



Branching Frequency and Markov Entropy of Repetition-Free Languages

Elena A. Petrova[✉] and Arseny M. Shur[✉]

Ural Federal University, Ekaterinburg, Russia
{elena.petrova,arseny.shur}@urfu.ru

Abstract. We define a new quantitative measure for an arbitrary factorial language: the entropy of a random walk in the prefix tree associated with the language; we call it Markov entropy. We relate Markov entropy to the growth rate of the language and the parameters of branching of its prefix tree. We show how to compute Markov entropy for a regular language. Finally, we develop a framework for experimental study of Markov entropy by modelling random walks and present the results of experiments with power-free and Abelian-power-free languages.

Keywords: Power-free language · Abelian-power-free language · Markov entropy · Prefix tree · Random walk

1 Introduction

Formal languages closed under taking factors of their elements (factorial languages) are natural and popular objects in combinatorics. Factorial languages include sets of factors of infinite words, sets of words avoiding patterns or repetitions, sets of minimal terms in algebraic structures, sets of palindromic rich words and many other examples. One of the main combinatorial parameters of factorial languages is their asymptotic growth. Usually, “asymptotic growth” means asymptotic behaviour of the function $C_L(n)$ giving the number of length- n words in the language L . (In algebra, the function that counts words of length n is more popular.)

In this paper we propose a different parameter of asymptotic growth, based on representation of factorial languages as *prefix trees*, which are diagrams of the prefix order on words. Given such an infinite directed tree, one can view each word as a walk starting at the root. We consider *random walks*, in which the next node is chosen uniformly at random among the children of the current node, and define their *entropy* to measure the expected uncertainty of a single step. As a random walk is a Markov chain, we call this parameter the Markov entropy of a language. This parameter was earlier considered for a particular subclass of regular languages in the context of antidictionary data compression

Supported by the Ministry of Science and Higher Education of the Russian Federation (Ural Mathematical Center project No. 075-02-2021-1387).

[4]. However, it seems that more general cases were not analysed up to now. Our interest in Markov entropy is twofold. First, it allows us to estimate growth properties of a language from statistics of experiments, if exact methods do not work. Second, it is related to a natural and efficient (at least theoretically) data compression scheme, which encodes the choices made during a walk in the prefix tree.

Our contribution is as follows. In Sect. 3 we define the order- n Markov entropy $\mu_n(L)$ of a language L in terms of length- n random walks in its prefix tree $\mathcal{T}(L)$ and the Markov entropy $\mu(L) = \lim_{n \rightarrow \infty} \mu_n(L)$. Then we relate Markov entropy to the exponential growth rate of L and to the parameter called *branching frequency* of a walk in $\mathcal{T}(L)$. In Sect. 4.1 we show how to compute Markov entropy for a regular language. Then in Sect. 4.2 we propose a model of random walk for an arbitrary factorial language through depth-first search and show how to recover branching frequency from observable parameters of a walk. Finally, in Sect. 5 we present algorithms used in the experimental study of Markov entropy for power-free and Abelian-power-free languages and the results of this study. All proofs are omitted; see [arXiv:2105.02750](https://arxiv.org/abs/2105.02750) for the full version.

2 Preliminaries

We study words and languages over finite alphabets; Σ^* denotes the set of all words over an alphabet $\Sigma = \{0, \dots, \sigma - 1\}$. Standard notions of prefix, suffix, factor are used. We use the array notation $w = w[1..n]$ for a word of length $n = |w|$; thus $w[i..i+k-1]$ stands for the length- k factor of w starting at position i . In particular, $w[i..i] = w[i]$ is the i th letter of w and $w[i..i-1]$ is the empty word, denoted by λ . A word w is *right extendable* in a language L if L contains infinitely many words with prefix w ; $\text{re}(L)$ denotes the set of all words which are right extendable in L .

A word w has *period* p if $w[1..|w|-p] = w[p+1..w]$. For an integer $k > 1$, the k -*power* of a word w is the concatenation of k copies of w . For an arbitrary real $\beta > 1$, the β -*power* (resp., the β^+ -*power*) of w is the prefix of length $\lceil \beta|w| \rceil$ (resp., $\lfloor \beta|w| + 1 \rfloor$) of the infinite word $w^\infty = ww \dots w \dots$. E.g., $(010)^{2^+} = (010)^{7/3} = 0100100$, $(010)^{5/2} = (010)^{(5/2)^+} = 01001001$. A word is β -*power-free* if it has no β -powers as factors; the k -ary β -*power-free language* $\text{PF}(k, \beta)$ consists of all β -power-free words over the k -letter alphabet. The same definitions apply to β^+ -powers. The crucial result on the power-free languages is the threshold theorem, conjectured by Dejean [7] and proved by efforts of many authors [3, 6, 16–18, 20]. The theorem establishes the boundary between finite and infinite power-free languages: the minimal infinite k -ary power-free languages are $\text{PF}(3, \frac{7}{4}^+)$, $\text{PF}(4, \frac{7}{5}^+)$, and $\text{PF}(k, \frac{k}{k-1}^+)$ for $k = 2$ and $k \geq 5$. These languages are called *threshold languages*.

The *Parikh vector* $\psi(w)$ of a word w is a length- σ vector such that $\psi(w)[i]$ is the number of occurrences of the letter i in w for each $i \in \Sigma$. Two words with equal Parikh vectors are said to be *Abelian equivalent*. A concatenation of k Abelian equivalent words is an *Abelian k th power*. Abelian k -power-free

words are defined in analogy with k -power-free words; Abelian square-free (resp., cube-free, 4-power-free) languages over four (resp., three, two) letters are infinite [8, 12].

