Theoretical Computer Science 412 (2011) 6507–6516

Contents lists available at SciVerse ScienceDirect

Theoretical Computer Science

journal homepage: www.elsevier.com/locate/tcs

# On the number of Dejean words over alphabets of 5, 6, 7, 8, 9 and 10 letters

Roman Kolpakov [a,*], Michaël Rao [b]

[a] Moscow State University, Leninskie Gory, 119992 Moscow, Russia

[b] LaBRI, Université Bordeaux 1, 351 cours de la libération, 33405 Talence, France

ARTICLE INFO

*Article history:*
Received 24 January 2010
Received in revised form 10 January 2011
Accepted 2 August 2011
Communicated by J. Karhumaki

*Keywords:*
Combinatorics on words
Repetitions
Growth rates

ABSTRACT

We give lower bounds on the growth rate of Dejean words, *i.e.* minimally repetitive words, over a $k$-letter alphabet, for $5 \leq k \leq 10$. Put together with the known upper bounds, we estimate these growth rates with the precision of 0.005. As a consequence, we establish the exponential growth of the number of Dejean words over a $k$-letter alphabet, for $5 \leq k \leq 10$.



## 1. Introduction

Let $w = a_1 \cdots a_n$ be a word over an alphabet $\Sigma$. The number $n$ is called the *length* of $w$ and is denoted by $|w|$. The symbol $a_i$ of $w$ is denoted by $w[i]$. A word $a_i \cdots a_j$, where $1 \leq i \leq j \leq n$, is called a *factor* of $w$ and is denoted by $w[i:j]$. For any $i = 1, \ldots, n$ the factor $w[1:i]$ ($w[i:n]$) is called a *prefix* (a *suffix*) of $w$. A positive integer $p$ is called a *period* of $w$ if $a_i = a_{i+p}$ for each $i = 1, \ldots, n-p$. If $p$ is the minimal period of $w$, the ratio $e(w) = n/p$ is called the *exponent* of $w$. Two words $w'$, $w''$ over $\Sigma$ are called *isomorphic* if $|w'| = |w''|$ and there exists a bijection $\sigma : \Sigma \longrightarrow \Sigma$ such that $w''[i] = \sigma(w'[i])$, $i = 1, \ldots, |w'|$. By $\mathcal{K}(w)$, we will denote the set of all words over $\Sigma$ which are isomorphic to the word $w$. We also denote by $|A|$ the number of elements of a finite set $A$. Let $|\Sigma| = k$. It is easy to note that $|\mathcal{K}(w)| = k!$ if $w$ contains at least $k-1$ different symbols of $\Sigma$.

Let $W$ be an arbitrary set of words. This set is called *factorial* if for any word $w$ from $W$ all factors of $w$ are also contained in $W$. We denote by $W(n)$ the subset of $W$ consisting of all words of length $n$. If $W$ is a factorial then it is not difficult to show (see, e.g., [3,1]) that there exists the limit $\lim_{n\to\infty} \sqrt[n]{|W(n)|}$ which is called the *growth rate* of words from $W$. For any words $u$, $v$ we denote by $W^{(v)}(n)$ the set of all words from $W(n)$ which contain $v$ as a suffix, and by $W^{(u,v)}(n)$ the set of all words from $W(n)$ which contain $v$ as a suffix and $u$ as a prefix.

One can mean by a repetition any word of exponent greater than 1. The best known example of repetitions is a *square*; that is, a word of the form $uu$, where $u$ is an arbitrary nonempty word. Avoiding ambiguity,[1] by the *period* of the square $uu$ we mean the length of $u$. In an analogous way, a *cube* is a word of the form $uuu$ for a nonempty word $u$, and the *period* of this cube is also the length of $u$. A word is called *square-free* (*cube-free*) if it contains no squares (cubes) as factors. It is easy to see that there are no binary square-free words of length larger than 3. On the other hand, by the classical results of Thue [20,21], there exist ternary square-free words of arbitrary length and binary cube-free words of arbitrary length. For ternary

* Corresponding author. Tel.: +7 495 939 42 68; fax: +7 495 932 89 52.
*E-mail addresses:* foroman@mail.ru (R. Kolpakov), rao@labri.fr (M. Rao).

1 Note that the period of a square is not necessarily the minimal period of this word.

0304-3975/$ – see front matter 
doi:10.1016/j.tcs.2011.08.006

square-free words this result was strengthened by Dejean in [9]. She found ternary words of arbitrary length which have no factors with exponents greater than 7/4. On the other hand, she showed that any long enough ternary word contains a factor with an exponent greater than or equal to 7/4. Thus, the number 7/4 is the minimal limit for exponents of avoidable factors which is universally called *the repetition threshold* in arbitrarily long ternary words. Dejean conjectured also that the repetition threshold in arbitrarily long words over a $k$-letter alphabet is equal to 7/5 for $k = 4$ and $k/(k-1)$ for $k \geq 5$. This conjecture is now proved for any $k$ through the work of several authors [5–8,13,12,15,16].

Denote the repetition threshold in arbitrarily long words over a $k$-letter alphabet by $\varphi_k$. In the paper we will call the words having no factors with exponents greater than $\varphi_k$ *minimally repetitive* words or *Dejean words*. By $S^{\langle k\rangle}(n)$ we denote the number of all minimally repetitive words of length $n$ over a $k$-letter alphabet. Note that the set of all minimally repetitive words is obviously factorial. So for any $k$ there exists the growth rate $\gamma^{\langle k\rangle} = \lim_{n\to\infty} \sqrt[n]{S^{\langle k\rangle}(n)}$.

The problem of estimating the number of repetition-free words has been investigated actively during the last decades (reviews of results on the estimations for the number of repetition-free words obtained before 2008 can be found in [2,10]). The most progress in this field has been made for the case of the binary alphabet. In this case Dejean words reduce to overlap-free words which are also a classical object for combinatorial investigations. It is proved in [17] that the growth of the number of binary overlap-free words is polynomial. Actually, binary overlap-free words of each length are counted by a 2-regular function [4].

In [11] we proposed a new approach for obtaining lower bounds on the number of repetition-free words. Using this approach, we obtained precise lower bounds for the growth rates of ternary square-free words, binary cube-free words, and ternary minimally repetitive words. This approach proved to be very effective. In particular, in [19] Shur proposed an interesting modification of our approach which allows to compute more effectively lower bounds for the growth rates of words which contain no repetitions of exponent greater than or equal to a given bound if this bound is not less than 2. The direction of our further investigations in this field is testing the proposed approach for "extreme" cases when the prohibitions imposed on words are maximal possible for the existence of words of arbitrary length avoiding these prohibitions. These cases are obviously the most difficult for obtaining lower bounds on the number of appropriate words. The case of minimally repetitive words is a natural example of such "extreme" cases. Moreover, the general case of minimally repetitive words over a $k$-letter alphabet for $k \geq 5$ when $\varphi_k = k/(k-1)$ is the most interesting for us. So this paper is devoted to obtaining lower bounds on $\gamma^{\langle k\rangle}$ for $k \geq 5$ by using the proposed approach. Note that the method proposed in [11] is not directly applicable to resolving this problem because of the huge size of required computer computations. In this paper we propose an improvement of this method which requires significantly fewer computer computations. Using this improvement, we obtain lower bounds on $\gamma^{\langle k\rangle}$ for $5 \leq k \leq 10$ which have the precision of 0.005. As an evident consequence of these results, we establish the exponential growth of the number of minimally repetitive words over a $k$-letter alphabet for $5 \leq k \leq 10$ (for $k = 3, 4$ this fact was proved by Ochem in [14]).

