Abelian powers and patterns in words: problems and perspectives

James D. Currie

University of Winnipeg

June 11, 2013

- Thue (1906): "We consider the question whether, given p and w, there always exists a nonerasing morphism $h : A \longrightarrow B$ such that $h(p)$ is a factor of w"

- Thue (1906): "We consider the question whether, given p and w, there always exists a nonerasing morphism $h : A \rightarrow B$ such that $h(p)$ is a factor of w"
- E. g. : If $p = xx$ and $w = 12342345$ we see that $h(p)$ is a factor of w if $h(p) = 234$.

- Thue (1906): "We consider the question whether, given p and w, there always exists a nonerasing morphism $h : A \to B$ such that $h(p)$ is a factor of w"
- E. g. : If $p = xx$ and $w = 12342345$ we see that $h(p)$ is a factor of w if $h(p) = 234$.

- Thue (1906): "We consider the question whether, given p and w, there always exists a nonerasing morphism $h : A \longrightarrow B$ such that $h(p)$ is a factor of w"
- E. g. : If $p = xx$ and $w = 12342345$ we see that $h(p)$ is a factor of w if $h(p) = 234$.
- We call p a **pattern** and $h(p)$ an **instance** of p.

- We study problems of **pattern avoidance**

- We study problems of **pattern avoidance**
- We say that w **encounters** p if $h(p)$ is a factor of w for some non-erasing h; otherwise w **avoids** p.

- We study problems of **pattern avoidance**
- We say that w **encounters** p if $h(p)$ is a factor of w for some non-erasing h; otherwise w **avoids** p.
- Consider the morphism $g : \{0,1,2\}^* \rightarrow \{0,1,2\}^*$ given by $g(0) = 1$, $g(1) = 20$, $g(2) = 210$.

- We study problems of **pattern avoidance**
- We say that w **encounters** p if $h(p)$ is a factor of w for some non-erasing h; otherwise w **avoids** p.
- Consider the morphism $g : \{0, 1, 2\}^* \rightarrow \{0, 1, 2\}^*$ given by $g(0) = 1$, $g(1) = 20$, $g(2) = 210$. It has a fixed point $s = g^\omega(2)$:

- We study problems of **pattern avoidance**
- We say that w **encounters** p if $h(p)$ is a factor of w for some non-erasing h; otherwise w **avoids** p.
- Consider the morphism $g : \{0,1,2\}^* \to \{0,1,2\}^*$ given by $g(0) = 1$, $g(1) = 20$, $g(2) = 210$. It has a fixed point $s = g^\omega(2)$:

$$s = 21020121012021020120210\cdots$$

- We study problems of **pattern avoidance**
- We say that w **encounters** p if $h(p)$ is a factor of w for some non-erasing h; otherwise w **avoids** p.
- Consider the morphism $g : \{0,1,2\}^* \rightarrow \{0,1,2\}^*$ given by $g(0) = 1$, $g(1) = 20$, $g(2) = 210$. It has a fixed point $s = g^\omega(2)$:

$$s = 210201210120210201202010\cdots$$

- Word s avoids xx.

- We study problems of **pattern avoidance**
- We say that w **encounters** p if $h(p)$ is a factor of w for some non-erasing h; otherwise w **avoids** p.
- Consider the morphism $g : \{0,1,2\}^* \to \{0,1,2\}^*$ given by $g(0) = 1$, $g(1) = 20$, $g(2) = 210$. It has a fixed point $s = g^\omega(2)$:

$$s = 210201210120210201202\!10\cdots$$

- Word s avoids xx.(The fixed point of a morphism is called a **D0L**.)

- Consider the morphism $f : \{0,1,2\}^* \rightarrow \{0,1\}^*$ given by $f(0) = 0$, $f(1) = 01$, $f(2) = 011$.

- Consider the morphism $f:\{0,1,2\}^*\rightarrow\{0,1\}^*$ given by $f(0)=0$, $f(1)=01$, $f(2)=011$. Let $t=f(s)$.

