An Improved Bound for an Extension of Fine and Wilf's Theorem and Its Optimality

Lila Kari and Shinnosuke Seki

August 23, 2010

Abstract

Considering two DNA molecules which are Watson-Crick (WK) complementary to each other "equivalent" with respect to the information they encode enables us to extend the classical notions of repetition, period, and power. WK-complementarity has been modelled mathematically by an antimorphic involution θ, i.e., a function θ such that $\theta(xy) = \theta(y)\theta(x)$ for any $x, y \in \Sigma^*$, and θ^2 is the identity. The WK-complementarity being thus modelled, any word which is a repetition of u and $\theta(u)$ such as uu, $u\theta(u)u$, and $u\theta(u)\theta(u)\theta(u)$ can be regarded repetitive in this sense, and hence, called a θ-power of u. Taking the notion of θ-power into account, the Fine and Wilf's theorem was extended as "given an antimorphic involution θ and words u, v, if a θ-power of u and a θ-power of v have a common prefix of length at least $b(|u|, |v|) = 2|u| + |v| - \gcd(|u|, |v|)$, then u and v are θ-powers of a same word." In this paper, we obtain an improved bound $b'(|u|, |v|) = b(|u|, |v|) - \lfloor \gcd(|u|, |v|)/2 \rfloor$. Then we show all the cases when this bound is optimal by providing all the pairs of words (u, v) such that they are not θ-powers of a same word, but one can construct a θ-power of u and a θ-power of v whose maximal common prefix is of length equal to $b'(|u|, |v|) - 1$. Furthermore, we characterize such words in terms of Sturmian words.

1 Introduction

This paper investigates an extension of Fine and Wilf's theorem in combinatorics of words. Recall that a positive integer p is called a *period* of a word w if the i-th and the $(i+p)$-th letters of w are the same for any $1 \leq i \leq |w| - p$. Fine and Wilf's theorem [12] states that if a word has two periods p, q and is of length at least $p + q - \gcd(p, q)$, then $\gcd(p, q)$ is also its period, where gcd denotes the greatest common divisor. A concise method to prove this result, [5], also proves that the lower bound is "strongly optimal" in the following sense, which was defined in [6], that for *any* pair (p, q) of integers with $p > q > \gcd(p, q)$, one can construct a word of length $p + q - \gcd(p, q) - 1$, with p and q as periods, but without $\gcd(p, q)$ as period (the set of all such words with p and q being coprime is denoted by PER). This theorem has several extensions: e.g., considering more than two periods [3], [4], [6], [13], based on abelian periods [7], for partial or bidimensional words [1], [2], [15].

Changing the focus from integers to words, this theorem can be reformulated as follows: "Given words u, v, if a power of u and a power of v have a common prefix of length at least $|u| + |v| - \gcd(|u|, |v|)$, then u and v are powers of a common word, i.e., they share their primitive root." This result was recently extended in [9], by generalizing the notion of power of a word as inspired by the characteristics of DNA-encoded information. Briefly, a DNA strand can be abstracted as a word over the four-letter alphabet $\{\mathtt{A}, \mathtt{C}, \mathtt{G}, \mathtt{T}\}$. Due to the so-called Watson-Crick (WK) complementarity A-T and C-G, two complementary DNA single strands with *opposite* orientations bind to each other to form the structure known as a DNA double strand. WK-complementarity has been modelled mathematically by an antimorphic involution θ, i.e., a function θ such that $\theta(xy) = \theta(y)\theta(x)$ for any $x, y \in \Sigma^*$ (antimorphism), and θ^2 is the identity (involution). An antimorphic involution captures the main features of WK-complementarity, namely that the WK-complement of a DNA single strand is the reverse (antimorphic property) complement (involution property) of the given strand. If we set the antimorphic involution on the four-letter DNA alphabet defined by $\theta(\mathtt{A}) = \mathtt{T}$ and $\theta(\mathtt{C}) = \mathtt{G}$, then for any word $w \in \{\mathtt{A}, \mathtt{C}, \mathtt{G}, \mathtt{T}\}^*$ representing a DNA single strand, the word $\theta(w)$ will represent its WK-complement. For example, using θ, we can calculate the WK-complement of AAC as $\theta(\mathtt{AAC}) = \theta(\mathtt{C})\theta(\mathtt{AA}) = \theta(\mathtt{C})\theta(\mathtt{A})\theta(\mathtt{A}) = \mathtt{GTT}$. We can say that two complementary DNA single strands are equivalent because one can be obtained from the other by θ. Based on this idea, for instance, the strand AACGTTGTT becomes a "power" of AAC because it consists of AAC followed by its WK-complement GTT twice. By using an antimorphic involution θ as a model of the WK-complementarity, a word in $u\{u, \theta(u)\}^*$ is called a *θ-power* of u [9]. With this extended notion of power, the Fine and Wilf's theorem was extended in [9] in the following way: "Given an antimorphic involution θ over an alphabet Σ, and given non-empty words u, v over Σ of lengths p, q with $p > q$, if a θ-power of u and a θ-power of v share a prefix of length at least $b(p, q) = 2p + q - \gcd(p, q)$, then u and v are θ-powers of a common word (in such case, we say that u and v share their *θ-primitive root*)." In [9] some examples of words u, v were provided with the property that such a common prefix of length $b(p, q) - 1$ is too short to force u and v to have the same θ-primitive root. However, these examples do not answer the question of whether $b(p, q)$ is strongly optimal or not, i.e., whether for *any* (p, q), we can find two words u, v of length p, q with different θ-primitive roots such that a θ-power of u and a θ-power of v share a prefix of length $b(p, q) - 1$.

The first contribution of this paper is to give the extended Fine and Wilf's theorem an improved bound $b'(p, q) = b(p, q) - \lfloor \gcd(p, q)/2 \rfloor$ in a constructive manner, which amounts to the negative answer to the above question. Specifically speaking, we design a pair (u, v) of words of lengths p, q with distinct θ-primitive roots in such a manner that one can construct a θ-power of u and a θ-power of v such that their common prefix is as long as possible relative to p and q. We prove that such a common prefix is of length at most $b'(p, q) - 1$, and hence, $b'(p,q)$ becomes the improved bound (Theorem 8). We call such a common prefix of length exactly $b'(p,q)-1$ a *boundary common prefix based on u and v*. Being constructive, our proof simultaneously characterizes the set of all pairs of words with distinct θ-primitive roots based on which one can construct a boundary common prefix. This characterization is the main contribution of this paper. Two corollaries of interest follow: First, there are (infinitely many) pairs of integers (p,q) such that there does not exist any boundary common prefix based on words of respective lengths p,q (Corollary 3), and hence, $b'(p,q)$ is not strongly optimal. Second, all the boundary common prefixes are homomorphic images of boundary common prefixes based on some binary words of coprime lengths. This is very similar to the fact that the words which verify the strong optimality of the bound for the Fine and Wilf's theorem are homomorphic images of a (binary) word in PER. de Luca and Mignosi in [11] proved that a word in PER is a finite Sturmian word, or more strongly, the set of all factors of words in PER is equal to the set of all finite Sturmian words. We will show that boundary common prefixes based on words of coprime lengths are also finite Sturmian words, but there exists a finite Sturmian word which never appears as a factor of such boundary common prefixes.

This paper is organized as follows: Section 2 introduces basic notions and notation as well as some known results used for our discussion. That is followed by the constructive proof of the improved bound $b'(p,q)$ in Section 3 with a few results stating that this bound is not strongly optimal. In Section 4, the relationship between boundary common prefixes and finite Sturmian words is discussed. Section 5 concludes this paper with some future directions of research.

2 Preliminaries

Let Σ be a finite alphabet containing at least two letters. Throughout this paper, elements of Σ (letters) will be denoted by a, b. By Σ^* we denote the set of all finite words over Σ. The empty word is denoted by λ and let $\Sigma^+ = \Sigma^* \setminus \{\lambda\}$. The *length* of a word $w \in \Sigma^*$ is denoted by $|w|$. For a set $X \subseteq \Sigma^+$, $X^* = \{x_1 x_2 \cdots x_n \mid x_i \in X \text{ for all } 1 \le i \le n\}$, and $X^+ = X^* \setminus \{\lambda\}$. For a word $w \in \Sigma^*$, a word $x \in \Sigma^*$ is called a *prefix* (*suffix*) of w if $w = xr$ (resp. $w = rx$) for some $r \in \Sigma^*$. Let $\mathrm{Pref}(w)$ and $\mathrm{Suff}(w)$ be the sets of all prefixes of w and of all suffixes of w, respectively. Also let $\mathrm{pref}_n(w)$ denote the prefix of w of length n. If $w = rxt$ for some $r, t \in \Sigma^*$, then x is called an *infix* of w, and if furthermore $r, t \neq \lambda$, x is called a *proper infix* of w. For $x, y \in \Sigma^*$, we denote by $x \wedge y$ the *maximal common prefix* of x and y.

A non-empty word $w \in \Sigma^+$ is said to be *primitive* if it cannot be written as a power of another word; that is, if $w = t^n$, then $n = 1$ and $w = t$. For a non-empty word $w \in \Sigma^+$, the shortest word $t \in \Sigma^+$ such that $w = t^n$ for some $n \ge 1$ is called the *primitive root* of w and is denoted by $\rho(w)$. With respect to the primitive root and the maximal common prefix, there is a result from [5] shown in a form that will be utilized in this paper.

Proposition 1 ([5]). *Let $X = \{r,t\} \subseteq \Sigma^+$, $x \in rX^*$, and $y \in tX^*$. If $|x \wedge y| \geq |rt|$, then $\rho(r) = \rho(t)$.*

A mapping $\theta : \Sigma^* \to \Sigma^*$ is called an *antimorphism* if for any words $x, y \in \Sigma^*$, $\theta(xy) = \theta(y)\theta(x)$; an *involution* if θ^2 is the identity function. Throughout this paper, θ is assumed to be an antimorphic involution on Σ unless otherwise noted explicitly. The mirror image (or *mirror involution*), which maps a word to its reverse, is a typical antimorphic involution. A word $w \in \Sigma^*$ is called a *θ-palindrome* if $w = \theta(w)$, see [10]. The next two lemmas on θ-palindromes play significant roles in this paper.

Lemma 1. *For θ-palindromes $x, y \in \Sigma^*$ of the same length d, if $\mathrm{pref}_{\lceil d/2 \rceil}(x) = \mathrm{pref}_{\lceil d/2 \rceil}(y)$, then $x = y$.*

Lemma 2. *Let $x, y \in \Sigma^+$ be two θ-palindromes with $d = \gcd(|x|, |y|)$ and $|x| + |y| \geq 3d$. For any integer $i \geq 1$, if $|xy \wedge y^i x| \geq |xy| - 2d$, then $\rho(x) = \rho(y)$.*

Proof. The first case is when $|x| = d$. Due to the hypothesis on $|x| + |y|$, in this case we have $|y| \geq 2d$. Then the overlap between xy and $y^i x$ implies that y begins with x. If $|y| = 2d$, then $y = x^2$ due to $x = \theta(x)$ and $y = \theta(y)$; otherwise ($|y| \geq 3d$), the overlap implies $|xy \wedge y| \geq 2d$, and hence, $x^2 \in \mathrm{Pref}(y)$. Because of $x = \theta(x)$ and $y = \theta(y)$, y has x^2 also as its suffix. Combining these together yields $xy = yx$. Using Proposition 1, we get $\rho(x) = \rho(y)$.

The second case is when $|x| \geq 2d$ and $|y| = d$. Let $x_p = \mathrm{pref}_{|x|-d}(x)$. Under this length condition, the overlap between xy and $y^i x$ implies that $x_p \in \mathrm{Pref}(y^i x_p)$. Since the length of x_p is a multiple of d, this means that x_p is a power of y and y is a prefix of x_p, i.e., $y \in \mathrm{Pref}(x)$. This is equal to $y \in \mathrm{Suff}(x)$ and actually now we have that x is a power of y.

