Characterization of morphic words

Pascal Ochem

CNRS, LIRMM, Université de Montpellier

With Golnaz Badkobeh, Matthieu Rosenfeld, Ľubomíra Dvořáková, Daniela Opočenská, Aseem Baranwal, James Currie, Lucas Mol, Narad Rampersad, and Jeffrey Shallit

One world combinatorics on words seminar 2024

Avoiding squares, 010, and 212

These infinite ternary words avoid squares, 010, and 212.

- **tm3**= 01202101210201202102012101202101210...
- $w_1 =$ 02101210201202102012101202101210...
- $w_2 =$ 102101210201202102012101202101210...

tm3 is the fixed point of $0 \mapsto 012$, $1 \mapsto 02$, $2 \mapsto 1$.

- $F(w_1) = F(\textbf{tm3})$.
- $1021 \in F(w_2) \setminus F(\textbf{tm3})$.

Avoiding $\{AA, \texttt{010}, \texttt{212}\}$

False statements:

- If u is finite and avoids $\{AA, \texttt{010}, \texttt{212}\}$, then $u \in F(\textbf{tm3})$.
- If w is right-infinite and avoids $\{AA, \texttt{010}, \texttt{212}\}$, then $F(w) = F(\textbf{tm3})$.

True statement:

- If w is bi-infinite and avoids $\{AA, \texttt{010}, \texttt{212}\}$, then $F(w) = F(\textbf{tm3})$.

We say that **tm3** essentially avoids $\{AA, \texttt{010}, \texttt{212}\}$ or that $\{AA, \texttt{010}, \texttt{212}\}$ characterizes **tm3**.

- **tm3** essentially avoids $\{AA, \texttt{1021}, \texttt{1201}\}$.

Patterns

- Pattern p: finite word in $\{A, B, C, \ldots\}^*$.
- Occurrence of p: any word such that $A, B, C, \ldots$ are replaced by a non-empty word.
- `1201020210021` contains the occurrence $A \mapsto \mathtt{0}$, $B \mapsto \mathtt{10}$, $C \mapsto \mathtt{2}$ of $ABCACB$.

Patterns

- Pattern p: finite word in $\{A, B, C, \ldots\}^*$.
- Occurrence of p: any word such that $A, B, C, \ldots$ are replaced by a non-empty word.
- $\texttt{1201020210021}$ contains the occurrence $A \mapsto \texttt{0}$, $B \mapsto \texttt{10}$, $C \mapsto \texttt{2}$ of $ABCACB$.

Formulas

- Consider the pattern $AABBCABBA$.
- `0000` is not forbidden.
- `0110011` is not forbidden.

Replace every isolated variable by a dot to get the formula $AABB.ABBA$.
$AABB$ and $ABBA$ are the *fragments* of the formula.

- Consider the formula $AABB.ABBA$.
- `0000` is forbidden ($A \mapsto \texttt{0}$, $B \mapsto \texttt{0}$).
- `0110011` is forbidden ($A \mapsto \texttt{0}$, $B \mapsto \texttt{1}$).
- If p is recurrent, then w avoids p iff w avoids f.

Fixed points with a characterization

name	morphism	forbidden	critical exponent	slope
fib	`01/0`	$AAAA$, `11`, `000`, `10101`	$(5+\sqrt{5})/2 \simeq 3.618$	1
p	`01/21/0`	AAA, `00`, `11`, `22`, `20`, `212`, `0101`, `02102`, `121012`, `01021010`, `21021012102`	$5X^3 - 26X^2 + 43X - 23$ $X \simeq 2.48$	2
tm2	`01/10`	$ABABA$, `000`, `111`	2^+	10/3
tm3	`012/02/1`	AA, `010`, `212`	2	10/3
fp5	`01/23/4/21/0`	AA, `02`, `03`, `13`, `14`, `20`, `24`, `31`, `32`, `40`, `41`, `43`, `121`, `212`, `304`, `3423`, `4234`	$(5+\sqrt{5})/4 \simeq 1.809$ $(*)$	4
fp4	`01/21/03/23`	$AB.BA.AC.CA.BC$, `02`, `010`, `12101`, `103230121`	2	4
pd	`01/00`	$AAAA$, $AAABABAA$, `11`, `1001`	4	5/3 $(*)$