- Consider the morphism $f : \{0, 1, 2\}^* \rightarrow \{0, 1\}^*$ given by $f(0) = 0$, $f(1) = 01$, $f(2) = 011$. Let $t = f(s)$.

$$t = 0110100110010110100101100110100 1 \cdots$$

- Consider the morphism $f : \{0,1,2\}^* \rightarrow \{0,1\}^*$ given by $f(0) = 0$, $f(1) = 01$, $f(2) = 011$. Let $t = f(s)$.

$$t = 01101001100101101001011001101001\cdots$$

- The word t is called the **Thue-Morse word**.

- Consider the morphism $f : \{0,1,2\}^* \rightarrow \{0,1\}^*$ given by $f(0) = 0$, $f(1) = 01$, $f(2) = 011$. Let $t = f(s)$.

$$t = 0110100110010110100101100110 1001 \cdots$$

- The word t is called the **Thue-Morse word**. Thue showed that it avoids xxx and $xyxyx$.

- Consider the morphism $f : \{0,1,2\}^* \to \{0,1\}^*$ given by $f(0) = 0$, $f(1) = 01$, $f(2) = 011$. Let $t = f(s)$.

$$t = 0110100110010110100101100110100 1 \cdots$$

- The word t is called the **Thue-Morse word**. Thue showed that it avoids xxx and $xyxyx$.
- The image of a D0L under a morphism is called an **HD0L**.

Fix a pattern p. We say that p is

Fix a pattern p. We say that p is

- **avoidable** if for some finite alphabet Σ there are arbitrarily long words over Σ which avoid p

Fix a pattern p. We say that p is

- **avoidable** if for some finite alphabet Σ there are arbitrarily long words over Σ which avoid p
- k-**avoidable** if there are arbitrarily long words over $\{1, 2, \ldots, k\}$ which avoid p

Fix a pattern p. We say that p is

- **avoidable** if for some finite alphabet Σ there are arbitrarily long words over Σ which avoid p
- k-**avoidable** if there are arbitrarily long words over $\{1, 2, \ldots, k\}$ which avoid p
- **D0L-avoidable (HD0L-avoidable)** if there is a D0L (HD0L) which avoids p

Fix a pattern p. We say that p is

- **avoidable** if for some finite alphabet Σ there are arbitrarily long words over Σ which avoid p
- k-**avoidable** if there are arbitrarily long words over $\{1, 2, \ldots, k\}$ which avoid p
- **D0L-avoidable (HD0L-avoidable)** if there is a D0L (HD0L) which avoids p
- k-**D0L-avoidable (**k-**HD0L-avoidable)** if there is a D0L (HD0L) on $\{1, 2, \ldots, k\}$ which avoids p

Fix a pattern p. We say that p is

- **avoidable** if for some finite alphabet Σ there are arbitrarily long words over Σ which avoid p
- k-**avoidable** if there are arbitrarily long words over $\{1, 2, \ldots, k\}$ which avoid p
- **D0L-avoidable (HD0L-avoidable)** if there is a D0L (HD0L) which avoids p
- k-**D0L-avoidable** (k-**HD0L-avoidable**) if there is a D0L (HD0L) on $\{1, 2, \ldots, k\}$ which avoids p

$$k - \text{D0L avoidable} \quad \Longrightarrow$$

$$k - \text{D0L avoidable} \implies$$
$$k - \text{HD0L avoidable} \implies$$

$$
\begin{aligned}
k - \text{D0L avoidable} &\quad \Longrightarrow \\
k - \text{HD0L avoidable} &\quad \Longrightarrow \\
k - \text{avoidable} &\quad \Longrightarrow
\end{aligned}
$$

$$
\begin{aligned}
k-\text{D0L avoidable} &\implies \\
k-\text{HD0L avoidable} &\implies \\
k-\text{avoidable} &\implies \\
\text{avoidable} &
\end{aligned}
$$

Consider the words Z_n where

Consider the words Z_n where

$$Z_1 = 1$$

Consider the words Z_n where

$$Z_1 = 1$$

$$Z_n = Z_{n-1} n Z_{n-1}, n > 1.$$

Consider the words Z_n where

$$Z_1 = 1$$

$$Z_n = Z_{n-1} n Z_{n-1}, n > 1.$$

Theorem

(Bean, Ehrenfeuch & McNulty; Zimin, 1979) Pattern p on $\{1, 2, \ldots k\}$ is avoidable iff Z_k avoids p