A language $L \subseteq \Sigma^*$ is *factorial* if it contains all factors of each its element. Power-free and Abelian-power-free languages are obviously factorial. The relation “to be a prefix (resp., a suffix, a factor)” is a partial order on any language. The diagram of the prefix order of a factorial language L is a directed tree $\mathcal{T}(L)$ called the *prefix tree*. Prefix trees are the main objects of study in this paper. For convenience, we assume that an edge of the form (w, wa) in $\mathcal{T}(L)$ is labeled by the letter a ; in this way, the path from the root to w is labeled by w . For regular languages we use deterministic finite automata with partial transition function (PDFA), viewing them as labelled digraphs. We assume that all states of a PDFA are reachable from the initial state; since we study factorial languages, we also assume that all states are final (so a PDFA accepts a word iff it can read it). When a PDFA \mathcal{A} is fixed, we write $q.w$ for the state of \mathcal{A} obtained by reading w starting at the state q .

Combinatorial complexity (or *growth function*) of a language $L \subseteq \Sigma^*$ is a function counting length- n words in L : $C_L(n) = |L \cap \Sigma^n|$. The *growth rate* $\text{gr}(L) = \limsup_{n \rightarrow \infty} (C_L(n))^{1/n}$ describes its asymptotic growth. The combinatorial complexity of factorial languages is submultiplicative: $C_L(m+n) \leq C_L(m)C_L(n)$; by Fekete’s lemma [9], this implies $\text{gr}(L) = \lim_{n \rightarrow \infty} (C_L(n))^{1/n} = \inf_{n \in \mathbb{N}} (C_L(n))^{1/n}$. A survey of techniques and results on computing growth rates for regular and power-free languages can be found in [27].

Infinite Trees. We consider infinite k -ary rooted trees: the number of children of any node is at most k . Nodes with more than one child are called *branching points*. The *level* $|u|$ of a node u is the length of the path from the root to u . A subtree \mathcal{T}_u of a tree \mathcal{T} consists of the node u and all its descendants. The tree \mathcal{T} is p -*periodic* (resp., p -*subperiodic*) if there exists a function f on the set of nodes such that each subtree \mathcal{T}_u is an isomorphic copy (resp., is a subgraph) of the subtree $\mathcal{T}_{f(u)}$ and $|f(u)| \leq p$. The prefix tree of any factorial language L is 0-subperiodic, since suffixes of elements of L are also in L . Furthermore, $\mathcal{T}(L)$ is p -periodic for some p iff L is regular (p -periodicity means exactly that L has finitely many quotients, which is equivalent to regularity).

There are two widely used parameters of growth for infinite trees; see, e.g., [15]. “Horizontal” growth is measured by the *growth rate* $\text{gr}(\mathcal{T}) = \lim_{n \rightarrow \infty} (T_n)^{1/n}$, where T_n is the number of nodes of level n , whenever this limit exists. Hence, $\text{gr}(\mathcal{T}(L)) = \text{gr}(L)$. “Vertical” growth is measured by the *branching number* $\text{br}(\mathcal{T})$, which is usually defined using the notion of network flow. However, Furstenberg’s theorem [10] says that $\text{br}(\mathcal{T}) = \text{gr}(\mathcal{T})$ for subperiodic trees, so for prefix trees we have only one parameter. In Sect. 3, we propose one more parameter of growth using the notion of entropy.

Entropy. Let $\xi = (x_{1|p_1}, \dots, x_{n|p_n})$ be a discrete finite-range random variable, where $p_i, i = 1, \dots, n$, is the probability of the outcome x_i . The *entropy* of ξ is the average amount of information in the outcome of a single experiment: $H(\xi) =$

$-\sum_{i=1}^k p_i \log p_i$ (throughout the paper, \log stands for the binary logarithm). Lemma 1 below contains basic properties of entropy, established by Shannon [23].

- Lemma 1.** (1) For a random variable $\xi = (x_{1|p_1}, \dots, x_{n|p_n})$, $H(\xi) \leq \log n$; equality holds for the uniform distribution only.
 (2) For a random vector (ξ, η) , $H(\xi, \eta) \leq H(\xi) + H(\eta)$; equality holds iff ξ and η are independent.

3 Entropy Characteristics of Prefix Trees

Let $\mathcal{T} = T(L)$ be a prefix tree. The entropy characteristics introduced below measure the expected uncertainty of a single letter in a random word from L . By *order- n general entropy* $H_n(L)$ we mean the entropy of a random variable uniformly distributed on the set $|L \cap \Sigma^n|$ (or on the set of level- n nodes of \mathcal{T}), divided by n . By Lemma 1(1), $H_n(L) = \frac{\log C_L(n)}{n}$. The fact that L is factorial guarantees the existence of the *general entropy* of L , which is by definition the limit

$$H(L) = \lim_{n \rightarrow \infty} H_n(L) = \lim_{n \rightarrow \infty} \log(C_L(n))^{1/n} = \log \text{gr}(L).$$

A different notion of entropy stems from consideration of random walks in \mathcal{T} . As usual in graph theory, by *random walk* we mean a stochastic process (Markov chain), the result of which is a finite or infinite walk in the given graph. The process starts in the initial state (either fixed or randomly chosen from some distribution) and runs step by step, guided by the following rule: visiting the node u , choose an outgoing edge of u uniformly at random¹ and follow it to reach the next node. The walk stops if u has no outgoing edges. Note that all walks in \mathcal{T} are directed paths; we refer to the walks starting at the root as *standard*. Let η_n be the random variable with the range $|L \cap \Sigma^n|$ such that the probability of a word $w \in L$ is the probability that a random standard walk in \mathcal{T} , reaching the level n , visits w . The *order- n Markov entropy* of L is $\mu_n(L) = \frac{H(\eta_n)}{n}$. The following lemma is immediate from definitions and Lemma 1(1).

- Lemma 2.** For any factorial language L and any n , one has $\mu_n(L) \leq H_n(L)$.

