Remarks on Pansiot encodings

James D. Currie, University of Winnipeg

July 27, 2020
One World Combinatorics on Words Seminar

Overview

Let $\Sigma_n = \{0, 1, 2, \ldots, n-1\}$. Consider an infinite square-free word

$$\mathbf{w} = a_0 a_1 a_2 a_3 a_4 a_5 \cdots$$

where the $a_i \in \Sigma_3$. The Pansiot encoding of $\mathbf{w}$ is the sequence

$$\pi(\mathbf{w}) = b_0 b_1 b_2 b_3 b_4 b_5 \cdots$$

where

$$b_i = \begin{cases} 0, & a_i = a_{i+2} \\ 1, & \text{otherwise} \end{cases}$$

For example, consider vtm, the fixed point of (012,02,1).

$$\mathtt{vtm} = 012021012102012021020121\cdots$$

The Pansiot encoding is in red:

For example, consider vtm, the fixed point of (012,02,1).

$$\texttt{vtm} = \overset{\color{red}{1}}{\underset{\uparrow}{0}}1\underset{\uparrow}{2}0210121020120210201 21\cdots$$

The Pansiot encoding is in red:

For example, consider $\mathtt{vtm}$, the fixed point of (012,02,1).

$$\begin{array}{rl} & \color{red}{1\ 1} \\ \mathtt{vtm} = & 012021012102012021020121\cdots \\ & \ \ \uparrow\ \ \uparrow \end{array}$$

The Pansiot encoding is in red:

For example, consider $\mathtt{vtm}$, the fixed point of (012,02,1).

$$\mathtt{vtm} = \overset{\color{red}1}{0}\overset{\color{red}1}{1}\underset{\uparrow}{\overset{\color{red}0}{2}}0\underset{\uparrow}{2}1012102012021020121\cdots$$

The Pansiot encoding is in red:

For example, consider vtm, the fixed point of (012,02,1).

$$\texttt{vtm} = \overset{\color{red}{1\,1\,0\,1}}{0120210121020120210201 21}\cdots$$

The Pansiot encoding is in red:

For example, consider vtm, the fixed point of (012,02,1).

1 1 0 1 1

$$\mathtt{vtm} = 012021012102012021020121\cdots$$

The Pansiot encoding is in red:

For example, consider vtm, the fixed point of (012,02,1).

$$\mathtt{vtm} = \overset{\color{red}{1\,1\,0\,1\,1\,0}}{0120210}\underset{\uparrow\ \ \uparrow}{12}10201202102012 1\cdots$$

The Pansiot encoding is in red:

For example, consider vtm, the fixed point of (012,02,1).

$$\begin{array}{l} \phantom{\mathtt{vtm} = {}} 1\,1\,0\,1\,1\,0\,1\,0\,1\,1\,0\,1\,1\,1\,0\,1\,1\,1\,0\,1\,1\,0 \\ \mathtt{vtm} = 012021012102012021020121\cdots \end{array}$$

The Pansiot encoding is in red:

A ternary square-free word can be recovered from its Pansiot encoding and its first two letters:

$$\begin{array}{l} \phantom{\mathtt{vtm} = 01}1\,1\,0\,1\,1\,0\;\cdots \\ \mathtt{vtm} = 01 \\ \phantom{\mathtt{vtm} = 01}\uparrow \end{array}$$

A ternary square-free word can be recovered from its Pansiot encoding and its first two letters:

$$\begin{array}{l} \quad\quad\;\; 1\,1\,0\,1\,1\,0\ \cdots \\ \mathtt{vtm} = 01 \\ \quad\quad\quad\quad\;\; \uparrow \end{array}$$

The 1 of the Pansiot encoding shows that the indicated digit is not a 0.

A ternary square-free word can be recovered from its Pansiot encoding and its first two letters:

$$\begin{array}{l} \phantom{\mathtt{vtm} = 01}{\color{blue}1}\,{\color{red}1\,0\,1\,1\,0\ \cdots} \\ \mathtt{vtm} = 011 \\ \phantom{\mathtt{vtm} = 01}\uparrow \end{array}$$

The 1 of the Pansiot encoding shows that the indicated digit is not a 0. It cannot be a 1, or we get the square 11.

A ternary square-free word can be recovered from its Pansiot encoding and its first two letters:

$$\begin{array}{l} \qquad\quad {\color{blue}1}\,{\color{red}1\,0\,1\,1\,0}\ \ {\color{red}\cdots} \\ \mathtt{vtm} = 012 \\ \qquad\qquad\quad\uparrow \end{array}$$

The 1 of the Pansiot encoding shows that the indicated digit is not a 0. It cannot be a 1, or we get the square 11. Therefore it is a 2.