Consider the words Z_n where

$$Z_1 = 1$$

$$Z_n = Z_{n-1} n Z_{n-1}, n > 1.$$

Theorem

(Bean, Ehrenfeuch & McNulty; Zimin, 1979) Pattern p on $\{1, 2, \ldots k\}$ is avoidable iff Z_k avoids p

E. g. Pattern $p = 12121$ is avoided by $Z_2 = Z_1 2 Z_1 = 121$. Thus 12121 is avoidable.

Consider the words Z_n where

$$Z_1 = 1$$

$$Z_n = Z_{n-1} n Z_{n-1}, n > 1.$$

Theorem

(Bean, Ehrenfeuch & McNulty; Zimin, 1979) Pattern p on $\{1, 2, \ldots k\}$ *is avoidable iff* Z_k *avoids p*

E. g. Pattern $p = 12121$ is avoided by $Z_2 = Z_1 2 Z_1 = 121$. Thus 12121 is avoidable. In fact we have seen that T-M word t avoids 12121 (a.k.a. *xyxyx*)

- Cassaigne (1994) gave a decision procedure to determine whether a given HD0L (with mild restrictions) avoided a given pattern.

- Cassaigne (1994) gave a decision procedure to determine whether a given HD0L (with mild restrictions) avoided a given pattern.
- Cassaigne determined the smallest alphabets on which each of the avoidable binary patterns could be avoided.

- Cassaigne (1994) gave a decision procedure to determine whether a given HD0L (with mild restrictions) avoided a given pattern.
- Cassaigne determined the smallest alphabets on which each of the avoidable binary patterns could be avoided.
- Ochem (1999) determined the smallest alphabets on which each of the avoidable ternary patterns could be avoided.

- Open problem: Given a pattern p is it k-avoidable?

- Open problem: Given a pattern p is it k-avoidable?
- Open problem: Is every k-avoidable pattern k-HD0L-avoidable?

- Open problem: Given a pattern p is it k-avoidable?
- Open problem: Is every k-avoidable pattern k-HD0L-avoidable?
- Open problem: Which patterns are k-D0L-avoidable?

- Open problem: Given a pattern p is it k-avoidable?
- Open problem: Is every k-avoidable pattern k-HD0L-avoidable?
- Open problem: Which patterns are k-D0L-avoidable?
- Open problem: Clark (2006) exhibited a pattern which is 5-avoidable but not 4-avoidable.

- Open problem: Given a pattern p is it k-avoidable?
- Open problem: Is every k-avoidable pattern k-HD0L-avoidable?
- Open problem: Which patterns are k-D0L-avoidable?
- Open problem: Clark (2006) exhibited a pattern which is 5-avoidable but not 4-avoidable. Is there a pattern which is 6-avoidable but not 5-avoidable?

Back to Thue-Morse:

Back to Thue-Morse:

- An instance of xx is a word of length $2q$ and period q.

Back to Thue-Morse:

- An instance of xx is a word of length $2q$ and period q.
- For example, *hotshots* has length 8, period 4, hence is an instance of xx.

Back to Thue-Morse:

- An instance of xx is a word of length $2q$ and period q.
- For example, *hotshots* has length 8, period 4, hence is an instance of xx.
- A k-power is a word of length kq, period q for some q. Thus *hotshots* is a 2-power while *onion* is a $5/3$-power.

Back to Thue-Morse:

- An instance of xx is a word of length $2q$ and period q.
- For example, *hotshots* has length 8, period 4, hence is an instance of xx.
- A k-power is a word of length kq, period q for some q. Thus *hotshots* is a 2-power while *onion* is a 5/3-power.
- The Thue-Morse word t avoids all k-powers with k greater than 2.

- In Abelian versions, we only require identity up to permutation.

- In Abelian versions, we only require identity up to permutation.
- An example of an Abelian square (Abelian 2-power, Abelian xx) is $abcacb$ since abc is an anagram of acb (We write $abc \sim acb$.)

- In Abelian versions, we only require identity up to permutation.
- An example of an Abelian square (Abelian 2-power, Abelian xx) is *abcacb* since *abc* is an anagram of *acb* (We write $abc \sim acb$.)
- Erdös asked (1961) whether Abelian squares were avoidable.