Similar to the case of the general entropy, the limit value exists:

- Lemma 3.** Let L be a factorial language. Then there exists a limit $\mu(L) = \lim_{n \rightarrow \infty} \mu_n(L) = \inf_{n \in \mathbb{N}} \mu_n(L)$.

We call $\mu(L)$ the *Markov entropy* of L . We want to estimate $\mu(L)$ for different languages; so our first goal is to relate $H(\eta_n)$, and thus $\mu_n(L)$, to the parameters of the tree \mathcal{T} . Let $\text{ch}(w)$ denote the number of children of the node w in \mathcal{T} and $P(w)$ be the probability of visiting the word w by a random standard walk.

¹ Non-uniform distributions are also used in many applications but we do not consider them here.

Lemma 4. $P(w) = \left(\prod_{i=0}^{|w|-1} \text{ch}(w[1..i]) \right)^{-1}$.

In general, $P(w)$ may underestimate the probability assigned to w by $\eta_{|w|}$; this is the case if some prefix of w has a child which generates a finite subtree with no nodes of level $|w|$. To remedy this, we consider trimming of prefix trees. By n -trimmed version of \mathcal{T} , denoted by $\mathcal{T}_{[n]}$, we mean the tree obtained from \mathcal{T} by deletion of all finite subtrees \mathcal{T}_u which have no nodes of level n (and thus of bigger levels). In other words, a node $w \in L$ is deleted iff L contains no length- n word with the prefix w .

Example 1. Let $L = \text{PF}(2, 3)$, $\mathcal{T} = \mathcal{T}(L)$. If $n \geq 9$, then $\mathcal{T}_{[n]}$ does not contain $u = 00100100$ because $u0, u1$ end with cubes; if $n \geq 15$, then $\mathcal{T}_{[n]}$ does not contain $v = 0100101001010$, because $v1, v00$, and $v01$ end with cubes.

The trimmed version of \mathcal{T} , denoted by \mathcal{T}_{\square} , is obtained from \mathcal{T} by deletion of all finite subtrees. The next lemma follows from definitions.

Lemma 5. (1) $\mathcal{T}_{\square} = \bigcap_{n \in \mathbb{N}} \mathcal{T}_{[n]}$. (2) \mathcal{T}_{\square} is the prefix tree of $\text{re}(L)$.

We write $\text{ch}_{[n]}(w)$ ($\text{ch}_{\square}(w)$) for the number of children of w in $\mathcal{T}_{[n]}$ (resp., \mathcal{T}_{\square}) and $P_{[n]}(w)$ ($P_{\square}(w)$) for the probability of visiting w by a random standard walk in $\mathcal{T}_{[n]}$ (resp., \mathcal{T}_{\square}). As in Lemma 4, one has

$$P_{[n]}(w) = \left(\prod_{i=0}^{|w|-1} \text{ch}_{[n]}(w[1..i]) \right)^{-1} \text{ and } P_{\square}(w) = \left(\prod_{i=0}^{|w|-1} \text{ch}_{\square}(w[1..i]) \right)^{-1}. \quad (1)$$

Lemma 6. Let $w \in L$, $|w| = n$. Then η_n assigns to w the probability $P_{[n]}(w)$.

Definitions and Lemma 6 imply $H(\eta_n) = -\sum_{w \in L \cap \Sigma^n} P_{[n]}(w) \log P_{[n]}(w)$.

Given an arbitrary tree \mathcal{T} , we assign to each internal node u its weight, equal to the logarithm of the number of children of u . Branching frequency of standard walk ending at a node w , denoted by $\text{bf}(\mathcal{T}, w)$, is the sum of weights of all nodes in the walk, except for w , divided by the length of the walk (=level of w). The use of branching frequency for prefix trees can be demonstrated as follows. For a language L , a natural problem is to design a method for compact representation of an arbitrary word $w \in L$. A possible solution is to encode the standard walk in $\mathcal{T} = \mathcal{T}(L)$, ending at w . We take $|w|$ -trimmed version of \mathcal{T} and encode consecutively all choices of edges needed to reach w . For each predecessor u of w we encode the correct choice among $\text{ch}_{[\square|w|]}(u)$ outgoing edges. The existence of asymptotically optimal entropy coders, like the arithmetic coder [21], allows us to count $\log \text{ch}_{[\square|w|]}(u)$ bits for encoding this choice. Thus w will be encoded by $\sum_{i=0}^{|w|-1} \log(\text{ch}_{[\square|w|]}(w[1..i]))$ bits, which is exactly $\text{bf}(\mathcal{T}_{[\square|w|]}, w)$ bits per symbol.

Remark 1. The proposed method of coding generalizes the antidictionary compression method [4] for arbitrary alphabets. Antidictionary compression works as follows: given $w \in L \subseteq \{0, 1\}^*$, examine each prefix $w[1..i]$; if it is the only child of $w[1..i-1]$ in the prefix tree of L , delete $w[i]$. In this way, the remaining

bits encode the choices made in branching points during the standard walk to w . The compression ratio is the fraction of branching points among the predecessors of w : any branching point contributes 1 to the length of the code, while other nodes in the walk contribute nothing.

The following theorem relates branching frequencies to Markov entropy.

Theorem 1. *For a factorial language L and a positive integer n , the order- n Markov entropy of L equals the expected branching frequency of a length- n random walk in the prefix tree $\mathcal{T}(L)$.*

4 Computing Entropy

4.1 General and Markov Entropy for Regular Languages

Let L be a factorial regular language, \mathcal{A} be a PDFA, recognizing L . The problem of finding $\text{gr}(L)$, and thus $H(L)$, was solved by means of matrix theory as follows. By the Perron–Frobenius theorem, the maximum absolute value of an eigenvalue of a non-negative matrix M is itself an eigenvalue, called the *principal* eigenvalue. A folklore theorem (see [27, Th. 2]) says that $\text{gr}(L)$ equals the principal eigenvalue of the adjacency matrix of \mathcal{A} . This eigenvalue can be approximated² with any absolute error δ in $O(|\mathcal{A}|/\delta)$ time [24, Th. 5]; see also [27, Sect. 3.2.1].

