Factors and Subwords of Rich Partial Words

R. Krishna Kumari Member, IAENG and R. Arulprakasam Member, IAENG *

Abstract—**Many classes of finite words have noticeable properties with reference to their palindromic factors and one among them are the words having zero palindromic defect i.e., words rich in palindromes. In this paper we introduce rich partial word and discuss its combinatorial properties. We show that the palindromic richness of a partial word can be studied by including the positions of the missing symbols in that word. The significant difference between rich and rich partial word is that a rich word of length n contains exactly $n+1$ distinct palindromic factors whereas a rich partial word of length n contains at least $n+1$ distinct palindromic factors. These factors differ from the classical palindromes due to the presence of holes.**

Keywords: palindromes, rich words, factors, partial words, primitivity.

1 Introduction

In the study of the various properties of words with finite length [8] such as structural and combinatorial properties, palindromes are natural objects which play a vital role in word combinatorics, automata theory and formal languages. Palindromes often occur in DNA and are extensively present in human cancer cells [11]. In biological context, complement DNA characters are considered by palindromes. By identification of these segments of DNAs, the instability of genomes could be understood. Biologists believe that palindromes play an major role in cell processes and other regulation gene activity because these are frequently noticed near introns, promoters and specific untranslated regions. So, locating palindromic factors in any genome sequence is vital. Also for comparison study, locating common palindromes in two genome sequences can be a major criterion. A palindromic word is a word when taken in reverse order gives the same word. Many classes of words have prominent properties with regard to their palindromic factors [7]. Algorithmic and combinatorial studies of palindromes are considered as a favorable tool to construct linear-time recognizable languages [3, 15]. In the study of palindromes, one of the recent topics of interest concerns an interesting class of finite words termed as rich words. Words comprising the greatest number of distinct and palindromic factors are rich words and are called as words with zero palindromic defect [2, 12, 13, 14, 17]. In [18], X. Droubay et al. showed that a finite word x of length $|x|$ has maximum $|x|$ distinct palindromic factors, excluding the empty word. Characterized by this palindromic richness property in [3], the authors launched a unified study of words with finite and infinite length. Accordingly we say that a finite word x is rich if and only if it has exactly $|x|+1$ distinct palindromic factors. In various contexts, rich words have appeared such as complementation-symmetric sequences, episturmian words and a certain class of words associated with β-expansions where β represents a simple Parry number. The number of rich binary words of length n can be referred in https://oeis.org/A216264.

Partial words are nothing but words with holes and are considered in gene comparisons [1, 9, 16]. For instance, alignment of two DNA sequences which are genetic information carriers can be regarded as formation of two compatible partial words. The DNA sequence is treated as a finite word in DNA computations, and is used to encode information. When encoding information, some parts of the information may be hidden or not visible, which are revealed by using a partial word which represents the position of a missing symbol in a word. Initial research on partial words was initiated by Berstel and Boasson [10] and later expanded by Blanchet-Sadri [4, 5, 6]. Partial words and palindromicity of words are classical topics in molecular biology and language theory which inspired and initiated a unified study of rich words and partial words. The hole(s) present in partial words is not a character of the alphabet but survives as a back-up symbol for the unknown letter. Since it is compatible to any of the letter(s) in the alphabet, if a hole in a rich partial word over the alphabet is replaced by a letter in the alphabet, the rich partial word turns out to be a rich word. On the other hand, since holes do not belong to the alphabet, we study the palindromic richness of a partial word by including the positions of the missing symbols in that word. This paper introduces rich partial words and study some combinatorial properties. We initially recall in Section 2 the fundamental notions and properties. We define rich partial words and discuss some properties based on

*Manuscript received September 21, 2022; January 28, 2023.
R. Krishna Kumari is an Assistant Professor in the Department of Career Development, SRM Institute of Science and Technology, Kattankulathur, Chennai-603203, Tamilnadu, India
E-mail: (krishnar10@srmist.edu.in).
R. Arulprakasam is an Assistant Professor in the Department of Mathematics, SRM Institute of Science and Technology, Kattankulathur, Chennai-603203, Tamilnadu, India (Corresponding author). E-mail: (arulprar@srmist.edu.in).

their palindromic richness in Section 3. In Section 4 we discuss the relation between partial palindromic perfect factors and partial palindromic perfect subwords of rich partial words followed by conclusion in Section 5.