## 2. Estimation for the number of minimally repetitive words

### 2.1. General

For obtaining a lower bound on $\gamma^{\langle k\rangle}$ we will consider the alphabet $\Sigma_k = \{a_1, a_2, \ldots, a_k\}$ where $k \geq 5$. We denote the set of all minimally repetitive words over $\Sigma_k$ by $\mathcal{F}$. By a prohibited factor we mean a factor with an exponent greater than $k/(k-1)$. Let $m$ be a natural number, $m > k$, and $w'$, $w''$ be two words from $\mathcal{F}(m)$. We call the word $w''$ a *descendant* of the word $w'$ if $w'[2:m] = w''[1:m-1]$ and $w'w''[m] = w'[1]w'' \in \mathcal{F}(m+1)$. The word $w'$ is called in this case an *ancestor* of the word $w''$. We introduce a notion of closed words in the following inductive way. A word $w$ from $\mathcal{F}(m)$ is called *right closed* (*left closed*) if and only if this word satisfies one of the two following conditions:

(a) **Basis of induction.** $w$ has no descendants (ancestors);

(b) **Inductive step.** All descendants (ancestors) of $w$ are right closed (left closed).

A word is *closed* if it is either right closed or left closed. We denote by $\hat{\mathcal{F}}(m)$ the set of all words from $\mathcal{F}(m)$ which are not closed. By $\mathcal{L}_m$ we denote the set of all words over $\Sigma_k$ such that the length of these words is not less than $m$ and all factors of length $m$ in these words belong to $\hat{\mathcal{F}}(m)$. We also denote by $\mathcal{F}_m$ the set of all minimally repetitive words from $\mathcal{L}_m$. Note that a word $w$ is closed if and only if any word isomorphic to $w$ is also closed. So we have the following obvious fact.

**Proposition 1.** *For any isomorphic words $w'$, $w''$ and any $n \geq |w'|$ the equality $|\mathcal{F}_m^{(w')}(n)| = |\mathcal{F}_m^{(w'')}(n)|$ holds.*

A word will be called *rarefied* if the distance between any two different occurrences of the same symbol in this word is not less than $k-1$.

**Proposition 2.** *Any word from $\mathcal{L}_m$ is rarefied.*

**Proof.** Let $w$ be an arbitrary word from $\mathcal{L}_m$. Assume that $w[i] = w[j]$ where $j < i \leq j + (k-2)$. Consider the factor $f = w[j:i]$. Since $|f| = i - j + 1 \leq k - 1 < m$, in $w$ the factor $f$ is contained in some factor $f'$ of length $m$. By the definition of $\mathcal{L}_m$ we have $f' \in \mathcal{F}(m)$, so $f \in \mathcal{F}$. On the other hand, $f$ has the period $|f| - 1$, so

$$e(f) \geq \frac{|f|}{|f|-1} = \frac{i-j+1}{i-j} \geq \frac{k-1}{k-2} > \frac{k}{k-1}$$

which contradicts the definition of $\mathcal{F}(m)$. □

A word $w$ of length $n \geq k-1$ over $\Sigma_k$ will be called *trimmed* if $w[n-(k-1)+j] = a_j$ for $j = 1, \ldots, k-1$. We denote by $\hat{\mathcal{F}}'(m)$ the set of all trimmed words from $\hat{\mathcal{F}}(m)$. Taking into account Proposition 2, it is not difficult to note that for any word from $\hat{\mathcal{F}}(m)$ there exists a single word from $\hat{\mathcal{F}}'(m)$ which is isomorphic to this word, and for any word from $\hat{\mathcal{F}}'(m)$ there exist exactly $k!$ different words from $\hat{\mathcal{F}}(m)$ which are isomorphic to this word. Thus $|\hat{\mathcal{F}}(m)| = k!|\hat{\mathcal{F}}'(m)|$. Let $w'$, $w''$ be two words from $\hat{\mathcal{F}}'(m)$. We call the word $w''$ a *quasi-descendant* of the word $w'$ if $w''$ is isomorphic to some descendant of $w'$. The word $w'$ is called in this case a *quasi-ancestor* of the word $w''$.

Let $\hat{s} = |\hat{\mathcal{F}}(m)|$ and $s = |\hat{\mathcal{F}}'(m)|$. Without loss of generality we can assume that $\hat{\mathcal{F}}(m) = \{w_1, w_2, \ldots, w_{\hat{s}}\}$ where $\hat{\mathcal{F}}'(m) = \{w_1, w_2, \ldots, w_s\}$. For any word $w$ from $\hat{\mathcal{F}}(m)$ we will denote by $\iota(w)$ the serial number of $w$ in $\hat{\mathcal{F}}(m)$, i.e. $\iota(w) = i$ if $w = w_i$ for some $i = 1, 2, \ldots, \hat{s}$. We define a matrix $\hat{\Delta}_m = (\hat{\delta}_{ij})$ of size $\hat{s} \times \hat{s}$ in the following way: $\hat{\delta}_{ij} = 1$ if and only if $w_i$ is an ancestor of $w_j$; otherwise $\hat{\delta}_{ij} = 0$. For any natural $t$ by $\hat{\Delta}_m^{(t)} = (\hat{\delta}_{ij}^{(t)})$ we will denote the $t$-th power of the matrix $\hat{\Delta}_m$, i.e.

$$\hat{\Delta}_m^{(t)} = \underbrace{\hat{\Delta}_m \times \hat{\Delta}_m \times \cdots \times \hat{\Delta}_m}_{t}.$$

Further we use the following evident fact.

**Proposition 3.** *For any $i, j = 1, 2, \ldots, \hat{s}$ and any $n > m$ the equality $|\mathcal{L}_m^{(w_i,w_j)}(n)| = \hat{\delta}_{ij}^{(n-m)}$ is valid.*

We also define a matrix $\Delta_m = (\delta_{ij})$ of size $s \times s$ in the following way: $\delta_{ij} = 1$ if and only if $w_i$ is a quasi-ancestor of $w_j$; otherwise $\delta_{ij} = 0$. Note that $\Delta_m$ is a nonnegative matrix, so, by the Perron–Frobenius theorem, for $\Delta_m$ there exists some maximal in modulus eigenvalue $r$ which is a nonnegative real number. Moreover, we can find some eigenvector $\tilde{x} = (x_1; \ldots; x_s)$ with nonnegative components which corresponds to $r$. Assume that $r > 1$ and all components of $\tilde{x}$ are positive. Then we denote by $\mu$ the ratio $\max_{i=1,\ldots,s} x_i / \min_{i=1,\ldots,s} x_i$, and for $n \geq m$ we define $S_m^{\langle k \rangle}(n) = \sum_{i=1}^{s} x_i \cdot |\mathcal{F}_m^{(w_i)}(n)|$. In an inductive way we estimate $S_m^{\langle k \rangle}(n+1)$ by $S_m^{\langle k \rangle}(n)$.

