Properties of Cover and Seed of Partial Words

R. Krishna Kumari, Member, IAENG, L. Jeyanthi, K. Janaki,
R. Arulprakasam, Member, IAENG and P. Madhusoodhanan *

Abstract—**We take into account the problem of identifying the repeated structure in a given ℘word $v_\Diamond$ of length l. We show that a ℘word $v_\Diamond$ is a cover of a ℘word $u_\Diamond$ if every letter of $u_\Diamond$ lies within an occurrence of $v_\Diamond$ in $u_\Diamond$ and $v_\Diamond$ is a border of $u_\Diamond$. Here, we examine string issues that are concerned with identifying recurring patterns in a given total word x. The total word's period p, a common regularity, captures x's repetition, since x is a prefix of a string created by concatenating p. We think about a challenge developed by expanding the scope of this repetitiveness idea by permitting overlaps between the segments that are repeated. We focus on a key issue in string processing: the compact representation of a word by its most frequent factors. The frequency cover, or the longest repeating factor, is a useful and simple form of quasi-periodicity in words that is proposed in this paper.**

Keywords: ℘words, cover, seeds, periodicity, border.

1 Introduction

Numerous branches of science such as combinatorics, system theory, coding and automata theory, formal language theory and molecular biology find regularities in total words (strings) [1, 9]. One of the common issues in pattern matching is how to effectively find repetitions in a given string. The biosequence analysis recently gave the practice of looking for repetitions in strings a new impetus. A common feature of many genomic structures, such as telomeric regions, is the existence of sequentially repeated pieces in DNA sequences that frequently carry significant biological information. The practical role of satellites and alu-repeats in chromosome analysis and genotyping makes them of great interest to genomic researchers. As a result, several biological investigations based on the examination of tandem repeats have been carried out and databases of tandem repeats in certain species have even been created.

Periodicity is the most fundamental concept that encompasses repetitiveness. We say that a string is periodic if it can be created by repeatedly concatenating its smaller component. By allowing superpositions in addition to concatenation, Apostolico and Ehrenfeucht [1] introduced the concept of quasiperiodicity, which broadened the field of periodicity. The basic terms of quasiperiodicity are the notions of cover and seed. In contrast to periods, which are defined solely by concatenations, the terms cover and seed are generalizations of periods in the sense that superpositions as well as concatenations are taken into account to define them.

Covers and their generalizations are an interesting extension of the idea of a repetition. They have potential applications in DNA sequence analysis. The ability to find repeats is helpful in a wide range of word manipulation-related applications. There are many well-known examples, including pattern recognition, computer vision, speech recognition, data compression, data communication, combinatorics, coding and automata theory, formal language theory, and system theory. Finding repeats can be used for text editing in general as well as for tasks like locating duplicate entries in databases. A cover of a word x is a string whose occurrences in x cover all positions of x, while a seed of x is a cover of some superstring of x. Various approximate variants of covers and seeds were studied [4, 10]. Iliopoulos et al. [7] have proposed a new notion of string regularity and an extension of the notions of period and cover called seed. The problem of all restricted seeds with the smallest Hamming distance is studied in [6]. In computation of covers, two problems have been considered in the literature. The shortest-cover problem (also known as the superprimitivity test) is that of computing the shortest cover of a given word of length n, and the al l-covers problem is that of computing all the covers of a given word.

*Manuscript received November 17, 2022; revised June 21, 2023.
R. Krishna Kumari is an Assistant Professor in the Department of Career Development Centre, College of Engineering and Technology, SRM Institute of Science and Technology, Kattankulathur, Chennai-603203, Tamilnadu, India (e-mail: krishrengan@gmail.com).
L. Jeyanthi is an Assistant Professor in the Department of Mathematics, Panimalar Engineering College, Varadharajapuram, Poonamallee, Chennai-600123, Tamilnadu, India (e-mail: jeyanthi.l@gmail.com).
K. Janaki is an Assistant Professor in the Department of Mathematics, Saveetha Engineering College, Saveetha Nagar, Thandalam Chennai-602105, Tamilnadu, India (e-mail: janu89lava@gmail.com).
R. Arulprakasam is an Assistant Professor in the Department of Mathematics, College of Engineering and Technology, SRM Institute of Science and Technology, Kattankulathur, Chennai-603203, Tamilnadu, India (corresponding author e-mail: r.aruljeeva@gmail.com).
P. Madhusoodhanan is the Head of the Department of Career Development Centre, College of Engineering and Technology, SRM Institute of Science and Technology, Kattankulathur, Chennai-603203, Tamilnadu, India (e-mail: madhupms1973@gmail.com).