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 $x = x[1 \ldots n]$ 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. A finite total word x is called a *palindrome* if $x = x^R$ where x^R is the *reversal (mirror image)* of x. A total word x is rich if it has exactly $|x| + 1$ distinct factors that are palindromic including the empty word λ. A non-empty factor x of a finite word u is *unioccurrent* in y if x has exactly one occurrence in y. If x has more than one occurrence in y, then there exists a factor z of y having exactly two distinct occurrences of x, one as a prefix and other as a suffix. Such a factor z is called a complete return to x in y. For example, *bbcacbb* is a complete return to bb in the rich word *bbcacbbba*. The sequence of symbols that contains a number of "do not know symbols" or "holes" denoted as $\lozenge$ is termed as a *finite partial word (or partial word)*.

The partial word of u denoted by $u_\lozenge$ is the total function $u_\lozenge : \{1, 2, \ldots, n\} \to \mathbb{A}_\lozenge = \mathbb{A} \cup \{\lozenge\}$ defined by

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

where $D(u)$ represents the domain set and $H(u)$ denotes the set of holes in u The set of all partial words over $\mathbb{A}_\lozenge$ is denoted as $\mathbb{A}_\lozenge^*$. $\mathbb{A}_\lozenge^+$ denotes the set of all partial words excluding the empty word. A partial language $L_\lozenge \subseteq \mathbb{A}_\lozenge^*$ is a set of all partial words over $\mathbb{A}_\lozenge$.

We note that,
(i) A total word is a partial word with zero holes and the empty word is not a partial word.
(ii) The symbol $\lozenge$ does not belong to the alphabet $\mathbb{A}$ but a standby symbol for the unknown letter.
(iii) The symbol $\lozenge$ is compatible to the letters of the alphabet $\mathbb{A}$.
(iv) The symbol $\lozenge$ alone of any length cannot exist as a word. In other words, the hole of any length is neither a total word nor a partial word.

A partial word $u_\lozenge = u_\lozenge[1 \ldots n]$ is *primitive (non-periodic)* if no word v exists such that $u_\lozenge \subset v^i$ with $i \geq 2$. Partial words that are not primitive are said to be *periodic* partial words. If $u_\lozenge$ and $v_\lozenge$ are two partial words of equal length and if all the elements in domain of $u_\lozenge$ are also in domain of $v_\lozenge$ with $u_\lozenge(i) = v_\lozenge(i)$ for all $i \in D(u_\lozenge)$, then $u_\lozenge$ is contained in $v_\lozenge$ and is denoted by $u_\lozenge \subset v_\lozenge$. Two partial words $u_\lozenge$ and $v_\lozenge$ are *compatible*, denoted by $u_\lozenge \uparrow v_\lozenge$ if $u_\lozenge(i) = v_\lozenge(i)$ for all $i \in D(u_\lozenge) \cap D(v_\lozenge)$. Equivalently, the partial words $u_\lozenge$ and $v_\lozenge$ are compatible if a partial word (or a total word) $w_\lozenge$ exists such that $u_\lozenge \subset w_\lozenge$ and $v_\lozenge \subset w_\lozenge$. A finite partial word $u_\lozenge$ is a palindrome if $u_\lozenge$ is compatible with its reversal (denoted by $u_\lozenge \uparrow u_\lozenge^R$). For instance $u_\lozenge = \lozenge ab \lozenge aba \lozenge$ is a palindrome.

3 Rich Partial Words

This section defines rich partial words in view of their palindromic richness and discusses their combinatorial properties. The empty word λ is regarded as a palindrome.

Definition 1. *A factor $p_\lozenge$ of a partial word $u_\lozenge$ over $\mathbb{A}_\lozenge$ is called a partial palindromic proper factor if $p_\lozenge$ is compatible with its reversal (denoted by $p_\lozenge \uparrow p_\lozenge^R$). The set of all non-empty partial palindromic proper factors of $u_\lozenge$ is denoted by $PPPF(u_\lozenge)$.*

Example 1. *Consider a partial word $u_\lozenge = baab\lozenge b$ over $\mathbb{A}_\lozenge = \{a, b\} \cup \{\lozenge\}$. The palindromic factors of $u_\lozenge$ are*

$$\{\lambda, a, b, aa, b\lozenge, \lozenge b, ab\lozenge, b\lozenge b, baab\}.$$

Here the factors $\{b\lozenge, \lozenge b, ab\lozenge, b\lozenge b\}$ are termed as partial palindromic factors.