A ternary square-free word can be recovered from its Pansiot encoding and its first two letters:

$$\mathtt{vtm} = \overset{1\,1\,0\,1\,1\,0\;\cdots}{012}$$

A ternary square-free word can be recovered from its Pansiot encoding and its first two letters:

$$\begin{array}{l} \phantom{\mathtt{vtm} = 0}1\,1\,0\,1\,1\,0\ \cdots \\ \mathtt{vtm} = 012 \\ \phantom{\mathtt{vtm} = 012}\uparrow \end{array}$$

The 1 of the Pansiot encoding shows that the indicated digit is not a 1.

A ternary square-free word can be recovered from its Pansiot encoding and its first two letters:

$$\begin{array}{l} \quad\quad\ 1\,1\,0\,1\,1\,0\ \ \cdots \\ \mathtt{vtm} = 0122 \\ \quad\quad\quad\quad\ \ \uparrow \end{array}$$

The 1 of the Pansiot encoding shows that the indicated digit is not a 1. It cannot be a 2, or we get the square 22.

A ternary square-free word can be recovered from its Pansiot encoding and its first two letters:

$$\begin{array}{l} \phantom{\mathtt{vtm} = 0}\,{\color{red}1}\,{\color{blue}1}\,{\color{red}0\,1\,1\,0\quad\cdots} \\ \mathtt{vtm} = 0120 \\ \phantom{\mathtt{vtm} = 012}{\color{blue}\uparrow} \end{array}$$

The 1 of the Pansiot encoding shows that the indicated digit is not a 1. It cannot be a 2, or we get the square 22. Therefore, it is a 0.

A ternary square-free word can be recovered from its Pansiot encoding and its first two letters:

$$\begin{array}{l} \phantom{\mathtt{vtm} = 01}1\,1\,0\,1\,1\,0\ \cdots \\ \mathtt{vtm} = 0120 \\ \phantom{\mathtt{vtm} = 0120}\uparrow \end{array}$$

The 0 of the Pansiot encoding shows that the indicated digit is a 2.

A ternary square-free word can be recovered from its Pansiot encoding and its first two letters:

$$\begin{array}{l} \phantom{\mathtt{vtm} = {}} \mathtt{\ \ }1\,1\,0\,1\,1\,0\ \cdots \\ \mathtt{vtm} = 01202 \\ \phantom{\mathtt{vtm} = 0120}\uparrow \end{array}$$

The 0 of the Pansiot encoding shows that the indicated digit is a 2.

Definition

A word of the form xyx, $xy \neq \epsilon$ is a k power, where $k = |xyx|/|xy|$. An $r+$ power is a k power, some $k > r$. Define the repetitive threshold function to be

$$\mathrm{RT}(n) = \sup\{r : \text{ every infinite word over } \Sigma_n \text{ contains an } r+ \text{ power}\}.$$

Theorem (Dejean's conjecture)

The repetitive threshold function is given by

$$RT(n) = \begin{cases} 2, & n = 2 \\ 7/4, & n = 3 \\ 7/5, & n = 4 \\ n/(n-1), & n \geq 5 \end{cases}$$

This was established via the work of many people, notably Carpi, who resolved all but finitely many cases.

Fix $n \geq 2$, and let $\mathbf{w}$ be a threshold word over Σ_n. Thus

$$\mathbf{w} = a_0 a_1 a_2 a_3 a_4 a_5 \cdots$$

where the $a_i \in \Sigma_n$, and $\mathbf{w}$ contains no $r+$ powers, where $r =$RT(n). The situation we saw with squarefree ternary words generalizes: Suppose ua is a factor of $\mathbf{w}$, where $|u| = n - 1$, and $a \in \Sigma_n$. Because u doesn't contain any $r+$ power, the letters of u must be distinct. Further, either a is the first letter of u, or a does not occur in u.

The Pansiot encoding of $\mathbf{w}$ is the sequence

$$\pi(\mathbf{w}) = b_0 b_1 b_2 b_3 b_4 b_5 \cdots$$

where

$$b_i = \begin{cases} 0, & a_i = a_{i+n-1} \\ 1, & \text{otherwise} \end{cases}$$

Word $\mathbf{w}$ can be recovered from its Pansiot encoding and its prefix of length $n-1$.

Pansiot encodings allowed Dejean's conjecture to be resolved via constructions on binary alphabets, rather than by working on Σ_n.

