Efficient identification of $k$-closed strings

Hayam Alamro$^1$  Mai Alzamel$^1$  Costas S. Iliopoulos$^1$  Solon P. Pissis$^1$  Wing-Kin Sung$^2$  Steven Watts$^1$

EANN 2017

$^1$Department of Informatics
King’s College London

$^2$Department of Computer Science
National University of Singapore
Outline

Background

New Problem

Algorithm

Summary
Background
Closed strings were introduced by Fici [1] as objects of combinatorial interest.

Closed strings have a relationship with palindromic strings [2].

Badkobeh et al. [3] factorised a string into a sequence of longest closed factors in time and space $O(n)$.

Badkobeh et al. [3] computed the longest closed factor starting at every position in a string in $O(n \frac{\log n}{\log \log n})$ time and $O(n)$ space.
Prefixes

Definition
A prefix of a string $x$ is a substring $p$ of length $m$, which occurs at the beginning of $x$, i.e. at index 0.

$p = x[0..m-1]$

A prefix is called a proper prefix if it does not correspond to the full string $x$, i.e. $|p| < |x|$.
**Definition**

A **suffix** of a string $x$ is a substring $s$ of length $m$, which occurs at the end of $x$, i.e. at index $n - m$, where $n$ is the length of $x$.

$s = x[n - m \ldots n - 1]$

A suffix is called a **proper suffix** if it does not correspond to the full string $x$, i.e. $|s| < |x|$. 
Definition

A bordered string is a string $x$ for which there exists a proper prefix $b$, which is simultaneously a proper suffix. We call such a $b$, a border.

$$x[0..b - 1] = x[n - b..n - 1]$$
**Closed Strings**

**Definition**

A closed string is a bordered string $x$ such that some border $b$ of $x$ occurs exactly twice in $x$. We call such a $b$, the closed border.

**Closed**

$$a b a g t a b t t t a b a a$$

$b$  

$b$

**Non-Closed**

$$a b a g t a b a a t a b a a$$

$b$  

$b$
New Problem
Goals

• Generalise closed strings to $k$-closed strings, where $k$ is a measure of approximation.

• Choose a natural definition of $k$-closed such that:
  closed $\implies$ 1-closed $\implies$ 2-closed $\implies$ 3-closed $\ldots$

• Develop an efficient algorithm to identify whether or not a string is $k$-closed.
Hamming Distance

We use **Hamming distance** (number of mismatched characters) as a measure of approximation between two strings or factors.

e.g. agtcta and agacga have Hamming distance 2.
Closed String: 2 Conditions

There are 2 conditions that must be satisfied for a string $x$ to be closed, both conditions can potentially be approximated individually or simultaneously by a parameter $k$:

1. Border Condition:
   $x$ has a border $b$.

2. No Internal occurrence Condition:
   $x$ has no internal occurrences of border $b$. 
Closed Definitions with Approximation

Closed (Already Defined)
Border Condition: Exact
No Internal occurrence Condition: Exact

\textit{k-Weakly-Closed}
Border Condition: Approximate
No Internal occurrence Condition: Exact

\textit{k-Strongly-Closed}
Border Condition: Exact
No Internal occurrence Condition: Approximate

\textit{k-Pseudo-Closed}
Border Condition: Approximate
No Internal occurrence Condition: Approximate
**k-Weakly-Closed Strings: Definition**

**Definition**
A string $x$ of length $n$ is called $k$-weakly-closed if and only if $n \leq 1$ or the following properties are satisfied:

1. There exists some proper prefix $u$ of $x$ and some proper suffix $v$ of $x$ of length $|u| = |v|$, such that $\delta_H(u, v) \leq k$.

2. Both factors $u$ and $v$ occur only as a prefix and suffix respectively within $x$, i.e. no internal occurrences of $u$ or $v$ exist in $x$.

We call such a pair $u$ and $v$ a $k$-weakly-closed border of $x$. In the case where $n \leq 1$, we assign $\varepsilon$ as the $k$-weakly-closed border.
k-Weakly-Closed Strings: Example \((k = 1)\)

**Border Condition:** Approximate

**No Internal occurrence Condition:** Exact

\[
\begin{align*}
\text{k-Weakly-Closed} & \quad \text{Non-k-Weakly-Closed} \\
\begin{array}{cccccccccccc}
\text{a} & \text{b} & \text{t} & \text{g} & \text{t} & \text{a} & \text{a} & \text{t} & \text{t} & \text{a} & \text{g} & \text{t} \\
\hline
\text{u} & \text{v} \\
\text{a} & \text{b} & \text{t} & \text{g} & \text{t} & \text{a} & \text{g} & \text{t} & \text{t} & \text{a} & \text{g} & \text{t} \\
\hline
\text{u} & \text{v}
\end{array}
\end{align*}
\]
Definition

A string $x$ of length $n$ is called $k$-strongly-closed if and only if $n \leq 1$ or the following properties are satisfied:

1. There exists some border $b$ of $x$.

2. There exists no factor $w$ of $x$ of length $|w| = |b|$ such that $\delta_H(b, w) \leq k$, except the prefix and suffix of $x$.