Definition 2. *Any partial word over $\mathbb{A}_\lozenge$ with length n is a rich partial word if it has at least n distinct partial palindromic proper factors.*

Example 2. *Consider a partial word $u_\lozenge = ba\lozenge aba$ over $\mathbb{A}_\lozenge = \{a, b\} \cup \{\lozenge\}$ with $|u_\lozenge| = 6$.*
The set of all distinct palindromic proper factors of $u_\lozenge$ are

$$\{\lambda, a, b, a\lozenge, \lozenge a, ba\lozenge, a\lozenge a, \lozenge ab, aba, ba\lozenge ab, a\lozenge aba\}.$$

Among the above set, the set of all distinct partial palindromic factors of $u_\lozenge$ are

$$\{a\lozenge, \lozenge a, ba\lozenge, a\lozenge a, \lozenge ab, ba\lozenge ab, a\lozenge aba\}.$$

Here the number of distinct partial palindromic proper factors is equal to $|u_\lozenge| + 1$. Hence $u_\lozenge$ is a rich partial word.

Example 3. *Consider a partial word $v_\lozenge = \lozenge ababb$ with length $|v_\lozenge| = 6$ over $\mathbb{A}_\lozenge = \{a, b\} \cup \{\lozenge\}$. Then the partial palindromic proper factors of $v_\lozenge$ are*

$$v_\lozenge = \{\lozenge a, \lozenge ab, \lozenge abab\}.$$

Algorithm 1: To determine whether the given partial word is rich

Input: S -Partial word of length n
Output: F- Partial Palindromic proper factors of S of length at least n, S is rich

```
Define a partial word with one hole S = s_1 s_2 s_3 ... s_n
for i in range n do
    for j in range (i+1, n+1) do
        all possible factors with one hole and of
        length at most equal to n
        if factors are palindromes then
            count++
        end
    end
end
if count ≥ n+1 then
    return Given partial word with one hole is
    rich
end
```

Here the number of distinct partial palindromic proper factors is less than $|v_\Diamond|$. Hence $v_\Diamond$ is not a rich partial word.

Remark 1. *Every factor of a rich word is rich but every factor of a rich partial word need not be rich.*

Remark 2. *A partial word $u_\Diamond$ is rich iff every prefix (resp. suffix) of $u_\Diamond$ has a unioccurent palindromic suffix (resp. prefix).*

Example 4. *Consider a rich partial word $u_\Diamond = ab\Diamond bba$ over $\mathbb{A}_\Diamond$. The set of all prefixes of $u_\Diamond$ with unioccurrent palindromic suffixes (underlined) are*

$$\{ab\underline{\Diamond bba}, a\underline{b\Diamond bb}, ab\underline{\Diamond b}, a\underline{b\Diamond}, a\underline{b}, \underline{a}\}.$$

The set of all suffixes of $u_\Diamond$ with unioccurrent palindromic prefixes (underlined) are

$$\{\underline{ab\Diamond} bba, \underline{b\Diamond bb}a, \underline{\Diamond bb}a, \underline{bb}a, \underline{b}a, \underline{a}\}.$$

Theorem 1. *For any partial word $u_\Diamond$ over $\mathbb{A}_\Diamond$, $u_\Diamond$ is rich iff each non-palindromic proper factor $r_\Diamond$ of $u_\Diamond$ is uniquely represented as a pair $p_\Diamond q_\Diamond$ of distinct palindromes such that*
(i)$p_\Diamond$ and $q_\Diamond$ are not equal;
(ii)$p_\Diamond$ and $q_\Diamond$ are not factors of one another;
(iii)$q_\Diamond$ is the palindromic suffix (denoted as pal_s) of $r_\Diamond$ with maximum length;
(iv)$p_\Diamond$ is the palindromic prefix (denoted as pal_p) of $r_\Diamond$ with maximum length.

Proof. Suppose $u_\Diamond$ is a rich partial word. By Remark 2, a non-palindromic factor $r_\Diamond$ of $u_\Diamond$ with prefix $p_\Diamond$ has a unioccurrent suffix $q_\Diamond$. Also $|r_\Diamond| \geq max\{|p_\Diamond|, |q_\Diamond|\}$. Inevitably this follows that $p_\Diamond$ and $q_\Diamond$ are not equal and also $p_\Diamond$ as well as $q_\Diamond$ are not factors of one another. Here the factors $p_\Diamond$ and $q_\Diamond$ are unioccurrent and $p_\Diamond \neq q_\Diamond$. Also $p_\Diamond$ and $q_\Diamond$ are the prefix and suffix of $u_\Diamond$ with maximum length and not factors of one another.