The last case is when $|x|, |y| \geq 2d$. In this case, the overlap gives $x \in \mathrm{Pref}(y^i x)$ and $y \in \mathrm{Pref}(xy)$. So, if $|y| \geq |x|$, then the latter prefix relation implies that $x \in \mathrm{Pref}(y)$, which is equivalent to $x \in \mathrm{Suff}(y)$. With $y \in \mathrm{Pref}(xy)$, this implies that $xy = yx$ so that $\rho(x) = \rho(y)$. Conversely, if $|y| < |x|$, then according to $x \in \mathrm{Pref}(y^i x)$, we can let $x = y^j y_p$ for some $j \geq 1$ and $y_p \in \mathrm{Pref}(y)$. Since x and y are θ-palindromes, $x = y^j y_p = \theta(y_p) y^j$ holds. This equality gives $y_p = \theta(y_p)$, and hence, imposes $\rho(y_p) = \rho(y) = \rho(x)$ due to Proposition 1. □

In [9], a special class of primitive words was proposed that takes into account the notion of antimorphic involution. For a non-empty word $t \in \Sigma^+$, we call a word in $t\{t, \theta(t)\}^*$ a *θ-power of t*. A non-empty word $w \in \Sigma^+$ is said to be *θ-primitive* if it cannot be written as a θ-power of another word, that is, for $t \in \Sigma^+$, $w \in t\{t, \theta(t)\}^*$ implies $w = t$. The *θ-primitive root* of w, denoted by $\rho_\theta(w)$, is the θ-primitive word t such that $w \in t\{t, \theta(t)\}^*$. The uniqueness of θ-primitive root was proved in [9] using Theorem 3 in Section 3.

Lemma 3 ([9]). *Let $w \in \Sigma^+$ be a θ-primitive word and $w_1, w_2, w_3, w_4 \in \{w, \theta(w)\}$. If $w_1 w_2 x = y w_3 w_4$ holds for some non-empty words $x, y \in \Sigma^+$ with $|x|, |y| < |w|$, then $w_2 \neq w_3$.*

From this lemma, the next theorem easily follows. This is an analogous result to the one stating that a primitive word cannot be a proper infix of its square.

Theorem 1 ([14]). *For a θ-primitive word $w \in \Sigma^+$, neither $w\theta(w)$ nor $\theta(w)w$ can be a proper infix of a word in $\{w, \theta(w)\}^3$.*

3 An Improved Bound for the Extension of Fine and Wilf's Theorem

Taking the θ-primitivity into account, an extension of the Fine and Wilf's theorem was proposed in [9], of the following two forms:

Theorem 2 ([9]). *For $u, v \in \Sigma^+$ with $|u| \geq |v|$, if a θ-power of u and a θ-power of v share a common prefix of length at least $2|u| + |v| - \gcd(|u|, |v|)$, then $\rho_\theta(u) = \rho_\theta(v)$, i.e., there exists a θ-primitive word $t \in \Sigma^+$ such that $u, v \in t\{t, \theta(t)\}^*$.*

Theorem 3 ([9]). *For $u, v \in \Sigma^+$, if a θ-power of u and a θ-power of v share a common prefix of length at least $\mathrm{lcm}(|u|, |v|)$, then $\rho_\theta(u) = \rho_\theta(v)$, where $\mathrm{lcm}(|u|, |v|)$ denotes the least common multiple of $|u|$ and $|v|$.*

These theorems give two bounds, and one can be larger than the other depending on the value of $\gcd(|u|, |v|)$ as $\mathrm{lcm}(|u|, |v|) < 2|u| + |v| - \gcd(|u|, |v|)$ if and only if $|v| \leq 2\gcd(|u|, |v|)$. Thus, for integers p, q with $p \geq q$, by letting

$$b(p, q) = \begin{cases} \mathrm{lcm}(p, q) & \text{if } q \leq 2\gcd(p, q); \\ 2p + q - \gcd(p, q) & \text{if } q \geq 3\gcd(p, q), \end{cases} \tag{1}$$

one can merge Theorems 2 and 3 into one theorem as follows.

Theorem 4. *For $u, v \in \Sigma^+$ with $|u| \geq |v|$, if a θ-power of u and a θ-power of v share a common prefix of length at least $b(|u|, |v|)$, then $\rho_\theta(u) = \rho_\theta(v)$.*

This theorem indicates the possibility of constructing two words u, v with $|u| > |v|$ such that a θ-power of u and a θ-power of v have a common prefix of length $b(|u|, |v|) - 1$, while at the same time $\rho_\theta(u) \neq \rho_\theta(v)$. Here we provide two of such examples, which were introduced in [9].

Example 1. Let $\theta : \{a, b\}^* \to \{a, b\}^*$ be the mirror involution, $u = a^2ba^3b$, and $v = a^2ba$. Then, u^3 and $v^2\theta(v)^2v$ have a common prefix of length $2|u| + |v| - \gcd(|u|, |v|) - 1$, but $\rho_\theta(u) \neq \rho_\theta(v)$. Figure 1 is a visualization of this example.

Example 2. Let $\theta : \{a, b\}^* \to \{a, b\}^*$ be the mirror involution, $u = ba^2baba$, and $v = ba^2ba$. Then $u\theta(u)^2$ and v^4 have a common prefix of length $2|u| + |v| - \gcd(|u|, |v|) - 1$, but $\rho_\theta(u) \neq \rho_\theta(v)$.

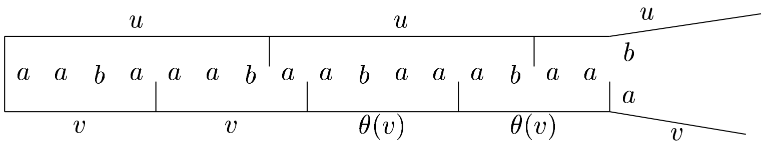

Figure 1: Even from words with distinct θ-primitive roots, it is possible to construct θ-powers whose maximal common prefix is shorter by 1 than the bound given in Theorem 4.

In [6], a sharp distinction was made between a "good" bound and an "optimal" bound. Following this distinction, we define the optimality of a bound in the context of the extended Fine and Wilf's theorem. For a pair of integers (p, q) with $p > q \geq 2\gcd(p, q)$[1], an integer k is called a *good bound for* (p, q) if for any antimorphic involution θ and for any words $u, v \in \Sigma^+$ with $|u| = p$ and $|v| = q$, once there exist a θ-power of u and a θ-power of v which share a prefix of length at least k, one has $\rho_\theta(u) = \rho_\theta(v)$. Based on this, k is an *optimal bound for* (p, q) if it is a good bound for (p, q) whereas $k - 1$ is not; i.e., there exist an antimorphic involution θ and words u, v of length p and q with $\rho_\theta(u) \neq \rho_\theta(v)$ from which one can construct a θ-power of u and θ-power of v whose maximal common prefix is of length $k - 1$. A bound $b(\cdot, \cdot)$ of two variables is said to be *strongly optimal* if for any (p, q) satisfying the inequality mentioned previously, $b(p, q)$ is optimal. Although the goodness, optimality, and strong optimality are defined here for the extended Fine and Wilf's theorem, these notions can be defined for any variant of this theorem.

Examples 1 and 2 prove the optimality of $b(p, q)$ for (p, q) equal to $(7, 4)$ and $(7, 5)$, respectively. The bound given by the Fine and Wilf's theorem is known to be strongly optimal (see [5]). A question, therefore, arises of whether $b(p, q)$ is *strongly optimal* or not. We will show that $b(p, q)$ is not strongly optimal by proving that $b'(p, q) = b(p, q) - \lfloor \gcd(p, q)/2 \rfloor$ is still a good bound, strictly smaller than $b(p, q)$ unless p and q are coprime.

Unlike the proof of Theorem 4 in [9], our proof in the following is constructive. More concretely speaking, we will search for words u and v based on which one can build a *boundary common prefix.* For words $u, v \in \Sigma^+$ with $|u| > |v|$ and $\rho_\theta(u) \neq \rho_\theta(v)$, we call a word $w \in \Sigma^+$ a *boundary common prefix based on* u *and* v if there exist a θ-power of u and a θ-power of v whose maximal common prefix is w and of length *at least* $b'(|u|, |v|) - 1$. By $\mathrm{BCP}_\theta(u, v)$, we denote the set of all boundary common prefixes based on u and v. Figure 1 illustrates a boundary common prefix $a^2ba^3ba^2ba^3ba^2$ based on the specific u and v given in Example 1. What we actually prove in the following is that the length of boundary common prefixes based on u and v is *exactly* $b'(|u|, |v|) - 1$.

[1] The first inequality can be assumed without loss of generality. The second one is reasonable in the context of Fine and Wilf's theorem because $q = \gcd(p, q)$ means that p is a multiple of q, and hence, the period p is not essential. Whenever we refer to p, q from now on, we implicitly assume that the inequality $p > q \geq 2\gcd(p, q)$ holds.

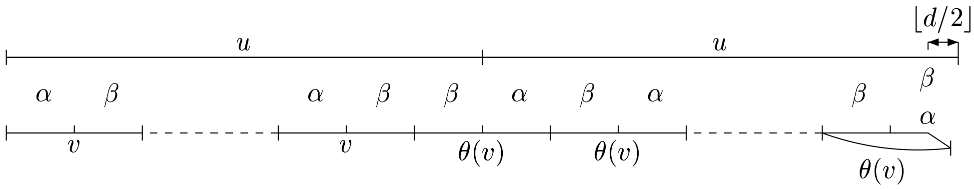

Figure 2: Two words u^2 and $v^{(n-1)/2}\theta(v)^{(n-1)/2+1}$ share a prefix of length $2|u| - \lfloor d/2 \rfloor$.

As shown in Eq.(1), $b(p,q)$ displays different behaviours depending on whether $q \leq 2\gcd(p,q)$ or not, and hence, so does $b'(p,q)$. As such, we will prove that $b'(p,q)$ is good for (p,q) with $p > q = 2\gcd(p,q)$ in Section 3.1, and for (p,q) with $p > q \geq 3\gcd(p,q)$ in Section 3.2. Note that we do not have to consider any (p,q) with $p > q = \gcd(p,q)$ as mentioned previously. In Section 3.3, we will combine these two results together to conclude that $b'(p,q)$ is good for any (p,q).

3.1 The case when $q = 2\gcd(p,q)$

Firstly we handle the case $q = 2\gcd(p,q)$ in Proposition 2. Its proof will suggest a construction of examples which verify the optimality of the new bound $b'(p,q)$ for any pair of integers (p,q) satisfying this condition.

Proposition 2. *Let $u, v \in \Sigma^+$ with $|u| > |v|$ and $|v| = 2\gcd(|u|,|v|)$. If a θ-power of u and a θ-power of v share a prefix of length $2|u| - \lfloor \gcd(|u|,|v|)/2 \rfloor$, then $\rho_\theta(u) = \rho_\theta(v)$.*

Proof. Let $d = \gcd(|u|,|v|)$. The length condition on $|u|$ and $|v|$ is equivalent to that $2|u| = n|v|$ holds for some odd integer $n \geq 3$. Let us translate the problem setting as: u_1u_2 and $v_1v_2\cdots v_n$ agree with each other up to their first $2|u| - \lfloor d/2 \rfloor$ letters, where $u_1 = u$, $u_2 \in \{u, \theta(u)\}$, $v_1 = v$, and $v_2, \ldots, v_n \in \{v, \theta(v)\}$ (see Figure 2). One can regard u_1 as a catenation of n words or '*blocks*' $w_1, w_2, \ldots, w_n$ of length d. In the similar fashion, one can let $u_2 = w_{n+1}\cdots w_{2n}$ for some words $w_{n+1}, \ldots, w_{2n}$ of length d. Then $v_i = w_{2i-1}w_{2i}$ holds for any i up to $n-1$. As for v_n, we can let $v_n = w_{2n-1}\mathrm{pref}_{\lceil d/2 \rceil}(w_{2n})x$ for some word x of length $\lfloor d/2 \rfloor$.

