Dejean's conjecture and letter frequency

Jérémie Chalopin and Pascal Ochem

LaBRI, Université Bordeaux 1
351 cours de la Libération, 33405 Talence Cedex, FRANCE
chalopin@labri.fr, ochem@labri.fr

Abstract. We prove two cases of a strong version of Dejean's conjecture involving extremal letter frequencies. The results are that there exist an infinite $\left(\frac{5}{4}^+\right)$-free word over a 5 letter alphabet with letter frequency $\frac{1}{6}$ and an infinite $\left(\frac{6}{5}^+\right)$-free word over a 6 letter alphabet with letter frequency $\frac{1}{5}$.

1 Introduction

We consider the extremal frequencies of a letter in factorial languages defined by an alphabet size and a set of forbidden repetitions. Given such a language L, we denote by $f_{\min}$ (resp. $f_{\max}$) the minimal (resp. maximal) letter frequency in an infinite word that belong to L. Letter frequencies have been mainly studied in [4, 6, 10, 9]. Let Σ_i denote the i-letter alphabet $\{0, 1, \ldots, i-1\}$. We consider here the frequency of the letter 0. Let $n(w)$ denote the number of occurrences of 0 in the finite word w. So the letter frequency in w is $\frac{n(w)}{|w|}$.

The *repetition threshold* is the least exponent $\alpha = \alpha(k)$ such that there exists an infinite (α^+)-free word over Σ_k. Dejean proved that $\alpha(3) = \frac{7}{4}$. She also conjectured that $\alpha(4) = \frac{7}{5}$ and $\alpha(k) = \frac{k}{k-1}$ for $k \geq 5$. In its full generality, this conjecture is still open, although Pansiot [8] proved that $\alpha(4) = \frac{7}{5}$ and Moulin-Ollagnier [5] proved that Dejean's conjecture holds for $5 \leq k \leq 11$. Recently, Currie and Mohammad-Noori [2] also proved the cases $12 \leq k \leq 14$. For more information, see [1].

In a previous paper, we proposed the following conjecture which implies Dejean's conjecture.

Conjecture 1. [6]

1. For every $k \geq 5$, there exists an infinite $\left(\frac{k}{k-1}^+\right)$-free word over Σ_k with letter frequency $\frac{1}{k+1}$.
2. For every $k \geq 6$, there exists an infinite $\left(\frac{k}{k-1}^+\right)$-free word over Σ_k with letter frequency $\frac{1}{k-1}$.

It is easy to see that the values $\frac{1}{k+1}$ and $\frac{1}{k-1}$ in Conjecture 1 are best possible. For $\left(\frac{5}{4}^+\right)$-free words over Σ_5, we obtained $f_{\max} < \frac{103}{440} = 0.23409090 \cdots < \frac{1}{4}$ [6]. That is why Conjecture 1.2 is stated with $k \geq 6$.

In this paper, we prove the first case of each part of Conjecture 1:

Theorem 1.

1. *There exists an infinite* $\left(\frac{5}{4}^+\right)$*-free word over* Σ_5 *with letter frequency* $\frac{1}{6}$.
2. *There exists an infinite* $\left(\frac{6}{5}^+\right)$*-free word over* Σ_6 *with letter frequency* $\frac{1}{5}$.

The C++ sources of the programs and the morphisms used in this paper are available at: `http://dept-info.labri.fr/~ochem/morphisms/`.

2 Structure and encoding

We consider an infinite $\left(\frac{k}{k-1}^+\right)$-free word w over Σ_k with letter frequency $\frac{1}{k+1}$ (resp. $\frac{1}{k-1}$). We easily check that 0's in w must be regularly spaced [1]. These words are the catenation of factors of size $k+1$ (resp. $k-1$) of the form $0\pi_1 \dots \pi_{k-1}\pi_1$ (resp. $0\pi_1 \dots \pi_{k-2}$), where π is a permutation of the elements $[1, \dots, k-1]$. Let Π denote the set of permutations of $[1, \dots, k-1]$. The word w can thus be encoded by the word $p \in \Pi^*$ consisting in the catenation of the permutations that correspond to the factors of size $k+1$ (resp. $k-1$) in w.

Let $p = p_0p_1p_2 \dots$ be the code of w and we suppose that p_0 is the identity. We now encode p by the word $c = c_0c_1c_2 \dots \in \Pi^*$ such that $p_{i+1} = c_i(p_i)$. Notice that whereas any permutation in Π may appear in p, only a small subset $S \subset \Pi$ of permutation can be used as letters in c. This is because the latter permutations rule the transition between two consecutive factors w_i and w_{i+1} in w, and then w_iw_{i+1} has to be $\left(\frac{k}{k-1}^+\right)$-free.

3 Sketch of proof of main result

Let us consider the possible transitions for $\left(\frac{5}{4}^+\right)$-free words over Σ_5 with letter frequency $\frac{1}{6}$. There are exactly two of them:

– 012341024312 correspond to the transition permutation 2431 (noted 0).
– 012341032143 correspond to the transition permutation 3214 (noted 1).