First we estimate $|\mathcal{F}_m^{(w)}(n+1)|$ for each $w \in \hat{\mathcal{F}}(m)$. It is obvious that

$$|\mathcal{F}_m^{(w)}(n+1)| = |\mathcal{G}^{(w)}(n+1)| - |\mathcal{H}^{(w)}(n+1)|, \tag{1}$$

where $\mathcal{G}^{(w)}(n+1)$ is the set of all words $v$ from $\mathcal{L}_m^{(w)}(n+1)$ such that $v[1:n], v[n-m+1:n+1] \in \mathcal{F}$, and $\mathcal{H}^{(w)}(n+1)$ is the set of all words from $\mathcal{G}^{(w)}(n+1)$ which contain some prohibited factor as a suffix. If $w \in \hat{\mathcal{F}}'(m)$ we denote by $\pi(w)$ the set of all quasi-ancestors of $w$. Taking into account Proposition 1, it is easy to see that

$$|\mathcal{G}^{(w)}(n+1)| = \sum_{v \in \pi(w)} |\mathcal{F}_m^{(v)}(n)|. \tag{2}$$

Therefore, using that $\tilde{x}$ is a eigenvector of $\Delta_m$ for the eigenvalue $r$, we obtain

$$\begin{aligned}
\sum_{i=1}^{s} x_i \cdot |\mathcal{G}^{(w_i)}(n+1)| &= \sum_{i=1}^{s} \Big( x_i \cdot \sum_{v \in \pi(w_i)} |\mathcal{F}_m^{(v)}(n)| \Big) \\
&= (x_1; x_2; \ldots; x_s) \begin{pmatrix} \delta_{11} & \delta_{21} & \ldots & \delta_{s1} \\ \delta_{12} & \delta_{22} & \ldots & \delta_{s2} \\ \vdots & \vdots & \ddots & \vdots \\ \delta_{1s} & \delta_{2s} & \ldots & \delta_{ss} \end{pmatrix} \begin{pmatrix} |\mathcal{F}_m^{(w_1)}(n)| \\ |\mathcal{F}_m^{(w_2)}(n)| \\ \vdots \\ |\mathcal{F}_m^{(w_s)}(n)| \end{pmatrix} \\
&= r \cdot (x_1; x_2; \ldots; x_s) \begin{pmatrix} |\mathcal{F}_m^{(w_1)}(n)| \\ |\mathcal{F}_m^{(w_2)}(n)| \\ \vdots \\ |\mathcal{F}_m^{(w_s)}(n)| \end{pmatrix} = r \cdot S_m^{\langle k \rangle}(n).
\end{aligned} \tag{3}$$

We now estimate $|\mathcal{H}^{(w)}(n+1)|$. For any word $v$ from $\mathcal{H}^{(w)}(n+1)$ we can find the minimal prohibited factor which is a suffix of $v$. We denote this factor by $h(v)$ and the minimal period of this factor by $\lambda(v)$. Since after removing the last symbol from $h(v)$ this factor cannot be prohibited, we have actually $|h(v)| = \lfloor k\lambda(v)/(k-1) \rfloor + 1$. Note that the value $\lambda(v)$ is not less than $p_0 = (m+1) - \lfloor (m+1)/k \rfloor$. Thus

$$|\mathcal{H}^{(w)}(n+1)| = \sum_{j \geq p_0} |\mathcal{H}_j^{(w)}(n+1)| \tag{4}$$

where $\mathcal{H}_j^{(w)}(n+1)$ is the set of all words $v$ from $\mathcal{H}^{(w)}(n+1)$ such that $\lambda(v) = j$.

### 2.2. Upper bound for $|\mathcal{H}_j^{(w)}(n+1)|$

To estimate $|\mathcal{H}_j^{(w)}(n+1)|$, let $\chi(j) = \lfloor j/(k-1) \rfloor + 1$ and let $t = j + \chi(j) + 1$. Recall that for any $v$ from $\mathcal{H}_j^{(w)}(n+1)$ the prohibited factor $h(v)$ is a word from $\mathcal{L}_m(j + \chi(j))$ with the minimal period $j$. Moreover, this word does not contain shorter prohibited factors and contains the word $w$ as a suffix.

Let $X_{j,t}^{(w)}$ be the set of words $v \in \mathcal{L}_m(t)$ such that $v[1 : t-1] \in \mathcal{F}(t-1)$, $v[3 : t] \in \mathcal{F}(t-2)$, $v[t-j-\chi(j)+1 : t-j] = v[t-\chi(j)+1 : t]$ and $w$ is a suffix of $v$. Note that for every $v \in X_{j,t}^{(w)}$, $v[t-j-\chi(j)] \neq v[t-\chi(j)]$, otherwise $v[1 : t-1]$ would have a forbidden factor. Suppose that $n+1 \geq t$ and let $u \in \mathcal{H}_j^{(w)}(n+1)$. Then

$$u[n_j'+1 : n+1] = u[n_j''+1 : n-j+1] \tag{5}$$

where $n_j' = n - \lfloor j/(k-1) \rfloor$ and $n_j'' = n - \lfloor kj/(k-1) \rfloor$. By definition of $\mathcal{H}_j^{(w)}(n+1)$, $u[n+1-t+1 : n+1] \in X_{j,t}^{(w)}$. Moreover $u[1 : n+1-t+m] \in \mathcal{F}(n+1-t+m)$. Thus we have

**Proposition 4.**

$$|\mathcal{H}_j^{(w)}(n+1)| \leq \sum_{v \in X_{j,t}^{(w)}} |\mathcal{F}^{(v[1:m])}(n+1-t+m)|.$$

Let $U_{j,t}^{(w)}$ be the multiset of all prefixes of size $m$ in words of $X_{j,t}^{(w)}$ (note that among words $U_{j,t}^{(w)}$ we can have identical words, i.e., the same word can be a prefix of different words of $X_{j,t}^{(w)}$ and so can be counted several times in $U_{j,t}^{(w)}$). Then Proposition 4 implies

$$|\mathcal{H}_j^{(w)}(n+1)| \leq \sum_{u \in U_{j,t}^{(w)}} |\mathcal{F}_m^{(u)}(n+1-t+m)|.$$

For $l = 1, \ldots, s$, denote by $\zeta_{j,t}^{(l)}(w)$ the number of occurrences of $w_l$ in the multiset $U_{j,t}^{(w)}$. Then

$$|\mathcal{H}_j^{(w)}(n+1)| \leq \sum_{u \in U_{j,t}^{(w)}} |\mathcal{F}_m^{(u)}(n+1-t+m)| = \sum_{l=1}^{s} \zeta_{j,t}^{(l)}(w) \cdot |\mathcal{F}_m^{(w_l)}(n+1-t+m)|. \tag{6}$$

### 2.3. *Weaker upper bound for* $|\mathcal{H}_j^{(w)}(n+1)|$

We can also obtain another estimation for $|\mathcal{H}_j^{(w)}(n+1)|$ where $w \in \hat{\mathcal{F}}'(m)$. This estimation is more rough in comparison with (6) but requires much fewer computer computations. To estimate $|\mathcal{H}_j^{(w)}(n+1)|$ by this way, we denote $\lfloor j/(k-1) \rfloor + 1$ by $\chi(j)$ and assume that $\chi(j) \geq k-1$ and $n \geq j+m$. Recall that for any $v$ from $\mathcal{H}_j^{(w)}(n+1)$ we have relation (5). We consider separately the two following cases: $\chi(j) \leq m$ and $\chi(j) > m$.

Let $\chi(j) \leq m$. For any $v$ from $\mathcal{H}_j^{(w)}(n+1)$ denote by $f'(v)$ the factor $v[n+2-j-m : n+1-j]$ of $v$. It follows from $v \in \mathcal{L}_m$ that $f'(v) \in \hat{\mathcal{F}}(m)$. Moreover, from (5) we obtain that $f'(v)$ and $w$ have the common suffix of length $\chi(j)$. Since $w \in \hat{\mathcal{F}}'(m)$ and $\chi(j) \geq k-1$, it implies that $f'(v) \in \hat{\mathcal{F}}'(m)$. Thus

$$|\mathcal{H}_j^{(w)}(n+1)| = \sum_{u \in W_j(w)} |\mathcal{I}_{j,u}^{(w)}(n+1)|$$

where $W_j(w)$ is the set of all words from $\hat{\mathcal{F}}'(m)$ which have the common suffix of length $\chi(j)$ with the word $w$, and $\mathcal{I}_{j,u}^{(w)}(n+1)$ is the set of all words $v$ from $\mathcal{H}_j^{(w)}(n+1)$ such that $f'(v) = u$. To estimate $|\mathcal{I}_{j,u}^{(w)}(n+1)|$, note that for any $v$ from $\mathcal{I}_{j,u}^{(w)}(n+1)$ we have $v[1 : n+1-j] \in \mathcal{F}_m^{(u)}(n+1-j)$ and $v[n+2-j-m : n+1] \in \mathcal{L}_m^{(u,w)}(j+m)$. Hence, using Proposition 3, we obtain

$$|\mathcal{I}_{j,u}^{(w)}(n+1)| \leq |\mathcal{F}_m^{(u)}(n+1-j)| \cdot |\mathcal{L}_m^{(u,w)}(j+m)| = |\mathcal{F}_m^{(u)}(n+1-j)| \cdot \hat{\delta}_{\iota(u),\iota(w)}^{(j)}.$$