In published work on Dejean's conjecture, to recover information about k powers in $\mathbf{w}$ from the Pansiot encoding, group theory is used.

$$\begin{array}{l} \phantom{\mathtt{vtm} = {}} \color{red}{1\,1\,0\,1\,1\,0\,1\,0\,1\,1\,0\,1\,1\,1\,0\,1\,1\,1\,0\,1\,1\,0} \\ \mathtt{vtm} = 012021012102012021020121\cdots \end{array}$$

Long repeated factors in **w** correspond to long repeated factors in the Pansiot encoding.

$$\begin{array}{l} \phantom{\mathtt{vtm} = {}} \color{red}{1\,1\,0\,1\,1\,0\,1\,0\,1\,1\,0\,1\,1\,1\,0\,1\,1\,1\,0\,1\,1\,0} \\ \mathtt{vtm} = \underline{0120210}121020\underline{1202102}0121\cdots \end{array}$$

Long repeated factors in **w** correspond to long repeated factors in the Pansiot encoding.

$$\textcolor{red}{\overline{11011}0101101\overline{11011}10110}$$

$$\mathtt{vtm} = \underline{0120210}12102\underline{0120210}20121\cdots$$

Long repeated factors in **w** correspond to long repeated factors in the Pansiot encoding.

$$\begin{array}{l} \overline{11011}010\overline{11011}101110110 \\ \mathtt{vtm} = 012021012102012021020121\cdots \end{array}$$

The reverse is not true.

$$\begin{array}{l} \phantom{\mathtt{vtm} = }\;\overline{1\,1\,0\,1\,1}\,0\,1\,0\,\overline{1\,1\,0\,1\,1}\,1\,0\,1\,1\,1\,0\,1\,1\,0 \\ \mathtt{vtm} = \underline{0120210}1\underline{2102012}021020121\cdots \end{array}$$

The reverse is not true.

$$\overline{1\,1\,0\,1\,1}\,0\,1\,0\,\overline{1\,1\,0\,1\,1}\,1\,0\,1\,1\,1\,0\,1\,1\,0$$

$$\mathtt{vtm} = \underline{0120210}1\underline{2102012}021020121\cdots$$

The reverse is not true. The factor in **w** is only repeated up to a permutation (switching 0 and 2).

Consider a length $n-1$ factor u of $\mathbf{w}$, $u = a_0 a_1 \cdots a_{n-2}$ where the a_i are distinct letters of Σ_n. Let a_{n-1} be the unique letter in $\Sigma_n - \{a_0, a_1, \ldots, a_{n-2}\}$. Consider the correspondence

$$u \xrightarrow{\varphi} \begin{pmatrix} 0 & 1 & \cdots & (n-2) & (n-1) \\ a_0 & a_1 & \cdots & a_{n-2} & a_{n-1} \end{pmatrix}$$

which matches words and permutations.

Let u_i be the length $n-1$ factor at index i in $\mathbf{w}$. As before, let b_i be the ith letter of $\pi(\mathbf{w})$, the Pansiot encoding of $\mathbf{w}$. Then

$$\varphi(u_{i+1}) = \sigma(b_i)\varphi(u_i),$$

where

$$\sigma(0) = \begin{pmatrix} 0 & 1 & \cdots & (n-3) & (n-2) & (n-1) \\ 1 & 2 & \cdots & (n-2) & 1 & (n-1) \end{pmatrix}$$

and

$$\sigma(1) = \begin{pmatrix} 0 & 1 & \cdots & (n-3) & (n-2) & (n-1) \\ 1 & 2 & \cdots & (n-2) & (n-1) & 1 \end{pmatrix}$$

In our example where $n = 3$ and $\mathbf{w} = \texttt{vtm} = \underline{01}2\underline{02}1012\cdots$, we have $u_0 = 01$, $u_3 = 02$. Also $\pi(\mathbf{w}) = 110110\cdots$, and

$$\sigma(0) = \begin{pmatrix} 0 & 1 & 2 \\ 1 & 0 & 2 \end{pmatrix} = (0\ 1); \sigma(1) = \begin{pmatrix} 0 & 1 & 2 \\ 1 & 2 & 0 \end{pmatrix} = (0\ 1\ 2)$$

and

$$\varphi(u_0) = \begin{pmatrix} 0 & 1 & 2 \\ 0 & 1 & 2 \end{pmatrix} = (); \varphi(u_3) = \begin{pmatrix} 0 & 1 & 2 \\ 0 & 2 & 1 \end{pmatrix} = (1\ 2)$$

Thus $\sigma(0)\sigma(1)\sigma(1)\varphi(u_0) = (0,1)(0,1,2)(0,1,2)() = (1\ 2) = u_3$.