A ℘word (or partial word) is nothing but a word with holes over the alphabet and is considered in gene comparisons [5, 8, 12]. For instance, the alignment of two DNA sequences which are genetic information carriers can be regarded as construction of two compatible ℘words. In DNA computation, DNA strands are considered as finite words and are utilized for encoding information. While encoding, some part of the information may be unseen or missing which are revealed by using ℘words that denote the positions of the missing symbols in a word. The study of ℘words was initiated by Berstel and Boasson [2] and later the work was extended by Blanchet Sadri [3]. ℘words have wide application in pattern matching. In this paper, we have presented results on quasiperiodicity, covers and seeds. Locating such a regularity can be useful in a wide area of applications, for example in molecular biology (study of the dosDNA microsatellites).

The paper has the following organization. We recall some basics in Section 2 and in Section 3 we introduce cover and seed of a ℘word and study their properties. Cover of a directed ℘ tree is also established. Finally, we conclude the paper in Section 4.

2 Preliminaries

Let the finite *alphabet* $\mathbb{A}$ represent a non-empty set of symbols (or letters). A *total word (or string)* is a sequence of letters over $\mathbb{A}$. The *length (or size)* of a total word $x = x[1 \dots n]$ is n. The length of a total word x is denoted by $|x|$. $Alph(x)$ denotes the set of all elements in x. λ denotes the empty word. $\mathbb{A}^*$ denotes the set of all total words from $\mathbb{A}$ including λ and $\mathbb{A}^+$ denotes the set of all total words from $\mathbb{A}$ excluding λ. A language L is a subset of $\mathbb{A}^*$. The total word x is a *subword (or factor)* of y if the total words u and v exists such that $y = uxv$. If $u, v \neq \lambda$ then x is a proper subword of y. If $u = \lambda$ then x is a prefix of y. If $v = \lambda$ then x is a suffix of y. If x, y and z are total words, with $x = yz$, then zy is a *conjugate* of x. A border of a non-empty word x is a proper factor of x that is both a prefix and a suffix of x.

A total word x of length n is a *cover* of a total word y of length $m \geq n$ if there exists a set of positions $k \subseteq \{0, \dots, m-n\}$ that satisfies the following two conditions; (i) $x[i \dots i+n-1] = y$ for all $i \in k$; (ii) $\bigcup_{i \in k}\{i, \dots, i+n-1\} = \{0, \dots, m-1\}$. A total word y is a *superword* of a total word x if there exists two total words p and q such that $y = pxq$. A total word z is a *seed* of the total word x if it is a cover of a superword of x. A string y is called *quasiperiodic* if it has a non-trivial cover, that is, there exists a shorter string x such that every position in y is inside one of the occurrences of x in y. The word xy is a concatenation of two words x and y. The concatenations of t copies of x is denoted by x^t. For two words $x = x_1 \dots x_n$ and $y = y_1 \dots y_m$ such that $x_{n-i+1} \dots x_n = y_1 \dots y_j$ for some $j \geq 1$, the word $x = x_1 \dots x_n y_{j+1} \dots y_m$ is a superposition of x and y with j overlaps. A word $w = w_1 \dots w_n$ is a circular rotation of $x = x_1 \dots x_n$ if $w_1 \dots w_n = x_j \dots x_n x_1 \dots x_{j-1}$ for some $1 \leq j \leq n$ (for $j : 1, w = x$).

