Theoretical Computer Science 412 (2011) 2955–2963

Contents lists available at ScienceDirect

Theoretical Computer Science

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

# Bounds for the generalized repetition threshold

Francesca Fiorenzi [a,*], Pascal Ochem [a,b], Elise Vaslet [c]

[a] *Laboratoire de Recherche en Informatique, UMR 8623, Université Paris-Sud, 91405 Orsay, France*

[b] *CNRS - Laboratoire J.-V.Poncelet, UMI 2615, Moscow, Russian Federation*

[c] *Institut de Mathématiques de Luminy, UMR 6206, Université Aix-Marseille II, Campus de Luminy, Case 907, 13288 Marseille Cedex 9, France*

ARTICLE INFO

*Keywords:*
Generalized repetition threshold

ABSTRACT

The notion of the *repetition threshold*, which is the object of Dejean's conjecture (1972), was generalized by Ilie et al. (2005) [8] to include the lengths of the avoided words. We give a lower and an upper bound on this generalized repetition threshold.



## 1. Introduction

The study of repetitions in words has been one of the main topics in combinatorics on words. Thue [13] showed the existence of an infinite square-free word on three letters, that is without concatenated occurrences of the same non-empty factor. This fact was actually implied by the existence of an infinite binary overlap-free word (i.e., without factors of the form $uvuvu$ with $u$ non empty).

A natural extension of this problem takes into account repetitions of fractional exponent, where the exponent of a non-empty finite word is the ratio between its length and its period. This notion has been introduced by Dejean [7] and Brandenburg [1]. Dejean proved the existence of an infinite word over a three-letter alphabet without repetitions of exponent strictly greater than 7/4. This bound is the best possible because every sufficiently long word over a three-letter alphabet contains a repetition of exponent at least 7/4. The least real number $\alpha > 1$ such that there exists an infinite word on $k$ letters avoiding repetitions of exponent strictly greater than $\alpha$ is called the repetition threshold on $k$ letters. Thus Thue's result implies that the repetition threshold on two letters is 2, while Dejean's result means that the repetition threshold on three letters is 7/4.

Dejean observed that for $k \geq 5$, the repetition threshold is not smaller than $\frac{k}{k-1}$, while for $k = 4$ it is not smaller than 7/5. She conjectured that these are the actual values of the repetition thresholds. This conjecture has been proved true for $k = 4$ by Pansiot [11]. For $k \geq 5$, the conjecture has been solved thanks to the contribution of many authors: Moulin-Ollagnier [10], Currie and Mohammad-Noori [3], Rao [12], Currie and Rampersad [6,4,5], and Carpi [2].

In [8] the authors generalize the repetition threshold of Dejean to handle avoidance of repetitions with sufficiently long period. They define $R(k, \ell)$ as the least real number $\alpha > 1$ such that there exists an infinite word on $k$ letters avoiding repetitions of exponent strictly greater than $\alpha$ and of period not smaller than $\ell$. This naturally extends the classical notion of the repetition threshold. Moreover, in [8] its value has been calculated in some particular cases and general lower and upper bounds have been given. In this paper we improve these bounds by studying the asymptotics of the generalized repetition threshold. Part of this paper was presented at the conference *WORDS 2009*.

## 2. Definitions

Let $\alpha > 1$ be a rational number, and let $\ell \geq 1$ be an integer. A word $w$ is an $(\alpha, \ell)$-*repetition* if $w = (uv)^n u$, where $|uv| = \ell$ and $\alpha = \frac{|w|}{\ell}$. In this case $\alpha$ is called *exponent* of the repetition and $\ell$ its *period*. Notice that in our definitions,

* Corresponding author. Tel.: +33 1 69 15 32 19.
*E-mail addresses:* fiorenzi@lri.fr (F. Fiorenzi), ochem@lri.fr (P. Ochem), vaslet@iml.univ-mrs.fr (E. Vaslet).

0304-3975/$ – see front matter 
doi:10.1016/j.tcs.2010.10.009

exponent and period of a repetition are not univocally defined. For instance, the word *aabcaabcaa* on the alphabet $\{a, b, c\}$ is a $(\frac{5}{2}, 4)$-repetition, a $(\frac{5}{4}, 8)$-repetition and a $(\frac{10}{9}, 9)$-repetition. We say a word that $(\alpha, \ell)$*-free* if it contains no factor that is an $(\alpha', \ell')$-repetition with $\alpha' \geq \alpha$ and $\ell' \geq \ell$. We say a word that $(\alpha^+, \ell)$*-free* if it is $(\alpha', \ell)$-free for all $\alpha' > \alpha$. Finally, a word is $\alpha$*-free* if it does not contain a repetition with exponent $\geq \alpha$ and is $\alpha^+$*-free* if it is $\alpha'$-free for all $\alpha' > \alpha$.

Let $\Sigma_k$ denote the $k$-letter alphabet $\{0, 1, \dots, k-1\}$. For integers $k \geq 2$ and $\ell \geq 1$, the *generalized repetition threshold* $R(k, \ell)$ is defined as the smallest real number $\alpha$ such that there exists an infinite $(\alpha^+, \ell)$-free word over $\Sigma_k$. Indeed, there always exists an infinite $(R(k, \ell)^+, \ell)$-free word over $\Sigma_k$. Nevertheless, as pointed out in [8], there is no known instance of a $(R(k, \ell), \ell)$-free infinite word. Finally, notice that by definition, $R(k+1, \ell) \leq R(k, \ell)$ and $R(k, \ell+1) \leq R(k, \ell)$.

The finiteness of $R(k, \ell)$ is due to the existence of an infinite binary overlap-free word. Ilie et al. [8] also obtained a lower bound on $R(k, \ell)$, namely

$$1 + \frac{\ell}{k^\ell} \leq R(k, \ell) \leq 2.$$

The aim of the paper is to improve the above inequalities.

The case $\ell = 1$ corresponds to the classical repetition threshold and the values of $R(k, 1)$ are now all determined. Moreover, the proof of our upper bound explicitly uses the fact that $R(k, 1) = \frac{k}{k-1}$ for $k \geq 5$.