Thus, in this case we get the estimation

$$|\mathcal{H}_j^{(w)}(n+1)| \leq \sum_{u \in W_j(w)} \hat{\delta}_{\iota(u),\iota(w)}^{(j)} \cdot |\mathcal{F}_m^{(u)}(n+1-j)|. \tag{7}$$

Let now $\chi(j) > m$. For any $v$ from $\mathcal{H}_j^{(w)}(n+1)$ denote by $f''(v)$ the factor $v[n_j''+1 : n_j''+m]$ of $v$. It follows from $v \in \mathcal{L}_m$ that $f''(v) \in \hat{\mathcal{F}}(m)$. Thus in this case

$$|\mathcal{H}_j^{(w)}(n+1)| = \sum_{u \in \hat{\mathcal{F}}(m)} |\mathcal{J}_{j,u}^{(w)}(n+1)|$$

where $\mathcal{J}_{j,u}^{(w)}(n+1)$ is the set of all words $v$ from $\mathcal{H}_j^{(w)}(n+1)$ such that $f''(v) = u$. To estimate $|\mathcal{J}_{j,u}^{(w)}(n+1)|$, consider an arbitrary word $v$ from $\mathcal{J}_{j,u}^{(w)}(n+1)$. Note that $v$ is determined uniquely by the prefix $v[1 : n_j'+m]$ which satisfies the following conditions: $v[1 : n''_j + m] \in \mathcal{F}_m^{(u)}(n''_j + m)$, $v[n''_j + 1 : n + 1 - j] \in \mathcal{L}_m^{(u,w)}(\chi(j))$, and $v[n + 2 - j - m : n'_j + m] \in \mathcal{L}_m^{(w,u)}(j + 2m - \chi(j))$. Hence, using Proposition 3, we obtain

$$|\mathcal{J}_{j,u}^{(w)}(n+1)| \le |\mathcal{F}_m^{(u)}(n''_j + m)| \cdot |\mathcal{L}_m^{(u,w)}(\chi(j))| \cdot |\mathcal{L}_m^{(w,u)}(j + 2m - \chi(j))| = |\mathcal{F}_m^{(u)}(n''_j + m)| \cdot \hat{\delta}_{\iota(u),\iota(w)}^{(\chi(j)-m)} \cdot \hat{\delta}_{\iota(w),\iota(u)}^{(j+m-\chi(j))}.$$

Thus, in this case we get the estimation

$$|\mathcal{H}_j^{(w)}(n+1)| \le \sum_{u \in \hat{\mathcal{F}}(m)} \hat{\delta}_{\iota(u),\iota(w)}^{(\chi(j)-m)} \cdot \hat{\delta}_{\iota(w),\iota(u)}^{(j+m-\chi(j))} \cdot |\mathcal{F}_m^{(u)}(n''_j + m)|.$$

Taking into account Proposition 1, we can rewrite this estimation in the form

$$|\mathcal{H}_j^{(w)}(n+1)| \le \sum_{u \in \hat{\mathcal{F}}'(m)} |\mathcal{F}_m^{(u)}(n''_j + m)| \left( \sum_{v \in \mathcal{K}(u)} \hat{\delta}_{\iota(v),\iota(w)}^{(\chi(j)-m)} \cdot \hat{\delta}_{\iota(w),\iota(v)}^{(j+m-\chi(j))} \right). \quad (8)$$

Note that, unlike estimation (6), estimations (7) and (8) can be computed in polynomial time.

### *2.4. Estimation of* $|\mathcal{H}^{(w)}(n+1)|$

We fix numbers $p_1, p_2$ such that $p_0 \le p_1 < p_2$ and $p_2 \ge 2k - 3$, and assume for convenience that $n > kp_2/(k-1)$. We present sum (4) in the form

$$|\mathcal{H}^{(w)}(n+1)| = \sum_{j=p_0}^{p_1} |\mathcal{H}_j^{(w)}(n+1)| + \sum_{j=p_1+1}^{p_2} |\mathcal{H}_j^{(w)}(n+1)| + |\hat{\mathcal{H}}^{(w)}(n+1)|$$

where $\hat{\mathcal{H}}^{(w)}(n+1) = \bigcup_{j>p_2} \mathcal{H}_j^{(w)}(n+1)$. Thus $\sum_{i=1}^{s} x_i |\mathcal{H}^{(w_i)}(n+1)|$ can be presented as

$$\sum_{j=p_0}^{p_1} \sum_{i=1}^{s} x_i |\mathcal{H}_j^{(w_i)}(n+1)| + \sum_{j=p_1+1}^{p_2} \sum_{i=1}^{s} x_i |\mathcal{H}_j^{(w_i)}(n+1)| + \sum_{i=1}^{s} x_i |\hat{\mathcal{H}}^{(w_i)}(n+1)|. \quad (9)$$

To estimate the first sum in (9), we use inequality (6)

$$\sum_{j=p_0}^{p_1} \sum_{i=1}^{s} x_i |\mathcal{H}_j^{(w_i)}(n+1)| \le \sum_{j=p_0}^{p_1} \sum_{i=1}^{s} x_i \sum_{l=1}^{s} \zeta_{j,\left\lfloor \frac{kj}{k-1} \right\rfloor + 2}^{(l)}(w_i) \cdot |\mathcal{F}_m^{(w_l)}\left(n - \left\lfloor \frac{jk}{k-1} \right\rfloor - 1 + m\right) = \sum_{d=\left\lfloor \frac{kp_0}{k-1} \right\rfloor + 2}^{\left\lfloor \frac{kp_1}{k-1} \right\rfloor + 2} \sum_{l=1}^{s} \eta'_l(d) \cdot \left| \mathcal{F}_m^{(w_l)}\left(n - \left\lfloor \frac{jk}{k-1} \right\rfloor - 1 + m\right) \right| \quad (10)$$

where $\eta'_l(d) = \sum_{i=1}^{s} x_i \cdot \zeta_{j,\left\lfloor \frac{kj}{k-1} \right\rfloor + 2}^{(l)}(w_i)$ if there is a $j$ such that $\left\lfloor \frac{kj}{k-1} \right\rfloor + 2 = d$, and $\eta'_l(d) = 0$ otherwise.