Now consider the computation of $\mu(L)$. By Lemma 3 and Theorem 1, $\mu(L)$ is the limit of expected branching frequencies of length- n random standard walks in the prefix tree $\mathcal{T} = \mathcal{T}(L)$. Standard walks in \mathcal{T} are in one-to-one correspondence with accepting walks in \mathcal{A} , so we can associate each node $w \in \mathcal{T}$ with the state $\lambda.w \in \mathcal{A}$ and consider random walks in \mathcal{T} as random walks in \mathcal{A} . We write $\text{deg}^{\rightarrow}(u)$ for the out-degree of the node u in \mathcal{A} .

We need the apparatus of *finite-state Markov chains*. Such a Markov chain with m states is defined by a row-stochastic $m \times m$ matrix A (row-stochastic means that all entries are nonnegative and all row sums equal 1). The value $A[i, j]$ is treated as the probability that the next state in the chain will be j given that the current state is i . Any finite directed graph G with no nodes of out-degree 0 represents a finite-state Markov chain. The stochastic matrix of G is built as follows: take the adjacency matrix and divide each value by the row sum of its row (see Fig. 1 below). Recall some results on finite-state Markov chains (see, e.g., [11, Ch. 11]). Let A be the $m \times m$ matrix of the chain. The process is characterized by the vectors $\mathbf{p}^{(n)} = (p_1^{(n)}, \dots, p_m^{(n)})$, where $p_i^{(n)}$ is the probability of being in state i after n steps; the initial distribution $\mathbf{p}^{(0)}$ is given as a part of description of the chain. The *stationary distribution* of A is a vector $\mathbf{p} = (p_1, \dots, p_m)$ such that $p_i \geq 0$ for all i , $\sum_{i=1}^m p_i = 1$ and $\mathbf{p}A = \mathbf{p}$. Every row-stochastic matrix has one or more stationary distributions; such a distribution is unique for the matrices obtained from strongly connected digraphs. The sequence $\{\mathbf{p}^{(n)}\}$ approaches some stationary distribution \mathbf{p} in the following sense:

² Note that it is not possible in general to find the roots of polynomials exactly.

(*) there exists an integer $h \geq 1$ such that $\mathbf{p} = \lim_{n \rightarrow \infty} \frac{\mathbf{p}^{(n)} + \mathbf{p}^{(n+1)} + \dots + \mathbf{p}^{(n+h-1)}}{h}$

(That is, the limit of the process is a length- h cycle and \mathbf{p} gives the average probabilities of states in this cycle. In practical cases usually $h = 1$ and thus $\mathbf{p} = \lim_{n \rightarrow \infty} \mathbf{p}^{(n)}$.)

Theorem 2. Let L be a factorial regular language, $\hat{\mathcal{A}}$ be a PDFA accepting $\text{re}(L)$. Suppose that $\hat{\mathcal{A}}$ has m states $1, \dots, m$ and $\mathbf{p} = (p_1, \dots, p_m)$ is the stationary distribution for a random walk in $\hat{\mathcal{A}}$, starting at the initial state. Then $\mu(L) = \sum_{i=1}^m p_i \log(\text{deg}^{-1}(i))$.

Example 2. Let $L \subset \{0, 1\}^*$ be the regular language consisting of all words having no factor 11. Its accepting PDFA, the corresponding matrices and entropy computations are presented in Fig. 1. Note that $\text{re}(L) = L$.

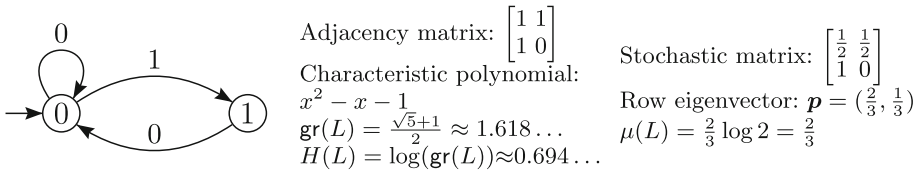


Fig. 1. Accepting PDFA and entropy computations for the language L (Example 2).

Computational Aspects. Computing $\hat{\mathcal{A}}$ from \mathcal{A} takes $O(|\mathcal{A}|)$ time, as it is sufficient to split \mathcal{A} into strongly connected components and traverse the acyclic graph of components. The vector \mathbf{p} can be computed by solving the size- m linear system $\mathbf{p}(\hat{\mathcal{A}} - I) = \mathbf{0}$, where $\hat{\mathcal{A}}$ is the adjacency matrix of $\hat{\mathcal{A}}$ and I is the identity matrix. This solution requires $\Theta(m^3)$ time and $\Theta(m^2)$ space, which is too much for large automata. More problems arise if the solution is not unique; but the correct vector \mathbf{p} still can be found by means of matrix theory (see [11, Ch. 11]). In order to process large automata (say, with millions of states), one can iteratively use the equality $\mathbf{p}^{(n+1)} = \mathbf{p}^{(n)}\hat{\mathcal{A}}$ to approximate \mathbf{p} with the desired precision. Each iteration can be performed in $O(m)$ time, because $\hat{\mathcal{A}}$ has $O(m)$ nonzero entries. One can prove, similar to [26, Th. 3.1], that under certain natural restrictions $O(\delta^{-1})$ iterations is sufficient to obtain \mathbf{p} within the approximation error δ .

