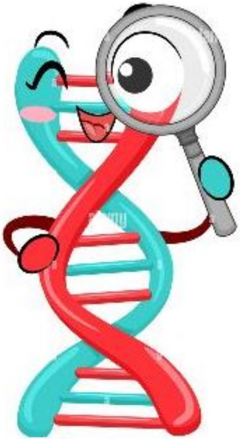
Lyndon Partial Arrays

Meenakshi Paramasivan

04 Oct 2023

Partial Words in DNA Sequences



Partial words are considered for finding good encodings for DNA computations.

Preliminaries

- $\Sigma \neq \emptyset$.
- Σ - a finite set of symbols (or letters) termed as alphabet.
- A total word (or word or (string)) is a sequence of letters over Σ.
- λ - the empty word.
- $|w|$ - the *length* of a word w ; $|\lambda| = 0$.
- Σ^* - the set of finite total words (or words) from Σ including the empty word λ.
- $\Sigma^+ = \Sigma^* \setminus \{\lambda\}$.

Primitive (Non-Periodic) Word

A word x is *primitive (non-periodic)* if there exists no word y such that $x = y^i$ with $i \geq 2$.

Note

- The empty word is NOT a primitive (non-periodic) word.

Preliminaries

- $\Diamond \notin \Sigma$ - a "hole" or a "wild card" symbol.
- $\Sigma_\Diamond = \Sigma \cup \{\Diamond\}$ - a partial alphabet.
- $\Sigma_\Diamond^*$ - the set of partial words from $\Sigma_\Diamond$ including the empty word λ.
- $\Sigma_\Diamond^+ = \Sigma_\Diamond^* \setminus \{\lambda\}$.
- $L_\Diamond \subseteq \Sigma_\Diamond^*$ - the set of all partial words over $\Sigma_\Diamond$ - (a partial language).

Partial Word (Unbordered)

A partial word $x_\Diamond$ is a total word x with $|x| \geq 2$ that has a number of "holes".
A partial word $x_\Diamond$ is unbordered if no non-empty words p, q, y exist such that $x_\Diamond \subset py$ and $x_\Diamond \subset qp$

Note

- A total word is a partial word with zero number of holes.
- The hole of any length is neither a total word nor a partial word.

Roger Conant Lyndon (1917 – 1988)

Lyndon Word

A finite *Lyndon word l* is a primitive word that is alphabetically (lexicographically) least among its rotations.

- M.J. Fischer and M.S. Paterson - Do not care symbols; string matching (1974)
- J. Berstel and L. Boasson - Partial Words (1999)
- F. Blanchet-Sadri - Algorithmic Combinatorics on Partial Words (2008)

Lyndon Words vs Lyndon Partial Words

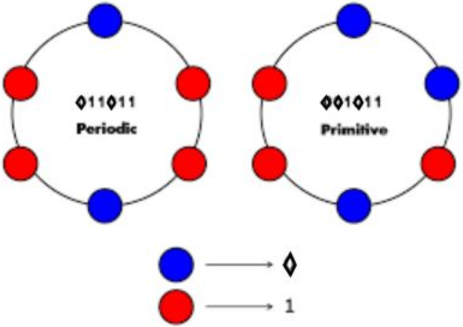
	Lyndon Words	*Lyndon Partial Words*
Ordered Alphabet	$\Sigma_k = \{a_1 < a_2 < \ldots < a_k\}, k > 1$	$\Sigma_\Diamond = \Sigma_k \bigcup \{\Diamond\}, k > 1$
Primitivity	*Conjugacy class of any primitive word contains a Lyndon word*	*Conjugacy class of any primitive partial word may not contains a Lyndon partial word*

