Repetition avoidance in products of factors

Pamela Fleischmann, Pascal Ochem, Kamellia Reshadi

Repetition avoidance in products of factors

Pamela Fleischmann* Pascal Ochem† Kamellia Reshadi‡

Abstract

We consider a variation on a classical avoidance problem from combinatorics on words that has been introduced by Mousavi and Shallit at DLT 2013. Let $\texttt{pexp}_i(w)$ be the supremum of the exponent over the products of i factors of the word w. The repetition threshold $\mathrm{RT}_i(k)$ is then the infimum of $\texttt{pexp}_i(w)$ over all words $w \in \Sigma_k^\omega$. Moussavi and Shallit obtained that $\mathrm{RT}_i(2) = 2i$ and $\mathrm{RT}_2(3) = \frac{13}{4}$. We show that $\mathrm{RT}_i(3) = \frac{3i}{2} + \frac{1}{4}$ if i is even and $\mathrm{RT}_i(3) = \frac{3i}{2} + \frac{1}{6}$ if i is odd and $i \geqslant 3$.

Keywords: Words; Repetition avoidance.

Acknowledgements: This work was partially supported by the ANR project CoCoGro (ANR-16-CE40-0005).

1 Introduction

A *repetition* in a word w is a pair of words p and e such that pe is a factor of w, p is non-empty, and e is a prefix of pe. If pe is a repetition, then its *period* is $|p|$ and its *exponent* is $\frac{|pe|}{|p|}$. A word is α^+*-free* (resp. α*-free*) if it contains no repetition with exponent β such that $\beta > \alpha$ (resp. $\beta \geqslant \alpha$).

Given $k \geqslant 2$, Dejean [2] defined the repetition threshold $\mathrm{RT}(k)$ for k letters as the smallest α such that there exists an infinite α^+-free word over a k-letter alphabet. Dejean initiated the study of $\mathrm{RT}(k)$ in 1972 for $k = 2$ and $k = 3$. Her work was followed by a series of papers which determine the exact value of $\mathrm{RT}(k)$ for any $k \geqslant 2$.

*Kiel University, Germany. fpa@informatik.uni-kiel.de
†LIRMM, Université de Montpellier, CNRS, Montpellier, France. ochem@lirmm.fr
‡Kiel University, Germany. kre@informatik.uni-kiel.de

- $\mathrm{RT}(2) = 2$ [2];
- $\mathrm{RT}(3) = \frac{7}{4}$ [2];
- $\mathrm{RT}(4) = \frac{7}{5}$ [7];
- $\mathrm{RT}(k) = \frac{k}{k-1}$, for $k \geqslant 5$ [1, 4, 8].

Mousavi and Shallit [5] have considered two notions related to the repetition threshold.

The first notion considers repetitions in conjugates of factors of the infinite word. A word is circularly r^+-free if it does not contain a factor pxs such that sp is a repetition of exponent strictly greater than r. Let $\Sigma_k = \{0, 1, \ldots, k-1\}$. The smallest real number r such that w is circularly r^+-free is denoted by $\mathtt{cexp}(w)$. Let $\mathrm{RTC}(k)$ be the minimum of $\mathtt{cexp}(w)$ over every $w \in \Sigma_k^\omega$.

The second notion considers repetitions in concatenations of a fixed number of factors of the infinite word. Let $\mathtt{pexp}_i(w)$ be the smallest real number r such that every product of i factors of w is r^+-free. Let $\mathrm{RT}_i(k)$ be the minimum of $\mathtt{pexp}_i(w)$ over every $w \in \Sigma_k^\omega$. Notice that $\mathrm{RT}_i(k)$ generalizes the classical notion of repetition threshold which corresponds to the case $i = 1$, that is, $\mathrm{RT}_1(k) = \mathrm{RT}(k)$ for every $k \geqslant 2$.

Our first result shows that the case $i = 2$ corresponds to the first notion of repetition avoidance in conjugates.

Theorem 1. $\mathrm{RT}_2(k) = \mathrm{RTC}(k)$ *for every* $k \geqslant 2$.

