

Sequences with long range exclusions

Kari Eloranta*

Institute of Mathematics
Aalto university
Finland

kari.eloranta@aalto.fi

Abstract

Given an alphabet S , we consider the size of the subsets of the full sequence space $S^{\mathbf{Z}}$ determined by the additional restriction that $x_i \neq x_{i+f(n)}$, $i \in \mathbf{Z}$, $n \in \mathbf{N}$. Here f is a positive, strictly increasing function. We review an other, graph theoretic, formulation and then the known results covering various combinations of f and the alphabet size. In the second part of the paper we turn to the fine structure of the allowed sequences in the particular case where f is a suitable polynomial. The generation of sequences leads naturally to consider the problem of their maximal length, which turns out highly random asymptotically in the alphabet size.

Keywords: symbolic dynamics, chromatic number, formal language, long range order, difference sets, Poincaré recurrence, Pascal's simplex, compound geometric distribution.

2010 AMS Classification: 37B20, 05C15, 11B05, 60F05

Running head: Sequences with long range exclusions

* Research partially supported by the Magnus Ehrnrooth foundation

0. Introduction

Given an **alphabet** $S = \{1, 2, 3, \dots, d\}$ on d symbols the **full shift** is the dynamical system $(S^{\mathbf{Z}}, \sigma)$. Here σ is the left shift: $\sigma(x)_i = x_{i+1}$, $\forall i \in \mathbf{Z}$. As the full shifts are rather simple objects the main study in topological dynamics usually concentrates on subshifts of finite type (SoFT), where one further restricts the sequences by listing all allowed neighboring symbols. This leads to the transfer matrix formulation which for 1-dimensional sequences has a pretty complete theory.

If the symbol interaction is not nearest neighbor but finite range, by using a larger alphabet (from blocks of S -symbols) one can still remain in the SoFT setup. For infinite range interactions this is not anymore possible and the transfer matrix formalism breaks down.

Suppose that we impose an infinite restriction, an **exclusion rule**, and define:

$$(0.1) \quad X_{(d,f)} = \{x \in S^{\mathbf{Z}} \mid x_i \neq x_{i+f(n)}, i \in \mathbf{Z}, n \in \mathbf{N}\},$$

where $f : \mathbf{N} \rightarrow \mathbf{N}$, ($\mathbf{N} = \{1, 2, 3, \dots\}$) is strictly increasing hence also unbounded. Note that the definition implies a symmetric rule (w.r. to any i). One could interpret this as some kind of exclusion principle: symbols on “shells” of radius f **block/exclude** their kind. Mike Keane proposed a model like this, with the particular choice $f(n) = n^2$ ([LM]). As a first question he asked for which S the space is nonempty. Turns out that even for the simplest choices of the **jump sequence** f answering this is fairly involved. Note that in symbolic dynamics context it would be most natural to expect a finite alphabet, $d < \infty$.

Alternatively one could view \mathbf{Z} as a **graph** where there is an edge between two distinct vertices/integers i and j iff $|i - j| = f(n)$ for some $n \in \mathbf{N}$. Coloring the graph by requiring that adjacent vertices have different colors leads to the natural question of the **chromatic number** of this (undirected) graph. This formulation seems to be due to Paul Erdős who was curious about the f for which the chromatic number is finite.

In this paper we will first present some examples and then review the known results. Through these the problem is then connected to more general dynamics and to additive number theory. It turns out that in some sense for most combinations of d and f the shift spaces are empty. This will lead to us to consider how the termination of the sequences comes about. Analysis of these generation algorithms is the main content of the latter half of the paper.

1. Infinite sequences

1.1. First observations

Together with the two-sided sequences of (0.1) we also deal with the one-sided version $X_{(d,f)}^+$ where the rule is restricted to $S^{\mathbf{N}}$ (or $S^{\mathbf{N}_0}$, whichever is more convenient). If we can show that $X_{(d,f)}^+$ is

small/empty then by $X_{(d,f)}|_{\mathbf{N}} \subset X_{(d,f)}^+$ (the restriction to positive integers) the same holds for the two-sided space. The inclusion holds since in $X_{(d,f)}^+$ the past haunts only back to the coordinate 0 whereas in $X_{(d,f)}$ the block can of course arise from a symbol arbitrarily far in the past. A second observation is even more basic but we state it nevertheless: $X_{(d,f)} \subset X_{(d',f)}$ if $d \leq d'$. This holds for one-sided sequences, too. If we can show the smallness for some alphabet size, it holds for all smaller alphabets.

Let us first consider some explicit cases to gain some insight into the contribution of the alphabet size and growth rate of the jump sequence. By **lexicographically generating** the one-sided sequences we mean that one proceeds from $x_1 = 1$ rightwards to the nearest coordinate always assigning the smallest allowed symbol.

Example 1.1. (linear f): Consider the case $S = \{1, 2\}$ and $f(n) = 2n$. Clearly if $x_0 = 1$ then $x_{2k} = 2, \forall k \neq 0$ but $x_2 = 2$ implies $x_{2m} = 1, \forall m \neq 1$, a contradiction. So $X_{(2,2n)} = \emptyset$.

On the other hand $X_{(2,2n-1)} \neq \emptyset$: it consists of the two periodic points $(12)^*$.

It is worth noticing that the emptiness is not due to too small S . Let $f(n) = kn, k \geq 2$, fixed integer. If $x_0 = 1$, then $x_{kn} = \neg 1, \forall n \geq 1$ ($\neg s$ meaning “any symbol but s ”). If then $x_k = 2$, we have $x_{lk} \notin \{1, 2\}, \forall l \geq 2$. Continuing like this considering symbol choices at coordinates that are multiples of k one exhausts any finite alphabet i.e. $X_{(d, kn)} = \emptyset$ for any finite d .

Example 1.2. ($d = 2$ and power f): Suppose $S = \{1, 2\}$ and $f(n) = n^r, r = 2, 3, \dots$. Let $x_0 = 1$ then $x_{2i} = 1, \forall i \in \mathbf{Z}$ so in particular $x_{2^r} = 1$ which generates a contradiction. Hence $X_{(2, n^r)} = \emptyset$. Here the minimal size of S is crucial: it strongly forces parity.

Example 1.3. (fast growing f): Suppose there is a natural number m which does not divide $f(n)$ for any $n = 1, 2, \dots$ (subsequently we use the notation $a|b$ meaning a divides b). Clearly for e.g. $n^r, n^n, n! \dots$ this isn't the case, but p^n has this property for prime p . For $d \geq m$ one can then have infinite periodic points. Hence for example $X_{(3, 2^n)} \neq \emptyset$ (at least 6 elements, $(123)^*$ and its permutations).