- In Abelian versions, we only require identity up to permutation.
- An example of an Abelian square (Abelian 2-power, Abelian xx) is *abcacb* since *abc* is an anagram of *acb* (We write $abc \sim acb$.)
- Erdös asked (1961) whether Abelian squares were avoidable.
- The answer is 'Yes' (Evdokimov, 1968).

- In Abelian versions, we only require identity up to permutation.
- An example of an Abelian square (Abelian 2-power, Abelian xx) is *abcacb* since *abc* is an anagram of *acb* (We write $abc \sim acb$.)
- Erdös asked (1961) whether Abelian squares were avoidable.
- The answer is 'Yes' (Evdokimov, 1968). The first construction was on a 25-letter alphabet

- In Abelian versions, we only require identity up to permutation.
- An example of an Abelian square (Abelian 2-power, Abelian xx) is $abcacb$ since abc is an anagram of acb (We write $abc \sim acb$.)
- Erdös asked (1961) whether Abelian squares were avoidable.
- The answer is 'Yes' (Evdokimov, 1968). The first construction was on a 25-letter alphabet
- In 1992, Keranen showed that Abelian squares are 4-avoidable

- In Abelian versions, we only require identity up to permutation.
- An example of an Abelian square (Abelian 2-power, Abelian xx) is $abcacb$ since abc is an anagram of acb (We write $abc \sim acb$.)
- Erdös asked (1961) whether Abelian squares were avoidable.
- The answer is 'Yes' (Evdokimov, 1968). The first construction was on a 25-letter alphabet
- In 1992, Keranen showed that Abelian squares are 4-avoidable – with a morphism where the images of letters are length 85.

Abelian fractional powers

- Avoidance of Abelian powers has implications for avoidance of other patterns.

Abelian fractional powers

- Avoidance of Abelian powers has implications for avoidance of other patterns.
- Abelian xx is 4-avoidable.

Abelian fractional powers

- Avoidance of Abelian powers has implications for avoidance of other patterns.
- Abelian xx is 4-avoidable.
- It follows that the (non-Abelian) pattern $xyzxzy$ is 4-avoidable.

- An Abelian r-power, $1 < r < 2$, is a word $xy\hat{x}$ where

- An Abelian r-power, $1 < r < 2$, is a word $xy\hat{x}$ where
  - $\hat{x}$ is a permutation of x

- An Abelian r-power, $1 < r < 2$, is a word $xy\hat{x}$ where
  - $\hat{x}$ is a permutation of x
  - $|xy\hat{x}|/|xy| = r$.

- An Abelian r-power, $1 < r < 2$, is a word $xy\hat{x}$ where
  - $\hat{x}$ is a permutation of x
  - $|xy\hat{x}|/|xy| = r$.
- For fixed k, $1 < k < 2$ we can avoid all Abelian r-powers with $r \geq k$ simultaneously. (C. & Cassaigne, 1995).

- An Abelian r-power, $1 < r < 2$, is a word $xy\hat{x}$ where
  - $\hat{x}$ is a permutation of x
  - $|xy\hat{x}|/|xy| = r$.
- For fixed k, $1 < k < 2$ we can avoid all Abelian r-powers with $r \geq k$ simultaneously. (C. & Cassaigne, 1995).
- Alphabet size is bounded by

$$\alpha 5^{(\alpha^2+1)^2} + 2\alpha^2, \alpha = 4 + 3/(k-1).$$

- An Abelian r-power, $1 < r < 2$, is a word $xy\hat{x}$ where
    - $\hat{x}$ is a permutation of x
    - $|xy\hat{x}|/|xy| = r$.
- For fixed k, $1 < k < 2$ we can avoid all Abelian r-powers with $r \geq k$ simultaneously. (C. & Cassaigne, 1995).
- Alphabet size is bounded by

$$\alpha 5^{(\alpha^2+1)^2} + 2\alpha^2, \alpha = 4 + 3/(k-1).$$

- Open problem: Find a reasonable bound on the alphabet size needed.

- An **Abelian instance** of pattern $p = p_1 p_2 \cdots p_n$ is a word $P_1 P_2 \cdots P_n$ where $P_i \sim P_j$ whenever $p_i \sim p_j$.