We call $b$ the $k$-strongly-closed border of $x$. In the case where $n \leq 1$, we assign $\varepsilon$ as the $k$-strongly-closed border.
**k-Strongly-Closed Strings: Example \((k = 1)\)**

**Border Condition:** Exact

**No Internal occurrence Condition:** Approximate

\[a b t g t a t b a a b t\]

\[\text{b} \quad \text{b}\]

\[a b t g t a t t t a a b t\]

\[\text{b} \quad \text{b}\]
k-Pseudo-Closed Strings: Definition

Definition

A string $x$ of length $n$ is called $k$-pseudo-closed if and only if $n \leq 1$ or the following properties are satisfied:

1. There exists some proper prefix $u$ of $x$ and some proper suffix $v$ of $x$ of length $|u| = |v|$, such that $\delta_H(u, v) \leq k$.

2. Except for $u$ and $v$, there exists no factor $w$ of $x$ of length $|w| = |u| = |v|$ such that $\delta_H(u, w) \leq k$ or $\delta_H(v, w) \leq k$.

We call such a pair $u$ and $v$ the $k$-pseudo-closed border of $x$. In the case where $n \leq 1$, we assign $\varepsilon$ as the $k$-pseudo-closed border.
**k-Pseudo-Closed Strings: Example** $(k = 1)$

**Border Condition:** Approximate

**No Internal occurrence Condition:** Approximate

$k$-Pseudo-Closed

```
  a  b  t  c  t  t  a  c  c  t  a  g  t
```

```
  u          v
```

Non-$k$-Pseudo-Closed

```
  a  b  t  c  t  t  a  b  c  t  a  g  t
```

```
  u          v
```
Finally, we define what we mean by a $k$-closed string:

**Definition**

A string $x$ of length $n$ is called $k$-closed if and only if $n \leq 1$ or $x$ is $k'$-pseudo-closed for some $0 \leq k' \leq k$:

The smallest $k'$ satisfying these conditions, has an associated $k'$-pseudo-closed border consisting of the pair $u$ and $v$. We call this pair the $k$-closed border of $x$. In the case where $n \leq 1$, we assign $\varepsilon$ as the $k$-pseudo-closed border.
Algorithm
Problem Statement

Problem

*Input:* A string $x$ of length $n$ and a natural number $k$, $0 < k < n$

*Output:* The $k$-closed border of $x$ or -1 if $x$ is not $k$-closed
Longest Prefix Match (LPM) and Longest Suffix Match (LSM)

$LPM_k(x)[j]$ is defined as the length of the longest factor of $x$ starting at index $j$, which matches the prefix of $x$ of the same length within $k$ errors.

$LSM_k(x)[j]$ is defined as the length of the longest factor of $x$ ending at index $j$, which matches the suffix of $x$ of the same length within $k$ errors.

Example for $k = 2$
The Longest Common Extension (LCE) of a string $X$ is defined as the length of the longest factor of $X$ starting at both $i$ and $j$, i.e. the longest $L$ such that $X[i..i+L-1] = X[j..j+L-1]$.

If no valid $L$ exists, the LCE equals 0.

<table>
<thead>
<tr>
<th>$j$</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
</tr>
</thead>
<tbody>
<tr>
<td>$x[j]$</td>
<td>b</td>
<td>b</td>
<td>a</td>
<td>a</td>
<td>b</td>
<td>a</td>
<td>a</td>
<td>b</td>
<td>a</td>
<td>b</td>
<td>a</td>
<td>b</td>
<td>b</td>
<td>b</td>
<td>a</td>
</tr>
</tbody>
</table>

$LCE(3,8) = 3$
Recursively Generating LPM and LSM

We may compute the $LPM_{k'+1}$ and $LSM_{k'+1}$ arrays from the $LPM_{k'}$ and $LSM_{k'}$ arrays, such that the arrays are progressively constructed:

\[
LPM_{k'+1}(x)[j] = p + 1 + LCE(p + 1, j + p + 1) \text{ of } x
\]

\[
LSM_{k'+1}(x)[j] = s + 1 + LCE(s + 1, n - j + s) \text{ of } x^R
\]

where $p = LPM_{k'}(x)[j]$ and $s = LSM_{k'}(x)[n - 1 - j]$.

