Reductions on Double Occurrence Words

Ryan Arredondo
rarredon@mail.usf.edu

University of South Florida
Tampa, Florida

03/08/2013
Table of Contents

1 Biological background
2 Double occurrence words
3 Reduction of a double occurrence word
4 Nesting index of a double occurrence word
5 Chord diagrams
6 Circle graphs
7 Questions and Conjectures
Biological motivation

Figure: A ciliate under microscope
Segments of ciliate DNA are eliminated or rearranged. ¹

Model for rearrangement

- 4-valent rigid vertex graphs with endpoints (assembly graphs) are used to model genome rearrangement \(^2\)

Figure: Assembly graph model for Actin I gene in *Oxytricha Nova*

Patterns in scrambled genes

A pattern observed in the scrambled genes of ciliates is related to its evolutionary complexity. ³

Patterns in scrambled genes

- A pattern observed in the scrambled genes of ciliates is related to its evolutionary complexity. ³

Goal:
  - Double occurrence words from assembly graphs

Patterns in scrambled genes

- A pattern observed in the scrambled genes of ciliates is related to its evolutionary complexity. ³

Goal:
- Double occurrence words from assembly graphs
- Define words that relate to the patterns observed

A pattern observed in the scrambled genes of ciliates is related to its evolutionary complexity.  

**Goal:**
1. Double occurrence words from assembly graphs
2. Define words that relate to the patterns observed
3. Define reduction operations on double occurrence words

---

Patterns in scrambled genes

- A pattern observed in the scrambled genes of ciliates is related to its evolutionary complexity. ³

Goal:

1. Double occurrence words from assembly graphs
2. Define words that relate to the patterns observed
3. Define reduction operations on double occurrence words
4. Define the nesting index as a measurement of complexity for scrambled genes

A pattern observed in the scrambled genes of ciliates is related to its evolutionary complexity.  

**Goal:**

1. Double occurrence words from assembly graphs
2. Define words that relate to the patterns observed
3. Define reduction operations on double occurrence words
4. Define the nesting index as a measurement of complexity for scrambled genes
5. Observe how the nesting index relates to other representations of double occurrence words (e.g. chord diagrams and circle graphs)

---

Table of Contents

1 Biological background

2 Double occurrence words

3 Reduction of a double occurrence word

4 Nesting index of a double occurrence word

5 Chord diagrams

6 Circle graphs

7 Questions and Conjectures
Double occurrence words

- A *double occurrence word* is a word over a finite alphabet \( \Sigma = \{1, 2, 3, \ldots, n\} \) in which each symbol appears exactly twice.
A *double occurrence word* is a word over a finite alphabet $\Sigma = \{1, 2, 3, \ldots, n\}$ in which each symbol appears exactly twice.

*Figure:* Assembly graph represented by word $v_1v_2v_1v_2$ or 1212
A *return word* is a word of the form 

\[ a_1 a_2 \cdots a_n a_n \cdots a_2 a_1, \quad a_i \in \Sigma \text{ for all } i, \text{ and } a_i \neq a_j \text{ for } i \neq j \]

(a) 123321 is a return word
Patterns translated

- A \textit{return word} is a word of the form
  \[ a_1 a_2 \cdots a_n a_n \cdots a_2 a_1, \quad a_i \in \Sigma \text{ for all } i, \text{ and } a_i \neq a_j \text{ for } i \neq j \]

- A \textit{repeat word} is a word of the form
  \[ a_1 a_2 \cdots a_n a_1 a_2 \cdots a_n, \quad a_i \in \Sigma \text{ for all } i, \text{ and } a_i \neq a_j \text{ for } i \neq j \]

(a) 123321 is a return word

(b) 12341234 is a repeat word
Table of Contents

1 Biological background

2 Double occurrence words

3 Reduction of a double occurrence word

4 Nesting index of a double occurrence word

5 Chord diagrams

6 Circle graphs

7 Questions and Conjectures
Reductions

- Two reduction operations

1. Removal of all maximal repeat and return words
2. Removal of a letter

A reduction $R$ of a word $w$ is a sequence of words $(w_0, w_1, w_2, ..., w_n = \epsilon)$ where $w_{i+1}$ is obtained from $w_i$ for $0 \leq i < n$ by one of the two reduction operations.
Reductions