4.2 Order- n Markov Entropy via Random Walks

Let $L \subseteq \Sigma^*$ be an arbitrary infinite factorial language such that the predicate $\mathcal{L}(w)$, which is true if $w \in L$ and false otherwise, is computable. There is little hope to compute $\mu(L)$, but one can use an oracle computing $\mathcal{L}(w)$ to build random walks in the prefix tree $\mathcal{T} = \mathcal{T}(L)$ and obtain statistical estimates of $\mu_n(L)$ for big n . We construct random walks by random depth-first search (Algorithm 1), executing the call $\text{DFS}(\lambda, n)$. The algorithm stops immediately

when level n is reached. When visiting node u , the algorithm chooses a non-visited child of u uniformly at random and visits it next. If all children of u are already visited, then u is a “dead end” (has no descendants at level n), and the search returns to the parent of u .

Algorithm 1. Random walk in $\mathcal{T}(L)$ by depth-first search

```

1: function DFS( $u, n$ )                                ▷  $u$ =node,  $n$ =length of walk
2: if  $|u| = n$  then break                               ▷ walk reached level  $n$ 
3:  $(a_1 a_2 \dots a_\sigma) \leftarrow$  random permutation of  $\Sigma$ 
4: for  $j = 1$  to  $\sigma$  do
5:   if  $\mathcal{L}(ua_j)$  then DFS( $ua_j, n$ )                 ▷ visit  $ua_j$  next
6: return                                               ▷  $u$  has no descendant at level  $n$ 

```

Lemma 7. DFS(λ, n) builds a length- n random standard walk in $\mathcal{T}(L)$.

Consider the values of the counter j in the instances DFS(λ, n), DFS($w[1], n$), ..., DFS($w[1..n-1], n$) at the moment when the search reaches level n . We define *profile* of the constructed walk as the vector $\mathbf{r} = (r_1, \dots, r_\sigma)$ such that r_i is the number of instances of DFS in which $j = i$. Note that different runs of Algorithm 1 may result in the same walk with different profiles (due to random choices made, depth-first search visits some dead ends and skips some of the others). Given a profile \mathbf{r} , one can compute the expected branching frequency $\text{bf}(\mathbf{r})$ of a walk with this profile: $\text{bf}(\mathbf{r}) = \frac{1}{n} \sum_{i=1}^\sigma c_i \log i$, where the parameters c_i are computed in Theorem 3 below.

Theorem 3. Let $\mathbf{r} = (r_1, \dots, r_\sigma)$ be a profile of a length- n random standard walk in a tree \mathcal{T} . For each $i = 1, \dots, \sigma$, let c_i be the expected number of nodes, having exactly i children in the tree $\mathcal{T}_{[n]}$, in a random standard walk with the profile \mathbf{r} . Then

$$(c_1, \dots, c_\sigma)P = \mathbf{r}, \text{ where } P[i, k] = \frac{\binom{\sigma-i}{k-1}}{\binom{\sigma}{k-1}} - \frac{\binom{\sigma-i}{k}}{\binom{\sigma}{k}} \text{ for } i, k = 1, \dots, \sigma. \quad (2)$$

Example 3. Let us solve (2) for $\sigma = 2$ (left) and $\sigma = 3$ (right):

$$\begin{aligned}
 (c_1, c_2) \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ 1 & 0 \end{bmatrix} &= (r_1, r_2) &
 (c_1, c_2, c_3) \begin{bmatrix} \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ \frac{2}{3} & \frac{1}{3} & 0 \\ 1 & 0 & 0 \end{bmatrix} &= (r_1, r_2, r_3) \\
 c_1 = 2r_2, c_2 = r_1 - r_2 & &
 c_1 = 3r_3, c_2 = 3r_2 - 3r_3, c_3 = r_1 - 2r_2 + r_3 & \\
 \text{bf}(\mathbf{r}) = \frac{r_1 - r_2}{r_1 + r_2} & &
 \text{bf}(\mathbf{r}) = \frac{3(r_2 - r_3) + (r_1 - 2r_2 + r_3) \log 3}{r_1 + r_2} &
 \end{aligned}$$

5 Experimental Results

With the goal of comparing general entropy and Markov entropy for power-free languages, we started with a side experiment. We took the ternary square-free language $\text{SF} = \text{PF}(3, 2)$, which is a well-studied test case. Its growth rate

$\text{gr}(\text{SF}) \approx 1.30176$ is known with high precision [27] from the study of its *regular approximations*. A k th regular approximation SF_k of SF is the language of all words having no squares of period $\leq k$ as factors. The sequence $\{\text{gr}(\text{SF}_k)\}$ demonstrates a fast convergence to $\text{gr}(\text{SF})$. So we wanted to (approximately) guess the Markov entropy $\mu(\text{SF})$ extrapolating the initial segment of the sequence $\mu(\text{SF}_k)$.

The results are as follows: we computed the values $\mu(\text{SF}_k)$ up to $k = 45$ with absolute error $\delta < 10^{-8}$ using the technique from Sect. 4.1. We obtained $\mu(\text{SF}_{45}) \approx 0.36981239$; the extrapolation of all obtained values gives $0.369810 < \mu(\text{SF}) < 0.369811$. At the same time we have $H(\text{SF}) = \log(\text{gr}(\text{SF})) \approx 0.380465$, so the two values are clearly distinct but close enough.

5.1 Random Walks in Power-Free Languages

To perform experiments with length- n random walks for a language L , one needs an algorithm to compute $\mathcal{L}(w)$ to be used with Algorithm 1. A standard approach is to maintain a data structure over the current word/walk w , which quickly answers the query “ $w \in L$?” and supports addition/deletion of a letter to/from the right. The theoretically best such algorithm for power-free words was designed by Kosolobov [14]: it spends $O(\log n)$ time per addition/deletion and uses memory of size $O(n)$. However, the algorithm is complicated and the constants under O are big. We developed a practical algorithm which is competitive for the walks up to the length of several millions. For simplicity, we describe it for square-free words but the construction is the same for any power-free language.