One iteration of the recursive formula requires $O(1)$ time for a single index (via standard operations on suffix trees) and thus $O(n)$ time for the whole array. Therefore, determining $LPM_{k'}$ and $LSM_{k'}$ for all $0 \leq k' \leq k$ requires $O(kn)$ time.
Once the $k$ LPM’s and LSM’s are known we can determine if $x$ is $k$-closed. This is done by finding some $j$ and $k'$ with $1 \leq j \leq n - 1$ and $0 \leq k' \leq k$ such that all the following 3 conditions are satisfied:

1. $j + \text{LPM}_{k'}(x)[j] = n$
2. $\forall i < j, \text{LPM}_{k'}(x)[i] < \text{LPM}_{k'}(x)[j]$
3. $\forall i > n - 1 - j, \text{LSM}_{k'}(x)[i] < \text{LSM}_{k'}(x)[n - 1 - j]$.

The length of the $k$-closed border is then $n - j$ for the smallest $k'$ for which there exists a $j$ satisfying the conditions.
Complete Example \((k = 2)\)

\[
\begin{array}{cccccccccccccc}
  j \\
 x[j] & 0 & 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 & 11 & 12 & 13 & 14 \\
  a & b & b & a & b & a & b & a & b & a & a & b & a & b & b \\
 LPM_2[j] & -1 & 3 & 4 & 7 & 2 & 10 & 4 & 4 & 7 & 2 & 5 & 4 & 3 & 2 & 1 \\
 LSM_2[j] & 1 & 2 & 3 & 4 & 5 & 2 & 7 & 6 & 2 & 10 & 2 & 5 & 7 & 2 & -1 \\
 Cond 1. & F & F & F & F & F & T & F & F & T & F & T & T & T & T \\
 Cond 2. & T & T & T & T & F & T & F & F & F & F & F & F & F & F \\
 Cond 3. & T & T & T & F & F & T & F & F & F & F & F & F & F & F \\
\end{array}
\]
Complete Example \((k = 2)\)

<table>
<thead>
<tr>
<th>(j)</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
</tr>
</thead>
<tbody>
<tr>
<td>(x[j])</td>
<td>a</td>
<td>b</td>
<td>b</td>
<td>a</td>
<td>b</td>
<td>a</td>
<td>b</td>
<td>a</td>
<td>b</td>
<td>a</td>
<td>b</td>
<td>a</td>
<td>b</td>
<td>a</td>
<td>b</td>
</tr>
<tr>
<td>LPM(_2[j])</td>
<td>-1</td>
<td>3</td>
<td>4</td>
<td>7</td>
<td>2</td>
<td>10</td>
<td>4</td>
<td>4</td>
<td>7</td>
<td>2</td>
<td>5</td>
<td>4</td>
<td>3</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>LSM(_2[j])</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>2</td>
<td>7</td>
<td>6</td>
<td>2</td>
<td>10</td>
<td>2</td>
<td>5</td>
<td>7</td>
<td>2</td>
<td>-1</td>
</tr>
</tbody>
</table>

Cond 1. F F F F F F T F F F T F T T T T T
Cond 2. T T T T F T F F F F F F F F F F F F
Cond 3. T T T F F T F F F F F F F F F F F F

2-Closed Border

\(\uparrow\)
Complete Example \((k = 2)\)

\[
\begin{array}{cccccccccccccccc}
\ \ \ \ \ j & 0 & 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 & 11 & 12 & 13 & 14 \\
\hline
x[j] & a & b & b & a & b & a & a & b & a & b & a & a & b & a & b \\
\hline
LPM_2[j] & -1 & 3 & 4 & 7 & 2 & 10 & 4 & 4 & 7 & 2 & 5 & 4 & 3 & 2 & 1 \\
LSM_2[j] & 1 & 2 & 3 & 4 & 5 & 2 & 7 & 6 & 2 & 10 & 2 & 5 & 7 & 2 & -1 \\
\hline
\text{Cond 1.} & F & F & F & F & F & F & T & F & F & F & T & T & T & T & T \\
\text{Cond 2.} & T & T & T & T & F & T & F & F & F & F & F & F & F & F & F \\
\text{Cond 3.} & T & T & T & F & F & T & F & F & F & F & F & F & F & F & F \\
\hline
\end{array}
\]
## Complete Example \((k=2)\)