The sequence of symbols that contains a number of "do not know symbols" or "holes" denoted as $\Diamond$ is termed as a ℘*word*. The ℘word of u denoted by $u_\Diamond$ is the total function $u_\Diamond : \{1, 2, \dots, n\} \to \mathbb{A}_\Diamond = \mathbb{A} \cup \{\Diamond\}$ defined by

$$u_\Diamond(i) = \begin{cases} u(i) & \text{if } i \in D(u) \\ \Diamond & \text{if } i \in H(u). \end{cases}$$

The set of all ℘words over $\mathbb{A}_\Diamond$ is denoted as $\mathbb{A}_\Diamond^*$. $\mathbb{A}_\Diamond^+$ denotes the set of all ℘words excluding the empty word. A partial language (℘ language) $L_\Diamond \subseteq \mathbb{A}_\Diamond^*$ is a set of all ℘ words over $\mathbb{A}_\Diamond$.

We note that,
(i) A total word is a ℘word with zero holes and the empty word is not a ℘word.

(ii) The symbol $\Diamond$ does not belong to the alphabet $\mathbb{A}$ but a standby symbol for the unknown letter.

(iii) The symbol $\Diamond$ is compatible to the letters of the alphabet $\mathbb{A}$.

(iv) The symbol $\Diamond$ alone of any length cannot exist as a word. In other words, the hole of any length is neither a total word nor a ℘word.

A ℘word $u_\Diamond = u_\Diamond[1 \dots n]$ is *primitive (non-periodic)* if no word v exists such that $u_\Diamond \subset v^i$ with $i \geq 2$. ℘words that are not primitive are said to be *periodic* ℘words. If $u_\Diamond$ is a ℘word then the period of $u_\Diamond$ is denoted as $\mathbf{p}(u_\Diamond)$. If $u_\Diamond$ and $v_\Diamond$ are two ℘words of equal length and if all the elements in domain of $u_\Diamond$ are also in domain of $v_\Diamond$ with $u_\Diamond(i) = v_\Diamond(i)$ for all $i \in D(u_\Diamond)$, then $u_\Diamond$ is contained in $v_\Diamond$ and is denoted by $u_\Diamond \subset v_\Diamond$. Two ℘words $u_\Diamond$ and $v_\Diamond$ are *compatible*, denoted by $u_\Diamond \uparrow v_\Diamond$ if $u_\Diamond(i) = v_\Diamond(i)$ for all $i \in D(u_\Diamond) \cap D(v_\Diamond)$. Equivalently, the ℘words $u_\Diamond$ and $v_\Diamond$ are compatible if a ℘word $w_\Diamond$ exists such that $u_\Diamond \subset w_\Diamond$ and $v_\Diamond \subset w_\Diamond$. A ℘word $u_\Diamond$ is bordered (denoted as $\mathrm{Bor}(u_\Diamond)$) if non-empty words p, q, v exist such that $u_\Diamond \subset pv$ and $u_\Diamond \subset qp$

3 Lseed and Rseed of ℘words

Definition 1. *A ℘word $v_\Diamond$ (or a set of compatible ℘words) of length l is a cover of a ℘word $u_\Diamond$ of length $m \geq l$ if there exists a set of positions $k \subseteq \{0, \dots, m-l\}$ that satisfies the following two conditions;*

1. $u_\Diamond[i \dots i+l-1] = v_\Diamond$ *for all* $i \in k$

2. $\bigcup_{i \in k}\{i, \dots, i+l-1\} = \{0, \dots, m-1\}$.

Note that a cover $v_\Diamond$ of a $\wp$word $u_\Diamond$ is proper if $v_\Diamond \neq u_\Diamond$.

Example 1. *Consider a $\wp$word $u_\Diamond = ba\Diamond baa\Diamond b\Diamond abaab$ over the alphabet $\mathbb{A}_\Diamond = \{a, b\} \cup \{\Diamond\}$. The set of $\wp$words $\{ba\Diamond b, baa\Diamond, b\Diamond ab, baab\}$ that are compatible to each other represent the cover of the $\wp$word $u_\Diamond$. Figure 1 shows the cover of $\wp$word $u_\Diamond$.*

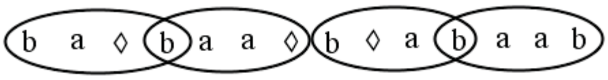

Figure 1: The cover of $\wp$word $u_\Diamond = ba\Diamond baa\Diamond b\Diamond abaab$

Remark 1. *If $v_\Diamond$ is a cover of the $\wp$word $u_\Diamond$, then $v_\Diamond$ is both a prefix and suffix of $u_\Diamond$.*