- Two reduction operations
  - Removal of all maximal repeat and return words

(a) 123324564561 $\rightarrow$ 11
Reductions

- Two reduction operations
  1. Removal of all maximal repeat and return words
  2. Removal of a letter

(a) \textcolor{red}{123324564561} \rightarrow \textcolor{red}{11}

(b) \textcolor{red}{123324564561} \rightarrow \textcolor{red}{2332456456}
Two reduction operations
1. Removal of all maximal repeat and return words
2. Removal of a letter

A reduction $R$ of a word $w$ is a sequence of words $(w = w_0, w_1, w_2, \ldots, w_n = \epsilon)$ where $w_{i+1}$ is obtained from $w_i$ for $0 \leq i < n$ by one of the two reduction operations.

(a) 123324564561 $\rightarrow$ 11
(b) 123324564561 $\rightarrow$ 2332456456
Example 1

(a) \( w_0 = 1234554231 \)
Example 1

(a) \( w_0 = 1234554231 \)

(b) \( w_1 = 123231 \)
Example 1

(a) $w_0 = 1234554231$

(b) $w_1 = 123231$

(c) $w_2 = 11$
Example 1

(a) $w_0 = 1234554231$

(b) $w_1 = 123231$

(c) $w_2 = 11$

(d) $w_3 = \epsilon$
Example 2

(a) $w_0 = 1234554231$
Example 2

(a) $w_0 = 1234554231$

(b) $w_1 = 124421$
Example 2

(a) $w_0 = 1234554231$

(b) $w_1 = 124421$

(c) $w_2 = \epsilon$
Table of Contents

1 Biological background
2 Double occurrence words
3 Reduction of a double occurrence word
4 Nesting index of a double occurrence word
5 Chord diagrams
6 Circle graphs
7 Questions and Conjectures
Nesting index

- *Nesting index* of a word $w$ is the least number of reduction operations that reduce $w$ to $\varepsilon$
- More formally,

$$NI(w) = \min\{n : (w = w_0, w_1, \ldots, w_n = \varepsilon) \text{ is a reduction of } w\}$$

<table>
<thead>
<tr>
<th></th>
<th>Example 1</th>
<th>Example 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>$w_0$</td>
<td>1234554231</td>
<td>1234554231</td>
</tr>
<tr>
<td>$w_1$</td>
<td>123231</td>
<td>12455421</td>
</tr>
<tr>
<td>$w_2$</td>
<td>11</td>
<td>$\varepsilon$</td>
</tr>
<tr>
<td>$w_3$</td>
<td>$\varepsilon$</td>
<td>N/A</td>
</tr>
</tbody>
</table>

*Table*: Reductions of $w = 12344321$

- No known efficient algorithm to compute Nesting index
A more complex case

Figure: Assembly graph for word 1, 2, 3, 4, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 11, 18, 12, 19, 20, 18, 21, 7, 22, 23, 16, 24, 25, 26, 3, 27, 27, 28, 28, 2, 1, 29, 30, 30, 29, 5, 26, 25, 24, 17, 23, 22, 6, 21, 31, 31, 8, 9, 10, 20, 19, 13, 14, 15

Nesting index = 15
Chord diagrams

(a) $w = 12341234$

(b) $w = 12344321$

If $w$ is a repeat word, then every pair of chords intersect.

If $w$ is a return word, then no pair of chords intersect.

And now for something completely different...

Word $w = 123213$ has chord diagram
Chord diagrams

(a) $w = 12341234$

(b) $w = 12344321$

- If $w$ is a repeat word, then every pair of chords intersect.
- If $w$ is a return word, then no pair of chords intersect.
Chord diagrams

(a) $w = 12341234$

(b) $w = 12344321$

- If $w$ is a repeat word, then every pair of chords intersect.
- If $w$ is a return word, then no pair of chords intersect.