Let σ be the antimorphism on Σ_2 generated by $\sigma(0), \sigma(1)$. If $\mathbf{b_m}$ is the prefix of $\pi(\mathbf{w})$ of length m, then

$$\varphi(u_m) = \sigma(\mathbf{b_m})\varphi(u_0).$$

A power xyx in $\mathbf{w}$, of period $|xy|$ and length xyx corresponds to a power $\pi(xyx)$ in $\pi(\mathbf{w})$, of period $|xy|$ and length $|xyx| - (n-1)$. If $|x| \geq n-1$, then $\sigma(\pi(xy))$ must be the identity permutation. We call $\pi(xyx)$ a kernel repetition in $\pi(\mathbf{w})$ (since $\pi(xy)$ is in the kernel of φ). Dejean's conjecture was solved by constructing binary sequences avoiding short repetitions (with $|x| < n-1$) and kernel repetitions.

If power xyx in $\mathbf{w}$ has period p and length m, then $\pi(xyx)$ has period p and length $m-(n-1)$. If $|x| \geq n-1$, then $\sigma(\pi(xy)) = ()$, the identity element of the permutation group $\mathcal{S}_n$.

For fixed n, rather than working in $\mathcal{S}_n$, we can consider the Cayley graph of $\mathcal{S}_n$ with generators $\sigma(0)$ and $\sigma(1)$.