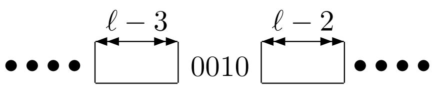
## 3. Lower bound

A natural way of obtaining a bound of the form $R(k, \ell) \geq \alpha$ is to show that there is no infinite word over $\Sigma_k^*$ which is $(\alpha, \ell)$-free. In this section we give lower bounds on $R(k, \ell)$. We treat separately the cases $k = 2, \ell \geq 2$ (Theorem 1), $k \geq 2, \ell = 2$ (Theorem 4) and $k, \ell \geq 3$ (Theorem 14).

**Theorem 1.** $R(2, \ell) \geq 1 + \frac{2}{\ell+2}$.

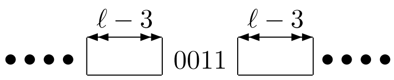
**Proof.** Suppose for the sake of contradiction that $R(2, \ell) < 1 + \frac{2}{\ell+2}$. That is, there exists an infinite binary word $w$ with no repetition of period $\geq \ell$ and exponent $\geq 1 + \frac{2}{\ell+2}$. In particular, repetitions $uvu$ such that $|u| = 2$ and $\ell \leq |uv| \leq \ell + 2$ cannot appear in $w$. Moreover, we can assume without loss of generality that $w$ is double-infinite, that is, infinite in both directions. We say that a factor is *forbidden* if it cannot appear in $w$.

First, we check that the factor 0010 is forbidden. Indeed, if 0010 was a factor of $w$, then $w$ would have a factor of the form $u_1u_2 0010 v_1v_2$, with $|u_1| = |v_2| = 4$, $|u_2| = \ell - 3$ and $|v_1| = \ell - 2$. The following picture illustrates the situation.




By previous considerations about the distances, we have that the blocks $u_1$ and $v_2$ coincide necessarily with 1111. This creates a repetition of period $2\ell+3$ and exponent $1 + \frac{4}{2\ell+3} > 1 + \frac{2}{\ell+2}$. Hence the factor 0010 and, by symmetry, the factors 0100, 1101, and 1011, are forbidden.

Similarly, we check that the factor 0011 is forbidden. Otherwise we had a factor of the form $u_1u_2 0011 v_1v_2$, with $|u_1| = |v_2| = 4$ and $|u_2| = |v_1| = \ell - 3$.




The block $u_1$ on the left of the factor 0011 must be of the form $\bullet 10 \bullet$ , then $110 \bullet$ , and finally 1100 since 1101 is forbidden. The block $v_2$ on the right of the factor 0011 must be of the form $\bullet 10 \bullet$ , then $\bullet 100$ , and finally 1100 since 0100 is forbidden. This creates a repetition of period $2\ell + 2$ and exponent $1 + \frac{2}{\ell+1} > 1 + \frac{2}{\ell+2}$.

The factor 001 is forbidden because 0010 and 0011 are forbidden. By symmetry, 100, 110, and 011 are also forbidden. Thus, the only remaining possibilities for $w$ are the words ${}^{\omega}0^{\omega}$, ${}^{\omega}1^{\omega}$, and ${}^{\omega}(01)^{\omega}$, which obviously contain repetitions of an arbitrarily great exponent and period. $\square$

Theorem 1 is certainly not optimal, since numerical evidences suggest that $R(2, \ell) = 1 + \frac{1}{\ell/3+1}$ for $\ell \geq 6, \ell = 0 \pmod 3$. We also mention that Kolpakov and Rao [9] have proved that $R(3, \ell) \geq 1 + \frac{1}{\ell}$, which would be a tight lower bound for the conjecture in [8] that $R(3, \ell) = 1 + \frac{1}{\ell}$ for $\ell \geq 2$.

We now consider the problem for a general $k$. The $(\alpha, \ell)$-freeness of a word imposes conditions on the length of some of its factors. Such conditions are described in the following, and will be used in the proof of Theorems 4 and 14.

**Definition 2.** Let $k, \ell$ be integers, let $m \geq \ell$ be a real number, and let $a$ be a letter in $\Sigma_k$. We denote by $\mathfrak{L}(k, \ell, m, a)$ the language of the words $w$ over $\Sigma_k$ such that the following conditions $(C_i^{\ell,m})$ hold for $1 \leq i \leq \ell$

$$(C_i^{\ell,m}): \text{ if } a^ixa^i \text{ is a subword of } w, \text{ then } |a^ix| > im \text{ or } |a^ix| < \ell.$$

We denote by $\mathfrak{L}(k, \ell, m)$ the language $\bigcap_{a\in\Sigma_k} \mathfrak{L}(k, \ell, m, a)$.

**Proposition 3.** *Let $k, \ell$ be integers, $m \geq \ell$ be a real number, and $w \in \Sigma_k^*$. If $w$ is $(1 + \frac{1}{m}, \ell)$-free, then, $w \in \mathfrak{L}(k, \ell, m)$.*

**Proof.** Let $a$ be a letter in $\Sigma_k$. Let $i \in \{1, \ldots, \ell\}$ be an integer. Suppose that $w$ does not satisfy condition $(C_i^{\ell,m})$, i.e., there exists a subword $a^ixa^i$ of $w$ such that $\ell \leq |a^ix| \leq im$. Then, $a^ixa^i$ is a repetition with period $|a^ix| \geq \ell$, and exponent

$$\frac{|a^ixa^i|}{|a^ix|} = 1 + \frac{i}{|a^ix|} \geq 1 + \frac{1}{m}.$$

Hence $w$ is not $(1 + \frac{1}{m}, \ell)$-free. $\square$

First we consider the case $\ell = 2$, and look for a lower bound of $R(k, 2)$.

**Theorem 4.** *Let $k \geq 2$. We have*

$$R(k, 2) \geq 1 + \frac{1}{m},$$

*where $m = 1 + \left\lfloor \frac{3}{2}(k-1) \right\rfloor$.*

**Proof.** As $m = 1 + \left\lfloor \frac{3}{2}(k-1) \right\rfloor$, we observe that $2m + 3 \geq 3k + 1$. Hence we have $(2m + 3)k > 3k^2 + (k - 1)$. Consider $r = 3k + 1$ and a word $w \in \Sigma_k^*$ such that $|w| = rk + 1$. Then, there exists a letter $a \in \Sigma_k$ such that

$$|w|_a \geq r + 1 = 3k + 2.$$

Suppose that $w$ is $(1 + \frac{1}{m}, 2)$-free. Then, as $m = 1 + \left\lfloor \frac{3}{2}(k-1) \right\rfloor \geq 2$, by Proposition 3, the word $w \in \mathfrak{L}(k, 2, m, a)$.