- An **Abelian instance** of pattern $p = p_1 p_2 \cdots p_n$ is a word $P_1 P_2 \cdots P_n$ where $P_i \sim P_j$ whenever $p_i \sim p_j$.
- E. g. , We can avoid $xyxzyxy$ in the Abelian sense:

- An **Abelian instance** of pattern $p = p_1 p_2 \cdots p_n$ is a word $P_1 P_2 \cdots P_n$ where $P_i \sim P_j$ whenever $p_i \sim p_j$.
- E. g. , We can avoid *xyxzyxy* in the Abelian sense:
- Let w be an infinite word over a finite alphabet containing no Abelian r-powers, $r \geq 3/2$.

- An **Abelian instance** of pattern $p = p_1 p_2 \cdots p_n$ is a word $P_1 P_2 \cdots P_n$ where $P_i \sim P_j$ whenever $p_i \sim p_j$.
- E. g. , We can avoid $xyxzyxy$ in the Abelian sense:
- Let w be an infinite word over a finite alphabet containing no Abelian r-powers, $r \geq 3/2$.
- Suppose w contains a factor $x_1 y_1 x_2 z y_2 x_3 y_3$ where x_1, x_2, x_3 are anagrams, y_1, y_2, y_3 are anagrams.

- An **Abelian instance** of pattern $p = p_1 p_2 \cdots p_n$ is a word $P_1 P_2 \cdots P_n$ where $P_i \sim P_j$ whenever $p_i \sim p_j$.
- E. g. , We can avoid $xyxzyxy$ in the Abelian sense:
- Let w be an infinite word over a finite alphabet containing no Abelian r-powers, $r \geq 3/2$.
- Suppose w contains a factor $x_1 y_1 x_2 z y_2 x_3 y_3$ where x_1, x_2, x_3 are anagrams, y_1, y_2, y_3 are anagrams.
    - If $|x_1| > |y_1|$, then $x_1 y_1 x_2$ is an Abelian r-power, $r \geq 3/2$.

- An **Abelian instance** of pattern $p = p_1 p_2 \cdots p_n$ is a word $P_1 P_2 \cdots P_n$ where $P_i \sim P_j$ whenever $p_i \sim p_j$.
- E. g. , We can avoid $xyxzyxy$ in the Abelian sense:
- Let w be an infinite word over a finite alphabet containing no Abelian r-powers, $r \geq 3/2$.
- Suppose w contains a factor $x_1 y_1 x_2 z y_2 x_3 y_3$ where x_1, x_2, x_3 are anagrams, y_1, y_2, y_3 are anagrams.
  - If $|x_1| > |y_1|$, then $x_1 y_1 x_2$ is an Abelian r-power, $r \geq 3/2$.
  - If $|x_1| \leq |y_1|$, then $y_2 x_2 y_3$ is an Abelian r-power, $r \geq 3/2$.

- An **Abelian instance** of pattern $p = p_1 p_2 \cdots p_n$ is a word $P_1 P_2 \cdots P_n$ where $P_i \sim P_j$ whenever $p_i \sim p_j$.
- E. g. , We can avoid $xyxzyxy$ in the Abelian sense:
- Let w be an infinite word over a finite alphabet containing no Abelian r-powers, $r \geq 3/2$.
- Suppose w contains a factor $x_1 y_1 x_2 z y_2 x_3 y_3$ where x_1, x_2, x_3 are anagrams, y_1, y_2, y_3 are anagrams.
  - If $|x_1| > |y_1|$, then $x_1 y_1 x_2$ is an Abelian r-power, $r \geq 3/2$.
  - If $|x_1| \leq |y_1|$, then $y_2 x_2 y_3$ is an Abelian r-power, $r \geq 3/2$.
- In either case, w contains an Abelian r-power, $r \geq 3/2$.

- Open problem: What patterns are avoidable in the Abelian sense?