Symmetries of **fp4**

fp4 $= \texttt{0121032101230321012103230123032101210}\ldots$

- **fp4** essentially avoids $\{AB.BA.AC.CA.BC, \texttt{02}, \texttt{010}, \texttt{12101}, \texttt{103230121}\}$.
- 24 permutations and reverse.
- $F\left(\mathbf{fp4}^R\right) = F(\pi(\mathbf{fp4}))$, where $\pi = (\texttt{02})(\texttt{13}) = \texttt{2}/\texttt{3}/\texttt{0}/\texttt{1}$.
- So there are 24 distinct versions of **fp4**.
- **fp4** essentially avoids $AB.BA.AC.CA.BC$ and 23 large factors.
- $\left\{\mathbf{fp4}_{(1)}, \mathbf{fp4}_{(2)}, \ldots, \mathbf{fp4}_{(24)}\right\}$ essentially avoids $AB.BA.AC.CA.BC$.

Erratum: symmetries of **fp4** are wrong in my papers, sorry.

The Fibonacci word

Proof that **fib** essentially avoids $\{AAAA, \texttt{11}, \texttt{000}, \texttt{10101}\}$:

- Suppose $w \in {}^{\omega}\{\texttt{0}, \texttt{1}\}^{\omega}$ avoids $\{AAAA, \texttt{11}, \texttt{000}, \texttt{10101}\}$.
- Since w avoids 11, then $w \in {}^{\omega}\{\texttt{01}, \texttt{0}\}^{\omega}$.
- So $w = h(v)$ with $h = \texttt{01}/\texttt{0}$.
    - If v contains $AAAA$, then $h(v)$ contains $AAAA$.
    - If v contains $\texttt{11}$, then $h(\texttt{11}x)$ contains $\texttt{000}$.
    - If v contains $\texttt{000}$, then $h(\texttt{000})$ contains $\texttt{10101}$.
    - If v contains $\texttt{10101}$, then v contains $\texttt{0101010}$ and $h(\texttt{0101010}x)$ contains $(\texttt{010})^4$.

The binary Thue-Morse word

- **tm2** essentially avoids $\{ABABA, \texttt{000}, \texttt{111}\}$.
- **tm2** essentially avoids $\{ABABA, AAA\}$.
- **tm2** essentially avoids $f_h = AABCAA.BCB$.
- **tm2** essentially avoids $f_e = AABCAAB.AABCAB.AABCB$.
  - $AAACAAA$ contains f_e via $A \mapsto A$, $B \mapsto A$, $C \mapsto C$.
  - $ABABACABABA$ contains f_e via $A \mapsto AB$, $B \mapsto A$, $C \mapsto C$.

$\{\mathbf{tm3}, \mathbf{tm3}', \mathbf{tm3}''\}$ essentially avoids 144 formulas

The hardest formulas

- $ABCA.BCAB.BCB.CBA$
- $ABCAB.BCB.AC$
- $ABCA.BCAB.ACB.BCB$
- $ABCA.BCAB.BCB.AC.BA$

The easiest formula

- $f_e = ABCAB.ABCBA.ACB.BAC$

$AACAA$ contains f_e via $A \mapsto A$, $B \mapsto A$, $C \mapsto C$.

$\{\mathbf{tm3}, \mathbf{tm3}', \mathbf{tm3}''\}$ essentially avoids other formulas

If f is such that

- $ABCA.ABA.ACA \preceq f \preceq ABCA.ABA.ACA.ACB.CBA$,
- $ABCA.ABA.BCB.AC \preceq f \preceq ABCA.ABA.ABCBA.ACB$, or
- $ABCA.ABA.BCB.CBA \preceq f \preceq ABCA.ABA.ABCBA.ACB$.

then $\{\mathbf{tm3}, \mathbf{tm3}', \mathbf{tm3}''\}$ essentially avoids f.