It is straightforward to see that $X_{(d, n!)}^+ = \emptyset$ at least for $d = 2, 3$. One simply sets $x_1 = 1$, lays down the blocks from it, sets $x_1 = 2$ and then the blocks that it generates etc. With $d = 2$ one is stuck in the second assignment round, with $d = 3$ in the sixth.

$X_{(4, n!)}^+$ is already considerably more interesting. With computer one can see that lexicographically generated, the sequence quickly becomes periodic with period $(1234123412312341231234123)$ (of length 25). However this period cannot persist indefinitely since starting with $x_1 = 1$ we would have $x_{7!+1} = 1$. The sequence does not terminate then though but instead there is a transient after which the original period is reinstated (in the lexicographic generation of the sequence). Since $25|n!, \forall n \geq 10$ we see that this periodicity would be interrupted infinitely often. Is there an infinite sequence in $X_{(4, n!)}^+$? In computer runs a sequence with the given period (except for short transients) will persist at least for 500.000 steps.

The existence or non-existence of a period leads naturally to a language theoretic characterization of these spaces. For a background on these we refer to [HU]. Let $|s|$ denote the length of the string/sequence s .

Theorem 1.1.: *If for any $m \in \mathbf{N}$ there is $n \in \mathbf{N}$ such that we have $m \mid f(n)$ then the words satisfying the exclusion rule of (0.1) do not form a context-free language. Hence the strings do not form a regular language (sofic shift) either.*

Proof: A context-free language necessarily satisfies a Pumping lemma: any sufficiently long string s , say $|s| \geq k$, can be written as $s = uvxyz$ such that (i) $|vxy| \leq k$, (ii) $|vy| \geq 1$ and (iii) uv^nxy^nz is an allowed string for all $n \in \mathbf{N}$. Here (ii) means that the “pumping” is indeed nontrivial action. If $\min\{|v|, |y|\} = 0$ but (ii) still holds, this more general Pumping lemma for context free languages is reduced to that for regular languages.

In our context the validity of a Pumping lemma would mean that there exists some period string that allows arbitrarily long repetitions always resulting in a legal string. But if this period is of length m and $m \mid f(n)$ for some $n \in \mathbf{N}$ then e.g. the first symbol in the block contradicts the identical symbol in the $(f(n)/m)^{\text{th}}$ block. ■

Remarks: The result ties the divisibility properties of $\{f(n)\}_{n \in \mathbf{N}}$ to the complexity of the sequences: if non-divisibility prevails periodic sequences exist, if not then the recipe for constructing infinite sequences must be more complicated. Examples like $X_{(3,2^n)}$ or $X_{(4,\text{primes})}^+$ containing periodic points are rare specialities. But the latter half of Example 1.3. suggests that forbidden periodicity is likely just part of the story.

1.2. Lacunary case

Considering one-sized sequences it is clear that for a finite d a faster growing f allows more freedom for the sequence generation. In particular P. Erdős seems to have formulated in the mid 80’s the question whether lacunary sequences are sufficiently fast growing for infinite sequences/finite chromatic number to prevail. Recall that $\{x_n\}$ is **lacunary** if there is $\epsilon > 0$ such that $x_{n+1}/x_n \geq 1 + \epsilon$, $\forall n$. This problem was solved by Y. Katznelson and published later in [K]:

Theorem 1.2.: *For any lacunary f there is a finite d such that $X_{(d,f)} \neq \emptyset$.*

Remarks 1. This is the sequence translation of the original solution which was in terms the chromatic number. The solution connected the problem to another one, in Diophantine approximation (also by Erdős), and solved them using a novel dynamical system formulation.

2. Katznelson didn’t provide an explicit bound for the chromatic number/number of symbols needed although one was implicit in his work. Other have later improved the bound the best being now $1 + c \log \epsilon / \epsilon$ where ϵ is as in the definition of lacunarity and c an absolute constant ([PS]). Note that while of course useful this still doesn’t resolve explicit cases like Example 1.3. i.e. whether $X_{(4,n!)}^+$ is non-empty.

It is natural to connect this to recurrence as follows (see [W] for more detail). Given (d, f) let X be the subset (0.1) and ρ the natural metric it inherits from the full shift.

Definition 1.3.: $D \subset \mathbf{N}$ is a **non-recurrence set** for the (topological) dynamical system (X, σ, ρ) if for some positive b and all $x \in X$ and $d \in D$ it holds that $\rho(\sigma^d x, x) \geq b$.

It is easy to show that the finiteness of the chromatic number is equivalent to D being a non-recurrence set ([W]). So in particular Theorem 1.2. shows that a lacunary D is a set of non-recurrence. Not much seems to be currently known about non-lacunary non-recurrence sets - very few examples are known.

For subsequent use will be to define D a **set of recurrence** if X has infinite chromatic number.

1.3. Sublacunary realm

1.3.1. Dynamics approach

We now chart a bit what is known about the sublacunary case. The resulting map will unfortunately be somewhat patchy, perhaps indication of the fact that not all right concepts are yet known. Clearly recurrence sets and non-recurrence sets partition the space of sequences but unfortunately there doesn't seem to be a definitive criterion for the recurrence sets. Within them most attention has been paid to Poincaré sequences.

Definition 1.4.: A sequence $D \in \mathbf{N}$ is a **Poincaré sequence** if for a (measurable) dynamical system (Y, \mathcal{B}, μ, T) and any $B \in \mathcal{B}$, $\mu(B) > 0$ it holds that $\mu(T^{-d_k} B \cap B) > 0$ for some $d_k \in D$.

Remarks: 1. The Poincaré Theorem concerns the case $D = \mathbf{N}$, showing this D to be a set of recurrence.

2. An ingenious example by I. Kriz shows that Poincaré sequences are not all of recurrence sets.

As the Definition above is not explicit, other forms have been searched for. For a subset of integers A , let the **difference set** be $A - A = \{a - a' \mid a, a' \in A\}$ and let $A^{(N)} = A \cap \{1, 2, 3, \dots, N\}$. By the **upper density of A** we mean $\limsup_{N \rightarrow \infty} \frac{|A^{(N)}|}{N}$. Betrand-Mathis established the equivalence: D is a Poincaré sequence if and only if $D \cap (A - A) \neq \emptyset$ for all A of positive upper density ([B-M]). Due to this formulation, in the literature such D is also called an **intersective sequence**.

