.

Throughout, we treat k as fixed, and hence are interested in the regime $\ell \rightarrow \infty$. It will be convenient to introduce, for a function $F: \mathbb{N} \rightarrow \mathbb{R}_{>0}$, the shorthand $\tilde{O}(F(\ell))$ to denote $O(\ell^{O(1)}F(\ell))$. Thus, for instance, addition or multiplication of two integers of size $O(k^\ell)$ can be performed using $\tilde{O}(1)$ operations.

At several points, we need to compute the values of $a(n)$ where $n = O(k^\ell)$. For a word $w \in \Sigma_k^*$ with length $|w| \leq \ell$, computing $\#(n, w)$ directly from the definition requires $\tilde{O}(1)$ operations. Since $|A| \leq k^\ell$, the values $\#(n, A)$ and $a(n)$ can be computed in time $\tilde{O}(k^\ell)$. Consequently, we can also compute $h(n)$ in time $\tilde{O}(k^\ell)$.

Following the steps in subsection 4.3, we compute $\gamma^{(r)}(1)$ for all $0 \leq r < k^\ell$. It takes $\tilde{O}(k^\ell)$ operations to write the values of r ($0 \leq r < k^\ell$) in an order consistent with $\nu(r)$. Note that each of the formulae (30), (31), (32) produces the corresponding value of $\gamma^{(r)}(1)$ using $\tilde{O}(1)$ arithmetic operations on rational numbers. One can also check by a simple inductive argument that all denominators and numerators that appear in these computations are bounded by $O(k^\ell)$, and hence each arithmetic operation takes only $\tilde{O}(1)$ basic operations. We also note that all the denominators take the form $(k \pm 1)k^\alpha$.

We next proceed to the computation of the sequences f_t ($t = 1, 2, 3, \dots$) in subsection 4.4. Strictly speaking, we compute the sequence $w^{(t)}$, which uniquely determine f_t via (33), and the auxiliary sequence q_t . For $t \leq k^\ell$, the explicit formula (34) allows us to compute $w^{(t)}$ and q_t with $\tilde{O}(k^{2\ell})$ operations (note that $w^{(t)} = (w_{r,e}^{(t)})_{r,e}$ has $k^{2\ell}$ entries, so this is the least number of operations possible).

Let us now consider the amount of computation required to compute f_t for $t > k^\ell$. Consider any u, v , as in the iterative procedure in second half of subsection 4.4. We note that the application of Lemma 4.4 used to compute w' in (35) requires no more than $\tilde{O}(k^{3\ell})$ arithmetic operations (for each of $O(k^\ell)$ summands in the decomposition of g , we substitute a sum of size $O(k^{2\ell})$). Once w' is computed, it only takes $\tilde{O}(k^{2\ell})$ operations to compute w'' . Then, for each of $O(k^\ell)$ values of q' , in order to verify if $g'_{q'}$ should be appended to the list f_1, f_2, \dots , we need to verify if the corresponding vector of coefficients belongs to a certain linear subspace of \mathbb{R}^{2k^ℓ} , see (38). Keeping track of how much the complexity increases in each step of the construction, we see that for each $t > k^\ell$, the entries of $w^{(t)}$ are rational numbers whose numerators are $\tilde{O}(k^{3t})$, and whose denominators are $O(k^t)$ and divide $(k^2 - 1)k^\alpha$ for some integer α . Thus, in (38) we may scale all of the relevant vectors by a factor of $(k^2 - 1)k^{O(u)}$, leaving us with the task of verifying if an integer-valued vector belongs to the span of other integer-valued vectors. The latter task is well-known to have polynomial complexity (with respect to dimensions and lengths of representations of entries), see e.g. [BCS97, Chpt. 16]. Hence, for each q' in order to decide if $g'_{q'}$ should be appended, we perform $\tilde{O}(k^{O(\ell)}) = k^{O(\ell)}$ operations. Consequently, the number of operations needed to process the step corresponding to the index u is $k^{O(\ell)}$.

Because of the linear independence conditions discussed at the end of subsection 4.4, the total number of the sequences f_1, f_2, \dots we construct is at most $2k^{2\ell+1}$. It follows that in total, we perform at most $k^{O(\ell)}$ operations.

