# A family of formulas with reversal of arbitrarily high avoidability index

Pascal Ochem

# A family of formulas with reversal of arbitrarily high avoidability index

Pascal Ochem[1]
LIRMM, CNRS, Université de Montpellier, Montpellier, France
ochem@lirmm.fr

**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 [3, 4] 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 $h(x)$. For example, the word $w = \texttt{20012210101122201}$ contains the occurrence $h : x \to \texttt{01},\ y \to \texttt{221}$ of the formula $xyx \cdot xy^R \cdot x^R$ because $h(xyx) = \texttt{0122101}$, $h(xy^R) = \texttt{01122}$, $h(x^R) = \texttt{10}$ are all factors of $w$. 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 [3] 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 $\lambda(\psi_1) = 4$, $\lambda(\psi_2) = \lambda(\psi_3) = \lambda(\psi_6) = 5$, $5 \leqslant \lambda(\psi_4) \leqslant 6$, $5 \leqslant \lambda(\psi_5) \leqslant 7$, $4 \leqslant \lambda(\psi_k) \leqslant 6$ if $k \geqslant 7$ and $k \not\equiv 0 \pmod 3$, and $4 \leqslant \lambda(\psi_k) \leqslant 5$ if $k \geqslant 9$ and $k \equiv 0 \pmod 3$. They conjecture that $\lambda(\psi_k) = 5$ for all $k \geqslant 2$. Computational experiments suggest that the upper bound $\lambda(\psi_k) \leqslant 5$ for $k \geqslant 3$ is witnessed by the image of every $\left(\frac{7}{4}^+\right)$-free ternary word under the following $(k+3)$-uniform morphism where $k = 3t + i$, $t \geqslant 1$, and $0 \leqslant i \leqslant 2$.

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

We do not try to prove that such words actually avoid $\psi_k$. Instead, we give a positive answer to their original question with Theorem 1 below. Consider the formula $\phi_k = x_0x_1 \cdot x_1x_2 \cdot \ldots \cdot x_{k-1}x_0 \cdot x_0^R \cdot x_1^R \cdot \ldots \cdot x_{k-1}^R$. That is, $\phi_1 = x_0x_0 \cdot x_0^R$, $\phi_2 = x_0x_1 \cdot x_1x_0 \cdot x_0^R \cdot x_1^R$, $\phi_3 = x_0x_1 \cdot x_1x_2 \cdot x_2x_0 \cdot x_0^R \cdot x_1^R \cdot x_2^R$, ...

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

This result contrasts with the situation of classical formulas (without reversal). Clark [1] has shown in 2001 that there exist formulas with index 5 (such as *ab.ba.ac.bc.cda.dcd*), but no avoidable classical formula with higher index is currently known.

Currie, Mol, and Rampersad [4] 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 following formulas are simultaneously 2-avoidable:*

- $xyzyx \cdot zyxy^Rz$
- $xyzyx \cdot zy^Rxyz$
- $xyzyx \cdot zy^Rxy^Rz$
- $xyzy^Rx \cdot zyxy^Rz$
- $xyzy^Rx \cdot zy^Rxyz$

**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, respectively.

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 $|h(x)| \leqslant d-1$.*

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 avoiding $xyx^Uy$. Notice that the notion of undirected avoidability recently considered by Currie and Mol [2] corresponds to the case where every occurrence of every variable of the formula is equipped with $-^U$.

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

The C code to find and check the morphisms in this paper is available at `http://www.lirmm.fr/~ochem/morphisms/reversal.htm`.

## 2. Formulas with unbounded avoidability index

Let us first show that for every $k \geqslant 2$, $\phi_k$ 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 $\phi_k$ is such that $|h(x_i)| = 1$ for every $0 \leqslant i < k$ 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 $\lambda(\phi_k) \leqslant k + 1$.

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 $1 \leqslant i \leqslant b$. 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 $\phi_k$ such that $h(x_j) = c_{j \bmod i}$ for every $0 \leqslant j < k$. 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. [5] 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, $xyzy^U x \cdot zy^U xy^U z$ is not 2-avoidable.

We use the method in [6] to show that the image of every $\left(\frac{7}{5}^+\right)$-free word over $\Sigma_4$ under the following 21-uniform morphism is $\left(\frac{22}{15}^+, 85\right)$-free. We also check that such a binary word is 11-directed.

$$\begin{array}{ll} 0 \rightarrow & 000010111000111100111 \\ 1 \rightarrow & 000010110011011110011 \\ 2 \rightarrow & 000010110001111010011 \\ 3 \rightarrow & 000010110001001101111 \end{array}$$

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

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

and

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

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

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

## 4. The formulas $xyzx \cdot yzxy \cdot z^R$ and $xyzx \cdot yz^R xy$

Notice that avoiding $xyzx \cdot yzxy \cdot z^R$ and $xyzx \cdot yz^R xy$ simultaneously is equivalent to avoiding $F = xyzx \cdot yz^U xy \cdot z^R$. We use the method in [6] to show that the image of every $\left(\frac{7}{5}^+\right)$-free word over $\Sigma_4$ under the following 9-uniform morphism is $\left(\frac{131}{90}^+, 28\right)$-free. We also check that such a ternary word is 4-directed.

$$\begin{array}{ll} 0 \rightarrow & 011122202 \\ 1 \rightarrow & 010121202 \\ 2 \rightarrow & 001112122 \\ 3 \rightarrow & 000101120 \end{array}$$

Consider an occurrence $h$ of $F$. Since $F$ contains $z \cdot z^R$, then $|h(z)| \leqslant 3$ by Remark 4. Suppose that $|h(xy)| \geqslant 27$. Then $h(xyzx)$ is a repetition with period $|h(xyz)| \geqslant 28$. This implies $\frac{|h(xyzx)|}{|h(xyz)|} \leqslant \frac{131}{90}$, which gives $|h(x)| \leqslant \frac{41}{49}|h(yz)|$. Since $|h(z)| \leqslant 3$, we deduce $|h(x)| \leqslant \frac{123}{49}+\frac{41}{49}|h(y)|$. Symmetrically, considering the repetition $h(yz^Uxy)$ gives $|h(y)| \leqslant \frac{123}{49}+\frac{41}{49}|h(x)|$. So

$$|h(x)| \leqslant \tfrac{123}{49}+\tfrac{41}{49}|h(y)| \leqslant \tfrac{123}{49}+\tfrac{41}{49}\left(\tfrac{123}{49}+\tfrac{41}{49}|h(x)|\right) = \tfrac{11070}{2401}+\tfrac{1681}{2401}|h(x)|$$

and

$$|h(x)| \leqslant \frac{\frac{11070}{2401}}{1-\frac{1681}{2401}} = \tfrac{123}{8} = 15.375.$$

So $|h(x)| \leqslant 15$ and, symmetrically, $|h(y)| \leqslant 15$.

In every case, $|h(xy)| \leqslant 30$ and $|h(z)| \leqslant 3$. Thus we can check exhaustively that $h$ does not exist.

## References


[1] R. J. Clark. *Avoidable formulas in combinatorics on words.* PhD thesis, University of California, Los Angeles, 2001. Available at `http://www.lirmm.fr/~ochem/morphisms/clark_thesis.pdf`

[2] J. Currie and L. Mol. The undirected repetition threshold and undirected pattern avoidance. *Theor. Comput. Sci.* **866** (2021), 56–69.

[3] 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.

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

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

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

**Declaration of interests**

☐ The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

There is no conflict of interest.