To estimate the second sum in (9), we use inequalities (7) and (8). In particular, in the case of $\chi(j) \le m$, using inequality (7) and taking into account that $u \in W_j(w)$ if and only if $w \in W_j(u)$, we obtain

$$\sum_{i=1}^{s} x_i |\mathcal{H}_j^{(w_i)}(n+1)| \le \sum_{i=1}^{s} \sum_{u \in W_j(w_i)} x_i \hat{\delta}_{\iota(u),i}^{(j)} \cdot |\mathcal{F}_m^{(u)}(n+1-j)| = \sum_{u \in \hat{\mathcal{F}}'(m)} |\mathcal{F}_m^{(u)}(n+1-j)| \left( \sum_{w_i \in W_j(u)} x_i \cdot \hat{\delta}_{\iota(u),i}^{(j)} \right) = \sum_{l=1}^{s} |\mathcal{F}_m^{(w_l)}(n+1-j)| \left( \sum_{w_i \in W_j(w_l)} x_i \cdot \hat{\delta}_{l,i}^{(j)} \right).$$

In the case of $\chi(j) > m$, using inequality (8), we have

$$\sum_{i=1}^{s} x_i |\mathcal{H}_j^{(w_i)}(n+1)| \le \sum_{i=1}^{s} \sum_{u \in \hat{\mathcal{F}}'(m)} x_i |\mathcal{F}_m^{(u)}(n''_j + m)| \left( \sum_{v \in \mathcal{K}(u)} \theta_{i,v}^{(j)} \right) = \sum_{u \in \hat{\mathcal{F}}'(m)} |\mathcal{F}_m^{(u)}(n''_j + m)| \sum_{i=1}^{s} x_i \left( \sum_{v \in \mathcal{K}(u)} \theta_{i,v}^{(j)} \right) = \sum_{l=1}^{s} |\mathcal{F}_m^{(w_l)}(n''_j + m)| \sum_{i=1}^{s} x_i \left( \sum_{v \in \mathcal{K}(w_l)} \theta_{i,v}^{(j)} \right)$$

where $\theta^{(j)}_{i,v} = \hat{\delta}^{(\chi(j)-m)}_{\iota(v),i} \cdot \hat{\delta}^{(j+m-\chi(j))}_{i,\iota(v)}$. Thus, defining $d(j) = j - 1$ for the case of $\chi(j) \le m$ and $d(j) = \lfloor kj/(k-1) \rfloor - m$ for the case of $\chi(j) > m$, we conclude that

$$\sum_{i=1}^{s} x_i |\mathcal{H}^{(w_i)}_j(n+1)| \le \sum_{l=1}^{s} \xi_l(j) \cdot |\mathcal{F}^{(w_l)}_m(n-d(j))|$$

where

$$\xi_l(j) = \begin{cases} \displaystyle\sum_{w_i \in W_j(w_l)} x_i \cdot \hat{\delta}^{(j)}_{l,i}, & \text{if } \chi(j) \le m; \\ \displaystyle\sum_{i=1}^{s} x_i \left( \sum_{v \in \mathcal{K}(w_l)} \theta^{(j)}_{i,v} \right), & \text{if } \chi(j) > m. \end{cases}$$

Hence

$$\sum_{j=p_1+1}^{p_2} \sum_{i=1}^{s} x_i |\mathcal{H}^{(w_i)}_j(n+1)| \le \sum_{j=p_1+1}^{p_2} \sum_{l=1}^{s} \xi_l(j) \cdot |\mathcal{F}^{(w_l)}_m(n-d(j))|.$$

We define $\xi'_l(d) = \xi_l(j)$ if there exists some $j$ such that $d(j) = d$, and $\xi'_l(d) = 0$ otherwise. Then

$$\sum_{j=p_1+1}^{p_2} \sum_{l=1}^{s} \xi_l(j) \cdot |\mathcal{F}^{(w_l)}_m(n-d(j))| = \sum_{d=d_2}^{d_3} \sum_{l=1}^{s} \xi'_l(d) \cdot |\mathcal{F}^{(w_l)}_m(n-d)| \tag{11}$$

where $d_2 = d(p_1+1)$, $d_3 = d(p_2)$.

Summing up (10) and (11), we get

$$\sum_{j=p_0}^{p_2} \sum_{i=1}^{s} x_i |\mathcal{H}^{(w_i)}_j(n+1)| \le \sum_{d=a}^{b} \sum_{l=1}^{s} \omega_l(d) \cdot |\mathcal{F}^{(w_l)}_m(n-d)|$$

where $\omega_l(d) = \eta'_l(d) + \xi'_l(d)$, $a = \min(d_2, \left\lfloor \frac{kp_0}{k-1} \right\rfloor + 2 - m - 1)$ and $b = \max(d_3, \left\lfloor \frac{kp_1}{k-1} \right\rfloor + 2 - m - 1)$.

We majorate this sum by some sum $\sum_{d=a}^{b} \rho_d \cdot S^{\langle k \rangle}_m(n-d)$ in the following way. We compute consecutively coefficients $\rho_d$ of this sum for $d = a, a+1, \ldots, b$. For each $d = a, a+1, \ldots, b-1$ together with the number $\rho_d$ we compute also numbers $\omega'_1(d+1), \ldots, \omega'_s(d+1)$ such that

$$\sum_{j=a}^{d+1} \sum_{l=1}^{s} \omega_l(j) \cdot |\mathcal{F}^{(w_l)}_m(n-j)| \le \sum_{l=1}^{s} \omega'_l(d+1) \cdot |\mathcal{F}^{(w_l)}_m(n-d-1)| + \sum_{j=a}^{d} \rho_j \cdot S^{\langle k \rangle}_m(n-j). \tag{12}$$

For $d = a$ we take $\rho_a = \min_{1 \le l \le s}(\omega_l(a)/x_l)$. Then

$$\sum_{l=1}^{s} \omega_l(a) \cdot |\mathcal{F}^{(w_l)}_m(n-a)| = \rho_a \cdot S^{\langle k \rangle}_m(n-a) + \sum_{l=1}^{s} \nu_l \cdot |\mathcal{F}^{(w_l)}_m(n-a)|$$

where $\nu_l = \omega_l(a) - \rho_a \cdot x_l$, $l = 1, \ldots, s$. Denote by $\tilde{\nu}$ the vector $(\nu_1; \ldots; \nu_s)$ and consider the vector $\tilde{\nu}' = \Delta_m \tilde{\nu}$. Let $\tilde{\nu}' = (\nu'_1; \ldots; \nu'_s)$. It follows from (1) and (2) that

$$|\mathcal{F}^{(w_l)}_m(n-a)| \le |\mathcal{G}^{(w_l)}(n-a)| = \sum_{v \in \pi(w_l)} |\mathcal{F}^{(v)}_m(n-a-1)|$$

for any $l = 1, \ldots, s$. Note also that $\nu_l \ge 0$ for $l = 1, \ldots, s$. Hence

$$\begin{aligned} \sum_{l=1}^{s} \nu_l \cdot |\mathcal{F}^{(w_l)}_m(n-a)| &\le \sum_{l=1}^{s} \left( \nu_l \cdot \sum_{v \in \pi(w_l)} |\mathcal{F}^{(v)}_m(n-a-1)| \right) \\ &= \sum_{l=1}^{s} \nu'_l \cdot |\mathcal{F}^{(w_l)}_m(n-a-1)|. \end{aligned}$$

Thus

$$\sum_{j=a}^{a+1} \sum_{l=1}^{s} \omega_l(j) \cdot |\mathcal{F}^{(w_l)}_m(n-j)| \le \rho_a \cdot S^{\langle k \rangle}_m(n-a) + \sum_{l=1}^{s} \omega'_l(a+1) \cdot |\mathcal{F}^{(w_l)}_m(n-a-1)| \tag{13}$$

where $\omega'_l(a+1) = \omega_l(a+1) + \nu'_l$. Assume now that for some $d$ such that $a < d < b$ we already computed the numbers $\rho_a, \ldots, \rho_{d-1}$ and $\omega'_1(d), \ldots, \omega'_s(d)$. Then we take $\rho_d = \min_{1 \le l \le s}(\omega'_l(d)/x_l)$, $\tilde{\nu} = (\omega'_1(d) - \rho_d \cdot x_1, \ldots, \omega'_s(d) - \rho_d \cdot x_s)$, and