The developments of the subject gained definitive momentum when Lazlo Lovász conjectured (apparently just verbally) that if one insists that $n^2 \notin A - A$ for all natural numbers n then A cannot have positive density. In the late seventies Furstenberg and Sárközy separately managed to prove (in [F] and [S]):

Theorem 1.5.: Given $\delta > 0$ there is $N_0(\delta)$ such that if $N \geq N_0(\delta)$ and $|A^{(N)}| \geq \delta N$ then there is natural n such that $n^2 \in A - A$.

Remark: The proofs were of different character; Furstenberg's ergodic theoretic, Sárközy's utilized Fourier analytic techniques enabling quantitative estimates for the density. This will have some implications in our further work.

Theorem 1.5. together with the Bertrand-Mathis -equivalence shows that $\{n^2\}$ is an example of a Poincaré/intersective sequence. Later on more examples of such sequences have been worked out by Furstenberg and many others. We are not going to list these but just highlight that e.g. all monomials n^k and more generally polynomials satisfying the hypothesis of Theorem 1.1. are intersective (the latter result is due to Kamae and Mendés France, [KMF]).

While Bertrand-Mathis alternative is conceptually valuable, a criterion often more usable is

Lemma 1.6.: *If for all $\alpha \in (0, 2\pi)$ $\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n e^{i\alpha s_k} = 0$ then $\{s_k\}$ is a Poincaré sequence.*

Remark: These are called **Weyl sums** due to his first use of them to show uniform distribution (for integers).

The results above have direct bearing on our original problem. Consider for simplicity its 1-sided version with d symbols and f an intersective polynomial. Let $A_i = \{j \mid x_j = i\}$, these are just the sets of coordinates where the i^{th} symbol is found. Suppose now that $A_i, i = 1, \dots, d$ partition \mathbf{N} i.e. the 1-sided sequence is uniquely defined everywhere on \mathbf{N} . Then Theorem 1.5. generalized as indicated above implies that if the exclusion is to hold for this sequence, for sufficiently large N the upper densities of A_i 's cannot add up to 1, a contradiction. Hence

Theorem 1.7.: *For intersective polynomials f the spaces $X_{(d,f)}^+$ and therefore all $X_{(d,f)}$ are empty.*

To gain some insight on what can be done from first principles and more importantly, how the sequence termination takes place, let us take for a moment a purely elementary view.

Example 1.4. ($d = 3, 4$ and n^2): To directly resolve the case $d = 3$ one can utilize simple identities like $3^2 + 4^2 = 5^2$: instead of dealing with all the entries upto a given distance left and right of origin, one only needs to check a rather thin subset of them. Let $x_0 = 1$ and generate its blocks on $\{-25, \dots, 25\}$. Then set $x_{16} = 2$ and generate the blocks from it. In third iteration one sets $x_{-9} = 3$ (or $x_{25} = 3$ just as well) since there is a double block at it already due to the given identity. Now the full alphabet is in use. If e.g. at the doubly blocked coordinate 7 one sets $x_7 = 1$ the following (fourth) assignment iterate succeeds but next one results in a full block at several sites on $\{-25, \dots, 25\}$. Hence $X_{(3,n^2)} = \emptyset$.

This procedure of trying to generate a contradiction in the construction of a two-sided sequences gets rather unwieldy for larger S even for this f . Suspecting the spaces might be empty one could switch instead to considering one-sided sequences i.e. aiming to show that $X_{(4,n^2)}^+ = \emptyset$. Indeed here one succeeds with a **computer assisted proof** (CAP). Checking all these sequences systematically is a manageable task and turns out they are finite, the longest being of length 47. $X_{(5,n^2)}^+$ is exponentially harder and CAP on a desktop machine does not seem to terminate (sequences of length about 170 exist).

The problem gets quickly rather intractable for higher monomials $n^r, r \geq 3$ one reason being that there are no simplifying identities like $k^r = n^r + m^r$ available then due to Fermat's theorem. For $r = d = 3$ a CAP program finds one-sided sequences of length at least 300. Note that by the

argument in Theorem 1.1. we immediately know that there are no periodic sequences in any of the spaces $X_{(d,n^r)}$.

1.3.2. Additive combinatorics

Sárközy's result inspired plenty of subsequent additive number theory. Some of these results shed light to our original problem. They were often parallel results to those in the previous section, while simultaneously being quantitative. We will now briefly highlight this development.

Sárközy's proof of Theorem 1.5. utilized the Hardy - Littlewood Circle Method. By honing it further Balog, Pelikán, Pintz and Szemerédi established in 1994

Theorem 1.8.: *Fix a natural $k \geq 2$. If $n^k \notin A - A$ for all n then*

$$\frac{|A^{(N)}|}{N} \ll \frac{1}{(\log N)^{c \log \log \log \log N}}.$$

Here \ll means “less than constant times”. Best c has been worked out to be $1/\log 3$. This in turn has been later been extended to **intersective** polynomials $f \in \mathbf{Z}[x]$ perhaps culminating to J. Lucier's result in 2006 ([L]):

Theorem 1.9.: *Suppose $f \in \mathbf{Z}[x]$ is an intersective polynomial of degree $k \geq 2$ with positive leading term. If $f(n) \notin A - A$ for all n with $f(n) > 0$ then*

$$\frac{|A^{(N)}|}{N} \ll \left(\frac{(\log \log N)^\mu}{\log N} \right)^{\frac{1}{k-1}}$$

for $\mu = 3$ if $k = 2$, $\mu = 2$ if $k > 2$ and the constant only depends on f .

Like in the dynamics approach, from these results one can readily deduce Theorem 1.7. Additionally here one gets a upper bound for the instant when the “total density” of the d subsequences drops below 1 i.e. the contradiction takes place. But as one can see from the formulas the bound will be huge. This is due to the method of proof, not the phenomenon at hand. This absence of reasonable bounds will motivate us to investigate in the next section how long sequences satisfying (0.1) in the sublacunary realm can actually prevail.

One could of course argue that in the sublacunary case there is nothing special about the polynomials. Perhaps so, in which case the findings above are just to indicate explicit cases, where the original problem can be solved. However there is a reason to believe that Theorem 1.7 is generic in the sense that for all sublacunary rates of growth, there are f such that the chromatic number is infinite i.e. $X_{(d,f)}$ is empty for all d . (see [AHK]).