<table>
<thead>
<tr>
<th>(j)</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
</tr>
</thead>
<tbody>
<tr>
<td>(x[j])</td>
<td>a</td>
<td>b</td>
<td>b</td>
<td>b</td>
<td>a</td>
<td>b</td>
<td>a</td>
<td>a</td>
<td>b</td>
<td>a</td>
<td>a</td>
<td>a</td>
<td>b</td>
<td>a</td>
<td>b</td>
</tr>
<tr>
<td>(\text{LPM}_2[j])</td>
<td>-1</td>
<td>3</td>
<td>4</td>
<td>7</td>
<td>2</td>
<td>10</td>
<td>4</td>
<td>4</td>
<td>7</td>
<td>2</td>
<td>5</td>
<td>4</td>
<td>3</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>(\text{LSM}_2[j])</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>2</td>
<td>7</td>
<td>6</td>
<td>2</td>
<td>10</td>
<td>2</td>
<td>5</td>
<td>7</td>
<td>2</td>
<td>-1</td>
</tr>
</tbody>
</table>

### Conditions

1. Cond 1. F F F F F F T F F T F T T T T T
2. Cond 2. T T T T T F T F F F F F F F F F
3. Cond 3. T T T T F F T F F F F F F F F F

### 2-Closed Border

\[\text{F F F F F F T F F F F F F F F F}\]
Complete Example ($k = 2$)

<table>
<thead>
<tr>
<th>$j$</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
</tr>
</thead>
<tbody>
<tr>
<td>$x[j]$</td>
<td>a</td>
<td>b</td>
<td>b</td>
<td>a</td>
<td>b</td>
<td>a</td>
<td>b</td>
<td>a</td>
<td>b</td>
<td>a</td>
<td>b</td>
<td>a</td>
<td>b</td>
<td>a</td>
<td>b</td>
</tr>
<tr>
<td>LPM$_2[j]$</td>
<td>-1</td>
<td>3</td>
<td>4</td>
<td>7</td>
<td>2</td>
<td>10</td>
<td>4</td>
<td>4</td>
<td>7</td>
<td>2</td>
<td>5</td>
<td>4</td>
<td>3</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>LSM$_2[j]$</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>2</td>
<td>7</td>
<td>6</td>
<td>2</td>
<td>10</td>
<td>2</td>
<td>5</td>
<td>7</td>
<td>2</td>
<td>-1</td>
</tr>
</tbody>
</table>

Cond 1. F F F F F F T F F T F T T T T
Cond 2. T T T T T T F T F F F F F F F
Cond 3. T T T F F T F F F F F F F F

2-Closed Border F F F F F T F F F F F F F F

▲
### Complete Example \((k = 2)\)

\[
\begin{array}{cccccccccccccccc}
\begin{array}{c}
\text{j} \\
x[j]
\end{array} & \begin{array}{cccccccccccccccc}
0 & 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 & 11 & 12 & 13 & 14 \\
\text{a} & \text{b} & \text{b} & \text{b} & \text{a} & \text{b} & \text{a} & \text{b} & \text{a} & \text{b} & \text{a} & \text{b} & \text{a} & \text{b} & \text{b} \\
\text{LPM}_2[j] & -1 & 3 & 4 & 7 & \textcircled{2} & 10 & 4 & 4 & 7 & 2 & 5 & 4 & 3 & 2 & 1 \\
\text{LSM}_2[j] & 1 & 2 & 3 & 4 & 5 & 2 & 7 & 6 & 2 & 10 & 2 & 5 & 7 & 2 & -1 \\
\text{Cond 1.} & \text{F} & \text{F} & \text{F} & \text{F} & \text{F} & \text{F} & \text{T} & \text{F} & \text{F} & \text{F} & \text{T} & \text{T} & \text{T} & \text{T} & \text{T} \\
\text{Cond 2.} & \text{T} & \text{T} & \text{T} & \text{T} & \text{T} & \text{F} & \text{T} & \text{F} & \text{F} & \text{F} & \text{F} & \text{F} & \text{F} & \text{F} & \text{F} \\
\text{Cond 3.} & \text{T} & \text{T} & \text{T} & \text{T} & \text{F} & \text{T} & \text{F} & \text{F} & \text{F} & \text{F} & \text{F} & \text{F} & \text{F} & \text{F} & \text{F} \\
\text{2-Closed Border} & \text{F} & \text{F} & \text{F} & \text{F} & \text{F} & \text{F} & \text{T} & \text{F} & \text{F} & \text{F} & \text{F} & \text{F} & \text{F} & \text{F} & \text{F} \\
\end{array}
\end{array}
\]
### Complete Example \((k = 2)\)