$\tilde{\nu}' = \Delta_m \tilde{\nu}$. We take also $\omega'_l(d+1) = \omega_l(d+1) + \nu'_l$ where $\nu'_l$ is the $l$-th component of the vector $\tilde{\nu}'$, $l = 1, \ldots, s$. Analogously to inequality (13), in this case we have the inequality

$$\sum_{l=1}^{s} \left(\omega'_l(d) \cdot |\mathcal{F}_m^{(w_l)}(n-d)| + \omega_l(d+1) \cdot |\mathcal{F}_m^{(w_l)}(n-d-1)|\right) \le \rho_d \cdot S_m^{\langle k\rangle}(n-d) + \sum_{l=1}^{s} \omega'_l(d+1) \cdot |\mathcal{F}_m^{(w_l)}(n-d-1)|.$$

This inequality implies that inequality (12) holds for every $d$. For $d = b$ we take $\rho_b = \max_{1\le l\le s}(\omega'_l(b)/x_l)$. Thus,

$$\sum_{d=a}^{b} \sum_{l=1}^{s} \omega_l(d) \cdot |\mathcal{F}_m^{(w_l)}(n-d)| \le \sum_{d=a}^{b} \rho_d \cdot S_m^{\langle k\rangle}(n-d)$$

which implies

$$\sum_{j=p_0}^{p_2} \sum_{i=1}^{s} x_i |\mathcal{H}_j^{(w_i)}(n+1)| \le \sum_{d=a}^{b} \rho_d \cdot S_m^{\langle k\rangle}(n-d). \tag{14}$$

### 2.5. Upper bound for $|\hat{\mathcal{H}}^{(w_i)}(n+1)|$

We estimate finally the sum $\sum_{i=1}^{s} x_i |\hat{\mathcal{H}}^{(w_i)}(n+1)|$. For this purpose we denote by $\hat{\mathcal{H}}(n+1)$ the set $\bigcup_{i=1}^{\hat{s}} \hat{\mathcal{H}}^{(w_i)}(n+1)$ and by $\hat{\mathcal{H}}'(n+1)$ the set $\bigcup_{i=1}^{s} \hat{\mathcal{H}}^{(w_i)}(n+1)$. Note that the sets $\hat{\mathcal{H}}^{(w_i)}(n+1)$ are non-overlapping, so $|\hat{\mathcal{H}}'(n+1)| = \sum_{i=1}^{s} |\hat{\mathcal{H}}^{(w_i)}(n+1)|$. Thus

$$\sum_{i=1}^{s} x_i |\hat{\mathcal{H}}^{(w_i)}(n+1)| \le |\hat{\mathcal{H}}'(n+1)| \cdot \max_{i=1,\ldots,s} x_i. \tag{15}$$

Moreover, since by Proposition 2 any word from $\hat{\mathcal{H}}(n+1)$ is rarified and $n+1 > k-1$, for any word from $\hat{\mathcal{H}}(n+1)$ there exists a single word from $\hat{\mathcal{H}}'(n+1)$ which is isomorphic to this word, and for any word from $\hat{\mathcal{H}}'(n+1)$ there exist exactly $k!$ different words from $\hat{\mathcal{H}}(n+1)$ which are isomorphic to this word. So $|\hat{\mathcal{H}}(n+1)| = k!|\hat{\mathcal{H}}'(n+1)|$.

Let $v$ be an arbitrary word from $\hat{\mathcal{H}}(n+1)$. Then for $v$ we have

$$v\left[n - \left\lfloor \frac{\lambda(v)}{k-1} \right\rfloor + 1 : n+1\right] = v\left[n' - \left\lfloor \frac{\lambda(v)}{k-1} \right\rfloor + 1 : n'+1\right]$$

where $n' = n - \lambda(v)$. Thus the word $v$ is determined uniquely by the number $\lambda(v)$ and the prefix $v[1 : n - \lfloor \lambda(v)/(k-1)\rfloor]$. We denote this prefix by $\tau(v)$. Further we use the following fact.

**Lemma 5.** *For any different $v', v'' \in \hat{\mathcal{H}}(n+1)$ the prefixes $\tau(v'), \tau(v'')$ are also different.*

**Proof.** Let $\tau(v') = \tau(v'') = u$ for some different $v', v'' \in \hat{\mathcal{H}}(n+1)$. Denote by $l$ the length of $u$. Note that $v', v'' \in \mathcal{L}_m$, so $v'$, $v''$ and $u$ are rarefied by Proposition 2. Thus without loss of generality we can assume that $u$ is trimmed, i.e.

$$a_j = u[l-(k-1)+j] = v'[l-(k-1)+j] = v''[l-(k-1)+j] \tag{16}$$

for $j = 1, \ldots, k-1$. As we noted above, the equalities $\tau(v') = \tau(v'')$ and $\lambda(v') = \lambda(v'')$ imply $v' = v''$. So $\lambda(v') \ne \lambda(v'')$. Without loss of generality we assume that $\lambda(v') > \lambda(v'')$. Since $n - l = \lfloor \lambda(v')/(k-1)\rfloor = \lfloor \lambda(v'')/(k-1)\rfloor$, we can assume moreover that $\lambda(v'') < \lambda(v') < \lambda(v'') + (k-1)$. Note also that the inequality $\lfloor \lambda(v')/(k-1)\rfloor \ge 2$ follows from $\lambda(v') \ge p_2 + 1 \ge 2k-2$. So $l = n - \lfloor \lambda(v')/(k-1)\rfloor \le n-2$. Recall that we have also

$$v'[l+1 : n+1] = v'[l-\lambda(v')+1 : n-\lambda(v')+1] = u[l-\lambda(v')+1 : n-\lambda(v')+1], \tag{17}$$
$$v''[l+1 : n+1] = v''[l-\lambda(v'')+1 : n-\lambda(v'')+1] = u[l-\lambda(v'')+1 : n-\lambda(v'')+1]. \tag{18}$$

Suppose $v'[l+1] = v''[l+1]$. Then by Eqs. (17) and (18) we obtain $u[l-\lambda(v')+1] = u[l-\lambda(v'')+1]$. Since

$$(l-\lambda(v'')+1) - (l-\lambda(v')+1) = \lambda(v') - \lambda(v'') \le k-2,$$

this contradicts that $u$ is rarefied. So $v'[l+1] \ne v''[l+1]$. Since $v', v''$ are rarefied, it is easy to note from (16) that $v'[l+1]$ and $v''[l+1]$ can be either $a_1$ or $a_k$. So we have only two possible cases: $v'[l+1] = a_1$, $v''[l+1] = a_k$ or $v'[l+1] = a_k$, $v''[l+1] = a_1$. We consider these cases separately.

Let $v'[l+1] = a_1$ and $v''[l+1] = a_k$. Then it is easy to note that the symbol $v'[l+2]$ can be only $a_k$. Thus, by Eqs. (17) and (18) we obtain $u[l-\lambda(v')+1] = a_1$, $u[l-\lambda(v')+2] = a_k$ and $u[l-\lambda(v'')+1] = a_k$. So $u[l-\lambda(v')+2] = u[l-\lambda(v'')+1]$. Since

$$(l-\lambda(v'')+1) - (l-\lambda(v')+2) = \lambda(v') - \lambda(v'') - 1 < k-1$$

and $u$ is rarefied, the only case we have to consider is $l-\lambda(v'')+1=l-\lambda(v')+2$, i.e. $\lambda(v')-\lambda(v'')=1$ (in this case $u[l-\lambda(v')+2]$ and $u[l-\lambda(v'')+1]$ are the same letter in $u$). Since $v''$ is rarefied, $v''[l+2]$ can be either $a_1$ or $a_2$. If $v''[l+2]=a_1$, then by (18) we obtain $a_1=u[l-\lambda(v'')+2]=u[l-\lambda(v')+3]$. Thus we have in this case that $u[l-\lambda(v')+1]=u[l-\lambda(v')+3]$ which contradicts that $u$ is rarefied since $2<k-1$. Let $v''[l+2]=a_2$. Then it is easy to note that the symbol $v''[l+3]$ can be only $a_1$. Therefore, $a_1=u[l-\lambda(v'')+3]=u[l-\lambda(v')+4]$ by (18). Thus we have that $u[l-\lambda(v')+1]=u[l-\lambda(v')+4]$ which contradicts again that $u$ is rarefied.