To prove the uniqueness, for any finite rich partial word $u_\Diamond$ with factors $v_\Diamond$ and $r_\Diamond$ having the same pal_p $p_\Diamond$ and same pal_s $q_\Diamond$ with maximum length. We have to show that $v_\Diamond = r_\Diamond$. Let us prove by contradiction. Suppose $v_\Diamond \neq r_\Diamond$ such that both $v_\Diamond$ and $r_\Diamond$ are not palindromes. Then $v_\Diamond$ and $r_\Diamond$ are not factors of one another and $p_\Diamond$ and $q_\Diamond$ are unioccurent in each of $v_\Diamond$ and $r_\Diamond$. Let $k_\Diamond$ be a factor of $u_\Diamond$ of least length. Let us assume that the factor $v_\Diamond$ is the prefix and the factor $r_\Diamond$ is the suffix of $k_\Diamond$. Then $p_\Diamond$(resp.$q_\Diamond$) occurs twice in $k_\Diamond$ as a prefix (resp.suffix) of each of $v_\Diamond$ and $r_\Diamond$ respectively. Since $p_\Diamond$ and $q_\Diamond$ are unioccurent in $v_\Diamond$ and $r_\Diamond$ respectively, we conclude that a factor say $l_\Diamond$ has a proper prefix (resp.suffix) starting with $v_\Diamond$(resp.$r_\Diamond$) and concluding with $r_\Diamond$(resp.$v_\Diamond$) which is a contradiction for the minimality of $k_\Diamond$. Hence $v_\Diamond = r_\Diamond$.

Conversely, to prove $u_\Diamond$ is a rich partial word, we have to verify that each prefix of $u_\Diamond$ has a unioccurrent pal_s. Consider the prefix of $u_\Diamond$ as $v_\Diamond$ and the pal_s of $u_\Diamond$ with maximum length as $q_\Diamond$. Suppose $v_\Diamond$ is palindromic then $v_\Diamond = q_\Diamond$ and thus $q_\Diamond$ is unioccurrent in $v_\Diamond$. Suppose $v_\Diamond$ is not palindromic, then let $p_\Diamond$ be the pal_p of $v_\Diamond$ with maximum length. If $q_\Diamond$ is not unioccurrent in $v_\Diamond$ then $v_\Diamond$ has a proper factor $r_\Diamond$ starting with $p_\Diamond$ and ending with $q_\Diamond$ where $p_\Diamond$ and $q_\Diamond$ are not factors of one another. Then $p_\Diamond$ is the pal_p of $r_\Diamond$ with maximum length. Similarly we can show that $q_\Diamond$ is the pal_s of $r_\Diamond$ with maximum length which contradicts our assumption. Hence $q_\Diamond$ is unioccurrent in $v_\Diamond$. □

Theorem 2. *If $u_\Diamond$ is a rich partial word over $\mathbb{A}_\Diamond$ and $u_\Diamond r_\Diamond$ has a unioccurent pal_s $q_\Diamond$ such that $r_\Diamond \in \mathbb{A}_\Diamond$ and $2|q_\Diamond| \geq |u_\Diamond r_\Diamond|$ then $u_\Diamond r_\Diamond$ is a rich partial word.*

Proof. Let us assume that $q_\Diamond$ is the pal_s of $u_\Diamond r_\Diamond$ with maximum length. Suppose $q_\Diamond$ is not unioccurent in $u_\Diamond r_\Diamond$ such as if $q_\Diamond$ has another occurrence in $u_\Diamond r_\Diamond$, then as $2|q_\Diamond| + 1 \geq |u_\Diamond r_\Diamond|$, the two occurences overlap each other or separated from each other by maximum of one letter of $\mathbb{A}_\Diamond$. Thus both the occurences form a pal_s of $u_\Diamond r_\Diamond$ such that they are strictly longer than $q_\Diamond$ which is a contradiction. Therefore $q_\Diamond$ is the unioccurent pal_s of $u_\Diamond r_\Diamond$ such that $u_\Diamond r_\Diamond$ is rich. Hence the proof. □

Theorem 3. *If the rich partial word $u_\Diamond$ over $\mathbb{A}_\Diamond$ is the product of two rich palindromic factors $p_\Diamond$ and $q_\Diamond$ and satisfies the conditions:*

$$(i)|u_\Diamond| - 4 \leq 2|q_\Diamond|$$
$$(ii)|u_\Diamond| - 4 \leq 2|p_\Diamond|,$$

then the products $p_\Diamond q_\Diamond p_\Diamond$ and $q_\Diamond p_\Diamond q_\Diamond$ are also rich partial words.