<table>
<thead>
<tr>
<th>(j)</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
</tr>
</thead>
<tbody>
<tr>
<td>(x[j])</td>
<td>a</td>
<td>b</td>
<td>b</td>
<td>a</td>
<td>b</td>
<td>a</td>
<td>b</td>
<td>a</td>
<td>b</td>
<td>a</td>
<td>b</td>
<td>a</td>
<td>b</td>
<td>a</td>
<td>b</td>
</tr>
<tr>
<td>(LPM_2[j])</td>
<td>-1</td>
<td>3</td>
<td>4</td>
<td>7</td>
<td>2</td>
<td>10</td>
<td>4</td>
<td>4</td>
<td>7</td>
<td>2</td>
<td>5</td>
<td>4</td>
<td>3</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>(LSM_2[j])</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>2</td>
<td>7</td>
<td>6</td>
<td>2</td>
<td>10</td>
<td>2</td>
<td>5</td>
<td>7</td>
<td>2</td>
<td>-1</td>
</tr>
</tbody>
</table>

**Cond 1.**

\[
\begin{array}{cccccccccccc}
& F & F & F & F & F & F & T & F & F & T & T & T & T \\
F & F & F & F & F & F & T & F & F & T & T & T & T & T \\
F & F & F & F & F & F & T & F & F & T & T & T & T & T \\
\end{array}
\]

**Cond 2.**

\[
\begin{array}{cccccccccccc}
& T & T & T & T & T & F & T & F & F & F & F & F & F & F \\
T & T & T & T & T & F & T & F & F & F & F & F & F & F \\
T & T & T & T & F & F & T & F & F & F & F & F & F & F \\
\end{array}
\]

**Cond 3.**

\[
\begin{array}{cccccccccccc}
& T & T & T & F & F & T & F & F & F & F & F & F & F & F \\
T & T & T & F & F & T & F & F & F & F & F & F & F & F \\
T & T & T & F & F & T & F & F & F & F & F & F & F & F \\
\end{array}
\]

**2-Closed Border**

\[
\begin{array}{cccccccccccc}
F & F & F & F & F & T & F & F & F & F & F & F & F & F \\
\end{array}
\]
### Complete Example ($k = 2$)

<table>
<thead>
<tr>
<th>$j$</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
</tr>
</thead>
<tbody>
<tr>
<td>$x[j]$</td>
<td>a</td>
<td>b</td>
<td>b</td>
<td>a</td>
<td>b</td>
<td>a</td>
<td>b</td>
<td>a</td>
<td>b</td>
<td>a</td>
<td>b</td>
<td>a</td>
<td>b</td>
<td>a</td>
<td>b</td>
</tr>
<tr>
<td>$LPM_2[j]$</td>
<td>-1</td>
<td>3</td>
<td>4</td>
<td>7</td>
<td>2</td>
<td>10</td>
<td>4</td>
<td>4</td>
<td>7</td>
<td>2</td>
<td>5</td>
<td>4</td>
<td>3</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>$LSM_2[j]$</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>2</td>
<td>7</td>
<td>6</td>
<td>2</td>
<td>10</td>
<td>2</td>
<td>5</td>
<td>7</td>
<td>2</td>
<td>-1</td>
</tr>
</tbody>
</table>

- **Cond 1.**
  - $F$ $F$ $F$ $F$ $F$ $F$ $T$ $F$ $F$ $F$ $T$ $T$ $T$ $T$ $T$

- **Cond 2.**
  - $T$ $T$ $T$ $T$ $T$ $F$ $T$ $F$ $F$ $F$ $F$ $F$ $F$ $F$ $F$

- **Cond 3.**
  - $T$ $T$ $T$ $F$ $F$ $T$ $F$ $F$ $F$ $F$ $F$ $F$ $F$ $F$ $F$

- **2-Closed Border**
  - $F$ $F$ $F$ $F$ $F$ $T$ $F$ $F$ $F$ $F$ $F$ $F$ $F$ $F$ $F$
**Complete Example** ($k = 2$)

<table>
<thead>
<tr>
<th>$j$</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
</tr>
</thead>
<tbody>
<tr>
<td>$x[j]$</td>
<td>a</td>
<td>b</td>
<td>b</td>
<td>a</td>
<td>b</td>
<td>a</td>
<td>b</td>
<td>a</td>
<td>b</td>
<td>a</td>
<td>b</td>
<td>a</td>
<td>b</td>
<td>a</td>
<td>b</td>
</tr>
<tr>
<td><strong>LPM$_2[j]$</strong></td>
<td>-1</td>
<td>3</td>
<td>4</td>
<td>7</td>
<td>2</td>
<td>10</td>
<td>4</td>
<td>4</td>
<td>7</td>
<td>2</td>
<td>5</td>
<td>4</td>
<td>3</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td><strong>LSM$_2[j]$</strong></td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>2</td>
<td>7</td>
<td>6</td>
<td>2</td>
<td>10</td>
<td>2</td>
<td>5</td>
<td>7</td>
<td>2</td>
<td>-1</td>
</tr>
</tbody>
</table>