$AAABAAACAAABAAA$ contains f via $A \mapsto A$, $B \mapsto ABA$, $C \mapsto ACA$.

And now the morphic words

Matthieu has found a dozen other fixed points, such as `010/12/10` with critical exponent $\simeq 2.414$ and slope $\simeq 2.78$, that fit in the critical exponent VS slope trade-off but do not seem to have a characterization.

Unsurprisingly, morphic words with a characterization are images of fixed points that have a characterization.

Images of **fib**

Consider $h = \texttt{0}/\texttt{01}$. Every bi-infinite $\frac{11}{3}$-free binary word that contains no pair of complementary factors of length 4 has the same factor set as $h(\mathbf{fib})$ or $\overline{h(\mathbf{fib})}$.

Consider $g = \texttt{01}/\texttt{11}$. Every bi-infinite $\frac{15}{4}$-free binary word that contains no antisquare other than `01` and `10` has the same factor set as $g(\mathbf{fib})$ or $\overline{g(\mathbf{fib})}$.

fib, $h(\mathbf{fib})$, and $g(\mathbf{fib})$ have critical exponent $\frac{5+\sqrt{5}}{2} \simeq 3.618$

No antisquare other than $\texttt{01}$ and $\texttt{10}$

Consider $g = \texttt{01}/\texttt{11}$. Every bi-infinite $\frac{15}{4}$-free binary word w that contains no antisquare other than $\texttt{01}$ and $\texttt{10}$ has the same factor set as $g(\mathbf{fib})$ or $\overline{g(\mathbf{fib})}$.

- W.l.o.g., w contains 11.
- $w = \cdots \texttt{01}^a\texttt{01}^b\texttt{01}^c\texttt{01}^d \cdots$ with $\cdots a, b, c, d \cdots \in \{1, 3\}$.
- $w \in {}^{\omega}\{\texttt{01}, \texttt{0111}\}^{\omega}$. $w \in {}^{\omega}\{\texttt{01}, \texttt{11}\}^{\omega}$. $w = g(v)$.
  - v avoids $AAAA$.
  - $g(\texttt{11}) = \texttt{1111}$ is a 4-power.
  - $g(\texttt{000}) = \texttt{010101}$ is an antisquare.
  - $g(x\texttt{0101010}) = x\texttt{10111011101101} = x(\texttt{1011})^{\frac{15}{4}}$.

$\left\{g(\mathbf{fib}), \overline{g(\mathbf{fib})}\right\}$ essentially avoids $\{AAAA, \texttt{0011}, \texttt{0110}, \texttt{1100}, \texttt{1001}, \texttt{010101}, \texttt{101010}, \texttt{1000101110}, \texttt{0111010001}, \texttt{10111011101101}, \texttt{01000100010010}\}$.

Images of $\mathbf{p}$

$\xi = \mathtt{01/0110/1}$.
$\rho = \mathtt{0110010110 1/0110010/011001}$.

Every bi-infinite $\frac{29}{11}$-free binary word that contains no pair of complementary factors of length 8 has the same factor set as either $\xi(\mathbf{p})$, $\overline{\xi(\mathbf{p})}$, $\xi(\mathbf{p})^R$, $\overline{\xi(\mathbf{p})^R}$, $\rho(\mathbf{p})$, $\overline{\rho(\mathbf{p})}$, $\rho(\mathbf{p})^R$, or $\overline{\rho(\mathbf{p})^R}$. ($\frac{5}{2}^+$-free).

$\psi = \mathtt{11001/0/01101}$.

Every bi-infinite $\frac{5}{2}$-free binary word that contains no pair of complementary factors of length 11 has the same factor set as either $\psi(\mathbf{p})$, $\overline{\psi(\mathbf{p})}$, $\psi(\mathbf{p})^R$, or $\overline{\psi(\mathbf{p})^R}$. ($X$-free, $X \simeq 2.48$).

Images of $\mathbf{p}$

$\mu = \texttt{011001/1001/0}$. $\nu = \texttt{011/0/01}$.