5. Dilation-invariant sequences

We now turn to the classification of dilation-invariant pattern sequences. Throughout, let $A \subseteq \Sigma_k^*$ be a set of patterns with no leading or trailing zeros, and let $a = a_A$ be the corresponding pattern sequence. We also retain the notation from Section 4, specifically the coefficients γ_r defined in (25). We let $\ell = \max_{v \in A} |v|$ denote the length of a , and we assume that $\ell \geq 2$.

The following condition turns out to be closely connected to the question of whether a is noncorrelated:

$$|A(ui_0)^{-1} \oplus A(ui_1)^{-1}| = \frac{k}{2} \text{ for each } u \in \Sigma_k^{\ell-2} \text{ and } i_0, i_1 \in \Sigma_k \text{ with } i_0 \neq i_1. \quad (\dagger)$$

Above, using the standard notation from semigroup theory, for a word $u \in \Sigma_k^*$ and a set $X \subseteq \Sigma_k^*$, we let $Xu^{-1} := \{v \in \Sigma_k^* \mid vu \in X\}$.

Remark 5.1. The condition (\dagger) can be stated in simpler terms when $k = 2$. Then, necessarily, $\{i_0, i_1\} = \{0, 1\}$ and since A has no trailing zeros, $A0^{-1} = \emptyset$. Hence, (\dagger) says that $|A(u1)^{-1}| = 1$ for all $u \in \Sigma_k^{\ell-2}$. Because all patterns in A have length $\leq \ell$, $A(u1)^{-1} \subseteq \{0, 1\}$; and because A has no leading zeros, $0 \notin A(u1)^{-1}$. Thus, (\dagger) reduces to the statement that $1u1 \in A$ for all $u \in \Sigma_k^{\ell-2}$, that is, $1\Sigma_2^{\ell-2}1 \subseteq A$. This is precisely the assumption that appears in Theorem C.

Remark 5.2. For general $k \geq 2$, it is not a priori clear if there exists a set of patterns A such that (\dagger) holds. Fix $u \in \Sigma_k^{\ell-2}$ and consider the matrix $M = (M_{i,j}^{(u)})_{i,j=0}^{k-1}$ where $M_{i,j}^{(u)} = -1$ if $iu j \in A$ and $M_{i,j}^{(u)} = +1$ otherwise. Then (\dagger) says that $M^T M = kI$, where I denotes the identity matrix, meaning that M is a Hadamard matrix. Additionally, $M_{i,j}^{(u)} = +1$ if $i = 0$ or $j = 0$, meaning that M is normalized. Conversely, given any normalized Hadamard matrix M' , one can easily reconstruct A so that $M = M'$ for each choice of $u \in \Sigma_k^{\ell-2}$. Thus, it is possible to satisfy the condition (\dagger) if and only if there is at least one Hadamard matrix of dimension k .

The question of existence of Hadamard matrices of a given dimension has long been investigated. They are easily constructed when k is a power of 2 through a tensor-power construction. More generally, given Hadamard matrices of dimensions k and k' one can construct a Hadamard matrix of dimension $k \cdot k'$. It is conjectured that Hadamard matrices exist for $k = 1, 2$ and all k divisible by 4. So far, this has been confirmed for $k < 668$. See e.g. [CD07, Chpt. V] for further discussion.

The main goal of this section is to prove a slightly more general variant of Theorem C. The second part of this theorem asserts that if a is noncorrelated, $k = 2$ and $\ell \leq 5$ then (\dagger) holds. This is verified by exhaustive search,² using the methods developed in Section 4.

² Code available from the author.

The remaining part of Theorem C follows from the following result, whose proof will occupy the remainder of this section.

Proposition 5.3. *Suppose that (†) holds. Then the sequence a is noncorrelated.*

From this point onwards, assume that (†) holds. Proceeding along similar lines as in Lemma 3.3 (or Section 4.3), we will compute $\gamma_r(m)$ for small values of $m \in \mathbb{N}_0$ ($0 \leq r < k^\ell$). The following lemma is the main consequence of (†) that we use.