Our knowledge of the sublacunary realm is hence left in a somewhat curious state. We do not know whether for all sublacunary growth rates there are non-empty $X_{(d,f)}$ coexisting with the empty

ones. Our Example 1.1. ($\{2n\}$ is recurrent, $\{2n - 1\}$ is non-recurrent) shows that this is possible for at least some rate. Is there an upper bound for the growth rate below which such coexistence is possible? Perhaps the inevitable conclusion from this state of affairs is that the growth rate is just too crude a measure for the problem. One should use something (explicit!) that genuinely captures the recurrence-properties of the f -sequence.

2. Finite sequences

2.1. Random generation

We now proceed to the random generation of the simpler one-sided subset of sequences compatible with the exclusion (0.1) and the analysis of the outcome of this algorithm. The main results here are aimed at showing that a probabilistic model gives surprisingly accurate information on the length distribution. While the methods are applicable to a wide variety of jump sequences $f = n^2$ is the only one that we give a statistical assessment.

Our algorithm generates random sequences obeying the restriction (0.1) on $S^{\mathbf{N}}$. Each coordinate is chosen independently and uniformly but in such a way as to respect the restrictions from all the coordinates on its left.

Algorithm v2.0:

0. set $M \geq 1$, let $S_j = S$ at each $j \in \{1, \dots, M\}$ and set $i = 1$.
1. if $S_i = \emptyset$ then **halt**,
else pick uniformly a random symbol $s \in S_i$.
2. update $S_j \leftarrow S_j \setminus \{s\}$ for all $j = i + f(n) \in \{i + 1, \dots, M\}$, $n \in \mathbf{N}$.
3. if $i = M$ **halt** and call **full length**,
else $i \leftarrow i + 1$ and go to 1.

So just halting means that the random procedure did not result in a legal string of symbols of length M but produced a contradiction earlier. “Full length” means that length M was reached and the sequence needs to be continued with larger M to further address its viability.

One might perhaps also want to think the the sequence $\{s_i\}$ as a *random walk in a random environment (RWRE)*. Since the walker itself generates the future obstacles one could call the walk in this sense *self-similar*.

Definition 2.1.: Given a site $j \geq 2$ call the set of sites $D_j = \{1, 2, \dots, j - 1\} \cap \{k \mid k = j - f(n) \geq 1, n \in \mathbf{N}\}$ the **dragnet** of j and let $D_1 = \emptyset$. The cardinality of the dragnet is an increasing step function. Let the interval between the $(d + i - 1)^{\text{st}}$ and $(d + i)^{\text{st}}$ steps span the i^{th} **interval**, $i \geq 1$. Its length is denoted by l_i .

If the sites in the dragnet D_j support the entire alphabet S , the site j has **full block**. The site where the first full block can materialize is the **first jump**, the next step after that is **second jump** etc.

Suppose that we take the particular choice $f(n) = n^2$. Then we have explicitly

Proposition 2.2.: *Given an alphabet of size d , the location of the i^{th} jump is at $(d + i)^2 + 1$, $i = 1, 2, \dots$. The i^{th} interval $\{(d + i - 1)^2 + 1, \dots, (d + i)^2\}$ is of length $l_i = 2(d + i) - 1$.*

The proof is simple calculation and we omit it. With this choice of jumps the first dragnet of size d materializes at $j = d^2 + 1$, the beginning of the first interval, which is also the first jump site. The first interval ends at $(d + 1)^2$. With the Algorithm one can always generate at least sequences of length d^2 , but after that i.e. at the beginning of the first interval there is the possibility that the sites in the dragnet generate a full block which terminates the generation.

2.2. A probabilistic model

We now proceed to set-up and investigate a simplified model for the halting of the Algorithm.

To state our first result we first recall that for non-negative integers k_i the **multinomial coefficient** is defined by the expression

$$\binom{k_1 + k_2 + \dots + k_d}{k_1 \ k_2 \ \dots \ k_d} = \frac{(k_1 + k_2 + \dots + k_d)!}{k_1! k_2! \dots k_d!}$$

where $0! = 1$ if need be. If $d = 2$ this is just the Binomial coefficient. The following basic property of multinomials will be needed subsequently:

The Multinomial Theorem: *For $n \geq d$*

$$(2.1) \quad \sum_{\substack{k_1, k_2, \dots, k_d \\ k_1 + \dots + k_d = n}} \binom{n}{k_1 \ k_2 \ \dots \ k_d} = d^n$$

where the sum is d -fold over the non-negative integers k_i .

The proof amounts to inductively rewriting $(\sum_1^d 1)^n$ in a more explicit form. It can be found in many combinatorics and statistics books (see e.g. [A]).

Theorem 2.3.: *Assume that all the symbols on $\{1, 2, \dots, j - 1\}$ have been laid out independently and uniformly from S i.e. they are distributed $\sim B(\frac{1}{d}, \frac{1}{d}, \dots, \frac{1}{d})$. Let B_j be the event that one has the first full block at j in the i^{th} interval. Then*

$$(2.2) \quad \mathbf{P}(B_j) = p_i = \frac{1}{d^{d+i-1}} \sum_{\substack{k_r \geq 1, \ r=1, \dots, d \\ k_1 + \dots + k_d = d+i-1}} \binom{d+i-1}{k_1 \ k_2 \ \dots \ k_d}$$

where the sum is d -fold over the given positive integers.

Proof: For j in the i^{th} interval the dragnet is of cardinality $d + i - 1$. Since the entries on the support of the dragnet (j_1, \dots, j_{d+1-1}) are independent, each elementary event $(s_{j_1}, \dots, s_{j_{d+1-1}})$ is of probability $d^{-(d+i-1)}$. A full block at j materializes if and only if each symbol of S appears at least once in this dragnet. Since the order of the symbols is immaterial the multiplicity of the elementary event is given by the multinomial coefficient

$$\binom{d+i-1}{k_1 k_2 \dots k_d} \quad \text{with } k_r \geq 1 \text{ for all } r = 1, \dots, d.$$

All these arrangements on the dragnet are disjoint, hence the sum (2.2) gives the total probability of the full block. ■

Remark: The independence assumption here could of course be a heavy one. For one thing identical subsequent symbols ($s_j = s_{j+1}$) are forbidden in the original model but not here. For large S one expects exclusions like this to be less significant i.e. our simplified model should then be more accurate. We return to this in Section 2.2. after the data comparison.