Every bi-infinite $\frac{15}{4}$-free binary word containing at most 18 palindromes has the same set of factors as either $\mu(\mathbf{p})$, $\overline{\mu(\mathbf{p})}$, $\mu(\mathbf{p})^R$, or $\overline{\mu(\mathbf{p})^R}$. ($\frac{28}{11}^+$-free).

Every recurrent $\frac{28}{11}$-free binary word containing at most 20 palindromes has the same set of factors as either $\nu(\mathbf{p})$, $\overline{\nu(\mathbf{p})}$, $\nu(\mathbf{p})^R$, or $\overline{\nu(\mathbf{p})^R}$. ($\frac{5}{2}^+$-free).

NB: $\nu(\mathbf{p})^R\texttt{010110}\nu(\mathbf{p})$ is bi-infinite, $\frac{5}{2}^+$-free and has 20 palindromes.

Images of **tm3**

tm2 $= w($**tm3**$)$ where $w = \texttt{011/01/0}$.

Polynomial binary formulas

- $g_x = \texttt{01110/0110/0}$
- $g_y = \texttt{0111/01/00}$
- $g_z = \texttt{0001/001/11}$
- $g_t = \texttt{01011011010/01011010/010}$

- $\left\{g_x(\mathbf{tm3}), g_y(\mathbf{tm3}), g_z(\mathbf{tm3}), \overline{g_z(\mathbf{tm3})}\right\}$ essentially avoids $AA.ABA.ABBA$.
- $\{g_x(\mathbf{tm3}), g_t(\mathbf{tm3})\}$ essentially avoids f, with $f \in \{ABA.AABB,\ BBA.ABA.AABB,\ AABA.AABB\}$.
- $g_x(\mathbf{tm3})$ essentially avoids $AABA.ABB.BBA$.

Images of **tm3**

Every bi-infinite binary word containing only the squares 00, 11, 001001, and 110110 has the same factor set as $g_4(\textbf{tm3})$.

$$\begin{aligned} g_4(\mathtt{0}) &= \mathtt{0001001100011101 1},\\ g_4(\mathtt{1}) &= \mathtt{00010011101 1},\\ g_4(\mathtt{2}) &= \mathtt{00111}. \end{aligned}$$

Images of **tm3**

Every bi-infinite binary word containing only the squares $\texttt{00}$, $\texttt{11}$, $\texttt{0000}$, $\texttt{0001100011}$, and $\texttt{1000010000}$ has the same factor set as $g_5(\textbf{tm3})$.

$$
\begin{aligned}
g_5(\texttt{0}) &= \texttt{0000100000111000011000111},\\
g_5(\texttt{1}) &= \texttt{000010000011000111},\\
g_5(\texttt{2}) &= \texttt{0000011}.
\end{aligned}
$$

The three words of Thue

$M_1 = \texttt{012/1/02/12/}\varepsilon$ and $M_2 = \texttt{02/1/0/12/}\varepsilon$

- **tm3** essentially avoids $\{AA, \texttt{010}, \texttt{212}\}$ or $\{AA, \texttt{1021}, \texttt{1201}\}$.
- M_1(**fp5**) essentially avoids $\{AA, \texttt{010}, \texttt{020}\}$.
- M_2(**fp5**) essentially avoids $\{AA, \texttt{121}, \texttt{212}\}$.

Distance constraints between occurrences of a letter:

- **tm3**: no two 1s at distance 3.
- M_1(**fp5**): no two 0s at distance 2.
- M_2(**fp5**): no two consecutive 0s at distance 4.

Characterization of $M_2(\mathbf{fp5})$

If $ABA.BCB.ACA \preceq f \preceq ABA.ABCBA.ACA.ACB.BCA$, then $\{M_2(\mathbf{fp5}), M_2(\mathbf{fp5})', M_2(\mathbf{fp5})''\}$ essentially avoids f.