We use arrays $\text{repeat}_i[1..n]$, $i = 0, \dots, \lfloor \log n \rfloor - 1$ to store previous occurrences of factors. If $|u| \geq j$ for the current word u , then $\text{repeat}_i[j]$ is the *last* position of the previous occurrence of the factor $u[j-2^i+1..j]$ or $-\infty$ if there is no previous occurrence. To delete a letter from u we just delete the entries $\text{repeat}_i[|u|]$; let us consider the procedure $\text{add}(u, a)$ (Algorithm 2) which adds the letter a to u , checks square-freeness of ua and computes $\text{repeat}_i[|u| + 1]$. The auxiliary array $\text{last}[1..\sigma]$ stores the rightmost position of each letter in the current word.

Algorithm 2. Online square detection: adding a letter

```

1: function add( $u, a$ )                                     ▷  $u$ =word,  $a$ =letter to add
2:  $\text{repeat}_0[|u| + 1] \leftarrow \text{last}[a]$ ;  $\text{last}[a] \leftarrow |u| + 1$    ▷ fill previous occurrence of  $a$ 
3:  $\text{free} \leftarrow \text{true}$                                        ▷ square-freeness flag
4: for  $i = 0$  to  $\lfloor \log n \rfloor - 1$  do
5:    $x \leftarrow \text{repeat}_i[|u| + 1]$ ;  $p = |u| + 1 - x$            ▷  $p$  is the possible period of a square
6:   if  $p \leq 2^{i+1}$  and  $\text{repeat}_i[x+2^i] = x + 2^i - p$  then  $\text{free} \leftarrow \text{false}$ ; break           ▷ Fig. 2
7:   if  $i = \lfloor \log n \rfloor - 1$  then break                       ▷ no more arrays to update
8:   compute  $\text{repeat}_{i+1}[|u| + 1]$                              ▷ from  $\text{repeat}_i$ 
9:   if  $\text{repeat}_{i+1}[|u| + 1] = -\infty$  then break             ▷ all repeated suffixes processed
10: return  $\text{free}$                                              ▷ the answer to “is  $ua$  square-free?”

```

Correctness. Recall that u is square-free, so the occurrences of a factor of u can neither overlap nor touch. Assume that ua ends with a square vv , $p = |v|$, $2^i < p \leq 2^{i+1}$. Then p will be found in line 5 as $|u| + 1 - \text{repeat}_i[|u|+1]$ (red arcs in Fig. 2 show the suffix of length 2^i and its previous occurrence). The condition in line 6 means exactly the equality of words marked by dash arcs in Fig. 2; thus, vv is detected and $\text{add}(u, a)$ returns false. For the other direction, if $\text{add}(u, a) = \text{false}$, then the condition in line 6 held true and thus a square was detected as in Fig. 2. The *time complexity* is $O(\sqrt{n})$ on expectation; the experiments confirm this estimate.

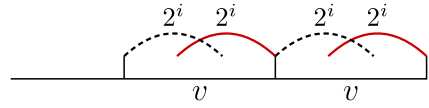


Fig. 2. Detecting a square by Algorithm 2.

Experiments. We studied the following languages: PF(2, 3) and PF(3, 2) as typical “test cases”, threshold languages over 3, . . . , 10, 20, 50, and 100 letters, and PF(2, $\frac{7}{3}^+$) as the smallest binary language of exponential growth. All languages from this list have “essentially binary” prefix trees: a letter cannot coincide with any of $(\sigma - 2)$ preceding letters, and so a node of level at least $\sigma - 2$ has at most two children. Hence we computed expected branching frequencies as in Example 3. For each language, we computed profiles of 1000 walks of length 10^5 and 100 walks of length 10^6 . The tables with the data are available at [29]. We briefly analysed the data. The most interesting findings, summarized below, are the same for each of the studied languages. Some figures are presented in Table 1.

1. The profiles of all walks in $\mathcal{T} = \mathcal{T}(L)$ are close to each other. To be precise, assume that $\text{bf}(\mathbf{r}) = \mu_n(L)$ for all constructed profiles. Then the number r_2 computed for a length- n random walk is the number of heads in c_1 tosses of a fair coin (among c_1 nodes with two children, in r_2 cases the dead end was chosen first). Hence the computed values of r_2 form a sample from the binomial distribution $B(c_1, \frac{1}{2})$. And indeed, the set of computed r_2 's looks indistinguishable from such a sample; see [29, stat100000]. This property suggests the mean value of $\text{bf}(\mathbf{r})$ over all experiments as a good approximation of $\mu_n(L)$.
2. The 99% confidence interval for the mean branching frequency $\text{bf}(\mathcal{T}_{[|w|]}, w)$ of the 1000 constructed walks of length 10^5 is of length $\sim 4 \cdot 10^{-4}$ and includes the mean value of $\text{bf}(\mathbf{r})$ for the walks of length 10^6 . For the language SF, this interval also includes the value $\mu(\text{SF})$ conjectured from the study of $\mu(\text{SF}_k)$. This property suggests that $\mu_n(L)$ for such big n is close to the Markov entropy $\mu(L)$.
3. As $\mu(L) \leq H(L) = \log(\text{gr}(L))$, the value of $\mu(L)$ can be converted to the lower bound for the growth rate of L . The values $2^{\text{mean}(\text{bf}(\mathbf{r}))}$ from our experiments differ from the best known upper bounds for the studied languages [27, Tbl. A1–A3] by the margin of 0.004–0.018. Such a bound is quite good for all

cases where specialized methods [13, 25] do not work. The results for threshold languages support the Shur–Gorbunova conjecture [28] that the growth rates of these languages tend to the limit $\alpha \approx 1.242$ as the size of the alphabet approaches infinity.