Periodic Partial Words vs Primitive Partial Words

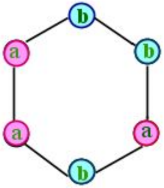
Example

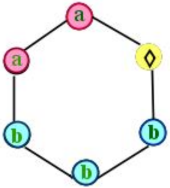
$(\Diamond 11)^2$

$\Diamond\Diamond 1 \Diamond 11, \Diamond 1 \Diamond 11 \Diamond, 1 \Diamond 11 \Diamond\Diamond, \Diamond 11 \Diamond\Diamond 1, 11 \Diamond\Diamond 1 \Diamond, 1 \Diamond\Diamond 1 \Diamond 1$



An Example





Lyndon Word

$\{bbabaa, babaab, abaabb, baabba, aabbab, abbaba\}$

Lyndon Partial Word

$\{a\Diamond bbba, \Diamond bbbaa, bbbaa\Diamond, bbaa\Diamond b, baa\Diamond bb, aa\Diamond bbb\}$

$\ell_\Diamond$ - Lyndon partial language

The set of all Lyndon partial words over $\Sigma_\Diamond$ with $\Sigma = \{a, b\}$

Length	*Lyndon partial words*
1	—
2	$a\Diamond$, $\Diamond b$
3	$aa\Diamond$, $a\Diamond b$, $\Diamond bb$
4	$aaa\Diamond$, $aa\Diamond b$, $a\Diamond ab$, $a\Diamond bb$, $ab\Diamond b$, $\Diamond bbb$
5	$aaaa\Diamond$, $aaa\Diamond b$, $aa\Diamond ab$, $aa\Diamond bb$, $aab\Diamond b$, $a\Diamond abb$, $a\Diamond bbb$, $ab\Diamond bb$, $\Diamond bbbb$
$\vdots$	$\ldots$

Results

Theorem 1 [Krishna Kumari et. al (2020)]

No proper subword exists as both prefix and a suffix of a Lyndon partial word.

Sample: $aa\Diamond bbb$

Proper Subwords: $aa\Diamond bb$, $a\Diamond bbb$, $aa\Diamond b$, $a\Diamond bb$, $\Diamond bbb$, $aa\Diamond$, $a\Diamond b$, $\Diamond bb$, bbb, aa, $a\Diamond$, $\Diamond b$, bb, a, b

Results

Theorem 2 [Krishna Kumari et. al (2020)]

A partial word $u_\Diamond$ over $\Sigma_\Diamond^+$ belongs to a set of Lyndon partial words $L_\Diamond$ iff $u_\Diamond < q_\Diamond$ for each proper suffix $q_\Diamond$ of $u_\Diamond$.

Sample: $u_\Diamond = aa\Diamond bb < b = q_\Diamond$

Proper Subwords: $a, aa, aa\Diamond, aa\Diamond b, a\Diamond, a\Diamond b, a\Diamond bb, \Diamond bb, \Diamond b, bb, b$

Proper Suffix Subwords: $a\Diamond bb, \Diamond bb, bb, b$

Results

Theorem 3 [Krishna Kumari et. al (2020)]

Consider $p_\Diamond \in L_\Diamond$ and $q_\Diamond \in L_\Diamond$. Then $p_\Diamond q_\Diamond \in L_\Diamond$ iff $p_\Diamond < q_\Diamond$.

Sample: $a\Diamond abb$

Consider $p_\Diamond = a\Diamond ab \in L_\Diamond$ and $q_\Diamond = b \in L_\Diamond$. $p_\Diamond q_\Diamond \in L_\Diamond$ iff $p_\Diamond < q_\Diamond$.

Theorem 4

Each Lyndon partial word $l_\Diamond$ over $\Sigma_\Diamond$ is unbordered but the converse is not true.

Sample: $aab\Diamond b$ and $bab\Diamond aa$

Here $aab\Diamond b$ is an unbordered Lyndon partial word but the unbordered partial word $bab\Diamond aa$ is not Lyndon.

Result

Theorem 5 [Krishna Kumari et. al (2020)]

Factorization Theorem Any partial word $u_{\Diamond}$ over the alphabet $\Sigma_{\Diamond}^{+}$ can be uniquely written as $u_{\Diamond} = l_{\Diamond}^{1}.....l_{\Diamond}^{r}$ with $l_{\Diamond}^{1}, ..., l_{\Diamond}^{r} \in L_{\Diamond}$ and $l_{\Diamond}^{1} \geq ...... \geq l_{\Diamond}^{r}$.