It is clear that when u_2 is $\theta(u)$, $v_{(n+1)/2} = w_nw_{n+1}$ becomes a θ-palindrome ($v = \theta(v)$) because it is located at the center of u_1u_2. Hence, $w_{n+1} = \theta(w_n)$, i.e., $v = \theta(v) = w_n\theta(w_n)$, and $u = v^{(n-1)/2}w_n$. These mean that $u, v \in w_n\{w_n, \theta(w_n)\}^*$ so that $\rho_\theta(u) = \rho_\theta(v)$.

Let us consider the other case when u_2 is u. Let $w_1 = \alpha$ and $w_2 = \beta$. Since $u_2 = u$ begins with $\alpha\beta$, $w_{n+1} = \alpha$ and $w_{n+2} = \beta$. If either $v_{(n+1)/2} = w_nw_{n+1}$ or $v_{(n+1)/2+1} = w_{n+2}w_{n+3}$ is v, then α overlaps with β and results in that $\alpha = \beta$. As a result, $u, v \in \alpha\{\alpha, \theta(\alpha)\}^*$, i.e., $\rho_\theta(u) = \rho_\theta(v)$. If neither holds, then we obtain $\alpha = \theta(\alpha)$, $\beta = \theta(\beta)$, and $w_n = \beta$; furthermore if $n + 3 \neq 2n$, then $w_{n+3} = \alpha$. According to the same argument, we can figure out that unless

$v_2 = \cdots = v_{(n+1)/2-1} = v$ and $v_{(n+1)/2} = \cdots = v_{n-1} = v_n = \theta(v)$, one has $\rho_\theta(u) = \rho_\theta(v)$. In this only one remaining case (illustrated in Figure 2), $w_{2n} = \beta$ and $\mathrm{pref}_{\lceil d/2 \rceil}(w_{2n})x = \alpha$. Thus, $\alpha = \beta$ due to Lemma 1, and hence, $\rho_\theta(u) = \rho_\theta(v)$. $\square$

This proof clarifies that the only pair of a θ-power of u and a θ-power of v which can share a prefix of length $2|u| - d$ without imposing $\rho_\theta(u) = \rho_\theta(v)$ is $(uu, v^{(n-1)/2}\theta(v)^{(n-1)/2+1})$, where n satisfies $2|u| = n|v|$. Since $2|u| - d \leq 2|u| - \lfloor d/2 \rfloor - 1$, the next result follows from this proof.

Corollary 1. *$|\mathrm{BCP}_\theta(u,v)| \leq 1$ for any $u, v \in \Sigma^+$ with $\rho_\theta(u) \neq \rho_\theta(v)$ and $|u| > |v| = 2\gcd(|u|, |v|)$.*

The proof of Proposition 2 and Figure 2 hint the possibility that if α and β are θ-palindromes of the same length d which disagree with each other for the first time at their center, i.e., their $\lceil d/2 \rceil$-th letter, then we can reach the new bound minus one while keeping $\rho_\theta(u) \neq \rho_\theta(v)$. For instance, let θ be the mirror involution on $\{a, b\}$, $\alpha = a^d$, and

$$\beta = \begin{cases} a^{\lceil d/2 \rceil - 1} b a^{\lceil d/2 \rceil - 1} & \text{if } d \text{ is odd} \\ a^{\lceil d/2 \rceil - 1} bb a^{\lceil d/2 \rceil - 1} & \text{if } d \text{ is even.} \end{cases} \tag{2}$$

For $u = (\alpha\beta)^{(n-1)/2}\beta$ and $v = \alpha\beta$, we have $|u^2 \wedge v^{(n-1)/2}\theta(v)^{(n-1)/2+1}| = 2|u| - \lfloor d/2 \rfloor - 1$. Since v contains at most two occurrences of b and they occur only in the latter half of it, v is θ-primitive. Hence, $\rho_\theta(u) \neq \rho_\theta(v)$.

Theorem 5. *$b'(p,q)$ is optimal for any pair (p,q) with $p > q = 2\gcd(p,q)$.*

Besides giving the verification of optimality to $b'(p,q)$, the proof enables us to enumerate all pairs of (u,v) with distinct θ-primitive roots, $|u| > |v| = 2\gcd(|u|,|v|)$, and $\mathrm{BCP}_\theta(u,v)$ is non-empty, i.e., $|\mathrm{BCP}_\theta(u,v)| = 1$ (Corollary 1). The way to construct (u,v) from (α,β) being known (see Figure 2), it suffices to provide the set of all possible values of (α,β). Note that it is insufficient for (α,β) to be a pair of two distinct θ-palindromes of the same length d and with the same prefix of length $\lceil d/2 \rceil - 1$. For instance, although $\alpha = a\theta(a)$ and $\beta = \theta(a)a$ satisfy these conditions, $u, v \in a\{a, \theta(a)\}^*$, i.e., $\rho_\theta(u) = \rho_\theta(v)$. Excluding these instances leaves the following three candidate sets:

$$\begin{aligned} T_1 &= \big\{(xa\theta(x), xb\theta(x)) \;\big|\; x \in \Sigma^*,\ a, b \in \Sigma \text{ such that } a \neq b,\ a = \theta(a),\ b = \theta(b)\big\}; \\ T_2 &= \big\{(xa\theta(a)\theta(x), xb\theta(b)\theta(x)) \;\big|\; x \in \Sigma^*,\ a, b \in \Sigma \text{ such that } a \neq b,\ a \neq \theta(b)\big\}; \\ T_3 &= \big\{(xa\theta(a)\theta(x), x\theta(a)a\theta(x)) \;\big|\; x \in \Sigma^+ \text{ such that } a \neq \theta(a),\ x \notin \{a, \theta(a)\}^+\big\}. \end{aligned}$$

Actually all of these sets serve our purpose, and hence, in the rest of this paper, we will use α, β only to denote a pair of words in $T_1 \cup T_2 \cup T_3$. In order to see that (α,β) makes the words $u = (\alpha\beta)^{(n-1)/2}\beta$ and $v = \alpha\beta$ have distinct θ-primitive roots, we just have to prove that there does not exist a word t such that $\alpha, \beta \in \{t, \theta(t)\}^+$. This is because $u, v \in \{\alpha, \beta\}^+$ and if $\rho_\theta(u) = \rho_\theta(v)$, then due to $d = \gcd(|u|,|v|)$, $t = \rho_\theta(u)$ is of length at most d, and hence, $\alpha, \beta \in \{t, \theta(t)\}^+$.

Proposition 3. *If $(\alpha, \beta) \in T_1 \cup T_2 \cup T_3$, then there does not exist $t \in \Sigma^+$ such that $\alpha, \beta \in \{t, \theta(t)\}^+$.*

Proof. Note that $\alpha \neq \beta$. Suppose the existence of such t and let $\alpha = t_1 \cdots t_k$ and $\beta = t'_1 \cdots t'_k$ for some $k \geq 1$ and $t_1, \ldots, t_k, t'_1, \ldots, t'_k \in \{t, \theta(t)\}$. If $(\alpha, \beta) \in T_1$, then the length of α (and β) is odd so that k is odd. Since $\alpha = \theta(\alpha)$, this means that $t = \theta(t)$, and hence, $\alpha = \beta$, which is a contradiction. Even if $(\alpha, \beta) \in T_2 \cup T_3$, an odd k causes the same problem.

Let us consider the case $(\alpha, \beta) \in T_2$ and k is even. Then $t_1 \cdots t_{k/2} = xa$, $t_{k/2+1} \cdots t_k = \theta(a)\theta(x)$, $t'_1 \cdots t'_{k/2} = xb$, and $t'_{k/2+1} \cdots t'_k = \theta(b)\theta(x)$. Hence, for some $y \in \mathrm{Suff}(x)$, $t_{k/2} = ya$, $t_{k/2+1} = \theta(a)\theta(y)$, $t'_{k/2} = yb$, and $t'_{k/2+1} = \theta(b)\theta(y)$. Since $a \neq b$, both $t_{k/2} \neq t'_{k/2}$ and $t_{k/2+1} \neq t'_{k/2+1}$ must hold. These four words are either t or $\theta(t)$ so that we have either $ya = \theta(a)\theta(y)$ and $yb = \theta(b)\theta(y)$ or $ya = \theta(b)\theta(y)$ and $yb = \theta(a)\theta(y)$. In the latter case, if y is empty, then $a = \theta(b)$; otherwise, these two equations imply that y begins with $\theta(b)$ and with $\theta(a)$ so that $\theta(b) = \theta(a)$; both contradict the condition on a, b in T_2. Even in the former case, unless y is empty, we reach this contradiction along the same argument. If y is empty, then $a = \theta(a)$, $b = \theta(b)$, and one of these has to be t and the other has to be $\theta(t)$. This is, however, impossible because a is assumed to be neither b nor $\theta(b)$.

The same but simpler argument works for $(\alpha, \beta) \in T_3$. Note that along this argument y should be non-empty because otherwise $t_{k/2} = a$, and hence, $x = t_1 \cdots t_{k/2-1} \in \{a, \theta(a)\}^+$, which is against the definition of T_3. ☐

Theorem 6. *Let $u, v \in \Sigma^+$ with $\rho_\theta(u) \neq \rho_\theta(v)$ and $|u| > |v| = 2\gcd(|u|, |v|)$. Then $\mathrm{BCP}_\theta(u, v) \neq \emptyset$ if and only if $u = (\alpha\beta)^{(n-1)/2}\beta$ and $v = \alpha\beta$ for some odd integer $n \geq 3$ and $(\alpha, \beta) \in T_1 \cup T_2 \cup T_3$.*

In the next subsection, we will prove that even when this length condition $|u| > |v| = 2\gcd(|u|, |v|)$ does not hold, the existence of boundary common prefix requires u and v to be described by two distinct θ-palindromes α, β of length d taken from T_1, T_2, or T_3, and hence, these three sets will completely characterize the boundary common prefixes.

3.2 The case when $q \geq 3\gcd(p, q)$

The proof of our improved bound $b'(p, q)$ continues here for (p, q) with $p > q \geq 3\gcd(p, q)$. Under this length condition, by definition, $b'(p, q) = 2p + q - \gcd(p, q) - \lfloor \gcd(p, q)/2 \rfloor$. Unlike the case considered in the previous subsection, this bound shall turn out not to be optimal for some such (p, q). A constructive way to find the optimal bound is to build an antimorphic involution θ and words u and v with distinct θ-primitive roots and $|u| > |v| \geq 3\gcd(|u|, |v|)$ such that the maximal common prefix between a θ-power of u and a θ-power of v gets as long as possible relative to $|u|$ and $|v|$. This informal description allows us to assume that u and v are θ-primitive, though in formal problem settings the validity of this assumption has to be verified (see Lemma 5.)