**Cond 1.**

| F | F | F | F | F | F | T | F | F | T | F | T | T | T | T |

**Cond 2.**

| T | T | T | T | T | F | T | T | F | F | F | F | F | F | F |

**Cond 3.**

| T | T | T | F | F | T | F | F | F | F | F | F | F | F | F |

**2-Closed Border**

| F | F | F | F | F | F | T | F | F | F | F | F | F | F | F |
**Complete Example \((k = 2)\)**

<table>
<thead>
<tr>
<th>(j)</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
</tr>
</thead>
<tbody>
<tr>
<td>(x[j])</td>
<td>a</td>
<td>b</td>
<td>b</td>
<td>a</td>
<td>b</td>
<td>a</td>
<td>b</td>
<td>a</td>
<td>b</td>
<td>a</td>
<td>a</td>
<td>b</td>
<td>a</td>
<td>b</td>
<td>a</td>
</tr>
<tr>
<td>LPM(_2)[j]</td>
<td>-1</td>
<td>3</td>
<td>4</td>
<td>7</td>
<td>2</td>
<td>10</td>
<td>4</td>
<td>4</td>
<td>7</td>
<td>2</td>
<td>5</td>
<td>4</td>
<td>3</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>LSM(_2)[j]</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>2</td>
<td>7</td>
<td>6</td>
<td>2</td>
<td>10</td>
<td>2</td>
<td>5</td>
<td>7</td>
<td>2</td>
<td>-1</td>
</tr>
</tbody>
</table>

**Cond 1.**
- F F F F F F T F F T F T T T T T T

**Cond 2.**
- T T T T T F T F F F F F F F F F F

**Cond 3.**
- T T T F F T T F F F F F F F F F F

**2-Closed Border**
- F F F F F F T F F F F F F F F F F

▲
## Complete Example ($k = 2$)

<table>
<thead>
<tr>
<th>$j$</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
</tr>
</thead>
<tbody>
<tr>
<td>$x[j]$</td>
<td>a</td>
<td>b</td>
<td>b</td>
<td>a</td>
<td>b</td>
<td>a</td>
<td>b</td>
<td>a</td>
<td>b</td>
<td>a</td>
<td>b</td>
<td>a</td>
<td>b</td>
<td>a</td>
<td>b</td>
</tr>
<tr>
<td>LPM$_2[j]$</td>
<td>-1</td>
<td>3</td>
<td>4</td>
<td>7</td>
<td>2</td>
<td>10</td>
<td>4</td>
<td>4</td>
<td>7</td>
<td>2</td>
<td>5</td>
<td>4</td>
<td>3</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>LSM$_2[j]$</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>2</td>
<td>7</td>
<td>6</td>
<td>2</td>
<td>10</td>
<td>2</td>
<td>5</td>
<td>7</td>
<td>2</td>
<td>-1</td>
</tr>
</tbody>
</table>

- **Cond 1.**
  - $F$ $F$ $F$ $F$ $F$ $F$ $T$ $F$ $F$ $T$ $F$ $T$ $T$ $T$ $T$

- **Cond 2.**
  - $T$ $T$ $T$ $T$ $T$ $F$ $T$ $F$ $F$ $F$ $F$ $F$ $F$ $F$ $F$

- **Cond 3.**
  - $T$ $T$ $T$ $F$ $F$ $T$ $F$ $F$ $F$ $F$ $F$ $F$ $F$ $F$ $F$

- **2-Closed Border**
  - $F$ $F$ $F$ $F$ $F$ $F$ $T$ $F$ $F$ $F$ $F$ $F$ $F$ $F$ $F$ $F$ $F$ $F$
### Complete Example ($k = 2$)