Sample: $aa\Diamond b$

Here $aa\Diamond b \in L_{\Diamond}$ with its subwords $\{a, a\Diamond, a\Diamond b, \Diamond b, b\} \in L_{\Diamond}$ and $aa\Diamond b$ can be uniquely written as non increasing product of a and $a\Diamond b$.

Position and Maximal Lyndon Partial Factor

Position	1	2	3	4	5	6	7
Lyndon partial word	a	a	b	a	b	◊	b
Maximal Lyndon partial factor	a	a	b	a	b	◊	b

Lyndon Partial Array

Position	1	2	3	4	5	6	7
Lyndon partial word	a	a	b	a	b	◊	b
Maximal Lyndon partial factor	a	a	b	a	b	◊	b
Lyndon partial array	5	3	3	5	5	7	7

Result

Theorem 6

If the positions i, j in $l_\Diamond[1..n]$ satisfy $1 \leq i < j \leq n$, then the two intervals $\langle i, l_\Diamond^A[i] \rangle$ and $\langle j, l_\Diamond^A[j] \rangle$ are not intersecting each other.

Sample: $aab\Diamond b$

The positions $i = 2$ and $j = 4$ in $l_\Diamond[1 \cdots 5] = aab\Diamond b$ satisfy $1 \leq i < j \leq n$ then $\langle 2, 3 \rangle$ and $\langle 4, 5 \rangle$ are not intersecting each other where $l_\Diamond^A[i] = 3$ and $l_\Diamond^A[j] = 5$.

Result

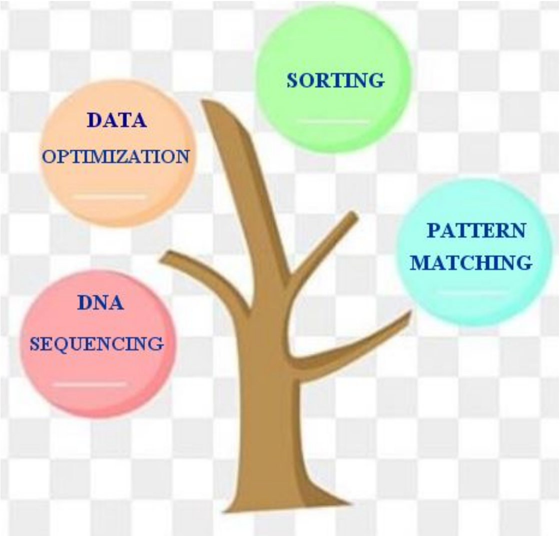
Theorem 7

Consider a Lyndon partial word $l_\Diamond[1..n]$ over the ordered alphabet $\Sigma_\Diamond$. Let $\text{suf}_{l_\Diamond}(i) = l_\Diamond[i..n]$ denote the suffix of $l_\Diamond$ beginning at position i. Then a lyndon partial factor $l_\Diamond[i \cdots j]$ is the maximal Lyndon partial factor of $l_\Diamond$ if and only if for any $1 < j \leq k$ $\text{suf}_{l_\Diamond}(i) < \text{suf}_{l_\Diamond}(k)$ and $\text{suf}_{l_\Diamond}(j+1) < \text{suf}_{l_\Diamond}(i)$.