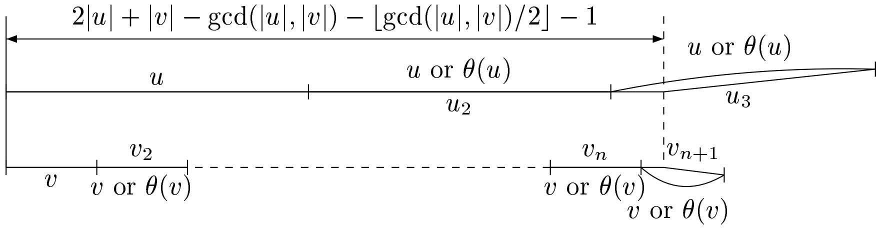

Figure 3: A boundary common prefix based on u and v. This shows how uu_2u_3 and $vv_2 \cdots v_nv_{n+1}$ overlap with each other when Condition (3) is satisfied.

First of all, we briefly mention how small the optimal bound for (p, q) can be relative to p and q. The following parameterized example proves that $2p$ is not a good bound for any such pair (p, q), that is, the optimal bound has to be bigger than $2p$.

Example 3. Let θ be the mirror involution on $\{a, b\}$. For a given (p, q) with $p > q \geq 3\gcd(p, q)$, let $v = a^{2p \pmod q} b^{-2p \pmod q}$ and $u\theta(u) = v^{\lfloor 2p/q \rfloor} a^{2p \pmod q}$. Then $|u\theta(u)u_3 \wedge v^{\lceil 2p/q \rceil}| = 2p$ regardless of the value of $u_3 \in \{u, \theta(u)\}$ because both u and $\theta(u)$ begin with a. In addition, $\rho_\theta(u) \neq \rho_\theta(v)$.

As a digression, this example can be easily modified to show that $2p + \lceil \gcd(p, q)/2 \rceil - 1$ is not a good bound for any such (p, q), either. Since $-2p \pmod q$ is a multiple of $\gcd(p, q)$, we can say that the suffix $b^{-2p \pmod q}$ of v consists of $\frac{-2p \pmod q}{\gcd(p,q)}$ blocks b^d. Replacing each of these blocks with β given in Eq. (2) verifies this point.

To return to our point, u and v are to be constructed so as for such a maximal common prefix to be of length at least $2|u|$ in light of Example 3. Hence, the common prefix is formalized with an integer n satisfying $(n-1)|v| < 2|u| < n|v|$ and words $u_1, u_2, u_3 \in \{u, \theta(u)\}$ and $v_1, \ldots, v_n, v_{n+1} \in \{v, \theta(v)\}$ as $u_1u_2u_3 \wedge v_1 \cdots v_nv_{n+1}$ with the following condition:

$$|u_1u_2u_3 \wedge v_1 \cdots v_nv_{n+1}| \geq 2|u| + k \text{ for some } k \geq 0 \tag{3}$$

Note that u_1 and v_1 are to be fixed to u and v without loss of generality. Figure 3 illustrates the maximal common prefix between a θ-power of u and a θ-power of v satisfying Condition (3) with $k = |v| - \gcd(|u|, |v|) - \lfloor \gcd(|u|, |v|)/2 \rfloor - 1$.

Lemma 4. *Let u, v be distinct θ-primitive words with $|u| > |v| \geq 3\gcd(|u|, |v|)$. If there exist an integer n satisfying $(n-1)|v| < 2|u| < n|v|$, and words $u_1 = u$, $u_2, u_3 \in \{u, \theta(u)\}$, $v_1 = v$, $v_2, \ldots, v_n, v_{n+1} \in \{v, \theta(v)\}$ satisfying Condition* (3)*, then one of the following two cases holds:*

1. *$u_2 = \theta(u)$ and $v_1 = \cdots = v_{n-1} = v$;*

2. *$u_2 = u$, $v_1 = \cdots = v_{\lceil n/2 \rceil - 1} = v$, and $v_{\lceil n/2 \rceil + 1} = \cdots = v_{n-1} = \theta(v)$.*

Proof. Let us consider the case when $u_2 = \theta(u)$ first. In this case, we have $u\theta(u) = v_1 \cdots v_{n-1}w$, where w is a non-empty prefix of v_n. Since $u\theta(u)$ is a θ-palindrome, $v_1 \cdots v_{n-1}w = \theta(w)\theta(v_{n-1}) \cdots \theta(v_1)$ holds. This means that $\theta(v_{n-1}) \cdots \theta(v_1)$ is a proper infix of $v_1 \cdots v_n$. Then we can apply Theorem 1 to conclude $\theta(v_{n-1}) = \cdots = \theta(v_1)$ because v is assumed to be θ-primitive.

Even for the second case when $u_2 = u$, the basic strategy is the same. Since the border between u_1 and u_2 is located on $v_{\lceil n/2 \rceil}$, one can let $v_{\lceil n/2 \rceil} = xy$ for some non-empty words x, y such that $u_1 = v_1 \cdots v_{\lceil n/2 \rceil - 1}x$ and $u_2 = yv_{\lceil n/2 \rceil + 1} \cdots v_{n-1}z$, where z is a non-empty prefix of v_n. Then we have $v_1 \cdots v_{\lceil n/2 \rceil - 1}x = yv_{\lceil n/2 \rceil + 1} \cdots v_{n-1}z$ because $u_1 = u_2$. This equation implies that $v_1 \cdots v_{\lceil n/2 \rceil - 1}$ is a proper infix of $v_{\lceil n/2 \rceil} \cdots v_n$ so that $v_1 = \cdots = v_{\lceil n/2 \rceil - 1} = v$. If $n \geq 4$, we can also determine the values of $v_{\lceil n/2 \rceil + 1}, \ldots, v_{n-1}$. Firstly, the value of $v_{\lceil n/2 \rceil + 1}$ is determined to be $\theta(v)$ by applying Lemma 3 to the overlap between $v_1 v_2$ and $v_{\lceil n/2 \rceil} v_{\lceil n/2 \rceil + 1}$. When $n \geq 6$, Theorem 1 is applied to that $v_{\lceil n/2 \rceil + 1} \cdots v_{n-1}$ being a proper infix of $v_1 \cdots v_{\lceil n/2 \rceil}$ to fix $v_{\lceil n/2 \rceil + 1} = \cdots = v_{n-1} = \theta(v)$. □

As suggested previously, an element of $\mathrm{BCP}_\theta(u, v)$ is characterized by Condition (3) with $k = |v| - \gcd(|u|, |v|) - \lfloor \gcd(|u|, |v|)/2 \rfloor - 1$. Thus, this condition is replaced by the next condition:

$$|u_1 u_2 u_3 \wedge v_1 \cdots v_n v_{n+1}| \geq 2|u| + |v| - \gcd(|u|, |v|) - \left\lfloor \frac{\gcd(|u|, |v|)}{2} \right\rfloor - 1. \quad (4)$$

Once this inequality proves not to hold strictly, $b'(p, q)$ becomes a good bound for an arbitrary pair (p, q). The next lemma verifies that the assumption of u, v being θ-primitive is valid when we consider $\mathrm{BCP}_\theta(u, v)$.

Lemma 5. *Let $u, v \in \Sigma^+$ such that $\rho_\theta(u) \neq \rho_\theta(v)$ and $|u| > |v| \geq 3\gcd(|u|, |v|)$. Unless both u and v are θ-primitive,* $\mathrm{BCP}_\theta(u, v) = \emptyset$.

Proof. Here we prove its contrapositive: if $\mathrm{BCP}_\theta(u, v) \neq \emptyset$, then both u and v are θ-primitive. For this purpose, suppose the non-emptiness and that u and v were not θ-primitive at the same time, and see that a contradiction is unavoidable. Let $r = \rho_\theta(u)$, $t = \rho_\theta(v)$, $d = \gcd(|u|, |v|)$, and $d' = \gcd(|r|, |t|)$. It suffices to show that $b(\max(|r|, |t|), \min(|r|, |t|)) \leq b'(|u|, |v|) - 1$ under this supposition, which would lead us to the contradictive conclusion $\rho_\theta(u) = \rho_\theta(v)$ due to Theorem 4.

If $\min(|r|, |t|) \leq 2d'$, then by definition $b(\max(|r|, |t|), \min(|r|, |t|)) = \mathrm{lcm}(|r|, |t|)$, and we have $\mathrm{lcm}(|r|, |t|) \leq 2\max(|r|, |t|) \leq 2|u| \leq b'(|u|, |v|) - 1$.

In the case $\min(|r|, |t|) \geq 3d'$, we claim that

$$2\max(|r|, |t|) + \min(|r|, |t|) - d' \leq b'(|u|, |v|) - 1 \quad (5)$$

holds if $r \neq u$ or $t \neq v$. To prove this claim, it is worth noting that $|t| \leq |u| - d$ holds because $|t| \leq |v|$ and $|v| \leq |u| - d$. Firstly, let us consider the case $|r| > |t|$. If $r \neq u$, then one easily obtains $2|r| + |t| - d' < 2|u| < b'(|u|, |v|) - 1$ because $r \neq u$ means $2|r| \leq |u|$. This inequality is exactly same as (5) when $|r| > |t|$. If $r = u$, then $t \neq v$ so that $|t| \leq |v|/2 \leq |v| - d - \lfloor d/2 \rfloor$ holds; the latter inequality follows from $|v| \geq 3d$. Thus, $|t| - d' \leq |t| - 1 \leq |v| - d - \lfloor d/2 \rfloor - 1$, and hence, we have $2|r| + |t| - d' \leq b'(|u|, |v|) - 1$. Conversely if $|r| < |t|$, then $|r| < |v|$ holds. Due to the inequality:

$$2|u| + |v| - 2d \leq 2|u| + |v| - d - \lfloor d/2 \rfloor - 1, \tag{6}$$

we have $2|t| + |r| - d' \leq 2(|u| - d) + |v| - d' \leq 2|u| + |v| - 2d \leq b'(|u|, |v|) - 1$. This is the same as (5) when $|r| < |t|$. Having proved the claim, now it suffices to note that the left-hand side of (5) is equivalent to $b(\max(|r|, |t|), \min(|r|, |t|))$. $\square$

Note that the inequality given in Eq. (6) will play a significant role throughout this paper.

Up to now, we have seen that the combinations of the values of $u_2, u_3, v_2, \ldots, v_{n+1}$ have been already severely-limited under the condition (3) due to Lemma 4. We will see in the following that some specific value of k in this condition further restricts the number of possible combinations.

Proposition 4 ([8])**.** *Let $u, v \in \Sigma^+$ such that v is θ-primitive, $u_2, u_3 \in \{u, \theta(u)\}$, and $v_2, \ldots, v_n \in \{v, \theta(v)\}$ for some integer $n \geq 3$. If $vv_2 \cdots v_n$ is a prefix of uu_2u_3 and $(n-1)|v| < 2|u| < n|v|$, then there are only two cases possible:*

1. *$u_2 = \theta(u)$ and $v_2 = \cdots = v_n = v$ with $u\theta(u) = (yx)^{n-1}y$ and $v = yx$ for some non-empty θ-palindromes x, y; and*
2. *$u_2 = u$, n is even, $v_2 = \cdots = v_{n/2} = v$, and $v_{n/2+1} = \cdots = v_n = \theta(v)$ with $v = r(tr)^i(rt)^{i+j}r$ and $u = v^{n/2-1}r(tr)^i(rt)^j$ for some $i \geq 0$, $j \geq 1$, and non-empty θ-palindromes r, t.*

This proposition is applicable to our problem when $v_1 \cdots v_n$ is a prefix of $u_1u_2u_3$, that is, when the border between v_n and v_{n+1} is at the left of the vertical dashed line in Figure 3. Since $2|u| - (n-1)|v|$ is a multiple of $\gcd(|u|, |v|)$, this condition is formalized as $2|u| - (n-1)|v| \geq 2\gcd(|u|, |v|)$. This always holds when n is odd because $|u| - \frac{n-1}{2}|v|$ is a multiple of $\gcd(|u|, |v|)$. On the contrary, $2|u| - (n-1)|v| = \gcd(|u|, |v|)$ may hold when n is even. Then $v_1 \cdots v_n$ disagrees with $u_1u_2u_3$ somewhere within the $(\lfloor \gcd(|u|, |v|)/2 \rfloor + 1)$ rightmost letters of v_n, as shown in the next example.