Lemma 5.4. *Let $u \in \Sigma_k^{\ell-2}$ and $j_0, j_1 \in \Sigma_k, j_0 \neq j_1$. Then*

$$\sum_{i=0}^{k-1} a([iuj_0]_k) a([iuj_1]_k) = 0. \tag{39}$$

Proof. Multiplying by $a([uj_0]_k) a([uj_1]_k)$, we see that (39) is equivalent to

$$\sum_{i=0}^{k-1} a([uj_0]_k) a([uj_1]_k) a([iuj_0]_k) a([iuj_1]_k) = 0. \tag{40}$$

Each pattern v in A of length $< \ell$ and each $i \in \Sigma_k$, considering the different positions where v can appear, one can check that

$$\#(v, uj_0) + \#(v, iuj_1) = \#(v, uj_1) + \#(v, iuj_0).$$

Conversely, if $v \in A$ and $|v| = \ell$ then

$$\#(v, uj_0) = \#(v, uj_1) = 0,$$

since $|uj_0|, |uj_1| < \ell$, and for each $i \in \Sigma_k$

$$\#(v, iuj_0) + \#(v, iuj_1) = \begin{cases} 1 & \text{if } v \in \{iuj_0, iuj_1\}, \\ 0 & \text{otherwise.} \end{cases}$$

Substituting the above identities into the sum on the left-hand side of (39) and applying (†) we conclude that

$$\begin{aligned} \sum_{i=0}^{k-1} a([iuj_0]_k) a([iuj_1]_k) &= \sum_{i=0}^{k-1} (-1)^{\#\{iuj_0, iuj_1\} \cap A} \\ &= k - 2 |A(uj_0)^{-1} \oplus A(uj_1)^{-1}| = 0. \quad \square \end{aligned}$$

Lemma 5.5. *Let $0 \leq r < k^\ell$ and $m \geq 0$. Put $j = r \bmod k$. Then*

$$\gamma_r(m) = \begin{cases} a(r)a(r+m) & \text{if } j+m < k \text{ and } m \neq 0, \\ 0 & \text{otherwise.} \end{cases}$$

Proof. Let us write $m = km' + i$ with $m' \geq 0$ and $i \in \Sigma_k$. Then by Lemma 4.4 (or, equivalently, by Lemma 3.2) we have

$$\gamma_r(m) = \gamma_r(km' + i) = \frac{h(r)h(r + km' + i)}{k} \sum_{r'} \gamma_{r'}(m' + e'), \tag{41}$$

where as usual $r' \in k^{\ell-1}\Sigma_k + \lfloor r/k \rfloor$ and $e' = \lfloor (i + j)/k \rfloor \in \{0, 1\}$. We consider several different cases.

Case 0: $m = 0$. It follows directly from the definition of γ_r that

$$\gamma_r(0) = 1 = a(r)^2 = a(r)a(r+m).$$

Case 1: $m \neq 0$ and $j + m < k$. Applying (41) and noticing that $i = m$, $m' = 0$, and $e' = 0$, we obtain

$$\gamma_r(m) = h(r)h(r+m) = a(r)a(r+m), \tag{42}$$

where the second equality holds because $\lfloor r/k \rfloor = \lfloor (r+m)/k \rfloor$.

In all of the remaining cases, we will show that $\gamma_r(m) = 0$. We start with the simplest situation where $e' = 1$.

Case 2: $m = 1$ and $j + m \geq k$, meaning that $j = k - 1$. Let $\nu(r)$ denote the first position where a digit distinct from $k - 1$ appears in the expansion of r , allowing $\nu(r) = \alpha$ if $r = k^\alpha - 1$. By (41),

$$\gamma_r(1) = \pm \frac{1}{k} \sum_{r'} \gamma_{r'}(1). \tag{43}$$

If $\nu(r) = 1$ then from the previously considered cases and Lemma 5.4 it follows that

$$\gamma_r(1) = \pm \frac{1}{k} \sum_{r'} a(r')a(r'+1) = 0.$$

If $1 < \nu(r) < \ell$ then $\nu(r') = \nu(r) - 1$ for all r' that enter the sum (43). Hence, reasoning by induction on $\nu(r)$ we conclude that $\gamma_r(1) = 0$. Finally, if $\nu(r) = \ell$ then $r = k^\ell - 1$, and $\nu(r') = \ell - 1$ for all r' that appear in the sum (43) except for $r' = r$. It follows that

$$\gamma_{k^\ell-1}(1) = \pm \frac{1}{k} \gamma_{k^\ell-1}(1),$$

which is only possible if $\gamma_{k^\ell-1}(1) = 0$.