And now for something completely different...

Word $w = 123213$ has chord diagram
**Theorem**

Every reduction of $w$ must use the second reduction operation if and only if the chord diagram of $w$ has the following sub-chord diagram.
Theorem

Every reduction of w must use the second reduction operation if and only if the chord diagram of w has the following sub-chord diagram.
Theorem

Every reduction of \( w \) must use the second reduction operation if and only if the chord diagram of \( w \) has the following sub-chord diagram.

Corollary

Let \( 2 \leq n \leq m \) be integers. If a word \( w \) has a chord diagram with the following as a sub-chord diagram, then \( \text{NI}(w) \geq n + 1 \).
Circle graphs

123456563412
Circle graphs

123456563412 →

[Diagram of a circle graph with numbers 1 to 6 connected by lines, showing a transformation process.]
Circle graphs

123456563412 →

1 3
2 4
5
6

1
2
3
4
5
6

1
2
3
4
5
6

Ryan Arredondo (USF)
Circle graphs

123456563412

121234345656
Circle graphs

123456563412 → 

1 3
2 4
5
6

121234345656 → 

1 3
2 4
5
6

Reductions on Double Occurrence Words

Ryan Arredondo (USF)
Fact

There exist words $w_1$ and $w_2$ over $2n$ letters, that have the same circle graph and

$$\text{NI}(w_1) = n, \quad \text{NI}(w_2) = 1$$
Questions and Conjectures

Questions involving circle graphs

- Can we characterize words with the same circle graph and the same nesting index?

Let $S$ be a set of words with the same circle graph and nesting index. Does there exist $N$ independent of $S$ such that $|S| \leq N$?

Minimal realization

How many letters are needed to construct a word with nesting index $n$?

Conjecture: $n + \# \text{nonzero squares less than } n$
Questions and Conjectures

Questions involving circle graphs

- Can we characterize words with same circle graph and same nesting index?
- Let $S$ be a set of words with the same circle graph and nesting index. Does there exist $N$ independent of $S$ such that $|S| \leq N$?
Questions and Conjectures

Questions involving circle graphs
- Can we characterize words with the same circle graph and same nesting index?
- Let $S$ be a set of words with the same circle graph and nesting index. Does there exist $N$ independent of $S$ such that $|S| \leq N$?

Minimal realization
- How many letters are needed to construct a word with nesting index $n$?
Questions and Conjectures

Questions involving circle graphs

- Can we characterize words with same circle graph and same nesting index?
- Let $S$ be a set of words with the same circle graph and nesting index. Does there exist $N$ independent of $S$ such that $|S| \leq N$?

Minimal realization

How many letters are needed to construct a word with nesting index $n$?

Conjecture:

$$n + \# \text{nonzero squares less than } n$$
Nesting index counts and conjectures

Table: Number of words of a given size and nesting index

<table>
<thead>
<tr>
<th>Size</th>
<th>Nesting Index</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>3</td>
<td>7</td>
</tr>
<tr>
<td>4</td>
<td>17</td>
</tr>
<tr>
<td>5</td>
<td>41</td>
</tr>
<tr>
<td>6</td>
<td>99</td>
</tr>
<tr>
<td>7</td>
<td>239</td>
</tr>
<tr>
<td>8</td>
<td>577</td>
</tr>
<tr>
<td>9</td>
<td>1393</td>
</tr>
<tr>
<td>10</td>
<td>2019</td>
</tr>
<tr>
<td>11</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td></td>
</tr>
</tbody>
</table>

- First column corresponds to the numerators of continued fraction convergents to $\sqrt{2}$.
Thanks!

Many thanks to Dr. Nataša Jonoska, Dr. Masahico Saito, Jonathan Burns, Sarah Croome, Egor Dolzhenko, Maja Milošević, Jamie Sprecher, and Timothy Yeatman.

For related work please visit http://knot.math.usf.edu

This work has been supported in part by NSF Grant DMS #0900671