Definition 2. *A frequency cover of $u_\Diamond$ is the longest of those repeating factors $v_\Diamond$ of $u_\Diamond$, $|v_\Diamond| > 1$, that occurs the maximum number of times in $u_\Diamond$.*

Example 2. *Consider a $\wp$word $u_\Diamond = a\Diamond a\Diamond a\Diamond a$, the factor $a\Diamond a$ is the frequency cover of $u_\Diamond$, occurring three times, as do the shorter factors $a\Diamond$ and $\Diamond a$.*

Remark 2. *A frequency cover of a $\wp$word is not unique.*

Example 3. *Consider a $\wp$word $u_\Diamond = a\Diamond a\Diamond c\Diamond c\Diamond$. It has two frequency covers $a\Diamond$ and $c\Diamond$.*

Remark 3. *Not all $\wp$words have a frequency cover.*

Example 4. *Consider a $\wp$word $u_\Diamond = a\Diamond cd\Diamond fgh$ does not have a frequency cover.*

It should be noted that we require the length of a $\wp$word's frequency covers to be greater than one because it is simple and quick to calculate the frequency of each distinct letter in a $\wp$word, at least for an ordered alphabet of manageable size (simply scan the string from left to right and count the number of occurrences of each distinct letter).

Theorem 1. *Suppose $v_\Diamond$ and $w_\Diamond$ are the longest and shortest frequency covers of $u_\Diamond$ respectively. Then $v_\Diamond$ always covers more positions in $u_\Diamond$ than $w_\Diamond$ does.*

Proof. Since both $v_\Diamond$ and $w_\Diamond$ are frequency covers, $f_{u_\Diamond, v_\Diamond} = f_{u_\Diamond, w_\Diamond}$. Observe that the shortest frequency cover $v_\Diamond$ will always be of size two; that is, $|v_\Diamond| = 2$. For if $|v_\Diamond| > 2$, any factor of $v_\Diamond$ of length two would have the same frequency as that of $v_\Diamond$ in $u_\Diamond$ and be shorter than $v_\Diamond$, thus contradicting the assumption that $v_\Diamond$ is the shortest frequency cover.

For $v_\Diamond$ to cover fewer positions than $w_\Diamond$ does, some occurrences of $v_\Diamond$ in $u_\Diamond$ must overlap. Note that the overlap between any two instances of $u_\Diamond$ cannot be greater than $\lfloor v_\Diamond/2 \rfloor$ as it would create a repetition in $v_\Diamond$ which leads to $v_\Diamond$ not being the frequency cover which is a contradiction. Therefore, $v_\Diamond = x_\Diamond b x_\Diamond$ (where b is a symbol). Additionally b is non-empty as otherwise it would create a repetition in $v_\Diamond$ which leads to $v_\Diamond$ not being the frequency cover which is a contradiction. If $|x_\Diamond| > 1$, then $x_\Diamond$ would be the frequency cover and not x. Therefore, $|v_\Diamond| = 3$. Note that the least positions covered by $v_\Diamond$ is when all occurrences of $v_\Diamond$ in $u_\Diamond$ overlap. However, assuming this worst case, $v_\Diamond$, where $|v_\Diamond| = 3$, still covers one more position in $u_\Diamond$ than $w_\Diamond$ does. Therefore, it is not possible for a shortest frequency cover to cover more positions than the positions covered by the longest frequency cover. □

Definition 3. *A border $v_\Diamond$ of a $\wp$word $u_\Diamond$ is an enriched cover of $u_\Diamond$, if the number of letters of $u_\Diamond$ which lie within some occurrence of $v_\Diamond$ in $u_\Diamond$ is a maximum over all borders of $u_\Diamond$.*

Example 5. *Consider a $\wp$word $u_\Diamond = ab\Diamond ab\Diamond abb\Diamond ab\Diamond ab$ over the alphabet $\mathbb{A}_\Diamond = \{a, b\} \cup \{\Diamond\}$. The $\wp$word $\{ab\Diamond ab\}$ represent the border of the $\wp$word $u_\Diamond$. Figure 5 shows the cover of $\wp$word $u_\Diamond$.*