Conditions $(C_1^{2,m})$ and $(C_2^{2,m})$ imply that in $w$ there are at least $m$ letters between two non-consecutive occurrences of the letter $a$ and at least $2m - 1$ letters between two occurrences of the factor $aa$.

We prove by induction on $s \in \mathbb{N}$ that any word in $\mathfrak{L}(k, 2, m, a)$ containing $3s+2$ occurrences of the letter $a$ is not shorter than $(2m + 3)s + 2$.

This is evident for $s = 0$. If the statement is true up to $s$, consider a word $v \in \mathfrak{L}(k, 2, m, a)$ containing $3(s + 1) + 2$ occurrences of the letter $a$. Then $v$ can be written as

$$v = v_1v_2 \ldots v_{s+2}$$

where $v_1$ contains exactly two occurrences of $a$ and for each $i \geq 2$, the factor $v_i$ begins with a letter $a$ and contains exactly three occurrences of this letter. Clearly we have that $|v_{s+2}| \geq 2m + 2$.

If the last letter of $v_{s+1}$ is different from $a$, we have that the last $m$ letters of $v_{s+1}$ are different from $a$ and we apply the inductive hypothesis on the prefix of $v$ of length $|v_1v_2 \ldots v_{s+1}| - m$. Thus $|v| \geq (2m + 3)s + 2 + m + 2m + 2 \geq (2m + 3)(s + 1) + 2$.

Otherwise, suppose that the last letter of $v_{s+1}$ is $a$. If for each $2 \leq i \leq s + 1$, the last letter of the factor $v_i$ is $a$, then this factor $v_i$ has the form $ax_iay_ia$ with $|x_i|_a = |y_i|_a = 0$. Thus

$$v = v_1a\,x_2\,a\,y_2\,aa\,x_3\,a\,y_3\,a\ \ldots\ a\,x_s\,a\,y_s\,aa\,x_{s+1}\,a\,y_{s+1}\,a\,v_{s+2}$$

and for each $i \geq 3$ we have $|x_i| \geq m$ and $|y_i| \geq m$. Moreover $|v_1| \geq m + 2$ and $|x_2| + |y_2| \geq 2m - 1$. This implies $|v| \geq m + 2 + 3 + 2m - 1 + (2m + 3)(s - 1) + 2m + 2 \geq (2m + 3)(s + 1) + 2$.

Finally, if for some $2 \leq i \leq s$ the factor $v_i$ does not end with $a$, we can apply the induction hypothesis on a proper prefix of $v$, and similar arguments on the lengths as before, to conclude that $|v| \geq (2m + 3)(s + 1) + 2$.

Hence

$$\begin{aligned} |w| &\geq (2m + 3)k + 2 \\ &> 3k^2 + k + 1 = rk + 1, \end{aligned}$$

which is a contradiction since $|w| = rk + 1$. Thus, there is no infinite $(1 + \frac{1}{m}, 2)$-free word over $\Sigma_k$, and

$$R(k, 2) \geq 1 + \frac{1}{m}. \quad \square$$

In Theorem 14, the result of Theorem 4 is generalized to any $\ell \geq 3$. The proof also uses considerations on letter frequencies in the words of $\mathfrak{L}(k, \ell, m)$. In the following, $\ell$ is a fixed integer greater than 2.

**Definition 5.** Let $a \in \Sigma_k$. A word $w = w_1 \dots w_{|w|} \in \Sigma_k^*$ is *of type $S_a$* if $1 \le |w| \le \ell$, $w_1 = w_{|w|} = a$, and if $b_1 b_2 \notin \mathrm{Fact}(w)$ for each $b_1, b_2 \in \Sigma_k \setminus \{a\}$.

Notice that any word of type $S_a$ is in $\mathfrak{L}(k, \ell, m, a)$, for any real number $m \ge \ell$.

**Definition 6.** Let $a \in \Sigma_k$. Let $w$ be a word of type $S_a$. The *weight $p_a(w)$ of $w$* is defined by

$$p_a(w) = \max\{j \in \mathbb{N} \mid a^j \in \mathrm{Fact}(w)\}.$$

Notice that $1 \le p_a(w) \le \ell$ for each word $w$.

**Proposition 7.** *Let $a \in \Sigma_k$. Let $w$ be a word of type $S_a$. We have*

$$|w|_a \le \ell - \left\lfloor \frac{\ell}{p_a(w)+1} \right\rfloor = \left\lceil \frac{\ell \cdot p_a(w)}{p_a(w)+1} \right\rceil.$$

**Proof.** Let us denote $p = p_a(w)$, and write $w$ as

$$w = z_1 z_2 \dots z_n z,$$

where $n \in \mathbb{N}$, $|z_i| = p + 1$, for $1 \le i \le n$, and $|z| = r \le p$. For each $i$, we have $|z_i|_a \le p$ (because $a^{p+1} \notin \mathrm{Fact}(w)$), and $|z|_a \le |z|$. Hence

$$|w|_a \le np + r = |w| - \left\lfloor \frac{|w|}{p+1} \right\rfloor \le \ell - \left\lfloor \frac{\ell}{p+1} \right\rfloor = \left\lceil \frac{\ell p}{p+1} \right\rceil. \quad \square$$

**Definition 8.** Let $a \in \Sigma_k$. A word $w = w_1 \dots w_{|w|} \in \Sigma_k^*$ is *of type $S'_a$* if $|w| \le \ell$, $w_1 = w_{|w|} = a$.

**Definition 9.** Let $a \in \Sigma_k$ and $m \ge \ell$ be a real number. A word $w \in \Sigma_k^*$ is *of type $L_a$* if $|w| > m - 1$ and $|w|_a = 0$.