Proof. Let us prove by contradiction. Consider the rich partial word $u_\Diamond = p_\Diamond q_\Diamond$ satisfying the condition $|u_\Diamond| - 4 \leq 2|q_\Diamond|$ such that $p_\Diamond q_\Diamond p_\Diamond$ is not rich. Let $r_\Diamond \in \mathbb{A}_\Diamond$, $s_\Diamond \in \{Alph(u_\Diamond)\}$ with $r_\Diamond s_\Diamond$ as the prefix of $p_\Diamond$ of minimum length such that $p_\Diamond q_\Diamond r_\Diamond s_\Diamond$ is not rich. Let $k_\Diamond$ be the pal_s of $p_\Diamond q_\Diamond r_\Diamond s_\Diamond$ with maximum length. Then as $s_\Diamond r_\Diamond^R q_\Diamond r_\Diamond s_\Diamond$ is the suffix of $p_\Diamond q_\Diamond r_\Diamond s_\Diamond$, we have $|q_\Diamond| + 2|r_\Diamond| + 2 \leq |k_\Diamond|$ which further infers that $|u_\Diamond| \leq |u_\Diamond| + 4|r_\Diamond| \leq 2|q_\Diamond| + 4|r_\Diamond| + 4 \leq |k_\Diamond|$. Thus by Theorem 2 we get $p_\Diamond q_\Diamond r_\Diamond s_\Diamond$ to be a rich partial word which contradicts our assumption. Therefore $p_\Diamond q_\Diamond p_\Diamond$ is a rich partial word only if $u_\Diamond = p_\Diamond q_\Diamond$ is rich and $|u_\Diamond| - 4 \leq 2|q_\Diamond|$. Similarly we can prove that $q_\Diamond p_\Diamond q_\Diamond$ is a rich partial word. ☐

Example 5. *Let $u_\Diamond = p_\Diamond q_\Diamond$ over $\mathbb{A}_\Diamond = \{a, b\} \cup \{\Diamond\}$ be a rich partial word with rich palindromic factors $p_\Diamond = ab\Diamond ba$ and $q_\Diamond = b$. Also*

$$\begin{aligned}(i)|u_\Diamond| - 4 &= 2 = 2|q_\Diamond| \\ (ii)|u_\Diamond| - 4 &= 2 < 2|p_\Diamond|.\end{aligned}$$

Then the products $p_\Diamond q_\Diamond p_\Diamond = ab\Diamond babab\Diamond ba$ and $q_\Diamond p_\Diamond q_\Diamond = bab\Diamond bab$ are also rich partial words.

Theorem 4. *For any non-empty rich partial word $u_\Diamond$ over $\mathbb{A}$, if $u_\Diamond u_\Diamond \uparrow v_\Diamond u_\Diamond w_\Diamond$ for some rich partial words $v_\Diamond, w_\Diamond$ such that $v_\Diamond = \lambda$ or $w_\Diamond = \lambda$ then $u_\Diamond$ is primitive.*

Proof. Let us assume that $u_\Diamond u_\Diamond \uparrow v_\Diamond u_\Diamond w_\Diamond$ such that $v_\Diamond = \lambda$ or $w_\Diamond = \lambda$. Suppose to the contrary that $u_\Diamond$ is not primitive then a non-empty rich word x exists such that $u_\Diamond \subset x^m$ where $m \geq 2$ is an integer. But then $u_\Diamond u_\Diamond \uparrow x^{m-1} u_\Diamond x$, and using our assumption we get $x^{m-1} = \lambda$ or $x = \lambda$, a contradiction. Therefore $u_\Diamond$ is a primitive rich partial word. ☐

Example 6. *Assume the rich partial words $u_\Diamond, v_\Diamond$ and $w_\Diamond$ over $\mathbb{A}_\Diamond = \{a, b, c\} \cup \{\Diamond\}$ such that $u_\Diamond = ac\Diamond ccb$, $v_\Diamond = \lambda$ and $w_\Diamond = acc\Diamond cb$. Then $u_\Diamond$ is primitive since*

$$u_\Diamond u_\Diamond = ac\Diamond ccbac\Diamond ccb \uparrow ac\Diamond ccbacc\Diamond cb = xuy.$$

Theorem 5. *Let $u_\Diamond$ and $v_\Diamond$ be non-empty rich partial words. If $u_\Diamond$ and $v_\Diamond$ are conjugate, then a rich partial word $w_\Diamond$ exists such that $u_\Diamond w_\Diamond \uparrow w_\Diamond u_\Diamond$. Also there exist rich partial words $p_\Diamond, q_\Diamond$ such that $u_\Diamond \subset p_\Diamond q_\Diamond$, $v_\Diamond \subset q_\Diamond p_\Diamond$ and $w_\Diamond \subset p_\Diamond (q_\Diamond p_\Diamond)^m$ for some $m \geq 1$.*