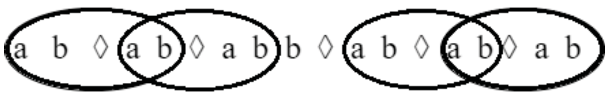

Figure 2: The cover of $\wp$word $u_\Diamond = ab\Diamond ab\Diamond abb\Diamond ab\Diamond ab\Diamond ab$

Definition 4. *A $\wp$word $v_\Diamond$ is the minimal enriched cover of a $\wp$word $u_\Diamond$, if $v_\Diamond$ is the shortest enriched cover of $u_\Diamond$.*

Theorem 2. *Any $\wp$word with minimal enriched cover is not periodic.*

Proof. Let $v_\Diamond$ be the minimal enriched cover of the $\wp$word $u_\Diamond$. Suppose $v_\Diamond$ is periodic with longest border $w_\Diamond$, then we have $\mathrm{Bor}(v_\Diamond)+\mathbf{p}(v_\Diamond) \geq 2\mathbf{p}(v_\Diamond)$. It follows that $|w_\Diamond| \geq \mathbf{p}(v_\Diamond) \geq |v_\Diamond|/2$ and so $w_\Diamond$ is a cover of $v_\Diamond$. Hence also the minimal enriched cover of $u_\Diamond$, which is a contradiction. □

Definition 5. *A $\wp$word $v_\Diamond$ is a super$\wp$word of a $\wp$word $u_\Diamond$ if there exists two $\wp$words $p_\Diamond$ and $q_\Diamond$ such that $v_\Diamond = p_\Diamond u_\Diamond q_\Diamond$. A $\wp$word $w_\Diamond$ (or a set of compatible $\wp$words) is a seed of the $\wp$word $u_\Diamond$ if it is a cover of a super$\wp$word of $u_\Diamond$.*

Example 6. *Consider a $\wp$word $u_\Diamond = bab\Diamond abb\Diamond ab\Diamond abba\Diamond b$ over the alphabet $\mathbb{A}_\Diamond = \{a, b\} \cup \{\Diamond\}$. The set of $\wp$words $\{ab\Diamond a, abb\Diamond, ab\Diamond a, abba\}$ that are compatible to each other represent the seed of the $\wp$word $u_\Diamond$ since the set is a cover of a super$\wp$word $v_\Diamond = ab\Diamond abb\Diamond ab\Diamond abba$ of $u_\Diamond$. Figure 6 shows the seed of $\wp$word $u_\Diamond$.*

Definition 6. *A left seed (denoted as Lseed) of a $\wp$word $u_\Diamond$ is a prefix of $u_\Diamond$ that exists as a cover of a super$\wp$word*

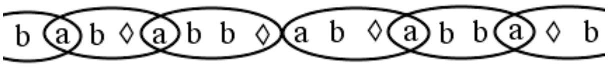

Figure 3: The seed of $\wp$word $u_\Diamond = bab\Diamond abb\Diamond ab\Diamond abba\Diamond b$

of $u_\Diamond$ in the form $u_\Diamond z$ where z is a possibly empty word.

Likewise a right seed [denoted as Rseed] of a $\wp$word $u_\Diamond$ is a suffix of $u_\Diamond$ that exists as a cover of a super$\wp$word of $u_\Diamond$ in the form $zu_\Diamond$ where z is a possibly empty word.

Example 7. *Consider a $\wp$word $u_\Diamond = aab\Diamond abaaaba\Diamond ba\Diamond b$ over the alphabet $\mathbb{A}_\Diamond = \{a,b\} \cup \{\Diamond\}$. The set of $\wp$words $\{aab\Diamond, \Diamond aba, aaba, a\Diamond ba\}$ that are compatible to each other represent the Lseed of the $\wp$word $u_\Diamond$ since the set is a prefix of $u_\Diamond$ and is a cover of a super$\wp$word $v_\Diamond = aab\Diamond abaaaba\Diamond ba$ of $u_\Diamond$. Figure 7 shows the Lseed of $\wp$word $u_\Diamond$.*

Figure 4: The Lseed of $\wp$word $u_\Diamond = aab\Diamond abaaaba\Diamond ba\Diamond b$