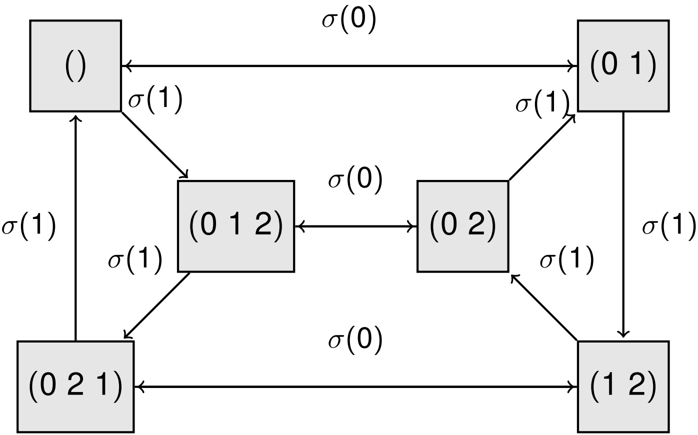

Figure: $\sigma(0)\sigma(1)\sigma(1)() = (1\ 2)$

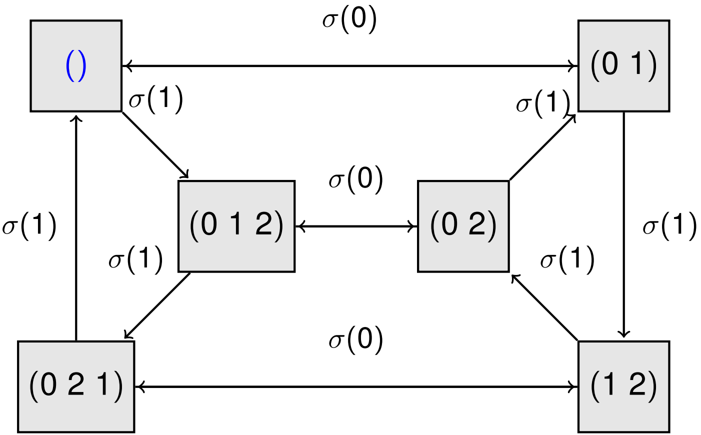

Figure: $\sigma(0)\sigma(1)\sigma(1)() = (1\ 2)$

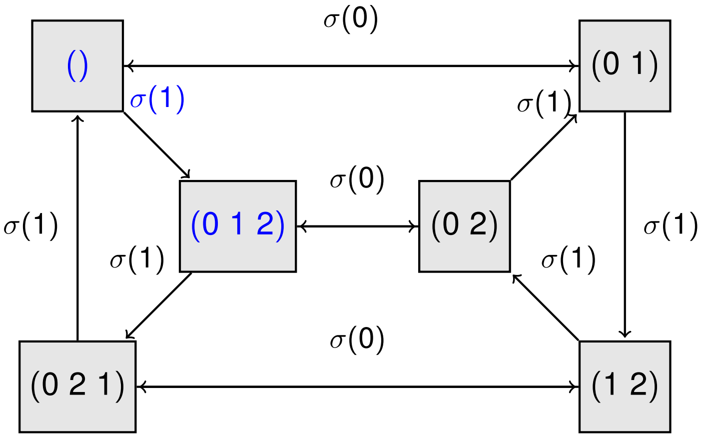

Figure: $\sigma(0)\sigma(1)\sigma(1)() = (1\ 2)$

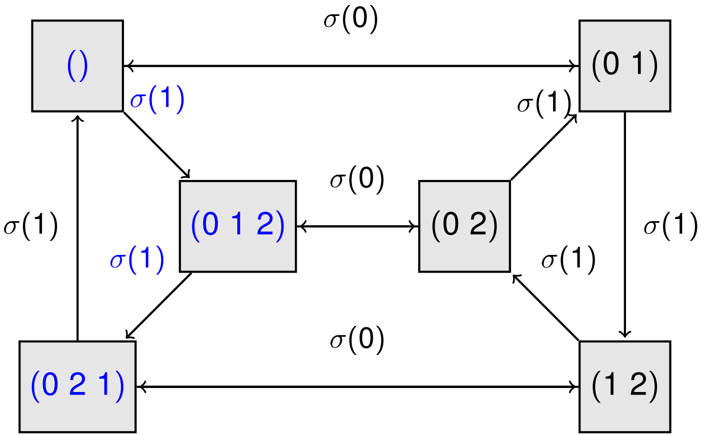

Figure: $\sigma(0)\sigma(1)\sigma(1)() = (1\ 2)$

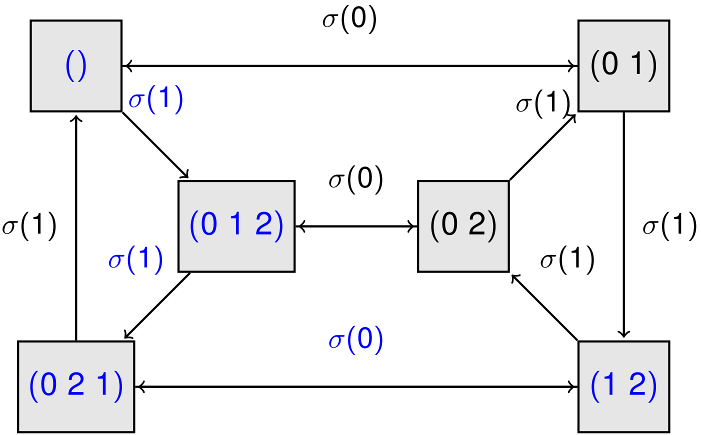

Figure: $\sigma(0)\sigma(1)\sigma(1)() = (1\ 2)$

If power xyx in $\mathbf{w}$ has period p and length m, then $\pi(xyx)$ has period p and length $m-(n-1)$. If $|x| \geq n-1$, $\sigma(\pi(xy))$ labels a closed walk in the Cayley graph. This Cayley graph is well-known to us under an alias; it is isomorphic to the De Bruijn graph of length 2 factors of $\mathbf{w}$.

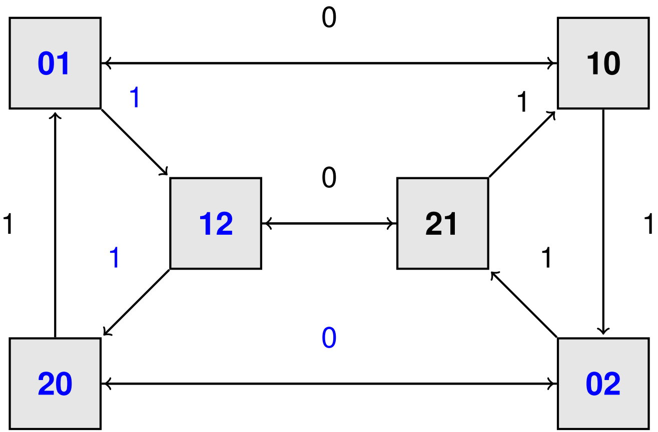

Figure: De Bruijn graph (with Pansiot encoding edge labels)

Parsing Pansiot encodings

One verifies that the Pansiot encoding of an infinite threshold word $\mathbf{w}$ cannot contain 00 or 1111 as a factor. It follows a final segment of $\pi(\mathbf{w})$ has the form $\pi(\mathbf{w}) = f(\mathbf{v})$ where

$$\begin{aligned} f(1) &= 01 \\ f(2) &= 011 \\ f(3) &= 0111 \end{aligned}$$

Parsing Pansiot encodings (continued)

We consider again the alphabet Σ_3. We form a graph $\mathcal{G}$ on the vertices of the previous graph: For each $\alpha \in S = \{1, 2, 3\}$, and for each vertex xy, introduce an edge from xy to zw, labeled by α, exactly when zw is the endpoint of the walk in the previous graph labeled by $f(\alpha)$ starting at xy. Thus $U \in S^*$ labels a closed walk on $\mathcal{G}$, exactly when $f(U)$ labels a closed walk on on the previous graph.