Table 1. Markov entropy for power-free languages: experiments

Language	Mean $\text{bf}(r)$ (10^5)	Mean $\text{bf}(r)$ (10^6)	$2^{\mu(L)}$	$\text{gr}(L)$
PF($2, \frac{7}{3}^+$)	0.27221	0.27220	≈ 1.20766	≈ 1.22064
PF(2, 3)	0.52562	0.52553	≈ 1.43956	≈ 1.45758
PF($3, \frac{7}{4}^+$)	0.30249	0.30251	≈ 1.23327	≈ 1.24561
PF(3, 2)	0.36988	0.36987	≈ 1.29223	≈ 1.30176
PF($4, \frac{7}{5}^+$)	0.09137	0.09151	≈ 1.06535	< 1.06951
PF($5, \frac{5}{4}^+$)	0.20279	0.20265	≈ 1.15092	< 1.15790
PF($6, \frac{6}{5}^+$)	0.28536	0.28526	≈ 1.21871	< 1.22470
PF($7, \frac{7}{6}^+$)	0.29753	0.29749	≈ 1.22903	< 1.23690
PF($8, \frac{8}{7}^+$)	0.28881	0.28867	≈ 1.22163	< 1.23484
PF($9, \frac{9}{8}^+$)	0.30716	0.30732	≈ 1.23727	< 1.24668
PF($10, \frac{10}{9}^+$)	0.29674	0.29669	≈ 1.22836	< 1.23931
PF($20, \frac{20}{19}^+$)	0.30002	0.29982	≈ 1.23099	< 1.24205
PF($50, \frac{50}{49}^+$)	0.30006	0.29970	≈ 1.23089	< 1.24210
PF($100, \frac{100}{99}^+$)	0.30047	0.29974	≈ 1.23093	< 1.24210

5.2 Random Walks in Abelian Power-Free Languages

Similar to Sect. 5.1, we need an algorithm checking Abelian power-freeness. Here we describe an algorithm detecting Abelian squares; its modification for other integer powers is straightforward. If a word $w[1..n]$ is fixed, we let $\psi_i = \psi(w[n-i+1..n]) - \psi(w[n-2i+1..n-i])$. A simple way to find whether w ends with Abelian square is to check $\psi_i = \mathbf{0}$ for all i . Since ψ_{i+1} can be obtained from ψ_i with a constant number of operations (add $w[n-i]$ twice, subtract $w[n-2i]$ and $w[n-2i-1]$), this check requires $\Theta(n)$ time. However, $\Theta(n)$ time per iteration appeared to be too much to perform experiments comparable with those for power-free languages, so we developed a faster algorithm. It maintains two length- n arrays for each letter $a \in \Sigma$: $d_a[i]$ is the position of i th from the left letter a in the current word w and $c_a[i]$ is the number of occurrences of a in $w[1..i]$ (i.e., a coordinate of $\psi(w[1..i])$). When a letter is added/deleted, these arrays are updated in $O(1)$ time (we regard σ as a constant). The function $\text{Asquare}(u)$ (Algorithm 3) checks whether the word w has an Abelian square as a suffix.

Correctness: see full version. *Complexity:* in experiments, Algorithm 3 checked $\Theta(\sqrt{n})$ suffixes of a length- n word, but we have no theoretical proof for this.

Algorithm 3. Online Abelian square detection

```

1: function Asquare( $u$ ) ▷  $u$ =word
2:  $l \leftarrow |u| - 1; i \leftarrow 1$  ▷ two counters
3:  $free \leftarrow false$  ▷ square-freeness flag; turns true when check finishes
4: while not free do
5:   for  $a \in \Sigma$  do
6:      $\psi[a] \leftarrow c_a[|u|] - c_a[|u|-i]$  ▷  $a$ -coordinate of  $\psi(u[|u|-i+1..|u|])$ 
7:     if  $\psi[a] > c_a[|u|-i]$  then  $free \leftarrow true$ ; break ▷ no squares possible
8:      $l \leftarrow \min\{l, d_a[c_a[|u|-i] - \psi[a] + 1]\}$ 
9:     if  $l = |u| - 2i + 1$  then break ▷  $u[|u| - 2i + 1..|u|]$  is an Abelian square
10:     $i \leftarrow \lceil (|u| - l + 1)/2 \rceil$ 
11: return free ▷ the answer to “is  $ua$  Abelian square-free?”

```

Experiments. The structure and growth of Abelian-power-free languages are little studied. We considered the 4-ary Abelian-square-free language ASF, the ternary Abelian-cube-free language ACF, and the binary Abelian-4-power-free language A4F; see Table 2. Our main interest was in estimating the actual growth rate of these languages. The upper (resp. lower) bounds for the growth rates are taken from [22] (resp., from [1, 2, 5]). For ASF and ACF we got profiles of 500 walks of length 10^5 and 100 walks of length $5 \cdot 10^5$; for A4F, 100 profiles of walks of length 10^5 were computed. The results suggest that the automata-based upper bounds for the growth rates of Abelian-power-free languages are quite imprecise, in contrast with the case of power-free languages. In addition, the experiments discovered the existence of very big finite subtrees on relatively low levels, which slow down the depth-first search. In fact, to obtain long enough words from A4F we modified the DFS function to allow “forced” backtracking if the length of the constructed word does not increase for a long time. Even with such a gadget, the time to build one walk of length 10^5 varied from 9 min to 4 h.

Table 2. Markov entropy for Abelian-power-free languages: experiments

Language	Mean $bf(r)$ (10^5)	Mean $bf(r)$ ($5 \cdot 10^5$)	$2^{\mu(L)}$	$gr(L)$
ASF	0.20475	0.20337	≈ 1.15138	$< 1.44435; > 1.00002$
ACF	1.08439	1.08418	≈ 2.12017	$< 2.37124; > 1.02930$
A4F	0.20736	–	≈ 1.15457	$< 1.37417; > 1.04427$

6 Conclusion and Future Work

In this paper we showed that efficient sampling of very long random words is a useful tool in the study of factorial languages. Already the first experiments allowed us to state a lot of problems for further research. To mention just a few:

- for which classes of languages, apart from regular ones, the Markov entropy can be computed (or approximated with a given error)?

