A family of formulas with reversal of arbitrarily high avoidability index

Pascal Ochem*

LIRMM, CNRS, Université de Montpellier

France

ochem@lirmm.fr

March 15, 2021

Abstract

We present a family of avoidable formulas with reversal whose avoidability index is unbounded. We also complete the determination of the avoidability index of the formulas with reversal in the 3-avoidance basis.

1 Introduction

The notion of formula with reversal [2, 3] is an extension of the notion of classical formula such that a variable x can appear both as x and x^R with the convention that in an occurrence h of the formula, $h(x^R)$ is the reverse (i.e., mirror image) of $h(x)$. The avoidability index $\lambda(F)$ of a formula with reversal F is the minimum number of letters contained in an infinite word avoiding F.

Currie, Mol, and Rampersad [2] have asked if there exist formulas with reversal with arbitrarily large avoidability index. They considered the formula $\psi_k = xy_1y_2 \cdots y_kx \cdot y_1^R \cdot y_2^R \cdots y_k^R$ and obtained that $\lambda(\psi_1) = 4$, $\lambda(\psi_2) = \lambda(\psi_3) = \lambda(\psi_6) = 5$, $5 \le \lambda(\psi_4) \le 6$, $5 \le \lambda(\psi_5) \le 7$, $4 \le \lambda(\psi_k) \le 6$ if $k \ge 7$

*The author was partially supported by the ANR project CoCoGro (ANR-16-CE40-0005).

and $k \not\equiv 0 \pmod 3$, and $4 \leq \lambda(\psi_k) \leq 5$ if $k \geq 9$ and $k \equiv 0 \pmod 3$. They conjecture that $\lambda(\psi_k) = 5$ for all $k \geq 2$. Computational experiments suggest that the upper bound $\lambda(\psi_k) \leq 5$ for $k \geq 3$ is witnessed by the image of every $\left(\frac{7}{4}^+\right)$-free ternary word by the following $(k+3)$-uniform morphism where $k = 3t + i$, $t \geq 1$, and $0 \leq i \leq 2$.

$$\begin{aligned} 0 &\to (012)^{t+1-i}(0123)^i \\ 1 &\to (013)^{t+1-i}(0134)^i \\ 2 &\to (014)^{t+1-i}(0142)^i \end{aligned}$$

We give a positive answer to their original question with Theorem 1 below.

We define the formula $\phi_k = x_0x_1 \cdot x_1x_2 \cdots x_{k-1}x_0 \cdot x_0^R \cdot x_1^R \cdots x_{k-1}^R$.

Theorem 1. *For every fixed b, there exists k such that $b < \lambda(\phi_k) \leq k + 1$.*

Currie, Mol, and Rampersad [3] have also determined the 3-avoidance basis for formulas with reversal, which contains the minimally avoidable formulas with reversal on 3 variables. They obtained several bounds on the avoidability index of the formulas with reversal in the 3-avoidance basis. The next two results finish the determination of the avoidability index of these formulas.

Theorem 2. *The formulas $xyzyx \cdot zyxy^Rz$, $xyzyx \cdot zy^Rxyz$, $xyzyx \cdot zy^Rxy^Rz$, $xyzy^Rx \cdot zyxy^Rz$, and $xyzy^Rx \cdot zy^Rxyz$ are simultaneously 2-avoidable.*

Theorem 3. *The formulas $xyzx \cdot yzxy \cdot z^R$ and $xyzx \cdot yz^Rxy$ are simultaneously 3-avoidable.*

Theorems 1 to 3 are proved in Sections 2 to 4.

A word w is *d-directed* if for every factor f of w of length d, the word f^R is not a factor of w.