Let now $v'[l+1]=a_k$ and $v''[l+1]=a_1$. Then it is easy to note that the symbol $v''[l+2]$ can be only $a_k$. Thus, by Eqs. (17) and (18) we obtain $u[l-\lambda(v')+1]=a_k$, $u[l-\lambda(v'')+1]=a_1$ and $u[l-\lambda(v'')+2]=a_k$. Since $u$ is rarefied, we have

$$(l-\lambda(v'')+2)-(l-\lambda(v')+1)=\lambda(v')-\lambda(v'')+1\geq k-1.$$

Thus $\lambda(v')-\lambda(v'')=k-2$ has to be valid in this case. Since $v'$ is rarefied, we have also that $v'[l+2]$ can be either $a_1$ or $a_2$. If $v'[l+2]=a_1$, then $u[l-\lambda(v')+2]=a_1$ by (17). Since $u[l-\lambda(v'')+1]=a_1$ and

$$(l-\lambda(v'')+1)-(l-\lambda(v')+2)=\lambda(v')-\lambda(v'')-1=k-3<k-1,$$

this contradicts that $u$ is rarefied. Let $v'[l+2]=a_2$. It is easy to note that in this case the symbol $v'[l+3]$ can be only $a_1$. Therefore, $u[l-\lambda(v')+3]=a_1$ by (17). Taking into account that $u[l-\lambda(v'')+1]=a_1$ and $k\geq 5$, we obtain again a contradiction with the fact that $u$ is rarefied, so the lemma is proved. □

Note that for any word $v\in\hat{\mathcal{H}}(n+1)$ we have $\tau(v)\in\mathcal{F}_m$ and $n-\lfloor n/k\rfloor\leq|\tau(v)|\leq n-\lfloor(p_2+1)/(k-1)\rfloor$, i.e. $\tau(v)\in\mathcal{Q}(n+1)=\bigcup_{j=n-\lfloor n/k\rfloor}^{n-\lfloor(p_2+1)/(k-1)\rfloor}\mathcal{F}_m(j)$. So from Lemma 5 we obtain that $|\mathcal{Q}(n+1)|\geq|\hat{\mathcal{H}}(n+1)|=k!|\hat{\mathcal{H}}'(n+1)|$. Denote by $\mathcal{Q}'(n+1)$ the set of all trimmed words from $\mathcal{Q}(n+1)$. Since by Proposition 2 any word from $\mathcal{Q}(n+1)$ is rarified and has the length greater than $p_2>k-1$, for any word from $\mathcal{Q}(n+1)$ there exists a single word from $\mathcal{Q}'(n+1)$ which is isomorphic to this word, and for any word from $\mathcal{Q}'(n+1)$ there exist exactly $k!$ different words from $\mathcal{Q}(n+1)$ which are isomorphic to this word. So $|\mathcal{Q}(n+1)|=k!|\mathcal{Q}'(n+1)|$. Thus $|\mathcal{Q}'(n+1)|\geq|\hat{\mathcal{H}}'(n+1)|$. Note that actually $\mathcal{Q}'(n+1)=\bigcup_{j=n-\lfloor n/k\rfloor}^{n-\lfloor(p_2+1)/(k-1)\rfloor}\bigcup_{i=1}^{s}\mathcal{F}_m^{(w_i)}(j)$ and, since all sets $\mathcal{F}_m^{(w_i)}(j)$ are non-overlapping,

$$|\mathcal{Q}'(n+1)|=\sum_{j=n-\lfloor\frac{n}{k}\rfloor}^{n-\lfloor\frac{p_2+1}{k-1}\rfloor}\sum_{i=1}^{s}|\mathcal{F}_m^{(w_i)}(j)|\leq\sum_{j=n-\lfloor\frac{n}{k}\rfloor}^{n-\lfloor\frac{p_2+1}{k-1}\rfloor}S_m^{\langle k\rangle}(j)/(\min_{i=1,\ldots,s}x_i).$$

Thus, taking into account (15), we obtain

$$\begin{aligned}\sum_{i=1}^{s}x_i|\hat{\mathcal{H}}^{(w_i)}(n+1)|&\leq|\hat{\mathcal{H}}'(n+1)|\cdot\max_{i=1,\ldots,s}x_i\leq|\mathcal{Q}'(n+1)|\cdot\max_{i=1,\ldots,s}x_i\\&\leq(\max_{i=1,\ldots,s}x_i)\sum_{j=n-\lfloor\frac{n}{k}\rfloor}^{n-\lfloor\frac{p_2+1}{k-1}\rfloor}S_m^{\langle k\rangle}(j)/(\min_{i=1,\ldots,s}x_i)=\mu\sum_{d=\lfloor\frac{p_2+1}{k-1}\rfloor}^{\lfloor\frac{n}{k}\rfloor}S_m^{\langle k\rangle}(n-d).\end{aligned}\tag{19}$$

*2.6. Getting a lower bound for $\gamma^{\langle k\rangle}$*

Summing up estimation (19) with relation (14), we conclude that

$$\sum_{i=1}^{s}x_i|\mathcal{H}^{(w_i)}(n+1)|\leq\sum_{d=a}^{b}\rho_d\cdot S_m^{\langle k\rangle}(n-d)+\mu\sum_{d=\lfloor\frac{p_2+1}{k-1}\rfloor}^{\lfloor\frac{n}{k}\rfloor}S_m^{\langle k\rangle}(n-d).\tag{20}$$

For the sake of convenience we denote by $\mathcal{P}(z)$ the polynomial $\sum_{d=a}^{b}\rho_d\cdot z^d$ in a variable $z$. Suppose for some $\alpha>1$ we have

$$S_m^{\langle k\rangle}(n)\geq\alpha^d\cdot S_m^{\langle k\rangle}(n-d)\tag{21}$$

for each $d=1,2,\ldots,n-m$. Then relation (20) implies that

$$\begin{aligned}\sum_{i=1}^{s}x_i|\hat{\mathcal{H}}^{(w_i)}(n+1)|&\leq S_m^{\langle k\rangle}(n)\sum_{d=a}^{b}\frac{\rho_d}{\alpha^d}+\mu S_m^{\langle k\rangle}(n)\sum_{d=\lfloor\frac{p_2+1}{k-1}\rfloor}^{\lfloor\frac{n}{k}\rfloor}\frac{1}{\alpha^d}\\&<S_m^{\langle k\rangle}(n)\left(\mathcal{P}\left(\frac{1}{\alpha}\right)+\mu\sum_{d=\lfloor\frac{p_2+1}{k-1}\rfloor}^{\infty}\frac{1}{\alpha^d}\right)\\&=S_m^{\langle k\rangle}(n)\left(\mathcal{P}\left(\frac{1}{\alpha}\right)+\frac{\mu}{\alpha^q(\alpha-1)}\right)\end{aligned}$$

where $q = \left\lfloor \frac{p_2+1}{k-1} \right\rfloor - 1$. Using this estimation and equalities (1) and (3), we obtain

$$\begin{aligned} S_m^{\langle k\rangle}(n+1) &= \sum_{i=1}^{s} x_i \cdot |\mathcal{G}^{(w_i)}(n+1)| - \sum_{i=1}^{s} x_i |\hat{\mathcal{H}}^{(w_i)}(n+1)| \\ &> S_m^{\langle k\rangle}(n) \cdot \left( r - \mathcal{P}\left(\frac{1}{\alpha}\right) - \frac{\mu}{\alpha^q(\alpha-1)} \right). \end{aligned}$$

Therefore, if $\alpha$ satisfy the inequality

$$r - \mathcal{P}\left(\frac{1}{\alpha}\right) - \frac{\mu}{\alpha^q(\alpha-1)} \geq \alpha,$$

we obtain the inequality $S_m^{\langle k\rangle}(n+1) \geq \alpha S_m^{\langle k\rangle}(n)$, and thus $S_m^{\langle k\rangle}(n+1) \geq \alpha^d \cdot S_m^{\langle k\rangle}(n+1-d)$ holds for any $d = 1, 2, \ldots, n-m+1$. If inequalities (21) hold for some $n'$, then inequalities (21) hold inductively in this case for every $n \geq n'$. Thus we have $S_m^{\langle k\rangle}(n) = \Omega(\alpha^n)$. Since, obviously, the order of growth of $S^{\langle k\rangle}(n)$ is not less than $S_m^{\langle k\rangle}(n)$, we then conclude that $S^{\langle k\rangle}(n) = \Omega(\alpha^n)$. Hence $\gamma^{\langle k\rangle} \geq \alpha$.