Proof. Let $u_\Diamond$ and $v_\Diamond$ be non-empty rich partial words. Suppose that $u_\Diamond$ and $v_\Diamond$ are conjugate and let $p_\Diamond, q_\Diamond$ be rich partial words such that $u_\Diamond \subset p_\Diamond q_\Diamond$ and $v_\Diamond \subset q_\Diamond p_\Diamond$. Then $u_\Diamond p_\Diamond \subset p_\Diamond q_\Diamond p_\Diamond$ and $p_\Diamond v_\Diamond \subset p_\Diamond q_\Diamond p_\Diamond$ and so for $w_\Diamond = p_\Diamond$ we have $u_\Diamond w_\Diamond \uparrow w_\Diamond u_\Diamond$.

☐

Theorem 6. *Let $u_\Diamond$ be a rich partial word over $\mathbb{A}_\Diamond$ and let x and y be two rich words over $\mathbb{A}$. If $u_\Diamond \subset xy$ and $u_\Diamond \subset yx$ then $xy = yx$.*

Proof. To prove the theorem, we consider $|x| \leq |y|$. Let $y = x'y'$ such that $|x'| = |x|$ where x' and y' are rich words. Also let $u_\Diamond = v_\Diamond w_\Diamond$ with $|x| = |v_\Diamond|$ where $v_\Diamond, w_\Diamond$ are rich partial words. Since $u_\Diamond \subset xy$, we have $v_\Diamond w_\Diamond \subset xy$ such that we get $v_\Diamond \subset x$ and $w_\Diamond \subset y$. Likewise $u_\Diamond \subset yx$ implies that $v_\Diamond w_\Diamond \subset yx$ which further implies that $v_\Diamond w_\Diamond \subset x'y'x$ such that we get $v_\Diamond \subset x'$ and $w_\Diamond \subset y'x$. Since $u_\Diamond = v_\Diamond w_\Diamond$ is a rich partial word, it has exactly one hole. Then the following two cases arises:
$Case(i)$: If $v_\Diamond$ is a rich partial word with zero hole and $w_\Diamond$ is a rich partial word with one hole. Then $v_\Diamond = x' = x$ and $w_\Diamond \subset y'x$. Also $w_\Diamond \subset y = x'y' = xy'$. Hence by induction process, $xy' = y'x$ which follows that $xy = yx$.
$Case(ii)$: If $v_\Diamond$ is a rich partial word with one hole and $w_\Diamond$ is a rich partial word with zero hole. Then $v_\Diamond \subset x' = x$ and $w_\Diamond = y'x = x'y' = y$. Then there exists two rich words p and q such that $x = pq, x' = qp$ and $y' = (qp)^m q$ for $m \geq 0$ where x and y' are conjugates to each other. Hence by induction process, $pq = qp$ which follows that $xy = yx$ since $v_\Diamond \subset pq$ and $v_\Diamond \subset qp$. ☐

3.1 Rich Palindromic Partial Words

A rich partial word is closed by factors and also under the operations of reversal and palindromic closures. Palindromic partial words help in encoding and decoding the information contained in DNA strands. The palindromic defect of rich partial words is zero; Most of the rich partial words are also palindromic which is not a necessary condition. In this section, the rich palindromic partial words are to be analyzed and examined to find the periodicity of possible elements in the $\Diamond$ positions of the partial word sequence.

Definition 3. *Rich partial words that are also palindromic are termed as rich palindromic partial words.*

Example 7. *Assume a partial word $u_\Diamond = aba\Diamond aba$ with $|u_\Diamond| = 5$ over $\mathbb{A}_\Diamond = \{a, b\} \cup \{\Diamond\}$. Here $u_\Diamond \uparrow u_\Diamond^R$ and so $u_\Diamond$ is a palindromic partial word. The set of all distinct partial palindromic proper factors of $u_\Diamond$ are*

$$\{ba\Diamond ab, a\Diamond aba, aba\Diamond a, ba\Diamond, a\Diamond a, a\Diamond, \Diamond a, \Diamond ab\}.$$

Here the number of distinct partial palindromic proper factors are more than $|u_\Diamond|$. Hence $u_\Diamond$ is a rich palindromic partial word.