Example 4. Let $u = abbab$, $v = abb$, and θ be the mirror image on $\{a, b\}$. Then $u\theta(u)^2$ and v^4 satisfy Condition (4), with $n = 4$, and $2|u| - (n-1)|v| = \gcd(|u|, |v|)$.

Lemma 6 ([8])**.** *Let $v \in \Sigma^+$ be a θ-primitive word, and $x, y \in \Sigma^+$ be words strictly shorter than v. For an integer $k \geq 1$, the solution to $v\theta(v)^k x = yv^{k+1}$ is characterized as $v = r(tr)^i(rt)^{i+j}r$, $y = r(tr)^i(rt)^j$, and $x = (rt)^{i+j}r$ for some $i \geq 0$, $j \geq 1$, and non-empty θ-palindromes r, t.*

Lemma 7. *Let u, v be distinct θ-primitive words with $|u| > |v| \geq 3\gcd(|u|, |v|)$. If there exist an integer n, and words $u_1 = u$, $u_2, u_3 \in \{u, \theta(u)\}$, $v_1 = v$, $v_2, \ldots, v_n, v_{n+1} \in \{v, \theta(v)\}$ satisfying Condition (4) and $2|u| - (n-1)|v| = \gcd(|u|, |v|)$, then $u_2 = \theta(u)$ and $v_2 = \cdots = v_n = v$. Moreover, $v = yx$ and $u\theta(u) = (yx)^{n-1}y$ for some θ-palindromes $y, x \in \Sigma^+$.*

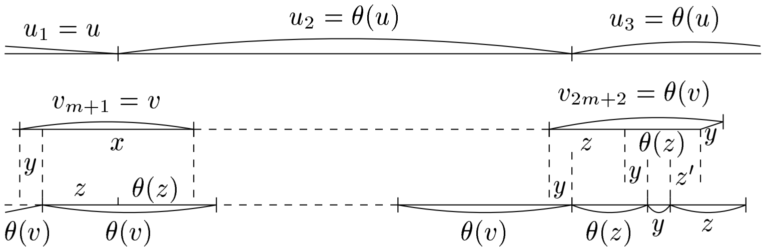

Figure 4: When $u_3 = \theta(u)$ and $|y| = d$, $v_{2m+2} = \theta(v)$ and the prefix $\theta(z)yz$ of $\theta(u)$ partially overlap as shown here.

Proof. Let $d = \gcd(|u|, |v|)$. As mentioned previously, in order for $2|u| - (n-1)|v| = d$ to hold, n has to be even. So, let $n = 2(m+1)$ for some $m \geq 1$.

Firstly, we investigate the case when $u_2 = \theta(u)$. In this case, Lemma 4 fixes all of $v_2, \ldots, v_{2m-1}$ to be equal to v_1, i.e., v. Then we can let

$$u_1 u_2 = u\theta(u) = v^{2m+1} y \tag{7}$$

for some $y \in \mathrm{Pref}(v_{2m+2})$. From Eq. (7) and the hypothesis of this lemma, $|y| = 2|u| - (2m+1)|v| = d$. Combining this relation and Condition (4) implies that yu_3 and v_{2m+2} share their prefix of length at least $|v| - d$. At any rate, since $u\theta(u)$ is a θ-palindrome, Eq. (7) gives $v^{2m+1}y = \theta(y)\theta(v)^{2m+1}$. This means that $vy \in \mathrm{Suff}(\theta(v)^2)$ because $m \geq 1$, and this suffix condition allows us to let $\theta(v) = xy$ for some $x \in \Sigma^+$. Substituting this back to the suffix condition results in $\theta(y)\theta(x)y \in \mathrm{Suff}(xyxy)$. From this, we can easily observe that $y = \theta(y)$ and $x = \theta(x)$. Now we have $v = yx$. Note that the relation $(2m+2)|v| - 2|u| = |x|$ results from this equation and Eq. (7) so that x is a θ-palindrome of even length; that is, we can let $x = z\theta(z)$ for some $z \in \Sigma^+$. Hence, $v = yz\theta(z)$, and by substituting this into Eq. (7), we can obtain that

$$u = v^m yz = (yx)^m yz = (yz\theta(z))^m yz. \tag{8}$$

According to this equation, the Euclidean algorithm gives $d = \gcd(|u|, |v|) = \gcd(|v|, |y|+|z|) = \gcd(|y|+|z|, |z|) = \gcd(|z|, |y|)$. This equation further implies $\gcd(|x|, |y|) = \gcd(2|z|, |y|) = d$ because $|y| = d$.

From now on, we will prove that v_{2m+2} has to be v. Suppose not, that is, $v_{2m+2} = \theta(v) = xy$. Recall that yu_3 and v_{2m+2} share their prefix of length at least $|x| + |y| - d$. In addition, $|x| + |y| - d \geq 2d$ according to the hypothesis $|v| \geq 3d$. Thus, if $|z| = d$, then this common prefix immediately gives $z = y$, and hence, v would be y^3. This contradicts the θ-primitivity of v so that z has to be of length at least $2d$. If $u_3 = u$, then $yu_3 = y(yx)^m yz$ holds due to Eq. (8), and hence, this common prefix implies that yyx and xy share their prefix of length $|x|+|y|-d$. As seen above, $\gcd(|x|, |y|) = d$ and $|v| = |yx| \geq 3d$. Thus, Lemma 2 is applicable to this common prefix, and results in $\rho(x) = \rho(y)$. This, however, contradicts $\rho_\theta(u) \neq \rho_\theta(v)$, and hence, u_3 has to be $\theta(u) = \theta(z)y(z\theta(z)y)^m$. See Figure 4. In this case, the common prefix between v_{2m+2} and yu_3 gives

$$z\theta(z) = y\theta(z)yz' \tag{9}$$

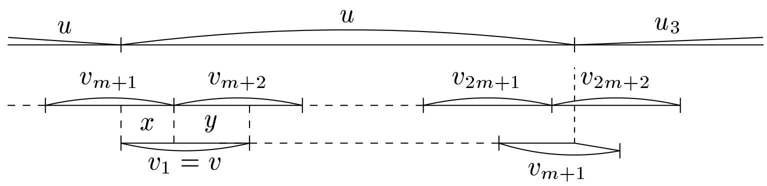

Figure 5: If $u_2 = u$, then $v_1 v_2 \ldots v_{m+1}$ overlaps with $v_{m+1} \ldots v_{2m+1}$ not depending on the value of u_3.

for some $z' \in \mathrm{Pref}(z)$. Eq. (9) implies that $z' \in \mathrm{Suff}(\theta(z))$, i.e., $\theta(z') \in \mathrm{Pref}(z)$, so that $z' = \theta(z')$. Eq. (9) also enables us to let $z = yz''$ for some $z'' \in \Sigma^*$. Substituting $\theta(z) = \theta(z'')y$ back into Eq. (9) yields $z\theta(z) = y\theta(z'')y^2 z'$, and hence, $\theta(z) = y^2 z'$, i.e., $z = z'y^2$. If $|z| = 2d$, then $z' = \lambda$; otherwise, by replacing Eq. (9) by these, we can obtain $z'y^2y^2z' = y^3z'yz'$, and hence, $z'y^3 = y^3z'$, which implies $\rho(z') = \rho(y)$. However, in both cases, we reach the contradiction with the θ-primitivity of $v = yz'y^2y^2z'$.

Now we have to prove that u_2 cannot be u. Suppose for the sake of contradiction that $u_2 = u$. Lemma 4 gives $v_1 = \cdots = v_m = v$ and $v_{m+2} = \cdots = v_{2m+1} = \theta(v)$. So,

$$u = v^m z = x\theta(v)^m z' \tag{10}$$

for some $z, x, z' \in \Sigma^+$ such that $v_{m+1} = zx$. Since $m \geq 1$, this allows us to let $v = xy$ for some $y \in \mathrm{Pref}(\theta(v))$ (see Figure 5). Substituting this into Eq. (10) gives $y = \theta(y)$. Eq. (10) also gives $|x| = (m+1)|v| - |u|$, and by combining this with the hypothesis $(2m+1)|v| < 2|u| < (2m+2)|v|$, we obtain $2|x| = (2m+2)|v| - 2|u| < |v| = |x| + |y|$, and hence, $|x| < |y|$. As done before, based on $u = v^m z$ and $|z| = |y|$, the Euclidean algorithm gives $d = \gcd(|u|, |v|) = \gcd(|x|, |y|)$. Thus, $|x| < |y|$ implies that $|y| \geq 2d$. With Eq. (10), this length condition results in

$$(m+1)|v| = |u| + |x| \leq |u| + |x| + |y| - 2d. \tag{11}$$

For our purpose, it suffices to prove that v_{m+1} can be neither v or $\theta(v)$. Suppose first that $v_{m+1} = \theta(v) = yx$. First of all, we can easily see that $z = y$ because v_{m+1} was let to be zx. Thus, Eq. (10) can be rewritten as $u = v^m y$. As illustrated in Figure 5, the overlap between $v_{2m+1} = \theta(v)$ and $v_{m+1} = \theta(v)$ implies that $x \in \mathrm{Pref}(\theta(v))$. This implies $x \in \mathrm{Pref}(y)$, i.e., $\theta(x) \in \mathrm{Suff}(y)$, because $\theta(v) = yx$ and $|x| < |y|$. As a result, $\theta(x) \in \mathrm{Suff}(u)$, and hence, x is a prefix of both u and $\theta(u)$. This means that $v^m\theta(v) \in \mathrm{Pref}(uu_3)$ regardless of whether u_3 is u or $\theta(u)$. Note that, by the hypothesis, $u_2u_3 = uu_3$ and $xv_{m+2}\cdots v_{2m+1}v_{2m+2}$ share their prefix of length at least $|u| + |v| - 2d$. Due to Condition (11), $v^m\theta(v) \in \mathrm{Pref}(xv_{m+2}\cdots v_{2m+2})$, that is, $v^m\theta(v)$ is an infix of $v_{m+1}\cdots v_{2m+2}$. However, since $m \geq 1$ and v is primitive, this contradicts Theorem 1. Thus, v_{m+1} cannot be $\theta(v)$ so that has to be v. If so, applying Lemma 6 to the overlap between $v_1 \cdots v_{m+1}$ and $v_{m+1} \cdots v_{2m+1}$ yields $v = r(tr)^i(rt)^{i+j}r$ and $u = v^m r(tr)^i(rt)^j$ for some $i \geq 0$, $j \geq 1$, and non-empty θ-palindromes r, t. One can easily check that $u^2 = v^{m+1}\theta(v)^m(rt)^j$ holds, and hence, $|(rt)^j| = d$ due to $2|u| - (2m+1)|v| = d$. On the contrary, from $\gcd(|u|,|v|) = d$, the Euclidean algorithm derives $\gcd(|r(tr)^i|, |(rt)^j|) = d$. However, $|r(tr)^i|$ cannot be a multiple of $|(rt)^j| = d$ because $r, t \neq \lambda$. $\square$

Lemma 8. *Let u, v be distinct θ-primitive words with $|u| > |v| \geq 3\gcd(|u|,|v|)$. If there exist an integer n, words $u_1, u_2, u_3 \in \{u, \theta(u)\}$, and $v_1, \ldots, v_n, v_{n+1} \in \{v, \theta(v)\}$ satisfying Condition (4), then $u_2 = u_3$.*

Proof. Let $d = \gcd(|u|,|v|)$. We will consider two cases depending on whether u_2 is $\theta(u)$ or u, and will prove that u_3 has to be equal to u_2.