Notice that any word of type $L_a$ is in $\mathfrak{L}(k, \ell, m, a)$.

**Lemma 10.** *Let $k, \ell$ be integers, let $m \ge 2\ell - 2$ be a real number, $a \in \Sigma_k$, and $w \in \mathfrak{L}(k, \ell, m, a)$. Then $w$ can be written as*

$$w = x s'_1 l_1 s'_2 l_2 \dots l_{r-1} s'_r y,$$

*where $r \in \mathbb{N}$ (if $r = 0$ then $w = x$), each word $s'_i$ is of type $S'_a$, each word $l_i$ is of type $L_a$ and $|x|_a = |y|_a = 0$.*

**Proof.** As $w \in \mathfrak{L}(k, \ell, m, a)$, $w$ satisfies condition $(C_1^{\ell,m})$: if $axa$ is a subword of $w$, then, $|axa| > m + 1$ or $|axa| < \ell + 1$.

If $|w|_a = 0$, we have the result with $x = w$ and $r = 0$.

Now, if $|w_a| > 0$, let us write $w = xw'y$, where $w'$ begins and ends with $a$, while $x$ and $y$ do not contain $a$. The maximal factors without $a$'s in $w'$ either have length $< \ell - 1$, or length $> m - 1$. Take the latter factors as $l_1, \dots, l_{r-1}$ to get the factorization $w = x s'_1 l_1 s'_2 \dots l_{r-1} s'_r y$. Each $s'_i$ begins and ends with $a$, all other blocks do not contain $a$, and by definition, $|l_i| > m - 1$ for all $i$. Thus, to prove the lemma, it suffices to show that $|s'_i| \le \ell$ for any $i$. If it is not the case, then for some $s'_i = aua$, we have $|s'_i| > m + 1$. Then $|u| > m - 1$, and $u$ contains the letter $a$ by construction. So we have $s'_i = au_1au_2a$. At least one of the words $au_1a$ and $au_2a$ has length $> m + 1$, because if both have length $> \ell - 1$, then $|s'_i| < 2\ell - 3 \le m + 1$. Thus, $u_1$ also contains $a$, and we repeat the argument to get a contradiction after a finite number of steps. $\square$

**Proposition 11.** *Let $k, \ell$ be integers, let $m \ge 2\ell - 2$ be a real number, $a \in \Sigma_k$ and $w \in \mathfrak{L}(k, \ell, m, a)$. There exists a word $\tilde{w}$ in $\mathfrak{L}(k, \ell, m, a)$, such that $|\tilde{w}| = |w|$, $|\tilde{w}|_a = |w|_a$ and $\tilde{w}$ can be written as*

$$\tilde{w} = x s_1 l_1 s_2 l_2 \dots l_{r-1} s_r y,$$

*where $r \in \mathbb{N}$ (if $r = 0$ then $\tilde{w} = x$), each word $s_i$ is of type $S_a$, each word $l_i$ is of type $L_a$, and $|x|_a = |y|_a = 0$.*

**Proof.** By Lemma 10, $w$ is of the form:

$$w = x' s'_1 l'_1 s'_2 \dots l'_{r-1} s'_r y',$$

where $r \in \mathbb{N}$, $|x'|_a = |y'|_a = 0$, $s'_i$ are words of type $S'_a$, and $l'_i$ are words of type $L_a$. For each $1 \le i \le r$, let us consider the factor $s'_i$, and iterate the rewriting rules $b_1 b_2 a \mapsto b_1 a b_2$, $\forall b_1, b_2 \in \Sigma_k \setminus \{a\}$, on it. After each iteration, with regard to the lexicographic order (where $a \prec b$, $\forall b \in \Sigma_k \setminus \{a\}$), we get a smaller word, because $b_1 a b_2 \prec b_1 b_2 a$. So the iteration finishes, and we get a word of the form $s_i x_i$, where $s_i$ is a word of type $S_a$, and where $|x_i|_a = 0$. We remark that $|s_i| \le |s'_i|$, $|s_i x_i| = |s'_i|$, $|s_i|_a = |s'_i|_a$, and $|s_i x_i|_a = |s'_i|_a$, since the iteration conserves the number of occurrences of each letter in $s'_i$.

Finally, let $\tilde{w} = x s_1 l_1 s_2 l_2 \dots l_{r-1} s_r y$, where $x = x'$, $y = x_r y'$, and $l_i = x_i l'_i$. It is clear that $|\tilde{w}| = |w|$ and $|\tilde{w}|_a = |w|_a$. It is then sufficient to prove that $\tilde{w} \in \mathfrak{L}(k, \ell, m, a)$. Condition $(C_1^{\ell,m})$ is clearly satisfied by definition of the types $S_a$ and $L_a$. Let now $j \in [2, \dots, \ell]$ and $z \in \Sigma_k^*$ be such that $a^j z a^j \in \mathrm{Fact}(\tilde{w})$.

Then, as $w \in \mathfrak{L}(k, \ell, m, a)$, we have $|a^j z| = |a^j z'| > jm + 1$, and condition $(C_j^{\ell,m})$ holds.

The letter $a$ moved by the iteration of the rewriting rules always has some letter $b$ on its immediate left. Hence the iteration cannot result in a new factor $a^j$. So, if $\tilde{w}$ contains $a^jza^j$, then $w$ contains $a^jz'a^j$, where $|z'| = |z|$. Then, as $w \in \mathfrak{L}(k, \ell, m, a)$, we have $|a^jz| = |a^jz'| > jm$, and condition $(C_j^{\ell,m})$ holds. $\square$

**Lemma 12.** *Let $p \geq 1$ be an integer. Let $v$ be a word in $\mathfrak{L}(k, \ell, m, a)$, of the form described in Proposition 11:*

$$v = xs_1l_1s_2l_2 \dots l_{r-1}s_ry,$$

*and where the words $s_i$ of type $S_a$ are such that $p_a(s_i) = p$ for each $i = 1, \dots, r$. Then,*

$$r < \begin{cases} 1 + \frac{|v|-1}{m} \text{ if } p = 1 \\ 1 + \frac{|v|-p}{p(m+1)-\ell} \text{ if } p > 1. \end{cases}$$