Theorem 7. *Let $u_\Diamond$ be a rich palindromic partial word. Then $u_\Diamond u_\Diamond^R$ and $u_\Diamond^R u_\Diamond$ are periodic partial words.*

Proof. Let the rich partial word $u_\Diamond \in \mathbb{A}_\Diamond^+$ be a palindrome. We have $u_\Diamond \uparrow u_\Diamond^R$. By the notion of compatibility, a total word x exists such that $u_\Diamond \subset x$ and $u_\Diamond^R \subset x$. Hence by the law of multiplication, $u_\Diamond u_\Diamond^R \subset x.x = x^2$. Thus $u_\Diamond u_\Diamond^R$ is periodic. Similarly it is easy to follow that $u_\Diamond^R u_\Diamond$ is a periodic partial word. ☐

Theorem 8. *For a rich partial word $u_\Diamond \in \mathbb{A}_\Diamond$, if $u_\Diamond^m$ is a rich palindromic partial word for $m > 0$ then $u_\Diamond$ is a rich palindromic partial word.*

Proof. We prove it by induction hypothesis. For $m = 1$, the assertion is accurate.. Let us assume that it is true for all $l < m$ that is if $u_\Diamond^l$ is a palindrome for all $l \leq m-1$, then $u_\Diamond$ is a palindrome. Now to prove it for m, assume that $u_\Diamond^m$ is a palindrome. We can write

$$u_\Diamond^m = u_\Diamond^{m-1} u_\Diamond = u_\Diamond^r u_\Diamond^{m-1}.$$

Now

$$u_\Diamond^m = u_\Diamond u_\Diamond^{m-1} \uparrow (u_\Diamond^m)^R = (u_\Diamond^R)^m = (u_\Diamond^R)(u_\Diamond^R)^{m-1}.$$

As $|u_\Diamond| = |u_\Diamond^R|$ and $u_\Diamond^{m-1} \uparrow (u_\Diamond^R)^{m-1}$, then by simplification we have $u_\Diamond \uparrow u_\Diamond^R$. Hence $u_\Diamond$ is a palindrome. □

4 Partial Palindromic Proper Subwords of Rich Partial Words

The study of subsequences (or subwords) in partial words involves a number of combinatorial complexities. One of them is the detection of palindromes, or subwords that are symmetric when reversed, in partial words. Since the last two decades, there has been interest in researching the characteristics of palindromes. In this section, we prove that the maximum number of partial palindromic perfect subwords in a partial word relies on both the length and the number of distinct letters in the partial word.

Definition 4. *A partial palindromic proper subword (or partial palindromic scattered proper subword) of a partial word $u_\Diamond$ over the alphabet $\mathbb{A}_\Diamond$ is a sequence that can be derived by deleting zero or more letters from it without altering the order of the remaining letters. The set of all non-empty partial palindromic proper subwords of $u_\Diamond$ is denoted by $PPPS(u_\Diamond)$.*

Example 8. *Consider a partial word $u_\Diamond = aab\Diamond ba$ over $\mathbb{A}_\Diamond = \{a, b\} \cup \{\Diamond\}$. The set of all distinct palindromic proper subwords of $u_\Diamond$ are*

$$\{aa, bb, a\Diamond, \Diamond a, b\Diamond, \Diamond b, aa\Diamond, aaa, a\Diamond a, ab\Diamond, b\Diamond b, \Diamond ba, aa\Diamond a, abba, a\Diamond ba, ab\Diamond ba, aab\Diamond a\}.$$

Among the above set, the set of all distinct partial palindromic proper subwords of $u_\Diamond$ are

$$\{a\Diamond, \Diamond a, b\Diamond, \Diamond b, aa\Diamond, a\Diamond a, ab\Diamond, b\Diamond b, \Diamond ba, aa\Diamond a, a\Diamond ba, ab\Diamond ba, aab\Diamond a\}.$$

Theorem 9. *Any rich partial word of length n with no three consecutive similar letters over $\mathbb{A}_\Diamond = \{a, b\} \cup \{\Diamond\}$ has a partial palindromic proper subword of length at least $\frac{2}{3}(n-2)$.*

Proof. Consider a rich partial word $u_\Diamond = u_\Diamond[1 \dots t]$ with no three consecutive similar letters over $\mathbb{A}_\Diamond = \{a, b\} \cup \{\Diamond\}$. Let each $u_\Diamond[i]$ be made up of only as or only bs and let two consecutive partial words $u_\Diamond[j]$ and $u_\Diamond[j+1]$ consist of different letters. Then we have length of each $u_\Diamond[j]$ as atmost 2. Now assume t to be even. Then at least one letter from each pair $u_\Diamond[j], u_\Diamond[t-j+1]$ together with all the letters from $u_\Diamond[\frac{t+1}{2}]$ results in a partial palindromic subword. Thus we get a partial palindromic subword of length at least $\frac{2}{3}(n-2)$. Hence the proof. □