The first case is when $u_2 = \theta(u)$. In this case, Proposition 4 and Lemma 7 imply that $v_2 = \cdots = v_n = v$, $v = yx$ and $u\theta(u) = (yx)^{n-1}y$ for some non-empty θ-palindromes x, y. The Euclidean algorithm yields $\gcd(2|u|,|v|) = \gcd(|y|,|v|) = \gcd(|x|,|y|)$, and hence, $\gcd(|x|,|y|)$ is either d or $2d$ because $d = \gcd(|u|,|v|)$. Suppose $u_3 = u$. This means that u_3 starts with yx, and hence, yx and xv_{n+1} share their prefix of length at least $|x| + |y| - d - \lfloor d/2 \rfloor - 1$. If $|x| + |y| = 2\gcd(|x|,|y|)$, then $|x| = |y| = d$ or $|x| = |y| = 2d$, but indeed only the latter is valid under the assumption $|v| = |x| + |y| \geq 3d$. This means that the common prefix is of length at least $|x|$ so that it implies $x = y$, which however contradicts the θ-primitivity of v. Conversely, if $|x| + |y| \geq 3\gcd(|x|,|y|)$, then $|x| + |y| - d - \lfloor d/2 \rfloor - 1 \geq |x| + |y| - 2d \geq |x| + |y| - 2\gcd(|x|,|y|)$ (here $d \leq \gcd(|x|,|y|)$ is used). Since v_{n+1} is either $v = yx$ or $\theta(v) = xy$, Lemma 2 is applicable to the common prefix to obtain $\rho(x) = \rho(y)$. Now that we have reached the same contradiction, we can conclude that the only possible choice of u_3 is $\theta(u)$.

The next case is when $u_2 = u$. Due to Proposition 4 and Lemma 7, n is even ($n = 2m+2$ for some $m \geq 1$), $v_2 = \cdots = v_{m+1} = v$ and $v_{m+2} = \cdots = v_{2m+2} = \theta(v)$, with $v = r(tr)^i(rt)^{i+j}r$ and $u = v^m r(tr)^i(rt)^j$ for some $i \geq 0$, $j \geq 1$, and non-empty θ-palindromes r, t. Then we have $v^{m+1}\theta(v)^{m+1} = uu(rt)^i r(rt)^i r$. The Euclidean algorithm derives $\gcd(|(rt)^i r|, |(tr)^j|) = d$ from $\gcd(|u|,|v|) = d$. Note that $|u_3 \wedge (rt)^i r(rt)^i r v_{2m+3}| = |v| - d - \lfloor d/2 \rfloor - 1 \geq |v| - 2d \geq |v| - 2|r(tr)^i| = |(tr)^j| \geq |rt|$. Consequently u_3 must not begin with t in light of Proposition 1, and hence, u_3 cannot be $\theta(u)$. $\square$

Now we are ready to prove that $b'(p,q)$ is an improved bound for the extended Fine and Wilf's theorem. Recall Eq. (6), which makes it possible to distinguish the cases in which boundary common prefixes are constructable.

Theorem 7. *Let $u, v \in \Sigma^+$ with $\rho_\theta(u) \neq \rho_\theta(v)$ and $|u| > |v| \geq 3\gcd(|u|,|v|)$. Then the length of a word in $\mathrm{BCP}_\theta(u,v)$ is $2|u| + |v| - \gcd(|u|,|v|) - \lfloor \gcd(|u|,|v|)/2 \rfloor - 1$. Moreover, $\mathrm{BCP}_\theta(u,v) \neq \emptyset$ if and only if one of the following two cases holds: for some $m \geq 1$, $i \geq 0$, and $(\alpha,\beta) \in T_1 \cup T_2 \cup T_3$ and*

1. *$u = (\alpha\beta(\beta\alpha)^i\beta)^m\alpha\beta$, $v = \alpha\beta(\beta\alpha)^i\beta$;*
2. *$u = [\alpha(\beta\alpha)^i(\alpha\beta)^{i+1}\alpha]^m\alpha(\beta\alpha)^i\alpha\beta$, $v = \alpha(\beta\alpha)^i(\alpha\beta)^{i+1}\alpha$.*

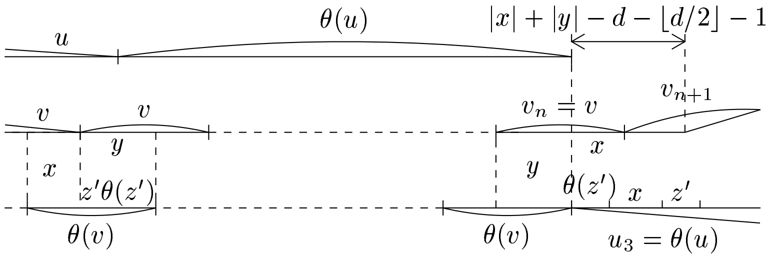

Figure 6: For an odd n, $u\theta(u)^2$ and $v^n v_{n+1}$ share the common prefix of length $2|u| + |v| - d - \lfloor d/2 \rfloor - 1$, where $d = \gcd(|u|, |v|)$.

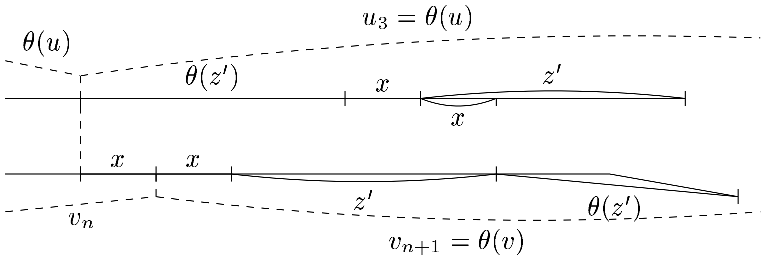

Figure 7: When n is odd and $|x| < |z'|$, u_2u_3 and v_nv_{n+1} overlap as shown here.

Proof. Let $d = \gcd(|u|, |v|)$ and assume that $\mathrm{BCP}_\theta(u, v)$ is not empty. Then Lemma 5 implies that both u and v are θ-primitive. Since an element of $\mathrm{BCP}_\theta(u, v)$ is characterized by Condition (4), Proposition 4, Lemmas 7 and 8 leave only two cases to be investigated:

1. $u_2 = u_3 = \theta(u)$, $v_2 = \cdots = v_n = v$, $v = yx$, and $u\theta(u) = v^{n-1}y$ for some non-empty distinct θ-palindromes x, y; and

2. $u_2 = u_3 = u$, $n = 2m + 2$ for some $m \geq 1$, $v_2 = \cdots = v_{m+1} = v$, $v_{m+2} = \cdots = v_{2m+2} = \theta(v)$, $v = r(tr)^i(rt)^{i+j}r$, and $u = v^m r(tr)^i(rt)^j$ for some $i \geq 0$, $j \geq 1$, and non-empty distinct θ-palindromes r, t.

Case 1: In this case, the parity of n matters so that we first consider the subcase when n is odd (see Figure 6). Then the border between u_1 and u_2 splits the prefix y of $v_{(n+1)/2}$ into half. Hence, we can let $y = z'\theta(z')$ for some $z' \in \Sigma^+$ and $u = v^{(n-1)/2}z'$. The Euclidean algorithm derives $\gcd(|z'|, |x|) = d$ from $\gcd(|u|, |v|) = d$. Focus to the right of border between u_2 and u_3. The rightmost dashed line in Figure 6, up to which $u_1u_2u_3$ agree with $v_1v_2 \cdots v_{n+1}$, is located on v_{n+1} because $|y| = 2|z'| \geq 2d$. Thus, the suffix x of v_n is a prefix of the prefix $\theta(z')x$ of u_3. So, if z' were of length d, then due to this prefix relation and $d = \gcd(|z'|, |x|)$, x would be a power of $\theta(z')$, which contradicts the θ-primitivity of v. Therefore, z' has to be of length at least $2d$. This means that the rightmost vertical dashed line in Figure 6 is on z' of the prefix $\theta(z')xz'$ of $u_3 = \theta(u)$, and hence, $\theta(z')x \in \mathrm{Pref}(xv_{n+1})$.

In what follows, we prove that in this subcase $\mathrm{BCP}_\theta(u, v) \neq \emptyset$ requires $v_{n+1} = v$ and $|z'| = 2d$. For the sake of contradiction, suppose that v_{n+1} were

$\theta(v) = xz'\theta(z')$. Then the prefix relation just mentioned is written as

$$\theta(z')x \in \mathrm{Pref}(xxz'\theta(z')). \tag{12}$$

First of all, $|z'| > |x|$ has to hold because otherwise Relation (12) would cause $\theta(z')x \in \mathrm{Pref}(x^2)$, that is, $\rho(z') = \rho(x)$ due to Proposition 1, which contradicts the θ-primitivity of v. With this condition, Relation (12) gives $x^2 \in \mathrm{Pref}(\theta(z')x)$. If $|z'| = 2d$ (i.e. $|x| = d$), then this prefix relation would result in $\theta(z') = x^2$ and lead us to the same contradiction. Otherwise ($|z'| \geq 3d$), as illustrated in Figure 7, $\mathrm{pref}_{|z'|-2d}(z') \in \mathrm{Pref}(x\theta(z'))$. Substituting this into the overlap between $\theta(z')x$ and xxz' implies either $\theta(z') \in \mathrm{Pref}(x^3\theta(z'))$ if $|x| = d$; or $\theta(z')x \in \mathrm{Pref}(x^3\theta(z'))$ otherwise. In the former case, $\theta(z')$ would be a power of x, whereas in the latter case Proposition 1 would imply $\rho(\theta(z')) = \rho(x)$. At any rate, we face the contradiction against the θ-primitivity of v. Consequently, v_{n+1} has to be v. Then the prefix relation $\theta(z')x \in \mathrm{Pref}(xv_{n+1})$ is rather equal to $\theta(z')x \in \mathrm{Pref}(xz'\theta(z')x)$, and we can immediately see that $\theta(z')x = xz'$. This is a well-known conjugacy equation and can be solved as

$$\theta(z') = rt, x = r(tr)^k, z' = tr \tag{13}$$

for some $k \geq 0$ and words r, t (see, e.g., [5]). The resulting equation $z' = tr$ implies $\theta(z') = \theta(r)\theta(t)$ and combining this with $\theta(z') = rt$ results in $r = \theta(r)$ and $t = \theta(t)$. These r, t have to be distinct and non-empty in light of the θ-primitivity of v.

Next we prove that, under the assumption $v_{n+1} = v$, $|z'|$ has to be $2d$. By applying Euclidean algorithm to Eq. (13), we can obtain $d = \gcd(|z'|, |x|) = \gcd(|r|, |t|)$. Recall that $xv_{n+1} (= xz'\theta(z')x)$ and $\theta(z')xz'$ share a prefix of length at least $|\theta(z')xz'| - 2d$. Removing the trivial common part $xz' = \theta(z')x$ from this prefix leaves us $|\theta(z') \wedge z'| \geq |z'| - 2d$, that is, $|rt \wedge tr| \geq |rt| - 2d$. So if $|z'| \geq 3d$, then Lemma 2 could be employed to give $\rho(r) = \rho(t)$, which would lead us to the contradiction with the θ-primitivity of v. Having successfully proved that $|z'| = 2d$, let us construct a boundary common prefix based on u and v. Based on the presentations of x and z' in Eq. (13), we can see $v = tr(rt)^{k+1}r$ and $u = v^{(n-1)/2}tr$. Due to $z' = 2d$ and $d = \gcd(|r|, |t|)$, we have $|t| = |r| = d$. By replacing (t, r) with $(\alpha, \beta) \in T_1 \cup T_2 \cup T_3$, we can get the first pair of presentations of u and v shown in the statement with $i \geq 1$. It is left to the readers to check that $|u\theta(u)^2 \wedge v^{n+1}| = 2|u| + |v| - d - \lfloor d/2 \rfloor - 1$.