The sequence $\{p_i\}$ is crucial in understanding our model. It is of course non-decreasing since the dragnet is not shrinking with increasing i . Indeed one can easily show that p_i is strictly increasing and $p_i \uparrow 1$. More specifically

Proposition 2.4.: *For an alphabet S of size d one has for all $i \geq 1$*

$$(2.3) \quad 1 - p_i < d \left(1 - \frac{1}{d}\right)^{d-1} \left(1 - \frac{1}{d}\right)^i.$$

Proof: Geometrically the Multinomial Theorem (2.1) is a statement about the total sum on the entries on the n^{th} level of Pascal's d -simplex i.e. on the simplex face perpendicular to $(1, 1, \dots, 1) \in \mathbb{N}^d$ at (lattice) distance n from the top of the pyramid (which is at the origin). For the size of $1 - p_i$ we need by Theorem 2.3 to estimate the boundary sum on that face:

$$\sum_{\substack{k_r=0 \text{ for some } r \\ k_1 + \dots + k_d = d+i-1}} \binom{d+i-1}{k_1 k_2 \dots k_d}$$

Without loss of generality assume that at least $k_d = 0$. Then by the Multinomial Theorem

$$(2.4) \quad \sum_{\substack{k_1, \dots, k_{d-1} \\ k_1 + \dots + k_{d-1} = d+i-1}} \binom{d+i-1}{k_1 k_2 \dots k_{d-1}} = (d-1)^{d+i-1}.$$

But d -simplex has $d + 1$ sides and its $(d + i - 1)^{\text{st}}$ level has d sides. For each of them (2.4) holds, hence for all boundary terms we have the bound

$$(2.5) \quad \sum_{\substack{k_r=0 \text{ for some } r \\ k_1 + \dots + k_d = d+i-1}} \binom{d+i-1}{k_1 k_2 \dots k_d} < d(d-1)^{d+i-1}$$

which is a strict inequality since corners are accounted multiple times and the multinomial is at least 1. But dividing (2.5) by d^{d+i-1} , the total sum over the $(d + i - 1)^{\text{st}}$ level, gives the result. ■

Remark: Although this is not subsequently used, it is interesting to note that in spite of the slight slack in (2.5), there seems to be no asymptotic gap in (2.3) (numerical check).

Theorem 2.5.: *Let the assumption on the sequence be as in Theorem 2.3. and let $f(n) = n^2$. Then the sequence generation halts at j i.e. the distribution of the full block is given by*

$$(2.6) \quad \mathbf{P}(\text{halts at } j) = \begin{cases} 0, & 1 \leq j \leq d^2; \\ (1 - p_1)^{j-d^2-1} p_1, & j \text{ in the first interval;} \\ \left(\prod_{k=1}^{i-1} (1 - p_k)^{l_k} \right) (1 - p_i)^{j-d^2-1-\sum_{k=1}^{i-1} l_k} p_i, & j \text{ in the } i^{\text{th}} \text{ interval, } i \geq 2. \end{cases}$$

Proof: Before $j = d^2 + 1$ the dragnet is too small to support all of S hence full block is impossible. On the first interval the distribution is geometric: with probability $1 - p_1$ no block and with probability p_1 a block and thereby halting. At each instant the termination is independent of the past.

At times $j = (d + i)^2 + 1$, $i = 2, 3, \dots$ the parameter changes to from p_i to p_{i+1} but otherwise the termination mechanism is still geometric. Compounding this to account the later intervals gives the general case in the expression (2.6). ■

Corollary 2.6.: *The halting time distribution has an exponential tail. The sequences generated are almost surely finite.*

Proof: Within the i^{th} interval the halting probability (2.6) is geometrically decreasing with the rate $1 - p_i$. Between the i^{th} and $(i + 1)^{\text{st}}$ intervals the probability jumps,

$$(2.7) \quad \frac{\mathbf{P}(\text{halts at } (d + i)^2 + 1)}{\mathbf{P}(\text{halts at } (d + i)^2)} = \frac{(1 - p_i)p_{i+1}}{p_i} = r_i .$$

For small i this expression can and will exceed 1. But by Proposition 2.4. $p_i \rightarrow 1$ and hence also $\frac{p_{i+1}}{p_i} \rightarrow 1$ so $\max\{r_i, 1 - p_i\}$ will eventually remain below $1 - \epsilon$ for some $\epsilon > 0$ which implies the result for i . The lengths l_i satisfy $l_{i+1} = l_i + 2$, so in particular $l_{i+1} \leq (1 + \delta)l_i$ for some $\delta > 0$ and the exponential decay holds in j as well.

If $A_j = \{\omega \mid \text{sequences of length } \geq j\}$ then by the above $\mathbf{P}(A_j) \leq e^{-cj}$ for some positive c . Hence $\sum_j \mathbf{P}(A_j)$ is summable and by the Borel-Cantelli Lemma $\mathbf{P}(A_j \text{ i.o.}) = 0$. ■

Remark: All the results above before Theorem 2.5. were independent of the particular choice $f(n) = n^2$. In this theorem the assumption was made just to have an explicit (but still complicated) formula. In the Corollary one needs to control the interval lengths well enough, a task difficult only for rather wildly behaving f .

2.3. Model versus data

To investigate the properties of the sequences generated by the Algorithm we collected samples for alphabets of sizes $d = 4, 5, 7, 10, 15$ and 20 . The upper limit arises simply from the computational effort. If we set $M = 2000$, a reasonable guess (after some playing around) for an upper bound for termination for $d = 20$, the sequences must be sieved out from 20^{2000} candidates using the Algorithm. Computed with *Mathematica* on a decent workstation these runs are timed in days. The sample sizes (number of sequences generated) that the following observations are based are indicated in Table I, last column.

On the other hand to assess the probabilistic model the main hurdle is to compute explicit values for p_i for sufficiently many $i = 1, 2, 3, \dots$ since these parametrize the halting distribution in (2.6). Here one can expedite things by computing the multinomials in the most efficient way. The key sum in (2.2),

$$\sum_{\substack{k_r \geq 1, r=1, \dots, d \\ k_1 + \dots + k_d = d+i-1}} \binom{d+i-1}{k_1 k_2 \dots k_d},$$

can be simplified. E.g. for $i = 4$ it can be computed as

$$\binom{d}{1} \binom{d+3}{1 \dots 1 4} + 2 \binom{d}{2} \binom{d+3}{1 \dots 1 2 3} + \binom{d}{3} \binom{d+3}{1 \dots 1 2 2 2},$$

an expression far easier to deal with than the original (\dots denotes here a constant string of 1's). The idea here is to keep record of the multiplicity of the string that appears inside the multinomial (like 23 in the middle expression - it can appear in $\binom{d}{2}$ different ways and in 2 different orders). But this too gets rather unwieldy for large d with i beyond the smallest values.