There are also exactly two possible transitions for $\left(\frac{6}{5}^+\right)$-free words over Σ_6 with letter frequency $\frac{1}{5}$:

– 0123405132 correspond to the transition permutation 51324 (noted 0).
– 0123405213 correspond to the transition permutation 52134 (noted 1).

In both cases, we have $|S| = 2$ and we construct a suitable infinite code c as the fixed point of the following binary endomorphisms:

[1] Notice that 0's cannot be regularly spaced if the letter frequency is $\frac{1}{k}$.

– For $\left(\frac{5}{4}^+\right)$-free words over Σ_5 with letter frequency $\frac{1}{6}$:
$0 \mapsto 010010010100101001001001010100101001001001010010101001001001010$
$1 \mapsto 100101001001010100101001001010100101001001001010010010010100101$

– For $\left(\frac{6}{5}^+\right)$-free words over Σ_6 with letter frequency $\frac{1}{5}$:
$0 \mapsto 001001010011100011010001 0$
$1 \mapsto 1000100111000100110100011$

These morphisms m satisfy the following properties:

1. m is q-uniform.
2. m is *synchronizing*, which means that for any $a, b, c \in \Sigma_2$ and $s, r \in \Sigma_2^*$, if $h(ab) = rh(c)s$, then either $r = \varepsilon$ and $a = c$ or $s = \varepsilon$ and $b = c$.
3. for all $i \in \Sigma_2$, $m(i) = il = fi$, where the transition corresponding to the factors l and f is the identity.

Let Φ denote the decoding function, we thus have $\Phi(0) = 0123405132$, $\Phi(1) = 0123405213$ and $\Phi(c = m^\omega(0)) = w$.

We have checked that for every factor x of c of size at most $2kq$, $\Phi(x)$ is $\left(\frac{k}{k-1}^+\right)$-free.

Let f be a smallest repetition in w of exponent strictly greater than $\frac{k}{k-1}$. This repetition in w implies that there is a repetition $r = is$ in c whose prefix i is an identity. Since $|s| \geq 2q$, s contains a full m-image. So $|i|$ and $|s|$ are multiples of q because m is synchronizing. Notice that a conjugate of an identity is an identity. By property 3, we can thus assume w.l.o.g that $|i|$ starts at the beginning of an m-image. Then our repetition is of the form $r = is = m(i')m(s') = m(i's') = m(r')$. By property 3, r' is a repetition whose prefix i' is an identity. Now, since m is a uniform morphism, the exponent of r' is equal to the exponent r. This is a contradiction because $\Phi(r')0$ is a repetition of exponent strictly greater than $\frac{k}{k-1}$ which must appear in w and is smaller than f.

4 Concluding remarks

Theorem 1 shows the existence of two types of infinite words, but it does not prove that there exist exponentially many such words (which is probably true). On the other hand, the growth rate of these words is smaller than those of $\left(\frac{k}{k-1}^+\right)$-free words. For example, the growth rate of $\left(\frac{5}{4}^+\right)$-free 5-ary words is about 1.159 [7], whereas 1.048 is a rough upper bound on the growth rate of $\left(\frac{5}{4}^+\right)$-free 5-ary words with letter frequency $\frac{1}{6}$.

Other cases of Conjecture 1 might be harder to settle. For $\left(\frac{6}{5}^+\right)$-free 6-ary words with letter frequency $\frac{1}{7}$, we have $|S| = 3$, and it is impossible to construct an infinite code using only two of these three transition permutations. We have not been able to find a $\Sigma_3^* \to \Sigma_3^*$ morphism with suitable properties for them.

References

1. C. Choffrut and J. Karhumäki. Combinatorics of words, In G. Rozenberg and A. Salomaa, editors, *Handbook of Formal Languages*, Vol. 1, pp. 329–438. Springer-Verlag, 1997.
2. J.D. Currie and M. Mohammad-Noori. Dejean's conjecture and sturmian words, *Manuscript*.
3. F. Dejean. Sur un théorème de Thue, *J. Combin. Theory. Ser. A* **13** (1972), 90–99.
4. R. Kolpakov, G. Kucherov, and Y. Tarannikov. On repetition-free binary words of minimal density, *Theoret. Comput. Sci.* **218** (1999), 161–175.
5. 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.
6. P. Ochem. Letter frequency in infinite repetition-free words, In *5th International Conference on Words (Words 2005)*, Montreal, Canada, September 13-17 2005. Publications du LaCIM, vol. 36
7. P. Ochem and T. Reix. Upper bound on the number of ternary square-free words, In *Workshop on Words and Automata (WOWA'06)*, St. Petersburg, Russia, June 7 2006.
8. J.-J. Pansiot. A propos d'une conjecture de F. Dejean sur les répétitions dans les mots, *Disc. Appl. Math.* **7** (1984), 297–311.
9. C. Richard and U. Grimm. On the entropy and letter frequencies of ternary square-free words, *Electron. J. Comb.* **11** (2004), #R14
10. Y. Tarannikov. The minimal density of a letter in an infinite ternary square-free word is 0.2746..., *J. Integer Sequences* 5(2):Article 02.2.2 (2002).