The second subcase of **Case 1** ($u_3 = \theta(u)$) is when n is even. Recall that x, y are θ-palindromes. In this subcase, x can be rather written as $x = z\theta(z)$ for some $z \in \Sigma^+$ (see Figure 8). As done before, one can obtain $\gcd(|y|, |z|) = d$ from $\gcd(|u|, |v|) = d$. The overlap between v_n and u_3 gives $z = \theta(z)$. Note that $v_nv_{n+1} = yz^2v_{n+1}$ and $yu_3 = yzyz^2$ share their prefix of length at least $|y| + |x| + |y| - d - \lfloor d/2 \rfloor - 1$. Hence, after reducing their common prefix yz, still zv_{n+1} and yz^2 share their prefix of length at least $|y| + |z| - 2d$. Since $v_{n+1} \in \{y, z\}^+$, if $|yz| \geq 3d$, then due to Lemma 2 this common prefix would give $\rho(y) = \rho(z)$ and we have reached the contradiction. Thus, yz has to be of length $2d$, i.e., $|y| = |z| = d$. Then by replacing (y, z) with $(\alpha, \beta) \in T_1 \cup T_2 \cup T_3$,

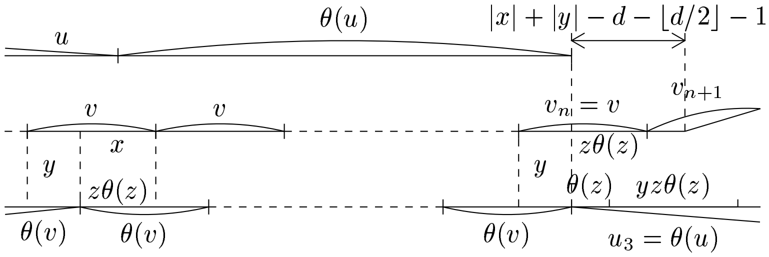

Figure 8: For an even n, $u\theta(u)^2$ and $v^n v_{n+1}$ share the common prefix of length $2|u|+|v|-d-\lfloor d/2\rfloor$, where $d=\gcd(|u|,|v|)$.

we obtain the first pair of presentations of u, v in the statement with $i=0$. The boundary common prefix based on u and v is constructed in the same manner as previous case.

(Case 2): Let us remind ourselves of the case: "$u_2=u_3=u$, $n=2m+2$ for some $m\geq 1$, $v_2=\cdots=v_{m+1}=v$, $v_{m+2}=\cdots=v_{2m+2}=\theta(v)$, $v=r(tr)^i(rt)^{i+j}r$, and $u=v^m r(tr)^i(rt)^j$ for some $i\geq 0$, $j\geq 1$, and non-empty distinct θ-palindromes r,t". Note that the following equations hold:

$$u^2 = v^{m+1}\theta(v)^m(rt)^j, \quad (14)$$
$$u^3 = v^{m+1}\theta(v)^{m+1}(tr)^j v^{m-1} r(tr)^i(rt)^j. \quad (15)$$

Due to Lemma 7, if $2|u|-(2m+1)|v|=d$, then $u_2=\theta(u)$. Since now we assume that $u_2=u$, $2|u|-(2m+1)|v|\geq 2d$ must hold. Combining this with Eq. (14) implies $|(rt)^j|\geq 2d$. With Eq. (15), this gives $|(tr)^j\wedge v_{2m+3}|\geq |(tr)^j|-d-\lfloor d/2\rfloor-1$. Note that v_{2m+3} begins with r regardless of whether it is v or $\theta(v)$.

The Euclidean algorithm derives $\gcd(|r(tr)^i|,|(tr)^j|)=d$ from $\gcd(|u|,|v|)=d$. Let $|(tr)^j|=kd$ for some $k\geq 2$. Here we shall see that unless $j=1$, we could not avoid a contradiction. Suppose $j\geq 2$. If $k\geq 4$, then $|tr|\leq \frac{1}{2}|(tr)^j|\leq |(tr)^j|-2d$. Thus, $|(tr)^j\wedge v_{2m+3}|\geq |tr|$, and Proposition 1 is applicable to this overlap to yield $\rho(r)=\rho(t)$. However, this contradicts the θ-primitivity of v. The same argument works for $k=3$ and $j\geq 3$. If $k=3$ and $j=2$, then $|trtr|=3d$. The Euclidean algorithm gives either $\gcd(|r|,2|t|)=d$ (if i is even) or $\gcd(2|r|,|t|)=d$ (otherwise). Combining these with $|trtr|=3d$ gives either $|r|=2|t|=d$ (if i is even) or $2|r|=|t|=d$ (otherwise). The overlap between $(tr)^j$ and v_{2m+3} is of length at least d, which is long enough to get $r=t^2$ (if i is even) or $t=r^2$ (otherwise.) In either case, we cannot accept such a conclusion in light of the θ-primitivity of v. As a result, the remaining case is $k=2$, i.e., $|(tr)^j|=2d$. Then $\gcd(|r(tr)^i|,|(tr)^j|)=d$ gives $\gcd(|r(tr)^{i \bmod j}|,|(tr)^j|)=d$, and further $\gcd(|r(tr)^{i \bmod j}|,|(tr)^{(-i \bmod j)-1}t|)=d$. This means that $|r(tr)^{i \bmod j}|=|(tr)^{(-i \bmod j)-1}t|=d$ because they are properly shorter than $|(tr)^j|=2d$. This further implies $i \bmod j=(-i \bmod j)-1$ and $|r|=|t|$. Hence, j has to be odd, i.e., $j\geq 3$. With the non-emptiness of r and t, $|(tr)^j|=2d$ now implies $d\geq 3$. As a result, $|(tr)^j\wedge v_{2m+3}|=d-\lfloor d/2\rfloor-1\geq d/j=|t|=|r|$, and thus $t=r$, the same contradiction.

Consequently, the only one possible value of j which may create a boundary common prefix is 1. Then the Euclidean algorithm yields $\gcd(|r|,|t|) = d$ from $\gcd(|r(tr)^i|,|tr|) = d$. If $|tr| \geq 3d$, then $|tr \wedge v_{2m+3}| \geq |tr| - 2d$ and the contradictory result $\rho(r) = \rho(t)$ would be obtained by Lemma 2. Thus, only the case $|tr| = 2d$, that is, $|t| = |r| = d$ remains valid. Actually in this case, substituting $(\alpha,\beta) \in T_1 \cup T_2 \cup T_3$ for (r,t) results in the second pair of presentations of (u,v) in the statement. One can easily check that $|u^3 \wedge v^{m+1}\theta(v)^{m+1}\alpha| = 2|u|+|v|-d-\lfloor d/2 \rfloor - 1$; note that α is a prefix of v_{2m+3} not depending on whether it is v or $\theta(v)$. □

Corollary 2. *$|\mathrm{BCP}_\theta(u,v)| \leq 1$ for any $u,v \in \Sigma^+$ with $\rho_\theta(u) \neq \rho_\theta(v)$ and $|u| > |v| \geq 3\gcd(|u|,|v|)$.*

Proof. As shown in the proof of Theorem 7, once u and v are given in one of the presentations present there, there is only one way to construct an element of $\mathrm{BCP}_\theta(u,v)$. Furthermore, v is of length $\gcd(|u|,|v|)$ times an odd number in the first presentation, whereas is of length $\gcd(|u|,|v|)$ times an even number in the second one. □

3.3 The improved bound and its optimality

Combining Proposition 2 and Theorem 7 completes our proof of the improved bound for the extended Fine and Wilf's theorem.

Theorem 8. *Let $u,v \in \Sigma^+$ with $|u| > |v| \geq 2\gcd(|u|,|v|)$. If a θ-power of u and a θ-power of v share a prefix of length $b'(|u|,|v|)$, then $\rho_\theta(u) = \rho_\theta(v)$.*

As opposed to the result mentioned in Theorem 5, $b'(p,q)$ is not optimal for all (p,q) with $p > q \geq 3\gcd(p,q)$. The presentations of u,v given in Theorem 7 make it possible to distinguish the non-optimal cases from the optimal cases.

Corollary 3. *For p,q with $d = \gcd(p,q)$ and $p > q \geq 3d$, $b'(p,q)$ is optimal for (p,q) if and only if $(p/d,q/d)$ is either $(m(2i+3)+2, 2i+3)$ or $(4m(i+1)+2i+3, 4(i+1))$ for some $m \geq 1$ and $i \geq 0$.*

Recall that the bound given by the classical Fine and Wilf's theorem is strongly optimal, i.e., for an arbitrary pair (p,q), there exists a word of length $p+q-\gcd(p,q)-1$ with periods p,q but without $\gcd(p,q)$ as its period; furthermore if p and q are coprime, then such a word is unique up to letter renaming. In contrast, the bound $b'(p,q)$ is not strongly optimal. Indeed, Corollary 3 says that there do not exist u,v of respective lengths $9,5$ with $\mathrm{BCP}_\theta(u,v) \neq \emptyset$. On the other hand, we can obtain an analogous result about the uniqueness of boundary common prefixes based on words of coprime lengths up to letter-renaming. For this purpose, let us construct all the boundary common prefixes according to Theorem 7 as well as its proof. The first presentation of u,v in this theorem is $u = (\alpha\beta(\beta\alpha)^i\beta)^m\alpha\beta$ and $v = \alpha\beta(\beta\alpha)^i\beta$ for some $m \geq 1$, $i \geq 0$, and $(\alpha,\beta) \in T_1 \cup T_2 \cup T_3$. The proof of this theorem says that the only boundary common prefix which can be generated based on u and v is the maximal common prefix between $u\theta(u)^2$ and $v^n v_{n+1}$, which is:

$$(\alpha\beta(\beta\alpha)^i\beta)^{2m+1}\alpha\beta x, \tag{16}$$

where x is the maximal common prefix between α and β (see the definition of T_1, T_2, T_3). In the similar fashion, for the second presentation in the theorem $u = (\alpha(\beta\alpha)^i(\alpha\beta)^{i+1}\alpha)^m\alpha(\beta\alpha)^i\alpha\beta$ and $v = \alpha(\beta\alpha)^i(\alpha\beta)^{i+1}\alpha$, $u^3 \wedge v^{m+1}\theta(v)^{m+1}v_{2m+3}$ is the only boundary common prefix constructable from u and v, and it is:

$$(\alpha(\beta\alpha)^i(\alpha\beta)^{i+1}\alpha)^{m+1}(\alpha(\beta\alpha)^{i+1}(\alpha\beta)^i\alpha)^{m+1}x, \tag{17}$$

where x is the maximal common prefix between α and β. Note that both presentations of u, v admit that $\gcd(|u|, |v|) = \gcd(|\alpha|, |\beta|) = |\alpha| = |\beta|$ due to the Euclidean algorithm. Therefore, all the boundary common prefixes which verify the optimality of $b'(p, q)$ for all the coprime pairs (p, q) for which $b'(p, q)$ is optimal can be obtained by choosing (α, β) in Eq. (16) and in Eq. (17) from

$$(\Sigma \times \Sigma) \cap (T_1 \cup T_2 \cup T_3) = \{(a, b) \mid a, b \in \Sigma, a \neq b, a = \theta(a), b = \theta(b)\}.$$

Consequently, for pairs of coprime integers, the next result holds, which is analogous to the uniqueness result just mentioned.