Case 3: $2 \leq m < k$ and $j + m \geq k$. By (41) and Case 2,

$$\gamma_r(m) = \pm \frac{1}{k} \sum_{r'} \gamma_{r'}(1) = 0. \tag{44}$$

Case 4: $k \leq m < k^2$. By (41),

$$\gamma_r(m) = \pm \frac{1}{k} \sum_{r'} \gamma_{r'}(m' + e') = 0. \tag{45}$$

Let $j' := \lfloor r/k \rfloor \bmod r$ and $i' := m' + e'$. Note that $r' \bmod k = j'$ for all r' in the sum in (45), where we are using the fact that $\ell \geq 2$. We have several subcases to consider. If $j' + i' < k$ then

$$\gamma_r(m) = \pm \frac{1}{k} \sum_{r'} a(r')a(r' + i') = 0$$

by Cases 0 and 1 and Lemma 5.4. If $j' + i' \geq k$ while $i' < k$ (i.e. $m' \neq k - 1$ or $e' \neq 1$) then $\gamma_{r'}(m' + e') = 0$ for all r' by Cases 2 and 3, and consequently also $\gamma_r(m) = 0$. Finally, if $j' + i' = k$ (i.e. $m' = k - 1$ and $e' = 1$) then

$$\gamma_r(m) = \pm \frac{1}{k} \sum_{r'} \gamma_{r'}(k \cdot 1 + 0) = 0$$

by the previously considered subcases.

Case 5: $m \geq k^2$. We reason by induction on m . By (41) and the inductive assumption,

$$\gamma_r(m) = \pm \frac{1}{k} \sum_{r'} \gamma_{r'}(m' + e') = 0 \tag{46}$$

since $k \leq m' + e' < m$. \square

Now that we have computed the values of the coefficients $\gamma_r(m)$, the remainder of the argument is straightforward.

Proof of Proposition 5.3. We need to show that

$$\gamma(m) = \frac{1}{k^\ell} \sum_{r=0}^{k^\ell-1} \gamma_r(m) = 0$$

for all $m \geq 1$. If $m \geq k$ there is nothing to prove since $\gamma_r(m) = 0$. Suppose now that $1 \leq m < k$. We may write arbitrary $0 \leq r < k^\ell - 1$ in the form $r = k^{\ell-1}i + ks + j$ where $i, j \in \Sigma_k$ and $0 \leq s < k^{\ell-2}$. Then, $\gamma_r(m) = 0$ if $j + m \geq k$ and $\gamma_r(m) = a(r)a(r + m)$ otherwise. It follows that

$$\gamma(m) = \sum_{j=0}^{k-m-1} \sum_{s=0}^{k^{\ell-1}-1} \sum_{i=0}^{k-1} a(k^{\ell-1}i + ks + j) a(k^{\ell-1}i + ks + j + m) = 0,$$

where the inner-most sum vanishes by Lemma 5.4. \square

Remark 5.6. Let $a' : \mathbb{N}_0 \rightarrow \{+1, -1\}$ be a sequence such that a'/a is $k^{\ell-1}$ -periodic. Then a' is pattern by Lemma 2.2. Defining γ' and γ'_r in analogy to γ and γ_r , with a' in place of a , by a direct computation we show for all $m \geq 0$ and $0 \leq r < k^\ell$ that

$$\gamma'_r(m) = \frac{a'(r)a'(r+m)}{a(r)a(r+m)}\gamma_r(m) = \pm\gamma_r(m). \tag{47}$$

It follows that $\gamma'_r(m) = 0$ for all $m \geq k$. In particular, $\gamma'(m) = 0$ for all $m \geq k$.

We check by exhaustive search that all noncorrelated binary pattern sequences of length ≤ 4 can arise as a' in the construction outlined above. It seems plausible that the same holds for all lengths. If this is the case, and if Conjecture 1.1 holds true, then the task of verifying if a given binary pattern sequence b' is noncorrelated can be split into two independent steps: First, check if the dilation-invariant sequence b obtained from b' in Lemma 2.9 satisfies (†); if not then b' is not noncorrelated.³ Second, check if the \pm signs in (the analogue of) (47) align in a way that ensures $\gamma_{b'}(1) = 0$. While the condition from the first step is quite conceptual, it appears that the second step relies mostly on arithmetic coincidence. This would provide an intuitive explanation for why the results in the dilation-invariant case are considerably more concise.

Appendix A. Supplementary material

Supplementary material related to this article can be found online at <https://doi.org/10.1016/j.jnt.2020.10.008>.