Figure 1., middle row, is joint plot of the sequence termination instant distributions of the data and the model for $d = 10$. One sees a reasonable overall match but also distinct discrepancies due to our simplified model. Both plots peak at the same spot, the start of the fifth interval ($i = 5$ so $(d+i-1)^2 + 1 = 197$) and their levels there are fairly close to each other (0.0140179 and 0.0150393, empirical and model resp.). The model favours slightly earlier halting as witnessed in the distributions: it both rises and sets earlier.

The model graph has by definition geometric slopes on the intervals and so indeed does the data: this is clear from the log-plot on the right of the same data (plot is $\ln(\text{number of samples} + 1)$). The experimental and model slope tilts are actually fairly close to each other as seen on the left, especially in the middle part of the distribution. The key observation here though is that the tail of the data distribution seems indeed majored by an exponential and tapers off around same range of j as it does for the model. This strongly suggests that the sequences don't get very far. The top row of Figure 1 present the same comparison for $d = 5$. The overall distribution match is actually quite remarkable. In the finer details like the slope roughness there is deviation from the model (although the sample size is here at its largest, see Table 1). We return to this in Section 2.3.

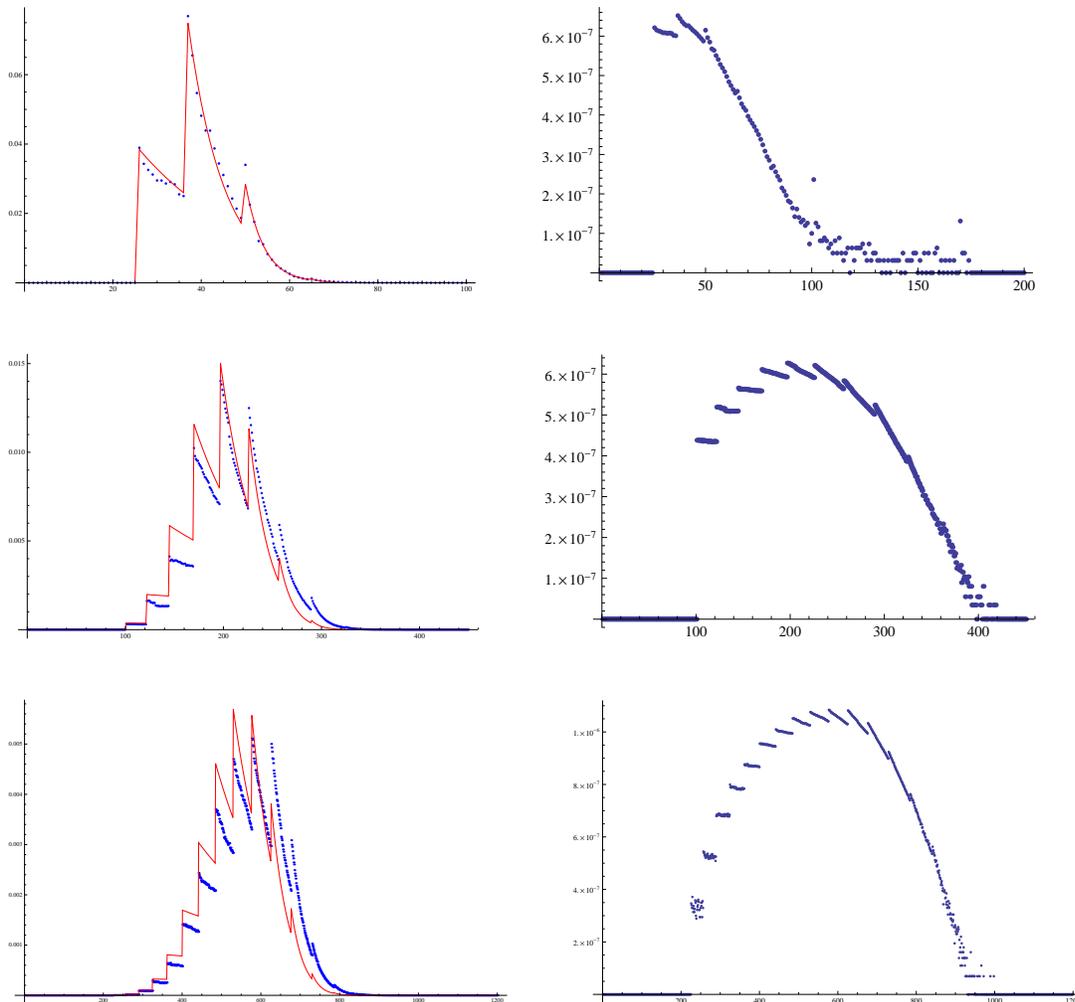


Figure 1: Empirical (blue/rough) and theoretical (red/continuous) halting distributions superimposed together (left) and corresponding log empirical distribution (right) for different alphabet sizes. From top to bottom row: $d = 5, 10$ and 15 .

The left plots of the bottom row (for $d = 15$) exhibits increasing symmetry in the distributions. It is also hinting at the high d -limit: with suitable scaling the halting distribution might be normal. Magnifying the slopes here shows that the mechanism that gave the top row data the “roughness” are not there anymore. One is tempted to think that this is due to randomness smoothing them out.

Perhaps a bit troubling facts in the mid and bottom plots on the left in Figure 1 are the vertical off-set between the slopes. There is a way to see a dependency mechanism in the sequences that contributes to them appearing. While this does not seem to provide a way to improve the model it is still worth understanding.

Consider the assignment sequence indicated in Figure 2. Here a, b, c and d are coordinates, $b = a + k^2$, $c = a + n^2$, $d = b + n^2$, $n > k$. Since $x_a = s$ we know that $x_b \neq s \neq x_c$ as indicated.

One should think of a dragnet being at d , hence b and c belong to it and a may or may not belong. The simplest case (and probably the one contributing most to the probabilities involved) is of course $k = 1$, $n = 2$. We are interested in the probability of a full block appearing at d .

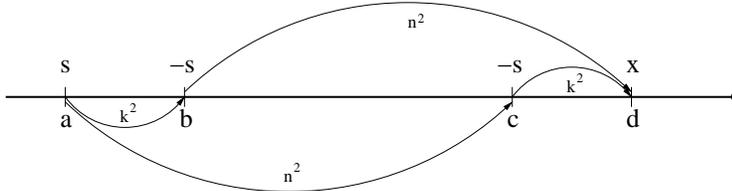


Figure 2: A dependency mechanism affecting the termination probability.