**Proof.** Consider two consecutive blocks $s_i$ and $s_{i+1}$, with $1 \leq i \leq r - 1$. As they have weight $p$, they both contain $a^p$ as a factor. So, $s_il_is_{i+1}$ has a factor $a^pua^p$, and by condition $(C_p^{\ell,m})$, we have $|a^pua^p| > p(m + 1)$. So $|s_il_is_{i+1}| > p(m + 1)$. But $|s_{i+1}| \leq \ell$, by definition of the type $S_a$. Thus, $\forall 1 \leq i \leq r - 1$,

$$|s_il_i| > p(m + 1) - \ell.$$

Moreover, $|s - r| \geq p$. Then,

$$|v| \geq \sum_{i=1}^{r-1} |s_il_i| + |s - r| > (r - 1)\,(p(m + 1) - \ell) + p,$$

and we deduce that

$$r < 1 + \frac{|v| - p}{p(m+1) - \ell}.$$

In the case $p = 1$, we can have a better bound: as $|l_i| > m - 1$ and $|s_i| \geq 1$, we have

$$|s_il_i| > m,$$

hence

$$r < 1 + \frac{|v| - 1}{m}. \quad \square$$

**Proposition 13.** *Let $k, \ell$ be integers, let $\xi \geq 2$ be a real number, let $m \geq \xi(\ell - 1)$ be a real number, $a \in \Sigma_k$, and $w \in \mathfrak{L}(k, \ell, m, a)$. We have*

$$\frac{|w|_a}{|w|} < \begin{cases} \frac{2}{|w|} + \frac{1}{m} \cdot \frac{3\xi-1}{2\xi-1} \text{ if } \ell = 2 \\ \frac{\ell}{|w|} + \frac{1}{m}\left(\ell - \frac{(\xi-1)(\ell-1)}{2(2\xi-1)} - \frac{\xi^2(\ell-2)}{(\xi\ell-1)(2\xi-1)}\right) \text{ if } \ell \geq 3. \end{cases}$$

**Proof.** By Proposition 11, we can assume without loss of generality that

$$w = xs_1l_1s_2l_2 \dots l_{r-1}s_ry,$$

where $r \in \mathbb{N}$, each word $s_i$ is of type $S_a$, each word $l_i$ is of type $L_a$, and $|x|_a = |y|_a = 0$. Words of type $L_a$ contain no occurrences of $a$ thus, by Proposition 7,

$$|w|_a = \sum_{i=1}^{r} |s_i|_a \leq \sum_{p=1}^{\ell} n_p(w) \left\lceil \frac{\ell p}{p+1} \right\rceil,$$

where

$$n_p(w) = \#\{i \in [1, \dots, r] \mid p_a(s_i) = p\}.$$

For any $1 \leq j \leq \ell$, we denote

$$q_j(w) = \sum_{p \geq j} n_p(w) = \#\{i \in [1, \dots, r] \mid p_a(s_i) \geq j\}.$$

For any $1 \leq p \leq \ell - 1$, we have $n_p(w) = q_p(w) - q_{p+1}(w)$, and $n_\ell(w) = q_\ell(w)$. Hence

$$\begin{aligned} |w|_a &\leq \sum_{p=1}^{\ell-1} \big(q_p(w) - q_{p+1}(w)\big) \left\lceil \frac{\ell p}{p+1} \right\rceil + q_\ell(w) \left\lceil \frac{\ell^2}{\ell+1} \right\rceil \\ &= \sum_{p=1}^{\ell} q_p(w) \left( \left\lceil \frac{\ell p}{p+1} \right\rceil - \left\lceil \frac{\ell(p-1)}{p} \right\rceil \right). \end{aligned}$$

Let $n = |w|$. Because of Proposition 11, it makes sense to speak of $q_p(v)$ for any word $v \in \mathfrak{L}(k, \ell, m, a)$. Thus for any $1 \le p \le \ell$, we have

$$q_p(w) \le \max\{q_p(v) \mid v \in \mathfrak{L}(k, \ell, m, a), \text{ and } |v| = n\}.$$

If $v$ reaches this maximum, by replacing some occurrences of $a$ by any other letter, we construct a word $v'$ such that $|v'| = |v|$ and $n_i(v') = 0$ for each $i \neq p$. So,

$$q_p(w) \le \max\{q_p(v) \mid v \in \mathfrak{L}(k, \ell, m, a),\ |v| = n, \text{ and } n_i(v) = 0, \forall i \neq p\}.$$

Finally, if $n_i(v) = 0$ for each $i \neq p$, then $q_p(v) = n_p(v)$, and we obtain

$$\begin{aligned} q_p(w) &\le \max\{n_p(v) \mid v \in \mathfrak{L}(k, \ell, m, a),\ |v| = n, \text{ and } n_i(v) = 0, \forall i \neq p\} \\ &< \begin{cases} 1 + \frac{n-1}{m} \text{ if } p = 1 \\ 1 + \frac{n-p}{p(m+1)-\ell} \text{ if } p > 1, \end{cases} \end{aligned}$$

by Lemma 12. Hence

$$\begin{aligned} |w|_a &< \sum_{p=1}^{\ell} \left( \left\lceil \frac{\ell p}{p+1} \right\rceil - \left\lceil \frac{\ell(p-1)}{p} \right\rceil \right) + \frac{(n-1)\left\lceil \frac{\ell}{2} \right\rceil}{m} + \sum_{p=2}^{\ell} \frac{(n-p)\left( \left\lceil \frac{\ell p}{p+1} \right\rceil - \left\lceil \frac{\ell(p-1)}{p} \right\rceil \right)}{p(m+1)-\ell} \\ &\le \ell + \frac{n\left(\ell - \left\lfloor \frac{\ell}{2} \right\rfloor\right)}{m} + \sum_{p=2}^{\ell} \frac{n\left( \left\lfloor \frac{\ell}{p} \right\rfloor - \left\lfloor \frac{\ell}{p+1} \right\rfloor \right)}{p(m+1)-\ell} \\ &\le \ell + \frac{n}{m}\left( \ell - \left\lfloor \frac{\ell}{2} \right\rfloor + \sum_{p=2}^{\ell} \frac{\left\lfloor \frac{\ell}{p} \right\rfloor - \left\lfloor \frac{\ell}{p+1} \right\rfloor}{p - \frac{1}{\xi}} \right). \end{aligned}$$