Images of **fp5**

- 1 overlap: $\mathtt{01010}$.
- 12 squares: $\mathtt{0}^2$, $\mathtt{1}^2$, $(\mathtt{01})^2$, $(\mathtt{10})^2$, $(\mathtt{010})^2$, $(\mathtt{0110})^2$, $(\mathtt{1001})^2$, $(\mathtt{100101})^2$, $(\mathtt{011001})^2$, $(\mathtt{100110})^2$, $(\mathtt{101001})^2$, $(\mathtt{1001011001101001})^2$.
- characterizes $h_{12}(\mathbf{fp5})$.

$$
\begin{array}{lll}
h_{12}(\mathtt{0}) = \mathtt{0011}, & M_2(\mathtt{0}) = \mathtt{02}, & \\
h_{12}(\mathtt{1}) = \mathtt{01}, & M_2(\mathtt{1}) = \mathtt{1}, & x_{12}(\mathtt{0}) = \mathtt{001}, \\
h_{12}(\mathtt{2}) = \mathtt{001}, & M_2(\mathtt{2}) = \mathtt{0}, & x_{12}(\mathtt{1}) = \mathtt{01}, \\
h_{12}(\mathtt{3}) = \mathtt{011}, & M_2(\mathtt{3}) = \mathtt{12}, & x_{12}(\mathtt{2}) = \mathtt{1}. \\
h_{12}(\mathtt{4}) = \varepsilon. & M_2(\mathtt{4}) = \varepsilon. &
\end{array}
$$

Notice that $h_{12} = x_{12} \circ M_2$, so that $h_{12}(\mathbf{fp5}) = x_{12}(M_2(\mathbf{fp5}))$.

Images of **fp5**

- 1 overlap: $\texttt{01010}$.
- 12 squares: $\texttt{0}^2$, $\texttt{1}^2$, $(\texttt{01})^2$, $(\texttt{10})^2$, $(\texttt{010})^2$, $(\texttt{0110})^2$, $(\texttt{1001})^2$, $(\texttt{100101})^2$, $(\texttt{011001})^2$, $(\texttt{100110})^2$, $(\texttt{101001})^2$, $(\texttt{1001011001101001})^2$.
- characterizes $h_{12}(\mathbf{fp5})$.

$$
\begin{array}{lll}
h_{12}(\texttt{0}) = \texttt{0011}, & M_2(\texttt{0}) = \texttt{02}, & \\
h_{12}(\texttt{1}) = \texttt{01}, & M_2(\texttt{1}) = \texttt{1}, & x_{12}(\texttt{0}) = \texttt{001}, \\
h_{12}(\texttt{2}) = \texttt{001}, & M_2(\texttt{2}) = \texttt{0}, & x_{12}(\texttt{1}) = \texttt{01}, \\
h_{12}(\texttt{3}) = \texttt{011}, & M_2(\texttt{3}) = \texttt{12}, & x_{12}(\texttt{2}) = \texttt{1}. \\
h_{12}(\texttt{4}) = \varepsilon. & M_2(\texttt{4}) = \varepsilon. &
\end{array}
$$

Notice that $h_{12} = x_{12} \circ M_2$, so that $h_{12}(\mathbf{fp5}) = x_{12}(M_2(\mathbf{fp5}))$.

Images of **fp5**

- 1 overlap: $\texttt{1001001}$.
- 14 squares: $\texttt{0}^2$, $\texttt{1}^2$, $(\texttt{01})^2$, $(\texttt{10})^2$, $(\texttt{001})^2$, $(\texttt{010})^2$, $(\texttt{100})^2$, $(\texttt{101})^2$, $(\texttt{0110})^2$, $(\texttt{1001})^2$, $(\texttt{100110})^2$, $(\texttt{0100110})^2$, $(\texttt{0110010})^2$, and $(\texttt{10010110})^2$.
- characterizes $g_{14}(\textbf{fp5})$.

$$
\begin{aligned}
g_{14}(\texttt{0}) &= \texttt{01}, \\
g_{14}(\texttt{1}) &= \texttt{00110}, \\
g_{14}(\texttt{2}) &= \texttt{1}, \\
g_{14}(\texttt{3}) &= \texttt{0010110}, \\
g_{14}(\texttt{4}) &= \texttt{0110}.
\end{aligned}
$$