- Open problem: What patterns are avoidable in the Abelian sense?
- A pattern p over $\{1, 2, 3\}$ is avoidable in the Abelian sense exactly when no Abelian instance of p appears in Z_3 (C. & Linek, 2001)

- Open problem: What patterns are avoidable in the Abelian sense?
- A pattern p over $\{1, 2, 3\}$ is avoidable in the Abelian sense exactly when no Abelian instance of p appears in Z_3 (C. & Linek, 2001)

Conjecture

A pattern p over $\{1, 2, \ldots k\}$ is avoidable in the Abelian sense exactly when no Abelian instance of p appears in Z_k

- Dekking introduced the morphism $h : 0 \to 0001,\ 1 \to 011$ and showed that the fixed point avoids Abelian $xxxx$.

- Dekking introduced the morphism $h : 0 \to 0001,\ 1 \to 011$ and showed that the fixed point avoids Abelian $xxxx$.
- Open problem: What patterns are avoided in the Abelian sense by the fixed point of h?

- Dekking introduced the morphism $h : 0 \rightarrow 0001,\ 1 \rightarrow 011$ and showed that the fixed point avoids Abelian $xxxx$.
- Open problem: What patterns are avoided in the Abelian sense by the fixed point of h?
- Open problem: What Abelian patterns are 2-avoidable?

- Dekking introduced the morphism $h : 0 \rightarrow 0001$, $1 \rightarrow 011$ and showed that the fixed point avoids Abelian $xxxx$.
- Open problem: What patterns are avoided in the Abelian sense by the fixed point of h?
- Open problem: What Abelian patterns are 2-avoidable?
- Long patterns ($|p| > 118$) are Abelian 2-avoidable (C. & Visentin, 2008)

- Dekking introduced the morphism $h : 0 \rightarrow 0001,\ 1 \rightarrow 011$ and showed that the fixed point avoids Abelian $xxxx$.
- Open problem: What patterns are avoided in the Abelian sense by the fixed point of h?
- Open problem: What Abelian patterns are 2-avoidable?
- Long patterns ($|p| > 118$) are Abelian 2-avoidable (C. & Visentin, 2008)
    - Machinery combines linear algebra and graph theory

- Keränen's proof that xx is Abelian 4-avoidable involves an 85-uniform morphism. This is not human digestible.

- Keränen's proof that xx is Abelian 4-avoidable involves an 85-uniform morphism. This is not human digestible.

Theorem

(C. & Rampersad, 2011) Let μ be a morphism on $\{1, 2, \ldots, m\}$ and M the frequency matrix of μ. Suppose that

- $\mu(1) = 1x$, *some* $x \in \Sigma^+$
- $|\mu(a)| > 1$, *for all* $a \in \Sigma$
- $|M| > 1$

It is decidable whether $\mu^\omega(1)$ avoids Abelian k-powers.

- Keränen's proof that xx is Abelian 4-avoidable involves an 85-uniform morphism. This is not human digestible.

Theorem

(C. & Rampersad, 2011) Let μ be a morphism on $\{1, 2, \ldots, m\}$ and M the frequency matrix of μ. Suppose that

- $\mu(1) = 1x$, *some* $x \in \Sigma^+$
- $|\mu(a)| > 1$, *for all* $a \in \Sigma$
- $|M| > 1$

It is decidable whether $\mu^\omega(1)$ avoids Abelian k-powers.

- I am now optimistic that a human-digestible alternate to Keränen's morphism may be found.

- Keränen's proof that xx is Abelian 4-avoidable involves an 85-uniform morphism. This is not human digestible.

Theorem

(C. & Rampersad, 2011) Let μ be a morphism on $\{1, 2, \ldots, m\}$ and M the frequency matrix of μ. Suppose that

- $\mu(1) = 1x$, *some* $x \in \Sigma^+$
- $|\mu(a)| > 1$, *for all* $a \in \Sigma$
- $|M| > 1$

It is decidable whether $\mu^\omega(1)$ avoids Abelian k-powers.

- I am now optimistic that a human-digestible alternate to Keränen's morphism may be found.
- This theorem is currently being extended to avoidance of arbitrary Abelian patterns.

- Open problem: What patterns are Abelian k-avoidable?

- Open problem: What patterns are Abelian k-avoidable?
- Open problem: What is the least power avoidable in the Abelian sense over a k-letter alphabet? (Abelianize Dejean's conjecture)

Thanks

Thanks for listening!