Note that for obtaining the bound $\gamma^{\langle k\rangle} \geq \alpha$ we have to prove initially that inequalities (21) holds for $n'$. For these purposes we compute the exact values of $S^{\langle k\rangle}(n)$ for $n \leq n_0$ by an enumeration of all Dejean's words of size at most $n_0$. The inequalities $S_m^{\langle k\rangle}(n+1) \geq \alpha S_m^{\langle k\rangle}(n)$ for $n_0 < n \leq kp_2/(k-1)$ could be verified in the same inductive way as described above with evident modifications following from the restriction $n \leq kp_2/(k-1)$.

## 3. Results

Using the described method of estimating $\gamma^{\langle k\rangle}$, we obtained lower bounds on $\gamma^{\langle k\rangle}$ for $5 \leq k \leq 10$. The obtained bounds together with the parameters $m$, $n_0$, $p_1$, $p_2$ used in the computer computations of these bounds are given in the following table. In this table we give also the upper bounds on $\gamma^{\langle k\rangle}$ we obtain with the method described in [18]. For the anti-dictionary $\mathcal{A}$, we take the set of all binary minimally forbidden words in the Pansiot's code (w.r.t. factor containment) of size at most $q$.

| $k$ | $m$ | $s$ | $n_0$ | $p_1$ | $p_2$ | Lower bound on $\gamma^{\langle k\rangle}$ | $q$ | $\lvert\mathcal{A}\rvert$ | Upper bound on $\gamma^{\langle k\rangle}$ |
|---|---|---|---|---|---|---|---|---|---|
| 5 | 50 | 5287 | 150 | 183 | 600 | 1.153811 | 158 | 12783585 | 1.157895 |
| 6 | 33 | 1926 | 100 | 125 | 500 | 1.223437 | 113 | 3946990 | 1.224695 |
| 7 | 28 | 318 | 100 | 126 | 600 | 1.236409 | 114 | 2958045 | 1.236899 |
| 8 | 18 | 31 | 100 | 119 | 600 | 1.234725 | 118 | 1399465 | 1.234843 |
| 9 | 20 | 42 | 100 | 123 | 600 | 1.246659 | 112 | 287646 | 1.246678 |
| 10 | 22 | 55 | 100 | 122 | 600 | 1.239287 | 115 | 65346 | 1.239308 |

Comparing the obtained lower bounds with the upper bounds on $\gamma^{\langle k\rangle}$ presented in the table, one can conclude that we have estimated $\gamma^{\langle k\rangle}$ for $5 \leq k \leq 10$ with the precision of 0.005.

## 4. Conclusion

In this paper we obtained lower bounds on $\gamma^{\langle k\rangle}$ for $5 \leq k \leq 10$, but we believe that by the method proposed in the paper lower bounds on $\gamma^{\langle k\rangle}$ could be computed for any fixed $k \geq 5$ (provided that $\gamma^{\langle k\rangle} > 1$). So we consider as an interesting problem for further investigations the question if the computations described in the paper can be generalized theoretically for obtaining theoretical lower bounds on $\gamma^{\langle k\rangle}$ valid for any $k \geq 5$.

## Acknowledgements

This work started when both authors were invited to LIAFA, University Paris Diderot (Paris-7), France, in June 2009. R. Kolpakov acknowledges the partial support of the Russian Foundation for Fundamental Research (Grant 11-01-00508) and the program for supporting Russian scientific schools (Grant NSh 5400.2006.1).

## References

[1] M. Baake, V. Elser, U. Grimm, The entropy of square-free words, Math. Comput. Modelling 26 (1997) 13–26.
[2] J. Berstel, Growth of repetition-free words — a review, Theoret. Comput. Sci. 340 (2005) 280–290.
[3] J. Brinkhuis, Nonrepetitive sequences on three symbols, Quart. J. Math. Oxford 34 (1983) 145–149.
[4] A. Carpi, Overlap-free words and finite automata, Theoret. Comput. Sci. 115 (1993) 243–260.
[5] A. Carpi, On Dejean's conjecture over large alphabets, Theoret. Comput. Sci. 385 (2007) 137–151.
[6] J. Currie, N. Rampersad, Dejean's conjecture holds for $n \geq 30$, Theoret. Comput. Sci. 410 (2009) 2885–2888.
[7] J. Currie, N. Rampersad, Dejean's conjecture holds for $n \geq 27$, RAIRO Theoret. Inform. Appl. 43 (2009) 775–778.
[8] J. Currie, N. Rampersad, A proof of Dejean's conjecture, Manuscript, 2009. http://arxiv.org/abs/0905.1129.

[9] F. Dejean, Sur un théorème de Thue, J. Combin. Theory, Ser. A 13 (1972) 90–99.
[10] U. Grimm, M. Heuer, On the entropy and letter frequencies of powerfree words, Entropy 10 (2008) 590–612.
[11] R. Kolpakov, Efficient lower bounds on the number of repetition-free words, J. Integer Seq. 10 (2007) Article 07.3.2.
[12] M. Mohammad-Noori, J. Currie, Dejean's conjecture and Sturmian words, European J. Combin. 28 (2007) 876–890.
[13] J. Moulin Ollagnier, Proof of Dejean's conjecture for alphabets with 5, 6, 7, 8, 9, 10 and 11 letters, Theoret. Comput. Sci. 95 (1992) 187–205.
[14] P. Ochem, A generator of morphisms for infinite words, in: Proceedings of Workshop on Word Avoidability, Complexity, and Morphisms, Turku, Finland, July 2004, pp. 9–14.
[15] J.J. Pansiot, A propos d'une conjecture de F. Dejean sur les répétitions dans les mots, Discrete Appl. Math. 7 (1984) 297–311.
[16] M. Rao, Last cases of Dejean's conjecture, Words 2009, Salerno, Italy, 2009.
[17] A. Restivo, S. Salemi, Overlap-free words on two symbols, Lecture Notes in Comput. Sci. 192 (1985) 198–206.
[18] A. Shur, I. Gorbunova, On the growth rates of complexity of threshold languages, in: 12th Mons Theoretical Computer Science Days, Mons, Belgium, 2008.
[19] A. Shur, Two-sided bounds for the growth rates of power-free languages, Lecture Notes in Comput. Sci. 5583 (2009) 466–477.
[20] A. Thue, Über unendliche Zeichenreihen, in: Norske Vidensk. Selsk. Skrifter. I. Mat.-Nat. Kl. 7, Christiania, 1906, pp. 1–22.
[21] A. Thue, Über die gegenseitige Lage gleicher Teile gewisser Zeichenreihen, in: Norske Vidensk. Selsk. Skrifter. I. Mat.-Nat. Kl. 10, Christiania, 1912, pp. 1–67.