Indeed, from $m \ge \xi(\ell - 1)$ follows $p(m+1) - \ell \ge (p - \frac{1}{\xi})m$. Then, if $\ell = 2$,

$$\frac{|w|_a}{|w|} < \frac{2}{n} + \frac{1}{m} \cdot \frac{3\xi - 1}{2\xi - 1},$$

and if $\ell \ge 3$ we have

$$|w|_a < \ell + \frac{n}{m}\left( \ell - \left\lfloor \frac{\ell}{2} \right\rfloor \cdot \frac{\xi - 1}{2\xi - 1} - \sum_{p=3}^{\ell} \left\lfloor \frac{\ell}{p} \right\rfloor \frac{1}{\left(p - \frac{1}{\xi}\right)\left(p - 1 - \frac{1}{\xi}\right)} \right).$$

Since

$$\begin{aligned} \sum_{p=3}^{\ell} \left\lfloor \frac{\ell}{p} \right\rfloor \frac{1}{\left(p - \frac{1}{\xi}\right)\left(p - 1 - \frac{1}{\xi}\right)} &\ge \sum_{p=3}^{\ell} \frac{1}{\left(p - \frac{1}{\xi}\right)\left(p - 1 - \frac{1}{\xi}\right)} \\ &= \frac{\xi^2(\ell - 2)}{(\xi\ell - 1)(2\xi - 1)}, \end{aligned}$$

we obtain

$$\frac{|w|_a}{|w|} < \frac{\ell}{n} + \frac{1}{m}\left( \ell - \frac{(\xi - 1)(\ell - 1)}{2(2\xi - 1)} - \frac{\xi^2(\ell - 2)}{(\xi\ell - 1)(2\xi - 1)} \right). \quad \square$$

**Theorem 14.** *For any $k \ge 3$, $\ell \ge 3$, we have $R(k, \ell) \ge 1 + \frac{1}{m}$, where*

$$m = \left( \frac{5}{6}\ell - \frac{1}{2} + \frac{2}{2\ell - 1} \right) k.$$

**Proof.** Let $N > 0$ be an integer and $m_N = \left(\frac{5}{6}\ell - \frac{1}{2} + \frac{2}{2\ell-1} + \frac{1}{N}\right)k$. Let $v \in \Sigma_k^*$. Suppose $v$ is $(1 + \frac{1}{m_N}, \ell)$-free, and $|v| \ge N \cdot \ell \cdot m_N$.

As $k \ge 3$, we have $m_N \ge \frac{5}{2}\ell - \frac{3}{2} \ge \ell$. Thus, by Proposition 3, $v \in \mathfrak{L}(k, \ell, m_N)$. Moreover, as $\Sigma_k$ is the $k$-letter alphabet, there exists a letter $a$ in $\Sigma_k$ such that $\frac{|v|_a}{|v|} \ge \frac{1}{k}$. Then, by Proposition 13 with $\xi = 2$ (notice that $m_N \ge \frac{5}{2}\ell - \frac{3}{2} \ge 2\ell - 2$), and since $|v| \geq N \cdot \ell \cdot m_N$, we have

$$\frac{1}{k} \leq \frac{|v|_a}{|v|} < \frac{\ell}{|v|} + \frac{1}{m_N}\left(\frac{5}{6}\ell - \frac{1}{2} + \frac{2}{2\ell-1}\right) \leq \frac{1}{m_N}\left(\frac{5}{6}\ell - \frac{1}{2} + \frac{2}{2\ell-1} + \frac{1}{N}\right).$$

Then, we have

$$m_N < \left(\frac{5}{6}\ell - \frac{1}{2} + \frac{2}{2\ell-1} + \frac{1}{N}\right)k.$$

Hence we have a contradiction because $m_N = \left(\frac{5}{6}\ell - \frac{1}{2} + \frac{2}{2\ell-1} + \frac{1}{N}\right)k$, and

$$R(k,\ell) \geq 1 + \frac{1}{m_N}.$$

Finally, as $m = \lim_{N\to+\infty} m_N$, we have

$$R(k,\ell) \geq 1 + \frac{1}{m}. \quad \square$$

**Remark 15.** For $\ell = 2$, the result of Proposition 13 and the same arguments as in the proof of Theorem 14 give $R(k, 2) \geq 1 + \frac{1}{m}$ with $m = \frac{5k}{3}$, which is not as good as the bound of Theorem 4.

**Remark 16.** Theorem 14 gives a general result and its proof is based on estimations used in Proposition 13. In order to have better results in some particular cases, one could improve these estimations as follows.

– For small values of $\ell$, we can explicitly compute the value of the sum

$$\sum_{p=3}^{\ell} \left\lfloor \frac{\ell}{p} \right\rfloor \frac{1}{\left(p - \frac{1}{\xi}\right)\left(p - 1 - \frac{1}{\xi}\right)}$$

in Proposition 13, instead of taking a lower bound for it. For example, for $\ell = 6$, we get $R(k, 6) \geq 1 + \frac{1}{m}$, with $m = \frac{1457k}{330}$.
– In the proof of Theorem 14, we choose $\xi = 2$ in applying Proposition 13. As it appears clear from the statement of this latter proposition, a greater $\xi$ would give a better upper bound on the frequency $\frac{|w|_a}{|w|}$. Though, we need $\xi \leq \frac{m}{\ell-1}$ and this implies a constraint on $k$. For example, if $k \geq 5$ one can choose $\xi = 3$ and get $m = (\frac{4}{5}\ell - \frac{2}{5} + \frac{3}{3\ell-1})k$ .

## 4. Upper bound

In this section we explicitly use the fact that Dejean's conjecture is proved for $K \geq 5$. We describe a morphism from a $K$-letter alphabet to a $k$-letter one that transforms an infinite $\left(\frac{K}{K-1}\right)^+$-free word into a word in which the sufficiently large repetitions are of exponent not much greater than $\frac{K}{K-1}$.