Images of **fp5**

- 2 overlaps: $\mathtt{0110110}$ and $\mathtt{1001001}$.
- 12 squares: $\mathtt{0}^2$, $\mathtt{1}^2$, $(\mathtt{01})^2$, $(\mathtt{10})^2$, $(\mathtt{001})^2$, $(\mathtt{010})^2$, $(\mathtt{011})^2$, $(\mathtt{100})^2$, $(\mathtt{101})^2$, $(\mathtt{110})^2$, $(\mathtt{01101001})^2$, $(\mathtt{10010110})^2$.
- characterizes $g_{12}(\mathbf{fp5})$.

$$
\begin{aligned}
g_{12}(\mathtt{0}) &= \mathtt{01},\\
g_{12}(\mathtt{1}) &= \mathtt{0},\\
g_{12}(\mathtt{2}) &= \mathtt{011},\\
g_{12}(\mathtt{3}) &= \varepsilon,\\
g_{12}(\mathtt{4}) &= \varepsilon.
\end{aligned}
$$

Images of **fp5**

- 2 overlaps: $\texttt{01010}$ and $\texttt{10101}$.
- 8 squares: $\texttt{0}^2$, $\texttt{1}^2$, $(\texttt{01})^2$, $(\texttt{10})^2$, $(\texttt{0110})^2$, $(\texttt{1001})^2$, $(\texttt{011001})^2$, $(\texttt{100110})^2$.
- characterizes $g_8(\mathbf{fp5})$.

$$
\begin{aligned}
g_8(\texttt{0}) &= \texttt{011},\\
g_8(\texttt{1}) &= \texttt{0},\\
g_8(\texttt{2}) &= \texttt{01},\\
g_8(\texttt{3}) &= \varepsilon,\\
g_8(\texttt{4}) &= \varepsilon.
\end{aligned}
$$

Images of **fp5**

Every bi-infinite binary word such that the only occurrences of *ABBA* are in $\{\mathtt{0000}, \mathtt{0110}, \mathtt{1001}, \mathtt{1111}, \mathtt{001100}, \mathtt{011110}, \mathtt{100001}, \mathtt{110011}\}$ has the same factor set as c(**fp5**).

$$\begin{aligned}
c(\mathtt{0}) &= \mathtt{0010111100},\\
c(\mathtt{1}) &= \mathtt{1101000011},\\
c(\mathtt{2}) &= \varepsilon,\\
c(\mathtt{3}) &= \mathtt{1101001100},\\
c(\mathtt{4}) &= \mathtt{0010110011}.
\end{aligned}$$

Polynomial complexity $\not\Longrightarrow$ characterization

Let $m_a(x) = x\texttt{01}$ and $m_b(x) = x\texttt{10}$.

Consider S:

- $\texttt{0} \in S$.
- If $v \in S$, then $m_a(v) \in S$ and $m_b(v) \in S$.
- If v' is a factor of $v \in S$, then $v' \in S$.

Consider $c(n) = \left|S \cup \Sigma_2^n\right|$:

- $c(n) = 2\left(c\left(\left\lfloor \frac{n}{3} \right\rfloor\right) + c\left(\left\lfloor \frac{n+1}{3} \right\rfloor\right) + c\left(\left\lfloor \frac{n+2}{3} \right\rfloor\right)\right)$ for $n \geqslant 8$.
- $c(n) = \Theta\left(n^{\ln 6/\ln 3}\right) = \Theta\left(n^{1+\ln 2/\ln 3}\right)$.

Let $f \in \{ABACA.ABCA,\ \ ABAC.BACA.ABCA\}$. A binary word u is recurrent in a binary word avoiding f if and only if $u \in S$.

Open problems

- Find a formula f such that $\left\{\mathbf{fib}, \overline{\mathbf{fib}}\right\}$ essentially avoids f.
- Find a formula f such that $\{M_1(\mathbf{fp5}), M_1(\mathbf{fp5})', M_1(\mathbf{fp5})''\}$ essentially avoids f.

Thank you