Example 8. *Consider a $\wp$word $u_\Diamond = ababb\Diamond ba\Diamond ab$ over the alphabet $\mathbb{A}_\Diamond = \{a,b\} \cup \{\Diamond\}$. The set of $\wp$words $\{aab\Diamond, \Diamond aba, aaba, a\Diamond ba\}$ that are compatible to each other represent the Rseed of the $\wp$word $u_\Diamond$ since the set is a suffix of $u_\Diamond$ and is a cover of a super$\wp$word $v_\Diamond = babb\Diamond ba\Diamond ab$ of $u_\Diamond$. Figure 8 shows the Lseed of $\wp$word $u_\Diamond$.*

Figure 5: The Rseed of $\wp$word $u_\Diamond = ababb\Diamond ba\Diamond ab$

Definition 7. *The minimal (maximal) Lseed of $u_\Diamond$ denoted as $L_{min}seed(u_\Diamond)$ ($L_{max}seed(u_\Diamond)$) is the prefix of $u_\Diamond$ with minimum (maximum) length such that it is a cover of a super$\wp$word of $u_\Diamond$. Likewise the minimal (maximal) Rseed of $u_\Diamond$ denoted as $R_{min}seed(u_\Diamond)$ ($R_{max}seed(u_\Diamond)$) is the suffix of $u_\Diamond$ with minimum (maximum) length such that it is a cover of a super$\wp$word of $u_\Diamond$.*

Theorem 3. *If a seed $w_\Diamond$ covers a $\wp$word $u_\Diamond$ by concatenation, then all the circular conjugates of $w_\Diamond$ cover $u_\Diamond$ by concatenations.*

Proof. Since the seed $w_\Diamond$ covers a $\wp$word $u_\Diamond$ by concatenation, a cover $w_\Diamond^m$ of $u_\Diamond$ by $w_\Diamond$ exists. Let $w'_\Diamond$ be a circular conjugate of $w_\Diamond$. Here $w_\Diamond^m$ is a factor of $w_\Diamond^{m+2}$. Then $w_\Diamond^{m+2}$ is a cover of $u_\Diamond$ and thus $w'_\Diamond$ covers $u_\Diamond$ by concatenations. Thus all the circular conjugates of $w_\Diamond$ cover $u_\Diamond$ by concatenations. □

Theorem 4. *A $\wp$word $v_\Diamond$ is a Lseed (Rseed) of a $\wp$word $u_\Diamond$ if and only if $v_\Diamond$ is a cover of the prefix (suffix) of $u_\Diamond$ with $|v_\Diamond| \le \mathbf{p}(u_\Diamond)$.*

Proof. Let us assume that a $\wp$word $v_\Diamond$ covers a prefix of $u_\Diamond$, say qr with $|v_\Diamond| \ge \mathbf{p}(u_\Diamond)$, such that $|q| = \mathbf{p}(u_\Diamond)$ and r is a possibly empty word. Consider a smallest integer t such that $u_\Diamond$ is a prefix of q^t. Then $v_\Diamond$ is a cover of $q^t r = u_\Diamond r$ for some word s, possibly empty. Hence $v_\Diamond$ is a Lseed of $u_\Diamond$.

Conversely, assume $v_\Diamond$ to be a Lseed of $u_\Diamond$. Then the following two cases occur:

1. If $|v_\Diamond| \le \mathrm{Bor}(u_\Diamond)$, then a suffix r of $v_\Diamond$, possibly empty is a prefix of the border. Now consider the Lseed which is a cover of $u_\Diamond[\mathbf{p}(u_\Diamond) - 1]$. Then $v_\Diamond$ is a cover of qr and also $|q| = \mathbf{p}(u_\Diamond)$.
2. Let us consider $|v_\Diamond| > \mathrm{Bor}(u_\Diamond)$. Suppose that $v_\Diamond$ with $|v_\Diamond| \ge \mathbf{p}(u_\Diamond)$, does not cover a prefix of $u_\Diamond$ then assume that r =Bor$(u_\Diamond)$ such that r is a factor of $v_\Diamond = qrs$, where q and r are non-empty. Now by considering the Lseed which is a cover of $u_\Diamond[\mathbf{p}(u_\Diamond) - 1]$, we get qr as a longest border of $u_\Diamond$, which is a contradiction.