Theorem 10. *For a given rich partial word $u_\Diamond$, $|u_\Diamond| \leq |PPPF(u_\Diamond)| \leq |PPPS(u_\Diamond)|$.*

Proof. It is clear from the notion of rich partial words that $|u_\Diamond| \leq |PPPF(u_\Diamond|$. Let $u_\Diamond = u_\Diamond[1 \dots n]$ be a rich partial word where $u_\Diamond[i] \in \mathbb{A}_\Diamond$. Let the prefix of length t of $u_\Diamond$ be $v_\Diamond = u_\Diamond[1 \dots t]$. We observe that on the concatenation of each $u_\Diamond[i]$ to $v_\Diamond[i-1]$, an additional partial palindromic perfect subword $u_\Diamond^s[i]$, where $s = |v_\Diamond[i]|_{u_\Diamond[i]}$ is formed. Hence, at least one partial palindromic perfect subword is always formed on the concatenation of each letter of $u_\Diamond$. Thus $|PPPS(u_\Diamond)| \geq |u_\Diamond|$. Also the set of all partial palindromic perfect factors is a subset of the set of all partial palindromic perfect subwords of $u_\Diamond$. Therefore $|PPPF(u_\Diamond)| \leq |PPPS(u_\Diamond)|$. Hence the result. □

5 Conclusion

In this paper we introduced rich partial words and studied the combinatorial properties. We also discussed the relation between partial palindromic perfect factors and partial palindromic perfect subwords of rich partial words.

References

[1] Aldo de Luca, On the Combinatorics of Finite Words, *Theor. Comput. Sci.*, vol. 218(1), pp. 13–39, 1999.

[2] Aldo de Luca and Amy Glen and Luca Q. Zamboni, Rich, Sturmian, and trapezoidal words, *Theor. Comput. Sci.*,vol. 407, pp. 569–573, 2008.

[3] Amy Glen and Jacques Justin and Steve Widmer and Luca Q. Zamboni, Palindromic richness, *Eur. J. Comb.*, vol. 30(2), pp. 510–531, 2009.

[4] F. Blanchet-Sadri, A Periodicity Result of Partial Words with One Hole, *Computers and Mathematics with Applications*, vol. 46, pp. 813–820, 2003.

[5] F. Blanchet-Sadri, Primitive partial words, *Discret. Appl. Math.*, vol. 148(3), pp. 195–213, 2005.

[6] F. Blanchet-Sadri, Algorithmic Combinatorics on Partial Words, *Discrete mathematics and its applications*, CRC Press, 2008.

[7] Groult. R, Prieur. E, Richomme. G, Counting distinct palindromes in a word in linear time, *Inf. Process. Lett.*, vol. 110 (20), pp. 908–912, 2010.

[8] 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, pp. 137–146, 2009.

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

[10] Jean Berstel and Luc Boasson, Partial Words and a Theorem of Fine and Wilf, *Theor. Comput. Sci.*, vol. 218(1), pp. 135–141, 1999.

[11] Jetro Vesti, Widespread and nonrandom distribution of DNA palindromes in cancer cells provides a structural platform for subsequent gene amplification, *Nature Genetics*, vol. 37(3), pp. 320–327, 2005.

[12] Jetro Vesti, Extensions of rich words, *Theor. Comput. Sci.*, vol. 548, pp. 14–24, 2014.

[13] Luke Schaeffer and Jeffrey O. Shallit, Closed, Palindromic, Rich, Privileged, Trapezoidal, and Balanced Words in Automatic Sequences, *Electron. J. Comb.*, 23(1), pp. 1–25, 2016.

[14] Michelangelo Bucci and Alessandro De Luca and Amy Glen and Luca Q. Zamboni, A connection between palindromic and factor complexity using return words, *Adv. Appl. Math.*, vol. 42(1), pp. 60–74, 2009.

[15] Pelantová. E and Starosta. Š, On Words with the Zero Palindromic Defect, Combinatorics on Words, *WORDS 2017. Lecture Notes in Computer Science, Springer*, vol. 10432, pp. 59–71, 2017.

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

[17] Stéphane Fischler, Palindromic prefixes and episturmian words, *J. Comb. Theory, Ser. A*, vol. 113(7), pp. 1281–1304, 2006.

[18] Xavier Droubay and Jacques Justin and Giuseppe Pirillo, Episturmian words and some constructions of de Luca and Rauzy, *Theor. Comput. Sci.*,vol. 255, pp. 539–553, 2001.