Mousavi and Shallit [5] have considered the binary alphabet and obtained that $\mathrm{RT}_i(2) = 2i$ for every $i \geqslant 1$. Our second result considers the ternary alphabet and gives the value of $\mathrm{RT}_i(3)$ for every $i \geqslant 1$. This extends the result of Dejean [2] that $\mathrm{RT}_1(3) = \frac{7}{4}$ and the result of Mousavi and Shallit [5] that $\mathrm{RT}_2(3) = \frac{13}{4}$.

Theorem 2.

- $\mathrm{RT}_i(3) = \frac{3i}{2} + \frac{1}{4}$ *if* $i = 1$ *or* i *is even.*
- $\mathrm{RT}_i(3) = \frac{3i}{2} + \frac{1}{6}$ *if* i *is odd and* $i \geqslant 3$.

2 Proofs

Proof of Theorem 1.
The language of words in Σ_k^* avoiding circular repetitions of exponent at least e (or strictly greater than e) is a factorial language. As it is well-known [3], if a factorial language is infinite, then it contains a uniformly recurrent word w. By Proposition 14 in [5], $\texttt{pexp}_2(w) = \texttt{cexp}(w)$. This implies that $\mathrm{RT}_2(k) = \mathrm{RTC}(k)$. $\square$

To obtain the two equalities of Theorem 2, we show the two lower bounds and then the two upper bounds.

Proof of $\mathrm{RT}_i(3) \geqslant \frac{3i}{2} + \frac{1}{4}$ *for every even* i.
Mousavi and Shallit [5] have proved that $\mathrm{RT}_2(3) = \frac{13}{4}$, which settles the case $i = 2$. We have double checked their computation of the lower bound $\mathrm{RT}_2(3) \geqslant \frac{13}{4}$. Suppose that i is a fixed even integer and that w_3 is an infinite ternary word. The lower bound for $i = 2$ implies that there exists two factors u and v such that $uv = t^e$ with $e \geqslant \frac{13}{4}$. Thus, the prefix t^3 of uv is also a product of two factors of w_3. So we can form the i-terms product $(t^3)^{i/2-1}uv$ which is a repetition of the form t^x with exponent $x = 3\left(\frac{i}{2} - 1\right) + e \geqslant 3\left(\frac{i}{2} - 1\right) + \frac{13}{4} = \frac{3i}{2} + \frac{1}{4}$. This is the desired lower bound. $\square$

Proof of $\mathrm{RT}_i(3) \geqslant \frac{3i}{2} + \frac{1}{6}$ *for every odd* $i \geqslant 3$.
Suppose that $i \geqslant 3$ is a fixed odd integer, that is, $i = 2j+1$. Suppose that w_3 is a recurrent ternary word such that the product of i factors of w_3 is never a repetition of exponent at least $\frac{3i}{2} + \frac{1}{6} = 3j + \frac{5}{3}$. First, w_3 is square-free since otherwise there would exist an i-terms product of exponent $2i$. Also, w_3 does not contain two factors u and v with the following properties:

- $uv = t^3$,
- $u = t^e$ with $e \geqslant \frac{5}{3}$.

Indeed, this would produce the i-terms product $(uv)^j u$ which is a repetition of the form t^x with exponent $x = 3j + e \geqslant 3j + \frac{5}{3}$.

So if a, b, and c are distinct letters, then w_3 does not contain both $u = abcab$ and $v = cabc$ and w_3 does not contain both $u = abcbabc$ and $v = babcb$. A computer check shows that no infinite ternary square-free word satisfies this property. This proves the desired lower bound. $\square$

Proof of $\mathrm{RT}_i(3) \leqslant \frac{3i}{2} + \frac{1}{4}$ *for every even* i.
Let i be any even integer at least 2. To prove this upper bound, it is sufficient to construct a ternary word w satisfying $\mathtt{pexp}_i(w) \leqslant \frac{3i}{2} + \frac{1}{4}$. The ternary morphic word used in [5] to obtain $\mathrm{RT}_2(3) \leqslant \frac{13}{4}$ seems to satisfy the property. However, it is easier for us to consider another construction. Let us show that the image of every $7/5^+$-free word over Σ_4 by the following 45-uniform morphism satisfies $\mathtt{pexp}_i \leqslant \frac{3i}{2} + \frac{1}{4}$.