Hence the result. □

Theorem 5. *Let $v_\Diamond$ be a cover of $\wp$word $u_\Diamond$ and let $w_\Diamond \neq v_\Diamond$ be a factor of $u_\Diamond$ with $|w_\Diamond| \le |v_\Diamond|$. Then $w_\Diamond$ is a cover of $u_\Diamond$ if and only if $w_\Diamond$ is a cover of $v_\Diamond$.*

Proof. If $w_\Diamond$ is a cover of $v_\Diamond$ and $v_\Diamond$ is a cover of $u_\Diamond$, then $w_\Diamond$ is a cover of $u_\Diamond$. Suppose if both $w_\Diamond$ and $v_\Diamond$ are covers of $u_\Diamond$, then $w_\Diamond = \mathrm{Bor}(v_\Diamond)$ since length of $w_\Diamond$ is less than or equal to $v_\Diamond$. Therefore $w_\Diamond$ must be a cover of $v_\Diamond$. Hence the result. □

Theorem 6. *For any $\wp$word $u_\Diamond$ with $|u_\Diamond| = m$, if $\mathbf{p}(u_\Diamond) = m$ then $L_{min}seed(u_\Diamond) = u_\Diamond$.*

Proof. By the notion of L$_{min}$seed, we get L$_{min}(u_\Diamond) \le m$. Let us assume that L$_{min}(u_\Diamond) < m$. Then in order to cover $u_\Diamond$, a non-empty prefix $v_\Diamond$ of L$_{min}(u_\Diamond)$ which is also a suffix of $u_\Diamond$ exists. Now let us consider the Lseed that covers $u_\Diamond[m-1]$. Then $m - |v_\Diamond|$ is a minimal period of $u_\Diamond$ which is a contradiction. Hence the L$_{min}$seed$(u_\Diamond)$ is equal to $u_\Diamond$. □

Theorem 7. *For any $\wp$word $u_\Diamond$ with $|u_\Diamond| = n$ and $\mathbf{p}(u_\Diamond) = m$, if $m = n$ then there is no longest right seed for $u_\Diamond$.*

Proof. Consider that $m = n$. Let us assume that $u_\Diamond[k \ldots n-1]$ is the longest right seed of $u_\Diamond$ with $0 < k \leq n-1$. Then to cover $u_\Diamond$, a non-empty suffix of $u_\Diamond[k \ldots n-1]$ say $v_\Diamond$ is a prefix of $u_\Diamond$. Then $n - |v_\Diamond|$ gives a shorter period for $u_\Diamond$ which contradicts our assumption. Therefore if $m = n$ then there is no longest right seed for $u_\Diamond$. □

Theorem 8. *For all $0 \leq i < m$, if $Lseed[i] = L_{min}seed(u_\Diamond[0 \ldots i])$, then for all $0 \leq i < m-1$ we get $Lseed[i] \leq Lseed[i+1]$.*

Proof. Let us prove by contradiction. Assume that $\mathrm{Lseed}[i] > \mathrm{Lseed}[i+1]$. By the notion of the L_{min}seed, we have $\mathrm{Lseed}[i]$ covers some super$\wp$word $u_\Diamond[0 \ldots i]r$, where r is a possibly empty word. Similarly $\mathrm{Lseed}[i+1]$ covers some super$\wp$word $u_\Diamond[0 \ldots i+1]s$, where s is a possibly empty word. This shows that $\mathrm{Lseed}[i+1]$ covers $u_\Diamond[0 \ldots i]u_\Diamond[i+1]s$. Thus by notion of L_{min}seed, $\mathrm{Lseed}[i+1]$ is the L_{min}seed of $u_\Diamond[0 \ldots i]$. But we get a shorter Lseed for $u_\Diamond[0 \ldots i]$ which is a contradiction. Hence the result. □

3.1 Cover of a directed $\wp$tree

Definition 8. *A $\wp$word $v_\Diamond$ is a cover of an edge labeled directed $\wp$tree τ if every edge of τ can be covered by some simple path with label $v_\Diamond$ such that all edges directed towards the parent node of τ.*