<table>
<thead>
<tr>
<th>$j$</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
</tr>
</thead>
<tbody>
<tr>
<td>$x[j]$</td>
<td>a</td>
<td>b</td>
<td>b</td>
<td>a</td>
<td>b</td>
<td>a</td>
<td>b</td>
<td>a</td>
<td>b</td>
<td>a</td>
<td>a</td>
<td>b</td>
<td>a</td>
<td>b</td>
<td>b</td>
</tr>
<tr>
<td>LPM$_2[j]$</td>
<td>-1</td>
<td>3</td>
<td>4</td>
<td>7</td>
<td>2</td>
<td>10</td>
<td>4</td>
<td>4</td>
<td>7</td>
<td>2</td>
<td>5</td>
<td>4</td>
<td>3</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>LSM$_2[j]$</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>2</td>
<td>7</td>
<td>6</td>
<td>2</td>
<td>10</td>
<td>2</td>
<td>5</td>
<td>7</td>
<td>2</td>
<td>-1</td>
</tr>
<tr>
<td>Cond 1.</td>
<td>F</td>
<td>F</td>
<td>F</td>
<td>F</td>
<td>F</td>
<td>F</td>
<td>T</td>
<td>F</td>
<td>F</td>
<td>F</td>
<td>T</td>
<td>T</td>
<td>T</td>
<td>T</td>
<td>T</td>
</tr>
<tr>
<td>Cond 2.</td>
<td>T</td>
<td>T</td>
<td>T</td>
<td>T</td>
<td>T</td>
<td>F</td>
<td>T</td>
<td>F</td>
<td>F</td>
<td>F</td>
<td>F</td>
<td>F</td>
<td>F</td>
<td>F</td>
<td>F</td>
</tr>
<tr>
<td>Cond 3.</td>
<td>T</td>
<td>T</td>
<td>T</td>
<td>F</td>
<td>F</td>
<td>T</td>
<td>F</td>
<td>F</td>
<td>F</td>
<td>F</td>
<td>F</td>
<td>F</td>
<td>F</td>
<td>F</td>
<td>F</td>
</tr>
</tbody>
</table>
| 2-Closed Border | F | F | F | F | F | T | F | F | F | F | F | F | F | F | F | ▲
### Complete Example \((k = 2)\)

<table>
<thead>
<tr>
<th>(j)</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
</tr>
</thead>
<tbody>
<tr>
<td>(x[j])</td>
<td>a</td>
<td>b</td>
<td>b</td>
<td>a</td>
<td>b</td>
<td>a</td>
<td>b</td>
<td>a</td>
<td>b</td>
<td>a</td>
<td>b</td>
<td>a</td>
<td>b</td>
<td>a</td>
<td>b</td>
</tr>
<tr>
<td>LPM(_2[j])</td>
<td>-1</td>
<td>3</td>
<td>4</td>
<td>7</td>
<td>2</td>
<td>10</td>
<td>4</td>
<td>4</td>
<td>7</td>
<td>2</td>
<td>5</td>
<td>(\textcircled{4})</td>
<td>3</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>LSM(_2[j])</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>2</td>
<td>7</td>
<td>6</td>
<td>2</td>
<td>10</td>
<td>2</td>
<td>5</td>
<td>7</td>
<td>2</td>
<td>-1</td>
</tr>
</tbody>
</table>

Cond 1.

\[
\begin{array}{cccccccccccccc}
  & F & F & F & F & F & T & F & F & T & T & T & T & T & T & T \\
\end{array}
\]

Cond 2.

\[
\begin{array}{cccccccccccccc}
  & T & T & T & T & F & T & F & F & F & F & F & F & F & F & F \\
\end{array}
\]

Cond 3.

\[
\begin{array}{cccccccccccccc}
  & T & T & T & F & F & T & F & F & F & F & F & F & F & F & F \\
\end{array}
\]

2-Closed Border

\[
\begin{array}{cccccccccccccc}
\end{array}
\]
**Complete Example** \((k = 2)\)

<table>
<thead>
<tr>
<th>(j)</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
</tr>
</thead>
<tbody>
<tr>
<td>(x[j])</td>
<td>a</td>
<td>b</td>
<td>b</td>
<td>a</td>
<td>b</td>
<td>a</td>
<td>b</td>
<td>a</td>
<td>b</td>
<td>a</td>
<td>b</td>
<td>a</td>
<td>b</td>
<td>a</td>
<td>b</td>
</tr>
<tr>
<td>LPM(_2)[(j)]</td>
<td>-1</td>
<td>3</td>
<td>4</td>
<td>7</td>
<td>2</td>
<td>10</td>
<td>4</td>
<td>4</td>
<td>7</td>
<td>2</td>
<td>5</td>
<td>4</td>
<td>3</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>LSM(_2)[(j)]</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>2</td>
<td>7</td>
<td>6</td>
<td>2</td>
<td>10</td>
<td>2</td>
<td>5</td>
<td>7</td>
<td>2</td>
<td>-1</td>
</tr>
</tbody>
</table>

Cond 1. F F F F F F T F F T F T T T T T
Cond 2. T T T T T F T F F F F T F F F F
Cond 3. T T T F F T F F F F F F F F F F

2-Closed Border

| F | F | F | F | F | F | T | F | F | F | F | F | F | F | F |

▲
### Complete Example \((k = 2)\)