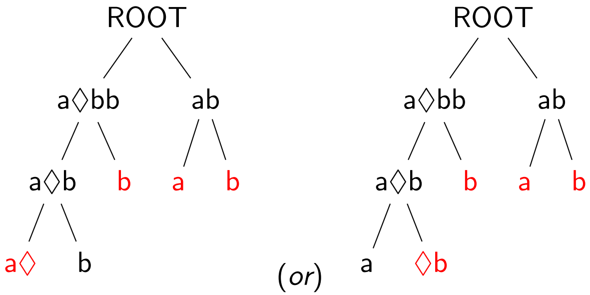
Sample: $aaba\Diamond bb$

Let $\text{suf}_{l_\Diamond}(i) = aba\Diamond bb$ beginning at position $i = 2$. A lyndon partial factor $l_\Diamond[i \cdots j] = ab$ with the position $j = 3$ is the maximal Lyndon partial factor of $l_\Diamond$ since for any position $k = 3$, $\text{suf}_{l_\Diamond}(i) = aba\Diamond bb < ba\Diamond bb = \text{suf}_{l_\Diamond}(k)$ and $\text{suf}_{l_\Diamond}(j+1) = a\Diamond bb < aba\Diamond bb = \text{suf}_{l_\Diamond}(i)$.

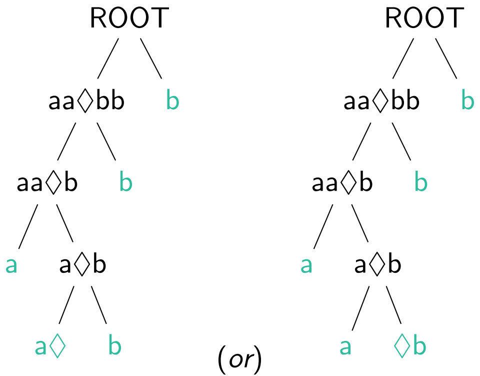
Trees vs Lyndon Partial Words



Tree - $a\Diamond bbab$



Tree - *aa◊bbb*



Tree

Tree associated with a Lyndon partial word

Formally, ζ associated with a Lyndon partial word is described as a tree that is alphabetically less than all of its rotations. $\Im$ denotes the trees set of $L_\Diamond$. A subtree of ζ is a tree with set of nodes as a subset of ζ.

Results

Theorem 8

No proper subtree exists as both initial and terminal of the tree ζ.

Theorem 9

ζ is a tree of a Lyndon partial word iff $\zeta = P + vQ, v \in \delta(P)$ where $\zeta, P, Q \in \Im$ and $P < Q$.

Theorem 10

Any tree ζ over the alphabet $\Sigma_\Diamond^+$ can be uniquely written as
$\zeta = P_0 + v_1 P_1 + v_2 P_2 + \dots . v_k P_k, v_m \in \delta(v_n P_n)$ for some $n \geq m$ such that
$P_0 \geq P_1 \geq P_2 \dots . \geq P_k$.

$\ell_\Diamond$- Morphism

$\ell_\Diamond$- morphism

A morphism that preserves the Lyndon partial words is defined as $\ell_\Diamond$- morphism.

Theorem 11

A non-empty morphism g on $\Sigma_\Diamond^+$ containing atleast two letters is a $\ell_\Diamond$- morphism iff g is an order preserving morphism such that for each $u \in \Sigma_\Diamond$, $g(u)$ is a Lyndon partial word.

Corollary

g is a $\ell_\Diamond$- morphism on $\Sigma_\Diamond = \{a, b\} \cup \{\Diamond\}$ iff $g(a)$ and $g(b)$ are lyndon partial words with $g(a) < g(b)$.

Two-dimensional Lyndon partial words

Two-dimensional Lyndon partial words

A horizontally primitive matrix which is least among its horizontal conjugates.

Sample: $A = \begin{matrix} a & \diamond & a \\ \diamond & b & a \\ b & b & \diamond \end{matrix}$

The horizontal conjugates of A are $\begin{matrix} a & \diamond & a \\ \diamond & b & a \\ b & b & \diamond \end{matrix}$, $\begin{matrix} \diamond & a & a \\ b & \diamond & a \\ b & \diamond & b \end{matrix}$, $\begin{matrix} a & a & \diamond \\ a & \diamond & b \\ \diamond & b & b \end{matrix}$. Here $\begin{matrix} a & a & \diamond \\ a & \diamond & b \\ \diamond & b & b \end{matrix}$ is a two-dimensional row Lyndon partial word.