Let $S_{k,t}$ be the set of words of length $t$ over $\Sigma_k$ of the form $0^e w$ where $e \geq 2$, $|w| \geq 1$, the first and the last letter of $w$ are different from 0, and $w$ does not contain 00 as a factor. For example, we have $S_{2,5} = \{00001, 00011, 00101, 00111\}$. Let $K = |S_{k,t}|$ and let $h$ be a $t$-uniform morphism $h : \Sigma_K^* \to \Sigma_k^*$ such that the set of $h$-images of letters in $\Sigma_K$ is $S_{k,t}$. Now, if $K \geq 5$, we consider the $h$-image of some infinite $\left(\frac{K}{K-1}\right)^+$-free word over $\Sigma_K$.

A uniform morphism $h : \mathcal{A}^* \to \mathcal{B}^*$ is said to be *comma-free* if for any $a, b, c \in \mathcal{A}$ and $s, r \in \mathcal{B}^*$, $h(ab) = rh(c)s$ implies that either $r = \varepsilon$ and $a = c$ or $s = \varepsilon$ and $b = c$.

**Remark 17.** A comma-free morphism $h$ is always injective (actually it is injective on the set $\mathcal{A}$ of monoid generators). Moreover, if it is $t$-uniform, then for each factor $u$ of a word in $h(\mathcal{A}^*)$ such that $|u| \geq 2t-1$, there exists a unique factorization $u = xh(u')y$ where $u' \in \mathcal{A}^*$ and $0 \leq |x|, |y| < t$.

**Lemma 18.** *The $t$-uniform morphism $h : \Sigma_K^* \to \Sigma_k^*$ defined above is comma-free.*

**Proof.** Suppose that the $h$ is not comma-free. Then there exist $a, b, c \in \Sigma_K$ and $s, r \in \Sigma_k^*$ such that $h(ab) = rh(c)s = w[1, \ldots, 2t]$ with $0 < |r| < t$. We obtain a contradiction for every possible value of $|r|$:

– if $|r| = 1$ or $|r| = 2$, then the letter $w[t + |r|]$ is 0 in $h(ab)$ and is not 0 in $rh(c)s$,
– if $|r| = t - 1$ or $|r| = t - 2$, then the letter $w[t]$ is not 0 in $h(ab)$ and is 0 in $rh(c)s$,
– if $2 < |r| < t - 2$, then $h(c)$ contains the factor $w[t, \ldots, t + 2]$. In $h(ab)$, this factor is of the form $x00$ with $x \neq 0$, whereas factors of this form do not exist in $h(c)$. $\square$

In order to get the mentioned repetition-freeness property in $h(\Sigma_K^*)$, we use the following lemma. For any real number $\ell \geq 1$, we write $(\alpha^+, \ell)$-free to mean $(\alpha^+, \lceil \ell \rceil)$-free and hence $R(k, \ell)$ to mean $R(k, \lceil \ell \rceil)$.

**Lemma 19.** *Let $\alpha, \beta \in \mathbb{R}$, $1 < \alpha < \beta < 2$. Let $h : \mathcal{A}^* \to \mathcal{B}^*$ be a comma-free $t$-uniform morphism. If $w \in \mathcal{A}^*$ is $\alpha^+$-free, then $h(w)$ is $(\beta^+, \frac{2t-2}{\beta-\alpha})$-free.*

**Proof.** Let $uvu$ be a $\beta'$-repetition in $h(w)$ with $\beta' > \beta$. Suppose $|u| \geq 2t - 1$. Hence $u$ contains an $h$-image and it can be uniquely written as $xh(u')y$ with $0 \leq |x|, |y| < t$. Thus this $h$-image $\bar{u} = h(u')$ appears at the same position of $u$ in $uvu$.

The factor $\bar{v} = yvx$ is an $h$-image. We have that $\bar{u}\bar{v}\bar{u}$ is the $h$-image of a repetition in $w$ and hence $\frac{|\bar{u}\bar{v}\bar{u}|}{|\bar{u}\bar{v}|} \leq \alpha$. Moreover, $\beta' = \frac{|\bar{u}\bar{v}\bar{u}|}{|\bar{u}\bar{v}|} + \frac{|x|+|y|}{|uv|}$ and $\beta' > \beta$ implies that $\frac{|x|+|y|}{|uv|} > \beta - \alpha$. Hence

$$|uv| < \frac{|xy|}{\beta - \alpha} \leq \frac{2t-2}{\beta - \alpha}.$$

Suppose now that $|u| \leq 2t - 2$. Hence $\beta' > \beta$ implies that $\frac{|u|}{|uv|} > \beta - 1$. Thus $\frac{|uv|}{|u|} < \frac{1}{\beta - 1}$ and

$$|uv| < \frac{|u|}{\beta - 1} \leq \frac{2t-2}{\beta - 1} < \frac{2t-2}{\beta - \alpha}. \quad \square$$

Recall that, by definition, if there exists a $(\beta^+, \ell)$-free infinite word over $\Sigma_k$ then $R(k, \ell) \leq \beta$.

**Corollary 20.** *Let $h : \Sigma_K^* \to \Sigma_k^*$ be a comma-free $t$-uniform morphism as above. If $\beta \in \mathbb{R}$, $\frac{K}{K-1} < \beta < 2$, and $K \geq 5$, then $R\left(k, \frac{2t-2}{\beta - \frac{K}{K-1}}\right) \leq \beta$.*

We now compute $K = K_{k,t}$. Consider the prefixes of length $(t-1)$ of the words in $S_{k,t}$:

- $K_{k,t-1}$ of them are such that the last letter is not 0,
- $K_{k,t-2}$ of them are such that the last letter is 0 and the penultimate letter is not 0,
- one them is the word $0^{t-1}$.

Each prefix can be extended by one of the $(k-1)$ letters distinct from 0 to get a word in $S_{k,t}$, so $K_{k,t}$ satisfies the recurrence relation

$$K_{k,1} = K_{k,2} = 0, \; K_{k,t} = (k-1)(K_{k,t-1} + K_{k,t-2} + 1).$$

Solving this relation, we obtain that

$$K_{k,t} = \frac{k-1}{(2k-3)\sqrt{(k-1)(k+3)}} \left(\lambda^t - \mu^t - (k-2)(\lambda^{t-1} - \mu^{t-1})\right) - \frac{k-1}{2k-3},$$

where $\lambda = \frac{(k-1)+\sqrt{(k-1)(k+3)}}{2}$ and $\mu = \frac{(k-1)-\sqrt{(k-1)(k+3)}}{2}$.