Example 9. *Consider a $\wp$word $v_\Diamond = a\Diamond b\Diamond bba\Diamond b\Diamond a$ over the alphabet $\mathbb{A}_\Diamond = \{a, b\} \cup \{\Diamond\}$. The set of $\wp$words $\{a\Diamond b\Diamond, \Diamond bba, \Diamond b\Diamond a\}$ that are compatible to each other and also compatible with abba represent the cover of the directed $\wp$tree τ. Figure 9 shows the cover of a directed tree τ.*

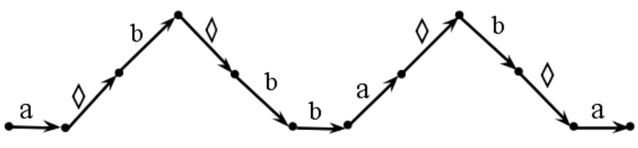

Figure 6: Cover of a directed tree τ

Remark 4. *If $v_\Diamond$ is a cover of a directed tree τ, then it is a cover of minimum one $\wp$word of τ corresponding to terminal nodes-to-parent node.*

Remark 5. *If $v_\Diamond$ is a cover of a directed tree τ, then $v_\Diamond$ is a prefix of all terminal nodes-to-parent node labels and also is a prefix of longest common prefix of all terminal nodes-to-parent node paths.*

4 Conclusion

In this paper, we have focused on the identification of various kinds of periodicities and other regularities in $\wp$words such as covers and seeds. The study is based on the maintenance of a new, simple but powerful data structure. For the future work, our immediate target is to investigate whether there exists $O(n)$-time algorithm for computing the minimal enriched cover. For certain applications, the notion of the minimal enriched cover might not be useful, since it primarily optimises the number of positions covered, while the length of the enriched cover cannot be managed. We can extend this notion by introducing the D-restricted enriched cover of $\wp$word $u_\Diamond$, which is the shortest border of $u_\Diamond$ of length not exceeding D which covers the largest number of positions among borders no longer than D. We would like to design an algorithm based on determinization of a suffix automaton which is appropriate for computation of all $\wp$word seeds with the smallest Hamming distance.

References

[1] Alberto Apostolico and Andrzej Ehrenfeucht, Efficient detection of quasiperiodicities in strings, *Theoretical Computer Science*, vol.119, (1993), pp.247–265.

[2] J. Berstel and L. Boasson, Partial Words and a Theorem of Fine and Wilf, *Theoretical Computer Science*, vol.218, (1999), pp.135–141.

[3] F. Blanchet-Sadri, Primitive Partial Words, *Discrete Applied Mathematics*, vol.148, (2005), pp.195–213.

[4] Christodoulakis, Manolis, Costas S. Iliopoulos, Kunsoo Park, and Jeong Seop Sim, Approximate Seeds of Strings, *Journal of Automata, languages and Combinatorics*, vol. 10(5), (2005), pp. 609–626.

[5] M.J. Fischer and M.S. Paterson, String Matching and other Products,*In: Karp, R.M. (ed.) Complexity of Computation. SIAM-AMS Proceedings*, (1974), pp.113–125.

[6] O. Guth and B. Melichar, Finite Automata Approach to Computing All Seeds of Strings with the Smallest Hamming Distance, *IAENG International Journal of Computer Science*, vol. 36, no. 2, (2009), pp. 137–146.

[7] C.S. Iliopoulos, D.W.G. Moore, K. Park, Covering a string, *Proc. 4th Ann. Symp. on Combinatorial Pattern Matching, Lecture Notes in Computer Science*, vol. 684, (1993), pp. 54–62.

[8] K. Janaki and R. Arulprakasam, Tiling systems and domino systems for partial array languages, *IAENG International Journal of Applied Mathematics*, vol. 52(2), (2022), pp. 418-425.

[9] M. Lothaire, Combinatorics on Words, *Cambridge University Press*, 1983.

[10] Mhaskar, Neerja, and William F. Smyth, String covering with optimal covers, *Journal of Discrete Algorithms* , vol. 51, (2018), pp. 26–38.

[11] Mhaskar, Neerja, and William F. Smyth, Frequency covers for strings, *Fundamenta Informaticae*, vol. 163(3), (2018), pp.275-289.

[12] K. Sasikala, V. Rajkumar Dare and D.G. Thomas, Learning of Partial Languages, *Engineering Letters*, vol.14(2), (2007), pp.72–80.