$$
\begin{aligned}
\mathsf{0} &\mapsto \mathsf{010201210212021012102010212012101202101210212}\\
\mathsf{1} &\mapsto \mathsf{010201210212012101202101210201021202101210212}\\
\mathsf{2} &\mapsto \mathsf{010201210120212012102120210121021201210120212}\\
\mathsf{3} &\mapsto \mathsf{010201210120210121021201210120212012102010212}
\end{aligned}
$$

Recall that a word is (β^+, n)-free if it does not contain a repetition with period at least n and exponent strictly greater than β. First, we check that such ternary images are $\left(\frac{202}{135}^+, 36\right)$-free using the method in [6]. By Lemma 2.1 in [6], it is sufficient to check this freeness property for the image of every $7/5^+$-free word over Σ_4 of length smaller than $\frac{2\times\frac{202}{135}}{\frac{202}{135}-\frac{7}{5}} < 32$. Since $\frac{202}{135} < \frac{3}{2}$, the period of every repetition formed from i pieces and with exponent at least $\frac{3i}{2}$ must be at most 35. Then we check exhaustively by computer that the ternary images do not contain two factors u and v such that

- $uv = t^e$,
- $e > 3$,
- $9 \leqslant |t| \leqslant 35$.

Thus, the period of every repetition formed from i pieces and with exponent strictly greater than $\frac{3i}{2}$ must be at most 8. So we only need to check that $\mathtt{pexp}_i \leqslant \frac{3i}{2} + \frac{1}{4}$ for i-terms products that are repetitions of period at most 8.

Now the period is bounded, but i can still be arbitrarily large, a priori. For every factor t of length at most 8, we define $\mathtt{pexp}_{i,t}$ as the length of a largest factor of t^ω that is a i-terms product, divided by $|t|$. We actually consider conjugacy classes, since if t' is a conjugate of t, then $\mathtt{pexp}_{i,t'} = \mathtt{pexp}_{i,t}$. Let t be such a factor. If, for some even j, we have $\mathtt{pexp}_{j+2,t} = \mathtt{pexp}_{j,t} + 3$, then it means that by appending a 2-terms product to a j-terms product that corresponds to a maximum factor of t^ω, that can only add a cube of period $|t|$. This implies that for every k, $\texttt{pexp}_{j+2k,t} = \texttt{pexp}_{j,t} + 3k$.

We have checked by computer that for every conjugacy class of words t of length at most 8, there exists a (small) even j such that $\texttt{pexp}_{j+2,t} = \texttt{pexp}_{j,t} + 3$. Thus we have $\texttt{pexp}_i \leqslant \frac{3i}{2} + \frac{1}{4}$ in all cases. $\square$

Proof of $\mathrm{RT}_i(3) \leqslant \frac{3i}{2} + \frac{1}{6}$ *for every odd* $i \geqslant 3$.
Let us show that the image of every $7/5^+$-free word over Σ_4 by the following 514-uniform morphism satisfies $\texttt{pexp}_i \leqslant \frac{3i}{2} + \frac{1}{6}$ for every odd $i \geqslant 3$.

0 $\mapsto$ 01020120210120102120210201210120102012021020121021201020121012
02102012102120210120102012102120102012021020121012010212021020
12102120102012021012010212021020121021202101201020121021201020
12101202102012102120210120102120210201210120102012021020121012
01021202102012102120102012101202102012102120102012021012010212
02102012101201020120210201210212021012010201210120210201210212
01020120210201210120102120210201210212010201210120210201210212
02101201021202102012101201020120210201210120102120210201210212
021012010201210212

1 $\mapsto$ 01020120210120102120210201210120102012021020121021201020121012
02102012102120102012021020121012010212021020121021201020120210
12010212021020121021202101201020121021201020121012021020121021
20210120102120210201210120102012021020121021201020121012021020
12101201021202102012102120210120102012102120102012021012010212
02102012101201020120210201210120102120210201210212010201210120
21020121021202101201021202102012101201020120210201210212021012
01020121012021020121021201020120210201210120102120210201210212
021012010201210212