Parsing Pansiot encodings (continued)

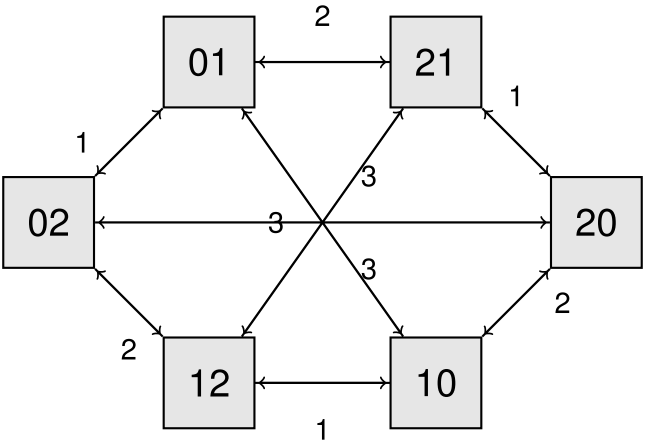

Figure: Graph $\mathcal{G}$

Thue studied finite, right-infinite, doubly-infinite, and circular words avoiding powers. The last two types of words are more structurally uniform than the first two, since there are no ends where ‘unsustainable’ behaviours can take place. Thue characterized the overlap-free circular binary words.
A lot of work has been done on ‘linear’ words avoiding patterns. Many questions are open regarding circular words.

Once an infinite word avoiding some pattern exists, finite linear words of every length obviously exist. The situation is different for circular words. For example, no circular binary overlap free word of length 5 exists.
The following conjecture appears to be due to R. Jamie Simpson:

Theorem

There are ternary circular square free words of length n for every n except $n = 5, 7, 9, 10, 14$ *and 17.*

This was solved in 2002, using an approach which has become typical in combinatorics on words: It was shown that such words exist for every $n \geq 180$, and then a computer search establishes the result.
In 2010, Shur gave a computer-free proof of this result, relying on a variation of Pansiot encoding.

Given a word $w = u_1 u_2 \cdots u_m$, $u_m \in \Sigma_n$, the circular word $[w]$ is the set of conjugates of w. It is natural to think of $[w]$ as consisting of the letters of w arranged in a circle or 'necklace'. Equivalently, we may consider the indices i of the letters of a circular word $[u] = [u_1 u_2 \cdots u_n]$ to belong to $\mathbb{Z}_n$, the integers modulo n. Thus $u_{n+1} = u_1$, for example.

0

1 1

2 2

0

Figure: Circular word [012021]

0

1 1

2 2

0

Figure: Circular word [012021]

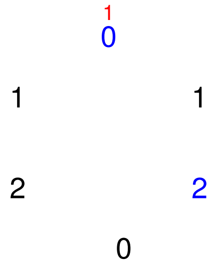

Figure: Circular word [012021]

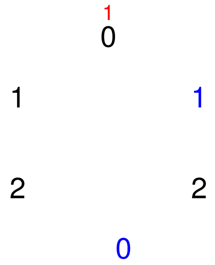

Figure: Circular word [012021]

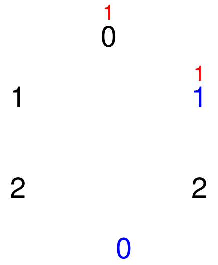

Figure: Circular word [012021]

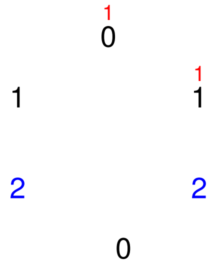

Figure: Circular word [012021]

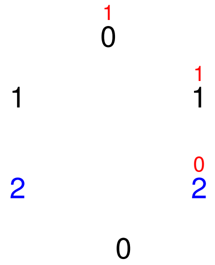

Figure: Circular word [012021]

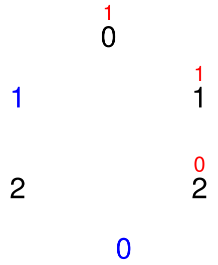

Figure: Circular word [012021]

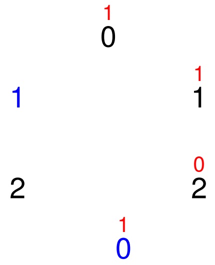

Figure: Circular word [012021]

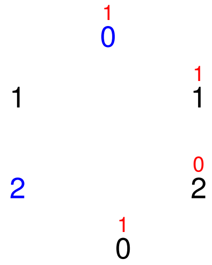

Figure: Circular word [012021]

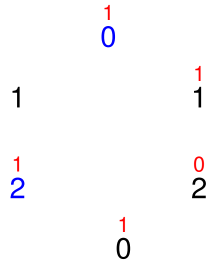

Figure: Circular word [012021]

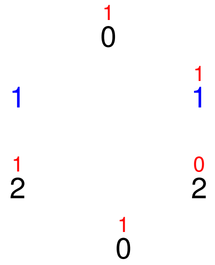

Figure: Circular word [012021]

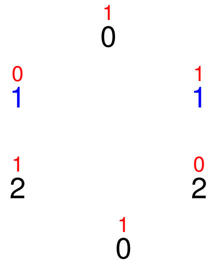

Figure: Circular word [012021]

The Pansiot encoding of 202101 is $011011 = f(22)$. In the circular word, $u_0u_1 = 01 = u_6u_7$, so that we have $\sigma(011011) = ()$. Put another way, 011011 gives a closed walk on our De Bruijn graph, and 22 labels a closed walk on $\mathcal{G}$. Parsing Pansiot encodings by f means we deal with words only 1/3 as long.