Remark 4. *If a d-directed word contains an occurrence h of $x \cdot x^R$, then $|h(x)| \leq d - 1$.*

In order to express the simultaneous avoidance of similar formulas, as in Theorems 2 and 3, we introduce the notation x^U to represent equality up to mirror image. That is, if $h(x) = w$, then $h(x^R) = w^R$ and $h(x^U) \in \{w, w^R\}$. For example, avoiding $xyxy$ and xyx^Ry simultaneously is equivalent to avoid xyx^Uy. Notice that the notion of undirected avoidability recently considered by Currie and Mol [1] corresponds to the case where every occurrence of every variable of the pattern/formula is equipped with $-^U$.

Recall that a word is (β^+, n)-free if it contains no repetition with exponent strictly greater than β and period at least n. Also, a word is (β^+)-free if it is $(\beta^+, 1)$-free.

2 Formulas with unbounded avoidability index

Let us first show that for every $k \geq 2$, ϕ_k is avoided by the periodic word $(\ell_0\ell_1 \cdots \ell_k)^\omega$ over $(k+1)$ letters. This word is 2-directed, so every occurrence h of ϕ_k is such that $|h(x_i)| = 1$ for every $0 \leq i < k$ by Remark 4. Without loss of generality, $h(x_0) = \ell_0$. This forces $h(x_1) = \ell_1$, $h(x_2) = \ell_2$, and so on until $h(x_{k-1}) = \ell_{k-1}$ and $h(x_0) = \ell_k$, which contradicts $h(x_0) = \ell_0$. Thus $\lambda(\phi_k) \leq k + 1$.

Let b be an integer and let w be an infinite word on at most b letters. Consider the Rauzy graph R of w such that the vertices of R are the letters of w and for every factor uv of length two in w, we put the arc $\overrightarrow{uv}$ in R. So R is a directed graph, possibly with loops (circuits of length 1) and digons (circuits of length 2). Since w is infinite, every vertex of R has out-degree at least 1. So R contains a circuit C_i of length i with $1 \leq i \leq b$. Let $c_0, c_1, \ldots, c_{i-1}$ be the vertices of C_i in cyclic order. Let k be the least common multiple of $1, 2, \ldots, b$. Since i divides k, w contains the occurrence h of ϕ_k such that $h(x_j) = c_{j \bmod i}$ for every $0 \leq j < k$. Thus $\lambda(\phi_k) > b$.

3 Formulas that flatten to $xyzyx \cdot zyxyz$

Notice that avoiding simultaneously the formulas in Theorem 2 is equivalent to avoiding $F = xyzy^Ux \cdot zy^Uxy^Uz \cdot y^R$. The fragment y^R is here to exclude the classical formula $xyzyx \cdot zyxyz$. Indeed, even though Gamard et al. [4] obtained that $\lambda(xyzyx \cdot zyxyz) = 2$, a computer check shows that $xyzyx \cdot zyxyz$ and F cannot be avoided simultaneously over two letters, that is, $xyzy^Ux \cdot zy^Uxy^Uz$ is not 2-avoidable.

We use the method in [5] to show that the image of every $\left(\frac{7}{5}^+\right)$-free word over Σ_4 by the following 21-uniform morphism is $\left(\frac{22}{15}^+, 85\right)$-free. We also check that such a binary word is 11-directed.

$$\begin{array}{rl} 0 \to & 00001011100011110011\,1 \end{array}$$

Consider an occurrence h of F. Since F contains $y \cdot y^R$, then $|h(y)| \leq 10$ by Remark 4. Suppose that $|h(xz)| \geq 83$. Then $h(xyzy^Ux)$ is a repetition with period $|h(xyzy)| \geq 85$. This implies $\frac{|h(xyzyx)|}{|h(xyzy)|} \leq \frac{22}{15}$, which gives $|h(x)| \leq \frac{7}{8}|h(yzy)|$. Since $|h(y)| \leq 10$, we deduce $|h(x)| \leq \frac{35}{2} + \frac{7}{8}|h(z)|$. Symmetrically, considering the repetition $h(zy^Uxy^Uz)$ gives $|h(z)| \leq \frac{35}{2} + \frac{7}{8}|h(x)|$. So