2 $\mapsto$ 01020120210120102120210201210120102012021020121021201020121012
02102012101201021202102012102120102012021012010212021020121021
20210120102012102120102012101202102012102120210120102120210201
21012010201202102012101201021202102012102120102012101202102012
10212010201202101201021202102012102120210120102012102120102012
02102012101201021202102012102120102012021012010212021020121012
01020120210201210212021012010201210212010201210120210201210212
02101201021202102012101201020120210201210120102120210201210212
021012010201210212

3 $\mapsto$ `010201202101201021202102012101201020120210201210212010201210120210201210120102120210201210212010201202101201021202102012102120210120102012101202102012101201021202102012102120102012021012010212021020121021202101201020121012021020121021201020121012021020121021202101201021202102012102120102012101202102012102120210120102120210201210120102012021020121012010212021020121021202101201020121012021020121021202101201020121012021020121021202101201020121012021020121021202101201020121012021020121021202101201020121012021020121021202101201020121012021020121021201020121012021020121021202101201020121012021020121021202101201020121012021020121021202101201020121012021020121021202101201020121012021020121021202101201020`

First, we check that such ternary images are $\left(\frac{3}{2}^+, 45\right)$-free using the method in [6]. By Lemma 2.1 in [6], it is sufficient to check this freeness property for the image of every $7/5^+$-free word over Σ_4 of length smaller than $\frac{2\times\frac{3}{2}}{\frac{3}{2}-\frac{7}{5}} = 30$. Thus, the period of every repetition formed from i pieces and with exponent strictly greater than $\frac{3i}{2}$ must be at most 44. Using the same argument as in the previous proof, we have checked by computer that for every conjugacy class of words t of length at most 44, there exists a (small) odd j such that $\mathtt{pexp}_{j+2,t} = \mathtt{pexp}_{j,t} + 3$. Thus we have $\mathtt{pexp}_i \leqslant \frac{3i}{2} + \frac{1}{6}$ in all cases. $\square$

3 Concluding remarks

The next step would be to consider the 4-letter alphabet. Obviously, $\mathrm{RT}_{i+1}(k) \geqslant \mathrm{RT}_i(k) + 1$ for every $i \geqslant 1$ and $k \geqslant 2$. Mousavi and Shallit [5] verified that $\mathrm{RT}_2(4) \geqslant \frac{5}{2}$, so that $\mathrm{RT}_i(4) \geqslant i + \frac{1}{2}$ for every $i \geqslant 2$. We conjecture that this is best possible, i.e., that $\mathrm{RT}_i(4) = i + \frac{1}{2}$ for every $i \geqslant 2$. However, a proof of an upper bound of the form $\mathrm{RT}_i(4) \leqslant i + c$ cannot be similar to the proof of the upper bounds of Theorem 2. The multiplicative factor of i, which drops from $\frac{3}{2}$ when $k = 3$ to 1 when $k = 4$, forbids that the constructed word is the morphic image of any (unspecified) Dejean word over a given alphabet.

References

[1] A. Carpi. On Dejean's conjecture over large alphabets. *Theoret. Comput. Sci.*, 385(1–3):137–151, 2007.

[2] F. Dejean. Sur un théorème de Thue, *J. Combin. Theory. Ser. A*, 13:90–99, 1972.

[3] Pytheas Fogg. Substitutions in Dynamics, Arithmetics and Combinatorics. Springer Science & Business Media, 2002.

[4] J. Moulin Ollagnier. Proof of Dejean's conjecture for alphabets with 5, 6, 7, 8, 9, 10, and 11 letters. *Theoret. Comput. Sci.*, 95(2):187–205, 1992.

[5] H. Mousavi and J. Shallit. Repetition avoidance in circular factors. *Developments in Language Theory 2013*, 384–395.

[6] P. Ochem. A generator of morphisms for infinite words. *RAIRO - Theor. Inform. Appl.*, 40:427–441, 2006.

[7] J.-J. Pansiot. A propos d'une conjecture de F. Dejean sur les répétitions dans les mots. *Discrete Appl. Math.*, 7(3):297–311, 1984.

[8] M. Rao. Last cases of Dejean's conjecture. *Theoret. Comput. Sci.*, 412(27):3010–3018, 2011.

[9] A. Thue. Über unendliche Zeichenreihen. *Norske Vid. Selsk. Skr. I. Mat. Nat. Kl. Christiania*, 7:1–22, 1906.