Ignoring all other symbols on the interval $\{a, \dots, d\}$ two cases should be distinguished:

1. if $k^2 + n^2$ is not a square then no symbol from the triple x_a, x_b and x_c forbids $x_d = s$ but x_b and x_c do block one or two non- s symbols at d .
2. if $k^2 + n^2$ is a square then x_a forbids $x_d = s$ and furthermore x_b and x_c block one or two non- s symbols at d .

On the basis of these we see immediately that

$$\mathbf{P}(\text{full block at } d \mid k^2 + n^2 \text{ not square}) < \mathbf{P}(\text{full block at } d) < \mathbf{P}(\text{full block at } d \mid k^2 + n^2 \text{ square}).$$

But as we vary k and n there are many more sites a where 1. instead of 2. holds. Hence one should expect the full block i.e. termination probabilities be majored by the independent probabilities of the model. This is indeed the case as seen in Figure 1. How to quantify this is another matter though, well worth further study.

To further characterize the sampled sequences we computed their expected lengths and standard deviations. These and the corresponding model data are presented in Table I for a few d values. For large d we already mentioned of the problem of computing p_i for high i . These are needed for an accurate tail estimate of the distribution. The asterisks in the Table are due to this complication.

The values of the two statistical indicators here just quantify aspects of the earlier distributional matching. A bit surprisingly the best match occurs at the value $d = 5$ indicating that small alphabets don't seem to suffer from the independence assumption in the model.

The decay rate of the density in Theorem 1.9. can be used to estimate termination times. These turn out truly staggering compared to the real ones above. This is of course due to the proof methods and suggests that other arguments truer to the nature of the problem should be found.

| Symbols d | Empirical mean | Empirical std. dev. | Model mean | Model std. dev. | Sequences |
|----------------|-------------------|------------------------|---------------|--------------------|-----------------|
| 4 | 27.2542 | 5.13374 | 23.992 | 5.23924 | $50 \cdot 10^6$ |
| 5 | 39.5672 | 8.28983 | 39.2172 | 8.22516 | $80 \cdot 10^6$ |
| 6 | 60.8247 | 13.5813 | 59.3666 | 11.9713 | $80 \cdot 10^6$ |
| 7 | 89.4687 | 18.5912 | 84.982 | 16.5113 | $30 \cdot 10^6$ |
| 10 | 209.315 | 38.2887 | 199.562 | 35.1369 | $20 \cdot 10^6$ |
| 15 | 566.87 | 92.2796 | 543.291 | 84.4349 | $10 \cdot 10^6$ |
| 20 | 1156.57 | 170.829 | * | * | $5 \cdot 10^6$ |

Table I. Some empirical and theoretical statistics of the one-sided sequences. Means and standard deviations of the halting times at various alphabet sizes. Asterisks refer to missing coefficients.

Analyzing the empirical data further reveals rather clearly certain exponents at play: the values in the second column grow almost exactly like $d^{5/2}$ and those of the third column roughly like $d^{15/7}$. Since there seems to be an increasingly random behavior in the halting of the sequences for higher d , we venture to

Conjecture 2.7.: Suppose $T^{(d)}$ is the halting instant of the Algorithm v2.0 with $f(n) = n^2$. For sufficiently rapidly growing $M(d)$ there are positive constants a and b such that as $d \rightarrow \infty$

$$\mathbf{P} \left(\frac{T^{(d)} - ad^{5/2}}{bd^{15/7}} \leq x \right) \rightarrow \Phi(x) \quad \forall x \in \mathbf{R}$$

where Φ is the cumulative distribution function of the standard normal $N(0, 1)$.

Remarks: 1. For the sequence length $M(d)$ we just need a rate that outgrows the off-set rate $d^{5/2}$. Clearly for the left threshold $(d^2 + 1 - ad^{5/2})/bd^{15/7} \rightarrow -\infty$ as desired.

2. A partial result would be to show the Central Limit Theorem for the model. For this one would need to analyze the asymptotics of $p_i(d)$ and then the scaling limit of (2.6).

3. The CLT for the algorithm would of course imply the a.s. emptiness of the sequence spaces.

2.4. Some fine detail

While generating the sequences some additional information was gleaned, too. To get some insight in how the haltings come about we recorded the pairs (i, n) at which the first terminal block was generated. With this one knows that while the sequence may be extended a bit (at least one step) it cannot be continued past $i + n^2$. Note that because of this the resulting distribution for i is not exactly the distribution of the halting instant j earlier, but slightly shifted to the left.

Algorithm v2.1:

0. set $M \geq 1$, let $S_j = S$ at each $j \in \{1, \dots, M\}$ and set $i = 1$.
1. pick uniformly a random symbol $s \in S_i$.
2. update $S_j \leftarrow S_j \setminus \{s\}$ for all $j = i + f(n) \in \{i + 1, \dots, M\}$, $n \in \mathbf{N}$
 if $S_j = \emptyset$ results for some j , record (i, n) , halt and call **empty**,
 if $i = M$ halt and call **full length**,
 else $i \leftarrow i + 1$ and go to 1.

With this modified algorithm one gets to see from where the terminal jumps originate. Figure 3. illustrates their distributions, call it the (i, n) -plot, for two d -values.

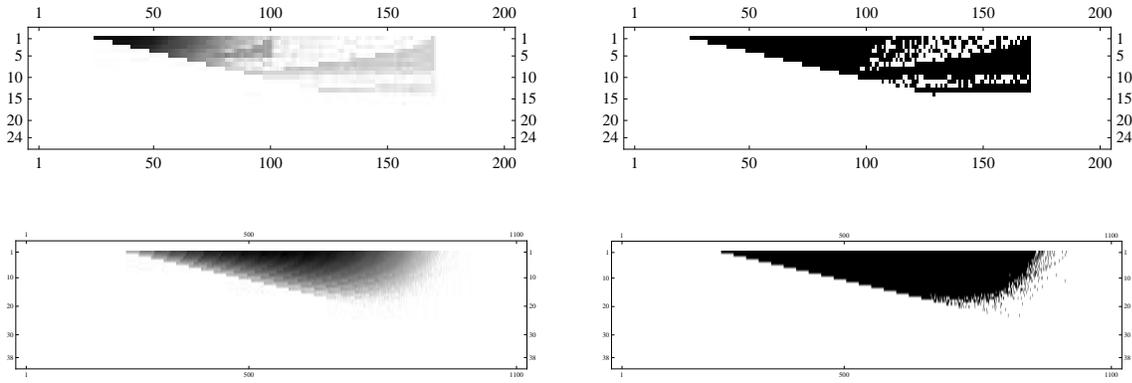


Figure 3: Termination distributions for $d = 5$ (top) and 15 (bottom). Iterate i runs rightwards and jump size n downwards. Same distributions in each row; on the left a log-plot, on the right the plain support (non-zero entries). Data from 20 and 10 million samples respectively. Note the different scales up and down.