References

[AL91] J.-P. Allouche, P. Liardet, Generalized Rudin-Shapiro sequences, *Acta Arith.* 60 (1) (1991) 1–27.
 [AS92] J.-P. Allouche, J. Shallit, The ring of k -regular sequences, *Theor. Comput. Sci.* 98 (2) (1992) 163–197.
 [AS99] J.-P. Allouche, J. Shallit, The ubiquitous Prouhet-Thue-Morse sequence, in: *Sequences and Their Applications*, Singapore, 1998, in: Springer Ser. Discrete Math. Theor. Comput. Sci., Springer, London, 1999, pp. 1–16.
 [AS03a] J.-P. Allouche, J. Shallit, Automatic sequences, in: *Theory, Applications, Generalizations*, Cambridge University Press, Cambridge, 2003.
 [AS03b] J.-P. Allouche, J. Shallit, The ring of k -regular sequences. II, *Theor. Comput. Sci.* 307 (1) (2003) 3–29, Words.
 [BCM89] D.W. Boyd, J. Cook, P. Morton, On sequences of ± 1 's defined by binary patterns, *Diss. Math. (Rozprawy Mat.)* 283 (1989) 64.

³ For the sake of simplicity, we work under the additional assumption that b and b' have equal lengths, which is not true in general.

- [BCS97] P. Bürgisser, M. Clausen, M.A. Shokrollahi, Algebraic Complexity Theory, Grundlehren der Mathematischen Wissenschaften (Fundamental Principles of Mathematical Sciences), vol. 315, Springer-Verlag, Berlin, 1997, With the collaboration of Thomas Lickteig.
- [CD07] C.J. Colbourn, J.H. Dinitz (Eds.), Handbook of Combinatorial Designs, Boca Raton, second edition, Discrete Mathematics and Its Applications, Chapman & Hall/CRC, Boca Raton, FL, 2007.
- [CKMF77] J. Coquet, T. Kamae, M. Mendès France, Sur la mesure spectrale de certaines suites arithmétiques, Bull. Soc. Math. Fr. 105 (4) (1977) 369–384.
- [Coq76] J. Coquet, Sur les fonctions q -multiplicatives pseudo-aléatoires, C. R. Acad. Sci. Paris Sér. A-B 282 (4) (1976) A175–A178.
- [DMR19] M. Drmota, C. Mauduit, J. Rivat, Normality along squares, J. Eur. Math. Soc. 21 (2) (2019) 507–548.
- [FM96] E. Fouvry, C. Mauduit, Méthodes de crible et fonctions sommes des chiffres, Acta Arith. 77 (4) (1996) 339–351.
- [Gel68] A.O. Gel'fond, Sur les nombres qui ont des propriétés additives et multiplicatives données, Acta Arith. 13 (1967/1968) 259–265.
- [Kon19] J. Konieczny, Gowers norms for the Thue–Morse and Rudin–Shapiro sequences, Ann. Inst. Fourier (Grenoble) 69 (4) (2019) 1897–1913.
- [MM89] P. Morton, W.J. Maurant, Paper folding, digit patterns and groups of arithmetic fractals, Proc. Lond. Math. Soc. (3) 59 (2) (1989) 253–293.
- [Mor90] P. Morton, Connections between binary patterns and paperfolding, J. Théor. Nombres Bordeaux 2 (1) (1990) 1–12.
- [MR10] C. Mauduit, J. Rivat, Sur un problème de Gelfond: la somme des chiffres des nombres premiers, Ann. Math. (2) 171 (3) (2010) 1591–1646.
- [MR15] C. Mauduit, J. Rivat, Prime numbers along Rudin–Shapiro sequences, J. Eur. Math. Soc. 17 (10) (2015) 2595–2642.
- [MS98] C. Mauduit, A. Sárközy, On finite pseudorandom binary sequences. II. The Champernowne, Rudin–Shapiro, and Thue–Morse sequences, a further construction, J. Number Theory 73 (2) (1998) 256–276.
- [Spi18] L. Spiegelhofer, The level of distribution of the Thue–Morse sequence, arXiv e-prints, arXiv:1803.01689, Mar 2018.
- [ZPK18] Y. Zheng, L. Peng, T. Kamae, Characterization of noncorrelated pattern sequences and correlation dimensions, Discrete Contin. Dyn. Syst. 38 (10) (2018) 5085–5103.