$$|h(x)| \leq \tfrac{35}{2} + \tfrac{7}{8}|h(z)| \leq \tfrac{35}{2} + \tfrac{7}{8}\left(\tfrac{35}{2} + \tfrac{7}{8}|h(x)|\right) = \tfrac{525}{16} + \tfrac{49}{64}|h(x)|$$

and

$$|h(x)| \leq \frac{\frac{525}{16}}{1 - \frac{49}{64}} = 140.$$

Symmetrically, $|h(z)| \leq 140$.

In every case, $|h(x)| \leq 140$, $|h(z)| \leq 140$, and $|h(y)| \leq 10$. Thus we can check exhaustively that h does not exist.

4 The formulas $xyzx \cdot yzxy \cdot z^R$ and $xyzx \cdot yz^Rxy$

Notice that avoiding $xyzx \cdot yzxy \cdot z^R$ and $xyzx \cdot yz^Rxy$ simultaneously is equivalent to avoiding $F = xyzx \cdot yz^Uxy \cdot z^R$. We use the method in [5] to show that the image of every $\left(\frac{7}{5}^+\right)$-free word over Σ_4 by the following 9-uniform morphism is $\left(\frac{131}{90}^+, 28\right)$-free. We also check that such a ternary word is 4-directed.

$$\begin{array}{rl} 0 \to & \texttt{011122202} \\ 1 \to & \texttt{010121202} \\ 2 \to & \texttt{001112122} \\ 3 \to & \texttt{000101120} \end{array}$$

Consider an occurrence h of F. Since F contains $z \cdot z^R$, then $|h(z)| \leq 3$ by Remark 4. Suppose that $|h(xy)| \geq 27$. Then $h(xyzx)$ is a repetition with period $|h(xyz)| \geq 28$. This implies $\frac{|h(xyzx)|}{|h(xyz)|} \leq \frac{131}{90}$, which gives $|h(x)| \leq$

$\frac{41}{49}|h(yz)|$. Since $|h(z)| \leq 3$, we deduce $|h(x)| \leq \frac{129}{49} + \frac{41}{49}|h(y)|$. Symmetrically, considering the repetition $h(yz^U xyz)$ gives $|h(y)| \leq \frac{129}{49} + \frac{41}{49}|h(x)|$. So

$$|h(x)| \leq \tfrac{129}{49} + \tfrac{41}{49}|h(y)| \leq \tfrac{129}{49} + \tfrac{41}{49}\left(\tfrac{129}{49} + \tfrac{41}{49}|h(x)|\right) = \tfrac{11610}{2401} + \tfrac{1681}{2401}|h(x)|$$

and

$$|h(x)| \leq \frac{\frac{11610}{2401}}{1 - \frac{1681}{2401}} = \tfrac{129}{8}.$$

So $|h(x)| \leq 16$ and, symmetrically, $|h(y)| \leq 16$.

In every case, $|h(xy)| \leq 32$ and $|h(z)| \leq 3$. Thus we can check exhaustively that h does not exist.

References

[1] J. Currie and L. Mol. The undirected repetition threshold and undirected pattern avoidance. arXiv:2006.07474

[2] J. Currie, L. Mol, and N. Rampersad. A family of formulas with reversal of high avoidability index. *International Journal of Algebra and Computation* **27(5)** (2017), 477–493.

[3] J. Currie, L. Mol, and N. Rampersad. Avoidance bases for formulas with reversal. *Theor. Comput. Sci.* **738** (2018), 25–41.

[4] G. Gamard, P. Ochem, G. Richomme, and P. Séébold. Avoidability of circular formulas. *Theor. Comput. Sci.*, 726:1–4, 2018.

[5] P. Ochem. A generator of morphisms for infinite words. *RAIRO - Theoret. Informatics Appl.*, 40:427–441, 2006.