Repetition avoidance in products of factors

Pamela Fleischmann

Institut für Informatik, Christian-Albrechts-Universität zu Kiel, Kiel, Germany

fpa@informatik.uni-kiel.de

Pascal Ochem

LIRMM, CNRS, Univ. Montpellier, Montpellier, France

ochem@lirmm.fr

Kamellia Reshadi

Institut für Informatik, Christian-Albrechts-Universität zu Kiel, Kiel, Germany

kre@informatik.uni-kiel.de

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 $\mathrm{RT}_i(3) \geqslant \frac{3i}{2} + \frac{1}{6}$ if i is odd and $i \geqslant 3$.

Keywords: Words; Repetition avoidance.

1 Main results

Mousavi and Shallit [2] have considered two generalizations of the avoidance of fractional repetitions in infinite words. A word is circularly r^+-power-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 r^+-power-free is denoted by $\texttt{cexp}(w)$. Let $\mathrm{RTC}(k)$ denote the minimum of $\mathtt{cexp}(w)$ over every $w \in \Sigma_k^\omega$. Similarly, $\mathtt{pexp}_i(w)$ is the smallest real number r such that every product of i factors of w is r^+-power-free word and $\mathrm{RT}_i(k)$ is the minimum of $\mathtt{pexp}_i(w)$ over every $w \in \Sigma_k^\omega$.

In this paper, we consider the ternary alphabet. We obtain bounds on $\mathrm{RT}_i(3)$ which extend the result of Mousavi and Shallit that $\mathrm{RT}_2(3) = \frac{13}{4}$.

Proposition 1. $\mathrm{RT}_2(k) = \mathrm{RTC}(k)$.

Proof. 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 [1], if a factorial language is infinite, then it contains a uniformly recurrent word w. By Proposition 14 in [2], $\mathtt{pexp}_2(w) = \mathtt{cexp}(w)$. This implies that $\mathrm{RT}_2(k) = \mathrm{RTC}(k)$. □

Proposition 2. *If i is even and $i \geqslant 2$, then* $\mathrm{RT}_i(3) \geqslant \frac{3i}{2} + \frac{1}{4}$.

Proof. Mousavi and Shallit [2] 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 2-terms product of 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. □

Proposition 3. *If i is odd and $i \geqslant 3$, then* $\mathrm{RT}_i(3) \geqslant \frac{3i}{2} + \frac{1}{6}$.

Proof. 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. □

Proposition 4. *If i is even and $i \geqslant 2$, then* $\mathrm{RT}_i(3) \leqslant \frac{3i}{2} + \frac{1}{4}$.

Proof. 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 [2] 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 $\mathtt{pexp}_i \leqslant \frac{3i}{2} + \frac{1}{4}$.

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

First, we check that such ternary images are $\left(\frac{202}{135}^+, 36\right)$-free using the method in [3]. 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 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. Finally, we check exhaustively that $\mathtt{pexp}_i \leqslant \frac{3i}{2} + \frac{1}{4}$ by considering only i-terms products that are repetitions of period at most 8. □

2 Concluding remarks

We conjecture that $\mathrm{RT}_i(3) = \frac{3i}{2} + \frac{1}{6}$ for every odd $i \geqslant 3$, based on numerical evidence. We hope to get a suitable morphism and a proof of this case in the near future. Then the next step would be to consider the 4-letter alphabet. A quick computer check shows that $\mathrm{RT}_i(4) \geqslant i + \frac{1}{2}$ for every $i \geqslant 2$ and we conjecture that this is best possible. However, a proof of an upper bound of the form $\mathrm{RT}_i(4) \leqslant i + c$ cannot be similar to the proof of Proposition 4. That is because 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 a morphic image of a Dejean word.

References

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

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

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

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