- are there natural classes of languages satisfying $\mu(L) = H(L)$? $\mu(L) \ll H(L)$?
- how the branching frequencies of walks in a prefix tree are distributed? which statistical tests can help to approximate this distribution?

Concerning the last questions, we note that though our experiments showed “uniformity” of branching frequencies in each of the studied languages, the frequencies of individual words can vary significantly. For example, the language $\text{PF}(2, 3)$ with the average frequency about 0.525 contains infinite words \mathbf{u} and \mathbf{v} satisfying $\text{bf}(\mathbf{u}) = 0.72$ and $\text{bf}(\mathbf{v}) < 0.45$ [19].

References

1. Aberkane, A., Currie, J.D., Rampersad, N.: The number of ternary words avoiding abelian cubes grows exponentially. *J. Integer Seq.* **7**(2) (2004)
2. Carpi, A.: On the number of Abelian square-free words on four letters. *Discrete Appl. Math.* **81**, 155–167 (1998)
3. Carpi, A.: On Dejean’s conjecture over large alphabets. *Theor. Comput. Sci.* **385**, 137–151 (1999)
4. Crochemore, M., Mignosi, F., Restivo, A., Salemi, S.: Data compression using antidictionaries. *Proc. IEEE* **88**(11), 1756–1768 (2000)
5. Currie, J.D.: The number of binary words avoiding abelian fourth powers grows exponentially. *Theor. Comput. Sci.* **319**, 441–446 (2004)
6. Currie, J.D., Rampersad, N.: A proof of Dejean’s conjecture. *Math. Comp.* **80**, 1063–1070 (2011)
7. Dejean, F.: Sur un théorème de Thue. *J. Combin. Theor. Ser. A* **13**, 90–99 (1972)
8. Dekking, F.M.: Strongly non-repetitive sequences and progression-free sets. *J. Combin. Theory. Ser. A* **27**, 181–185 (1979)
9. Fekete, M.: Über der Verteilung der Wurzeln bei gewissen algebraischen Gleichungen mit ganzzahligen Koeffizienten. *Math. Zeitschrift* **17**, 228–249 (1923)
10. Furstenberg, H.: Disjointness in ergodic theory, minimal sets, and a problem in diophantine approximation. *Math. Syst. Theor.* **1**, 1–49 (1967)
11. Grinstead, C.M., Snell, J.L.: *Introduction to Probability*. American Mathematical Society, Providence (1997)
12. Keränen, V.: Abelian squares are avoidable on 4 letters. In: Kuich, W. (ed.) *ICALP 1992*. LNCS, vol. 623, pp. 41–52. Springer, Heidelberg (1992). <https://doi.org/10.1007/3-540-55719-9.62>
13. Kolpakov, R., Rao, M.: On the number of Dejean words over alphabets of 5, 6, 7, 8, 9 and 10 letters. *Theor. Comput. Sci.* **412**, 6507–6516 (2011)
14. Kosolobov, D.: Online detection of repetitions with backtracking. In: Cicalese, F., Porat, E., Vaccaro, U. (eds.) *CPM 2015*. LNCS, vol. 9133, pp. 295–306. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-19929-0_25
15. Lyons, R., Peres, Y.: *Probability on Trees and Networks*, Cambridge Series in Statistical and Probabilistic Mathematics, vol. 42. Cambridge University Press, New York (2016)
16. Mohammad-Noori, M., Currie, J.D.: Dejean’s conjecture and Sturmian words. *Eur. J. Comb.* **28**, 876–890 (2007)
17. Moulin-Ollagnier, J.: Proof of Dejean’s conjecture for alphabets with 5, 6, 7, 8, 9, 10 and 11 letters. *Theor. Comput. Sci.* **95**, 187–205 (1992)

18. Pansiot, J.J.: A propos d'une conjecture de F. Dejean sur les répétitions dans les mots. *Discrete Appl. Math.* **7**, 297–311 (1984)
19. Petrova, E.A., Shur, A.M.: Branching densities of cube-free and square-free words. *Algorithms* **14**(4), 126 (2021)
20. Rao, M.: Last cases of Dejean's conjecture. *Theor. Comput. Sci.* **412**, 3010–3018 (2011)
21. Rissanen, J.J.: Generalized Kraft inequality and arithmetic coding. *IBM J. Res. Dev.* **20**, 198–203 (1976)
22. Samsonov, A.V., Shur, A.M.: On Abelian repetition threshold. *RAIRO Theor. Inf. Appl.* **46**, 147–163 (2012)
23. Shannon, C.E.: A mathematical theory of communication. *Bell Syst. Tech. J.* **27**(379–423), 623–656 (1948)
24. Shur, A.M.: Combinatorial complexity of regular languages. In: Hirsch, E.A., Razborov, A.A., Semenov, A., Slissenko, A. (eds.) *CSR 2008. LNCS*, vol. 5010, pp. 289–301. Springer, Heidelberg (2008). https://doi.org/10.1007/978-3-540-79709-8_30
25. Shur, A.M.: Two-sided bounds for the growth rates of power-free languages. In: Diekert, V., Nowotka, D. (eds.) *DLT 2009. LNCS*, vol. 5583, pp. 466–477. Springer, Heidelberg (2009). https://doi.org/10.1007/978-3-642-02737-6_38
26. Shur, A.M.: Growth rates of complexity of power-free languages. *Theor. Comput. Sci.* **411**, 3209–3223 (2010)
27. Shur, A.M.: Growth properties of power-free languages. *Comput. Sci. Rev.* **6**, 187–208 (2012)
28. Shur, A.M., Gorbunova, I.A.: On the growth rates of complexity of threshold languages. *RAIRO Theor. Inf. Appl.* **44**, 175–192 (2010)
29. Markov entropy of repetition-free languages—statistics (2021). <https://tinyurl.com/2j36b6j7>