Connections to image analysis

- Kolam is one of several types of street/home art practices performed as ephemeral designs on thresholds in India.
- In recent years, kolam figures have also attracted the attention of computer scientists interested in analysing and describing images with picture languages.

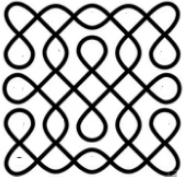
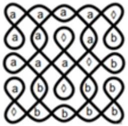
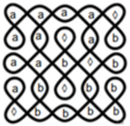
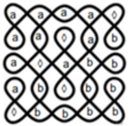


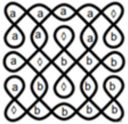
Figure: Kambi Kolam (5×5)

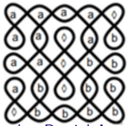
Kolams for 2D row Lyndon partial word



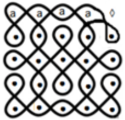


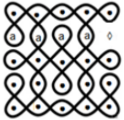


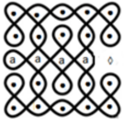


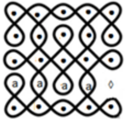


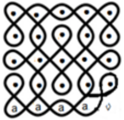
Kolam patterns to show sensitivity of $aaaa\Diamond$



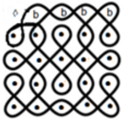
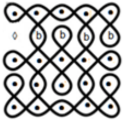
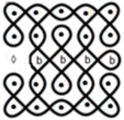
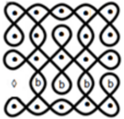
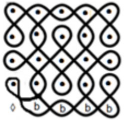








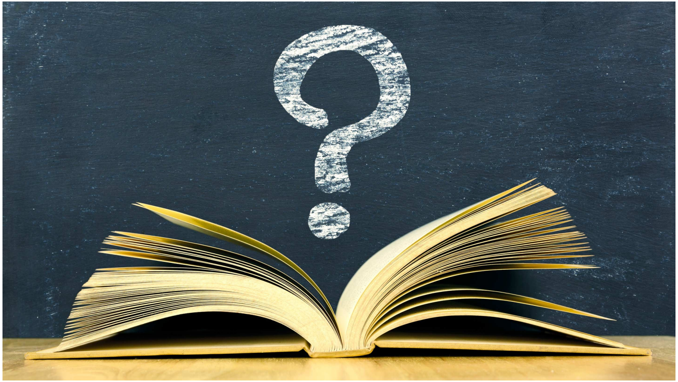
Kolam patterns to show sensitivity of $\lozenge bbbb$

Conclusion and Future Work

- We introduced Lyndon partial arrays and trees associated with Lyndon partial words.
- We learned to find out very basic connections of kolams to two-dimensional formal languages through Lyndon partial words and arrays.
- In near future, we will do a detailed study on several variants of 2D Lyndon partial words and their properties in an upgraded version of this paper.

Discussions



References

- F. Blanchet-Sadri, Algorithmic Combinatorics on Partial Words. Discrete mathematics and its applications, CRC Press, (2008)
- C. Hohlweg and C. Reutenauer, Lyndon words, permutations and trees, Theoret. Comput. Sci., 307 (2003), 173–178.
- F. Franek, A.S.M. Sohidull Islam, M. Sohel Rahman, and W.F. Smyth, Algorithms to compute the Lyndon array. In Proc. of Prague Stringology Conference 2016, (2016) 172–184.
- R.Krishna Kumari, R.Arulprakasam and V.R.Dare, Language of Lyndon Partial Words, International Journal of Mathematics and Computer Science, 15, (2020), 1173–1177.
- G.Richomme, Lyndon Morphisms, Bull.Belg.Math.Soc., 10(2003),761–785.
- R. Siromoney, K.G. Subramanian, V.R. Dare and D.G. Thomas, Some results on picture languages, Pattern Recognition, 32(2), (1999), 295-304.
- K.G.Subramanian, Rani Siromoney and Lisa Mathew, Lyndon Trees, Theoret. Comput. Sci., 106(1992),373–383.

Many Many Thanks!