The first distinct feature, the left hand “staircase” in the plots, can be readily explained.

Proposition 2.8.: *The staircase jumps in the (i, n) -plot are at $d^2 + 2(n - 1)d - 2(n - 1)$, $n \geq 1$ and the step length is $2(d - 1)$.*

Proof: If a terminal block is generated with a jump of size n^2 , then the previous $d - 1$ (non terminal) blocks into this future site have been generated with jumps of size at least $(n + 1)^2, (n + 2)^2, \dots, (n + d - 1)^2$. These particular jump sizes give the minimal i for which $i + n^2$ results in a termination. So necessarily $j = i + n^2 \geq (n + d - 1)^2 + 1$, from which we get that $i \geq (n + d - 1)^2 - n^2 + 1 = \dots = d^2 + 2((n - 1)d - n + 1)$. Evaluating this for $n \geq 1$ gives the step values and their difference gives the step size. ■

In a similar fashion to Definition 2.1. one can formulate locations and length of the horizontal $(l \times 1)$ -blocks inside the distributions at any given level n . Within these the distribution of terminations seems to be approximately geometric as expected from our earlier analysis.

A couple of things in these plots are far more intriguing though. The first is the internal and right end structure in the plots for $d = 5$. There seem to be some hard combinatorial constraints at play here. Hence the lace-like interior and the abrupt wall on the right (at 173). Note that the plots are based on a large data set of $20 \cdot 10^6$ samples. Furthermore, for the computation the sequence length attempted was $M = 600$ i.e. the shown graphics is a cutout without any boundary effect from 200. Also a blank i.e. no sample points is rather improbable event for low n in the interior of the plot unless there is a definite reason for the absence.

The lower plots clearly indicate random phenomena dominating for higher a alphabet size (here $d = 15$). Apart from the left staircase there are no traces of hard constraints elsewhere. The right hand side of the distribution is tapering off like the halting distribution earlier, most likely with exponential tails at all n -levels.

3. Conclusions

Here we have tried to chart of what still seems unexplored or at least rather hazy terrain. This of course applies to much of the study on long range order e.g. in Statistical Mechanics and related fields. To come up with an empty set is perhaps a bit disappointing but hopefully the methods here are useful in further work. At least we know how stringent even the most innocuous looking exclusion rules are if applied with unbounded reach (once the obvious case of periodicity has been ruled out). Moreover the large discrepancy between the termination times arising from proofs and the simple probabilistic model suggests that there should be a more natural proof to be found.

In the latter half we have restricted ourselves mostly for the jump sequence $f(n) = n^2$. The reason is two-fold: firstly, the original question was about this and secondly other nontrivial f 's get quickly out of hand computationally. Our probabilistic model should work verbatim for n^k , $k \geq 3$ but because of their jump sizes, computing the exact values of p_i for high i remains a problem. So does generating data: the range M in our algorithms has to be much larger hence a far bigger number of exclusions have to be recorded. The reason why at high d a sequence generated eventually terminates remains the same, the bulk accumulation of exclusions at a site (which is overcome only beyond the lacunary threshold). But a much larger store of excluded symbol values need to be kept in order to finally record a full block. Conjecture 2.7. is likely to be true more generally but because of lack of supporting statistical data we will not venture to guess it.

Acknowledgements

The author would like to thank Mike Keane, Rob Tijdeman and Benjamin Weiss for valuable comments and pointers that led to the right literature.

References

- [A] Aigner, M.: *A course in Enumeration*, Springer, 2007.
- [AHK] Ajtai, M., Havas, I, Komlós, J.: Every group admits a bad topology, *Studies in Pure Mathematics*, pp. 21-34, Birkhäuser, 1983.
- [BPPS] Balog, A., Pelikán, J., Pintz, J., Szemerédi, E.: Difference sets without k -th powers, *Acta Math. Hungar.*, **65** (2), pp. 165-187, 1994
- [B-M] Bertrand-Mathis, A.: Ensembles intersectifs et recurrence de Poincaré, *Colloque de Théorie Analytique des Nombres "Jean Coquet" (Marseille, 1985)*, pp. 55-72, Publ. Math. Orsay, 1988-2, Univ. Paris XI, Orsay, 1988.
- [CT] Chow, Y.S., Teicher, H.: *Probability Theory: Independence, Interchangeability, Martingales*, Springer, 1978.
- [F] Furstenberg, H.: Ergodic behavior of diagonal measures and a theorem of Szemerédi on arithmetic progressions, *J. d'Analyse Math.*, **71**, pp. 204-256, 1977.
- [HU] Hopcroft, J.E., Ullman, J.D.: *Introduction to Automata Theory, Languages, and Computation*, Addison-Wesley Publishing, Reading Massachusetts, 1979.
- [KMF] Kamae, T., Mendés France, M.: van der Corput difference theorem, *Israel J. Math.*, **31** (3-4), pp. 335-342, 1978.
- [K] Katznelson, Y.: Chromatic numbers of Cayley graphs on \mathbf{Z} and recurrence, *Combinatorica*, **21**, pp. 211-9, 2001.
- [L] Lucier, J.: Intersective sets given by a polynomial, *Acta Arith.*, **123**, pp. 57-95, 2006.
- [LM] Lind, D., Marcus, B.: *An introduction to Symbolic Dynamics and Coding*, Cambridge Univ. Press, 1995.
- [PS] Peres, Y., Schlag, W.: Two Erdős problems on lacunary sequences: chromatic number and diophantine approximation, [arxiv:math.CO:0706.0223v1](https://arxiv.org/abs/math.CO/0706.0223v1)
- [S] Sárközy, A.: On difference sets of sequences of integers, *Acta Math. Hungar.*, **31** (1-2), pp. 125-149, 1978.
- [W] Weiss, B.: *Single orbit dynamics*, CBMS Regional Conference Series in Mathematics, **95**, AMS, 2000.