<table>
<thead>
<tr>
<th>(x[j])</th>
<th>a</th>
<th>b</th>
<th>b</th>
<th>a</th>
<th>b</th>
<th>a</th>
<th>b</th>
<th>a</th>
<th>b</th>
<th>a</th>
<th>a</th>
<th>b</th>
<th>a</th>
<th>b</th>
</tr>
</thead>
<tbody>
<tr>
<td>(j)</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
<td>8</td>
<td>9</td>
<td>10</td>
<td>11</td>
<td>12</td>
<td>13</td>
</tr>
</tbody>
</table>

\[
\begin{align*}
LPM_2[j] & = -1 \ 3 \ 4 \ 7 \ 2 \ 10 \ 4 \ 4 \ 7 \ 2 \ 5 \ 4 \ 3 \ 2 \ 1 \\
LSM_2[j] & = 1 \ 2 \ 3 \ 4 \ 5 \ 2 \ 7 \ 6 \ 2 \ 10 \ 2 \ 5 \ 7 \ 2 \ -1 \\
\end{align*}
\]

Cond 1.  \(F\ F\ F\ F\ F\ T\ F\ F\ T\ F\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T\ T|
Complete Example \((k = 2)\)

<table>
<thead>
<tr>
<th>(j)</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
</tr>
</thead>
<tbody>
<tr>
<td>(x[j])</td>
<td>a</td>
<td>b</td>
<td>b</td>
<td>a</td>
<td>b</td>
<td>a</td>
<td>b</td>
<td>a</td>
<td>b</td>
<td>a</td>
<td>b</td>
<td>a</td>
<td>b</td>
<td>a</td>
<td>b</td>
</tr>
<tr>
<td>LPM(_2[j])</td>
<td>-1</td>
<td>3</td>
<td>4</td>
<td>7</td>
<td>2</td>
<td>10</td>
<td>4</td>
<td>4</td>
<td>7</td>
<td>2</td>
<td>5</td>
<td>4</td>
<td>3</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>LSM(_2[j])</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>2</td>
<td>7</td>
<td>6</td>
<td>2</td>
<td>10</td>
<td>2</td>
<td>5</td>
<td>7</td>
<td>2</td>
<td>-1</td>
</tr>
<tr>
<td>Cond 1.</td>
<td>F</td>
<td>F</td>
<td>F</td>
<td>F</td>
<td>F</td>
<td>F</td>
<td>T</td>
<td>F</td>
<td>F</td>
<td>T</td>
<td>F</td>
<td>T</td>
<td>T</td>
<td>T</td>
<td>T</td>
</tr>
<tr>
<td>Cond 2.</td>
<td>T</td>
<td>T</td>
<td>T</td>
<td>T</td>
<td>T</td>
<td>F</td>
<td>T</td>
<td>F</td>
<td>F</td>
<td>F</td>
<td>F</td>
<td>F</td>
<td>F</td>
<td>F</td>
<td>F</td>
</tr>
<tr>
<td>Cond 3.</td>
<td>T</td>
<td>T</td>
<td>T</td>
<td>F</td>
<td>F</td>
<td>T</td>
<td>F</td>
<td>F</td>
<td>F</td>
<td>F</td>
<td>F</td>
<td>F</td>
<td>F</td>
<td>F</td>
<td>F</td>
</tr>
</tbody>
</table>

2-Closed Border: \[ F \ F \ F \ F \ F \ F \ T \ F \ F \ F \ F \ F \ F \ F \ F \ F \]
Complexity Analysis

1. Preprocess $x$ (via a suffix tree) to allow for constant time LCE queries. 
   $O(n)$ time and $O(n)$ space.

2. Recursively generate $LPM_{k'}$ and $LSM_{k'}$ for $0 \leq k' \leq k$. 
   $k$ steps each requiring $O(n)$ time. Total of $O(n)$ space.

3. During each of the $k$ steps, determine the "peaks" of the 
   $LPM$ and $LSM$ arrays, then verify if the 3 conditions are 
   satisfied for some $j$ where $1 \leq j \leq n - 1$. 
   Requires additional $O(n)$ time for each of the $k$ steps.
Summary
Summary

- We have generalised closed strings to $k$-closed strings.
- We have an algorithm that identifies whether a string $x$ is $k$-closed, and determines the $k$-closed border, in $O(kn)$ time and $O(n)$ space.
- **Further Work:** Improvement in the construction of the LPM and LSM arrays, currently requiring $O(kn)$ time. Decreasing this time complexity appears to be a reasonable, however non-trivial, goal for any future work on this problem.
Appendix
Gabriele Fici

**A Classification of Trapezoidal Words**


Golnaz Badkobeh and Gabriele Fici and Zsuzsanna Lipták

**A Note on Words With the Smallest Number of Closed Factors**

*CoRR, 1305.6395, 2013.*

Golnaz Badkobeh and Hideo Bannai and Keisuke Goto and Tomohiro I and Costas S. Iliopoulos and Shunsuke Inenaga and Simon J. Puglisi and Shiho Sugimoto

**Closed factorization**

Thank you for listening 😊