Theorem (Shur 2010)

If w labels a closed walk on $\mathcal{G}$, then $[f(w)]$ is the Pansiot encoding of a square-free circular word if

- $[w]$ *has no factor 11, 222, 223, 322, or 333*
- $[w]$ *has no factor UxyU with Uxy a closed walk,* $|U| \geq 2$*,* $x, y \in \{1, 2, 3\}$ *.*

This is not quite if and only if: for example, word 0102 has Pansiot encoding 0101$= f(11)$

Call a word w over Σ_n level if $|w|_i - |w|_j \in \{-1, 0, 1\}$ for $i, j \in \Sigma_n$. Call a word w over Σ_2 an FS word if the only square factors of w are 00, 11 and 0101.

Jesse Johnson used Shur's approach to prove these results in his MSc thesis:

Theorem (Johnson, 2020)

There are level ternary circular square free words of length n for every n except $n = 5, 7, 9, 10, 14$ *and 17.*

Theorem (Johnson, 2020)

There are circular FS words of every length.

Remark

Level words are useful when we care about the length of constructed words: If $w \in \Sigma_n^{kn}$ is level, then for any morphism g, word $g(w)$ has length $k|g|$. Level words can be used analogously to k-uniform morphisms, where the image of word u has length $k|u|$.

Call xyz an undirected k power if $z \in \{x, x^R\}$, and $|xyx|/|xy| = k$.
Thus *reenter* is an undirected $7/5$ power, and so is *stalest*.
If we required instead that z was an anagram of x, we would have an **Abelian power**, in the sense used by Cassaigne & C. (1999). One can also see undirected powers as a common generalization of gapped repeats and gapped palindromes.

The undirected repetition threshold function is defined to be

$$URT(k) = \inf\{r : \text{ undirected } r\text{-powers are avoidable on } k \text{ letters.}\}$$

Theorem (Mol &C., 2019)

For $k = 4, 5, \ldots, 21$, we have

$$URT(k) = \frac{k-1}{k-2}.$$

Avoiding gapped palindromes xyx^R where $|x| > 1$ is rather easy; for example $(123)^\omega$ contains no such gapped palindrome. We will therefore focus on the issue of constructing words over Σ_k avoiding ordinary $(k-1)/(k-2)+$ powers.

Dejean's conjecture revisited

Suppose **w** is a word of Σ_k^ω that avoids $r+$ powers for some fixed r. Consider first the case of Dejean's conjecture, where we are interested in $r = k/(k-1)$. In this case, factors of **w** of length $k-1$ must contain $k-1$ distinct letters; otherwise **w** contains a power of exponent at least $(k-1)/(k-2) > k/(k-1)$.

Dejean's conjecture revisited (continued)

Suppose that $pa_1 a_2 \cdots a_k$ is a prefix of $\mathbf{w}$, where the $a_i \in \Sigma_k$. The letters of $\{a_2, a_3 \cdots a_k\}$ must be distinct. It follows that a_k is one of the two letters in $\Sigma_k - \{a_2, a_3, \ldots, a_{k-1}\}$. Of these two letters, the one occurring closest to the end of $pa_1 a_2 \cdots a_{k-1}$ is a_1 (because the letters of $\{a_1, a_2 \cdots a_{k-1}\}$ must be distinct).

The Pansiot encoding may be constructed by scanning prefixes of $\mathbf{w}$, starting with the length k prefix: When we scan $pa_1 a_2 \cdots a_k$, we afix 0 to the Pansiot encoding if a_k equals the letter of $\Sigma_k - \{a_2, a_3, \ldots, a_{k-1}\}$ closest to the end of $pa_1 a_2 \cdots a_{k-1}$; we afix 1 if a_k equals the letter of $\Sigma_k - \{a_2, a_3, \ldots, a_{k-1}\}$ second closest to the end.

A Pansiot-like encoding

Now consider the case $r = (k-1)/(k-2)$, as in the undirected powers problem. In this case, factors of **w** of length $k-2$ must contain $k-2$ distinct letters; otherwise **w** contains a power of exponent at least $(k-2)/(k-3) > (k-1)/(k-2)$.