We thus have

$$K_{k,t} = C_k\lambda^{t-1} - O(1), \text{ where } C_k = \frac{(k-1)(\sqrt{(k-1)(k+3)} - k + 3)}{2(2k-3)\sqrt{(k-1)(k+3)}}.$$

For each real number $\alpha$, $\lfloor \alpha \rceil$ denotes the nearest integer to $\alpha$.

**Theorem 21.** *$R(k, \ell) \leq 1 + \frac{2\ln \ell}{\ell \ln \lambda} + O\left(\frac{1}{\ell}\right)$ if $k$ is fixed and $\ell$ tends to infinity.*

**Proof.** Let us fix $t = \left\lfloor \frac{\ln \ell}{\ln \lambda} \right\rceil + 1$ and $\beta = 1 + \frac{2t-2}{\ell} + \frac{1}{K-1}$. For $\ell$ sufficiently large, we have $t \geq 6$ which ensures that $K \geq 5$, and we also have $\beta < 2$. We can thus use Corollary 20, which gives

$$R(k, \ell) = R\left(k, \frac{2t-2}{\beta - \frac{K}{K-1}}\right) \leq \beta = 1 + \frac{2t-2}{\ell} + \frac{1}{K-1}.$$

Since

$$\frac{2t-2}{\ell} = \frac{2\left\lfloor \frac{\ln \ell}{\ln \lambda} \right\rceil}{\ell} = \frac{2\ln \ell}{\ell \ln \lambda} + O\left(\frac{1}{\ell}\right)$$

and

$$\frac{1}{K-1} = \frac{1}{C_k\lambda^{t-1} - O(1)} = \frac{1}{C_k\lambda^{\left\lfloor \frac{\ln \ell}{\ln \lambda} \right\rceil} - O(1)} = O\left(\frac{1}{\ell}\right),$$

the result follows. $\square$

## 5. An example

Let us illustrate our results with a concrete example: $k = 8$ and $\ell = 100$. Theorem 14 gives $R(8, 100) \geq 1 + \left(\left(\frac{5}{6} \times 100 - \frac{1}{2} + \frac{2}{199}\right) \times 8\right)^{-1} = 1.001508871\ldots$. For the upper bound, we have to decide which morphism $h$ will be used, or equivalently to choose the value of the parameter $t$. For a given $t$, we can compute $K = K_{k,t}$ and then the bound $\beta = 1 + \frac{2t-2}{\ell} + \frac{1}{K-1}$. So, we have to choose $t$ so that $\beta$ is minimized. The choice of $t = \left\lfloor \frac{\ln \ell}{\ln \lambda} \right\rceil + 1$ in Theorem 21 is well-suited to get such an asymptotic result, but for a given pair $(k, \ell)$ like this example, it is better to make a specific case study.

- If $t = 3$, then $K = 7$ and $\beta = \frac{181}{150} = 1.20666666\ldots$.
- If $t = 4$, then $K = 56$ and $\beta = \frac{593}{550} = 1.07818181\ldots$.
- If $t = 5$, then $K = 448$ and $\beta = \frac{12094}{11175} = 1.08223713\ldots$.

Since $\beta$ gets bigger if $t > 5$, the minimum is reached at $t = 4$, whereas $\left\lfloor \frac{\ln \ell}{\ln \lambda} \right\rceil + 1 = 3$. We thus obtain $R(8, 100) \leq \frac{593}{550} \leq 1.078182$.

## 6. Conclusion

For $k$ fixed and $\ell$ tending to infinity, we know now in particular that the asymptotics of the generalized repetition threshold $R(k, \ell)$ is between $1 + \Omega(1/\ell)$ and $1 + O(\ln \ell/\ell)$. New ideas are needed to settle this and other questions about $R(k, \ell)$, such as good estimates for $R(k, 2)$ or $R(k, k)$. The case $1.001848 < R(8, 100) < 1.078182$ suggests that there is still room for improvement.

## References


[1] F.-J. Brandenburg, Uniformly growing $k$-th power-free homomorphisms, Theoret. Comput. Sci. 23 (1983) 69–82.
[2] A. Carpi, On Dejean's conjecture over large alphabets, Theoret. Comput. Sci. 35 (2007) 137–151.
[3] J.D. Currie, M. Mohammad-Noori, Dejean's conjecture and Sturmian words, European J. Combin. 28 (3) (2007) 876–890.
[4] J.D. Currie, N. Rampersad, Dejean's conjecture holds for $n \geq 27$, Theor. Inform. Appl. 43 (4) (2009) 775–778.
[5] J.D. Currie, N. Rampersad, Dejean's conjecture holds for $n \geq 30$, Theoret. Comput. Sci. 345 (2005) 359–369.
[6] J.D. Currie, N. Rampersad, A proof of Dejean's conjecture, Math. Comp. 80 (2011) 1063–1070.
[7] F. Dejean, Sur un théorème de Thue, J. Combin. Theory. Ser. A 13 (1972) 90–99.
[8] L. Ilie, P. Ochem, J.O. Shallit, A generalization of repetition threshold, Theoret. Comput. Sci. 345 (2005) 359–369.
[9] R. Kolpakov, M. Rao, Personal communication.
[10] 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.
[11] J.-J. Pansiot, A propos d'une conjecture de F. Dejean sur les répétitions dans les mots, Discrete Appl. Math. 7 (1984) 297–311.
[12] M. Rao, Last cases of Dejean's conjecture, Theoret. Comput. Sci. 412 (2011) 3010–3018.
[13] A. Thue, Über unendliche Zeichenreihen, Christiania Vidensk.-Selsk, Skrifter. I. Mat. Nat. Kl 7 (1906) 1–22. Reprinted in: T. Nagell (Ed.), Selected Mathematical Papers of Axel Thue, Universitetsforlaget, Oslo, 1977, pp. 139–158.