A Pansiot-like encoding (continued)

Suppose that $pa_1a_2\cdots a_{k-1}$ is a prefix of $\mathbf{w}$, where the $a_i \in \Sigma_k$. The letters of $\{a_2, a_3\cdots a_{k-1}\}$ must be distinct. It follows that a_{k-1} is one of the three letters in $\Sigma_k - \{a_2, a_3, \ldots, a_{k-2}\}$. Of these three letters, the one occurring closest to the end of $pa_1a_2\cdots a_{k-2}$ is a_1 (because the letters of $\{a_1, a_2\cdots a_{k-2}\}$ must be distinct).

A Pansiot-like encoding $\mathbf{b}$ may be constructed by scanning prefixes of $\mathbf{w}$, starting with the length $k-1$ prefix: When we scan $pa_1a_2\cdots a_{k-1}$, we afix 0 to the Pansiot encoding if a_{k-1} equals the letter of $\Sigma_k - \{a_2, a_3, \ldots, a_{k-2}\}$ closest to the end of $pa_1a_2\cdots a_{k-2}$; we afix 1 if a_k equals the letter of $\Sigma_k - \{a_2, a_3, \ldots, a_{k-2}\}$ second closest to the end; we afix 2 if a_k equals the letter of $\Sigma_k - \{a_2, a_3, \ldots, a_{k-2}\}$ third closest to the end.

A Pansiot-like encoding (continued)

There are some technical issues at the very start, where the question of ‘second closest’ versus ‘third closest’ is undefined. One can either introduce a tie-breaking convention, or alter the encoding, to start with the first prefix of **w** containing all letters of Σ_k.

Let the encoding be

$$\mathbf{b} = b_0 b_1 b_2 b_3 \cdots$$

Consider a prefix $P = pa_1a_2\cdots a_{k-1}$ of $\mathbf{w}$, where the a_i are distinct letters of Σ_k. Consider the correspondence

$$P \xrightarrow{\varphi} \begin{pmatrix} 0 & 1 & 2 & 3\cdots & (k-2) & (k-1) \\ r_3 & r_2 & a_1 & a_2\cdots & a_{k-3} & a_{k-2} \end{pmatrix}$$

where r_3 is the letter of $\Sigma_k - \{a_2, a_3, \ldots, a_{k-2}\}$ third closest to the end of $pa_1a_2\cdots a_k$, and r_2 is the letter of $\Sigma_k - \{a_2, a_3, \ldots, a_{k-2}\}$ second closest to the end of $pa_1a_2\cdots a_k$.

Let u_i be the prefix of $\mathbf{w}$ of length $i+k-1$. Then

$$\varphi(u_{i+1}) = \sigma(b_i)\varphi(u_i),$$

where

$$\sigma(0) = \begin{pmatrix} 0 & 1 & 2 & 3 & \cdots & (k-3) & (k-2) \\ 0 & 1 & 3 & 4 & \cdots & (k-2) & 2 \end{pmatrix},$$

$$\sigma(1) = \begin{pmatrix} 0 & 1 & 2 & 3 & \cdots & (k-3) & (k-2) \\ 0 & 2 & 3 & 4 & \cdots & (k-2) & 1 \end{pmatrix}$$

and

$$\sigma(0) = \begin{pmatrix} 0 & 1 & 2 & 3 & \cdots & (k-3) & (k-2) \\ 1 & 2 & 3 & 4 & \cdots & (k-2) & 0 \end{pmatrix}$$

Let σ be the antimorphism on Σ_3 generated by $\sigma(0)$,$\sigma(1)$, $\sigma(2)$. If $\mathbf{b_m}$ is the prefix of $\pi(\mathbf{w})$ of length m, then

$$\varphi(u_m) = \sigma(\mathbf{b_m})\varphi(u_0).$$

We are again led to avoiding short repetitions and kernel repetitions.

Pansiot encodings are unnecessary!

The constructions in the solutions to Dejean's conjecture use binary fixed points of D0Ls. These are Pansiot encodings, and are decoded into words of Σ_n^ω. From a formal point of view, the words witnessing the correctness of Dejean's conjecture are thus transductions of D0L sequences.

Theorem (Dekking, 1994 (Theorem 7.9.1 in Allouche & Shallit))

The transduction of an HD0L sequence is an HD0L sequence.

Thus witnesses to Dejean's Conjecture can be given directly by HD0Ls.

Points in favour of Pansiot encodings

- Attacking threshold problems via encodings unifies questions involving multiple alphabets.
- Carpi's contribution involved group theory in an important way, implying that the group viewpoint is necessary.
- In the case of squarefree words, encodings (and the parsing function f) reduce the depth of searches by 1/3.

Some open problems

1. Recast Carpi's group-theoretic work via De Bruijn graphs. Does this give further insight?
2. Prove that $URT(k) = \frac{k-1}{k-2}$ for all $k \geq 4$.
3. Generalize classical pattern avoidance results to circular words.

Thank you!