Corollary 4. *Let (p, q) be a pair of coprime integers with $p > q$. Then all the boundary common prefixes based on words of respective lengths p, q are equal up to renaming.*

Note that this uniqueness result does not hold any more once the coprime assumption is taken out. This is because the choice of x in Eq. (16) and in Eq. (17) is arbitrary and also even if $\gcd(p, q) = 2$, there are two choices about (α, β) from T_2 or from T_3.

We conclude this section by defining two respective sets of boundary common prefixes thus obtained from Eq. (16) and Eq. (17) by limiting the choice of (α, β) only from $(\Sigma\times\Sigma)\cap(T_1\cup T_2\cup T_3)$. Due to the uniqueness mentioned in Corollary 4, we can set $\alpha = a$ and $\beta = b$ without loss of generality. As such, these sets are rather defined as:

$$\begin{aligned} \mathrm{S_o} &= \{(ab(ba)^i b)^{2m+1}ab \mid i \geq 0, m \geq 1\} \\ \mathrm{S_e} &= \{(a(ba)^i(ab)^{i+1}a)^{m+1}(a(ba)^{i+1}(ab)^i a)^{m+1} \mid i \geq 0, m \geq 1\}. \end{aligned}$$

The aim of the next section is to discuss the relationship between the words in $\mathrm{S_o} \cup \mathrm{S_e}$ and Sturmian words.

4 Sturmian words

It is known that for an arbitrary pair of integers (p, q) with $p > q > \gcd(p, q)$, there is a word of length $p + q - \gcd(p, q) - 1$ which has p, q as its periods but

$\gcd(p,q)$ is not its period, and hence, the bound $p+q-\gcd(p,q)$ for the Fine and Wilf's theorem is strongly optimal. Furthermore, all of these words can be constructed based on a (binary) word with two coprime periods p,q whose length is $p+q-2$. It is known that the set of all such basic words, denoted by PER, is closely related to Sturmian words. (Infinite) Sturmian words are (one-sided) infinite words which are not ultimately-periodic, and whose number of factors of length n is minimal $(n+1)$ for any $n \geq 1$. *Finite Sturmian words* are any factors of an infinite Sturmian word. Let St be the set of finite Sturmian words. By $F(w)$ we denote the set of all infixes (factors) of w and we can extend this notation to the set of words L as $F(L) = \bigcup_{w \in L} F(w)$. de Luca and Mignosi proved in [11] that $St = F(\mathrm{PER})$, that is, a binary word with two coprime periods whose length is the sum of these two periods minus 2 is a finite Sturmian word.

The aim of this section is to characterize $\mathrm{S_o}$ and $\mathrm{S_e}$, which correspond to PER for the optimal bound of Fine and Wilf's theorem, by finite Sturmian words.

A finite Sturmian word is called *standard* if it appears as an intermediate product (see Definition 1) when constructing an infinite Sturmian word using a procedure called *standard method.*

Definition 1 ([11])**.** Let $\Sigma = \{a,b\}$. The infinite sequence of pairs of words (A_n, B_n), $n \geq 0$, is constructed in the following manner. Set $(A_0, B_0) = (a,b)$. For any $n \geq 0$, the pair (A_{n+1}, B_{n+1}) is obtained from (A_n, B_n) by using one of the following two rules:

1. $(A_{n+1}, B_{n+1}) = (A_n, A_nB_n)$, or
2. $(A_{n+1}, B_{n+1}) = (B_nA_n, B_n)$.

The elements of $\{A_n, B_n \mid n \geq 0\}$ are the standard finite Sturmian words.

A property, called R in [11], plays an important role here. A word $w \in \Sigma^+$ is said to satisfy R if there exist palindromes x, y, z such that $w = zab = xy$. It was proved that a word with the property R is a standard Sturmian word.

Proposition 5 ([11])**.** *If a word has the property R, then it is a standard Sturmian word.*

Lemma 9. *For a word w in $\mathrm{S_e}$, the words wab and wba satisfy R.*

Proof. Let $w = ((ab)^i a(ab)^{i+1}a)^{m+1}(a(ba)^{i+1}a(ba)^i)^{m+1}$ for some $i \geq 0$ and $m \geq 1$. Since any word in $\mathrm{S_e}$ is a palindrome, it is enough, for our purpose, to show that wab is a product of two palindromes. In fact, wab can be split into $((ab)^i a(ab)^{i+1}a)^{m+1}a(ba)^i$ and $baa(ba)^i(a(ba)^{i+1}a(ba)^i)^m ab$, which are palindromes. Thus, wab satisfies R. In the same fashion, $wba = ((ab)^i a(ab)^{i+1}a)^m(ab)^i a \cdot (ab)^{i+1}a(a(ba)^{i+1}a(ba)^i)^{m+1}ba$, and hence, wba satisfies R. □

Corollary 5. *For a word w in $\mathrm{S_e}$, wab and wba are standard Sturmian words.*

Thus, we can see that all words in $\mathrm{S_e}$ are finite Sturmian words. Combining Lemma 9 with the following result obtained in [11], which relates a word with the property R with PER, we can obtain a stronger result than this.

Lemma 10 ([11])**.** *Let $u = zab = xy$ for some palindromes x, y, z. If z contains at least two letters, then z has the periods $p = |x| + 2$ and $q = |y| - 2$ such that* $\gcd(p, q) = 1$.

Corollary 6. $\mathrm{S_e} \subseteq \mathrm{PER}$.

Having considered $\mathrm{S_e}$, now we turn our attention to $\mathrm{S_o}$. In a similar manner as above, we can prove that any element of $\mathrm{S_o}$ is a finite Sturmian word.

Lemma 11. *For a word w in $\mathrm{S_o}$, there exists a word $u \in \Sigma^+$ such that uw has the property R.*

Proof. Let $w = (ab(ba)^i b)^m ab$ for some $i \geq 0$ and $m \geq 1$. When $i = 0$, let $u = bb$. Then $uw = bb(abb)^m ab$. Since $bb(abb)^m$ is a palindrome and uw can be written as a product of b and $(bab)^{m+1}$, uw satisfies R.

When $i \geq 1$, let $u = (ba)^{i-1}b$ so that $uw = (ba)^{i-1}b(abba(ba)^{i-1}b)^m ab$. Note that uw has as prefix of length $|uw|-2$ a θ-palindrome, and of length $|uw|-2$, and it can be split into two palindromes $(ba)^{i-1}bab$ and $ba(ba)^{i-1}b(abba(ba)^{i-1}b)^{m-1}ab$. Thus, we can say that uw has the property R. $\square$

Corollary 7. *Words in $\mathrm{S_o}$ are finite Sturmian words.*

Now we know that all the words in $\mathrm{S_e} \cup \mathrm{S_o}$ are finite Sturmian words. We shall differentiate these two sets with respect to PER. Recall that any element of $\mathrm{S_e}$ is included in PER (Corollary 6). On the contrary, $\mathrm{S_o}$ and PER are disjoint. This is because an element of PER has been proved to be a palindrome [11], while any element of $\mathrm{S_o}$ is not.

Proposition 6. $\mathrm{S_o} \cap \mathrm{PER} = \emptyset$.

To summarize this discussion, Figure 9 clarifies the inclusion relations among the set of finite Sturmian words, PER, $\mathrm{S_e}$, and $\mathrm{S_o}$. Due to the fact that a factor of a word in St also belongs to St, $St \supseteq F(\mathrm{S_e} \cup \mathrm{S_o})$ holds, but this inclusion relation is in fact proper. For instance, a word $aaabaaaabaaa$ is of length 12 and has two periods 9 and 5 while $\gcd(9, 5)$ is not its period. Hence, this word is in $\mathrm{PER} \subseteq St$, while it is not in $F(\mathrm{S_e} \cup \mathrm{S_o})$ because no word in $\mathrm{S_e} \cup \mathrm{S_o}$ has a continuous run of the same four letters as its infix. Moreover, the infix $baaaab$ of this example word shows $\mathrm{PER} \cup (\mathrm{S_e} \cup \mathrm{S_o}) \subsetneq St$. For the reason mentioned above, it is clear that $baaaab \in F(\mathrm{PER})$ but $baaaab \notin \mathrm{S_e} \cup \mathrm{S_o}$. In addition, $baaaab$ has only one period which is strictly smaller than its length, and hence, $baaaab \notin \mathrm{PER}$.

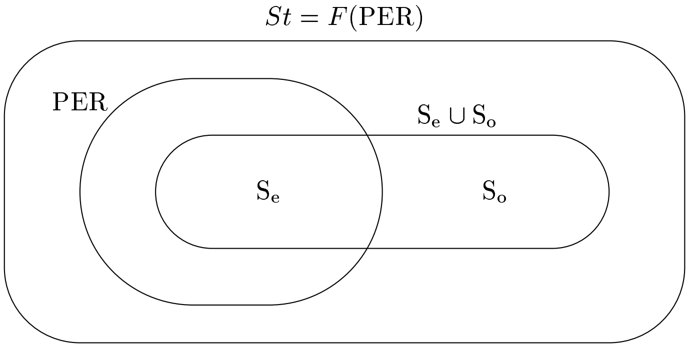

Figure 9: The set St of finite Sturmian words, PER, $\mathrm{S_e}$, and $\mathrm{S_o}$

5 Concluding remarks

In this paper, we improved the bound for the extension of the Fine and Wilf's theorem of [9] from $b(p,q)$ to $b'(p,q) = b(p,q) - \lfloor \gcd(p,q)/2 \rfloor$. The complete characterization of boundary common prefixes given here allows us to distinguish the case when this improved bound is optimal in terms of the lengths of given words. In particular, this improved bound is optimal for any (p,q) with $p > q = 2\gcd(p,q)$. We also discussed the relationship between finite Sturmian words and the boundary common prefixes.

One open case is finding optimal bound for a pair (p,q) with $d = \gcd(p,q)$ and $p > q \geq 3d$, for which the improved bound $b'(p,q) = 2p + q - d - \lfloor d/2 \rfloor$ is not optimal due to Corollary 3. Note that for such (p,q), the bound $b'(p,q) - 1$ remains good, while in Section 3.2, $2p + \lceil d/2 \rceil - 1$ was proved not to be good. Thus, the optimal bound for such (p,q) exists between $2p + \lceil d/2 \rceil$ and $b'(p,q) - 1$.

References

[1] J. Berstel and L. Boasson. Partial words and a theorem of Fine and Wilf. *Theoretical Computer Science*, 218(1):135–141, 1999.

[2] F. Blanchet-Sadri and R. A. Hegstrom. Partial words and a theorem of Fine and Wilf revisted. *Theoretical Computer Science*, 270:401–419, 2002.

[3] M. Gabriella Castelli, F. Mignosi, and A. Restivo. Fine and Wilf's theorem for three periods and a generalization of Sturmian words. *Theoretical Computer Science*, 218(1):83–94, 1999.

[4] S. Cautis, F. Mignosi, J. Shallit, M.-w. Wang, and Soroosh Yazdani. Periodicity, morphisms, and matrices. *Theoretical Computer Science*, 295:107–121, 2003.

[5] C. Choffrut and J. Karhumäki. Combinatorics of words. In G. Rozenberg and A. Salomaa, editors, *Handbook of Formal Languages*, volume 1, pages 329–438. Springer-Verlag, Berlin-Heidelberg-New York, 1997.

[6] S. Constantinescu and L. Ilie. Generalized Fine and Wilf's theorem for arbitrary number of periods. *Theoretical Computer Science*, 339(1):49–60, 2005.

[7] S. Constantinescu and L. Ilie. Fine and Wilf's theorem for Abelian periods. *Bulletin of the EATCS*, 89:167–170, June 2006.

[8] E. Czeizler, E. Czeizler, L. Kari, and S. Seki. An extension of the Lyndon Schützenberger result to pseudoperiodic words. In V. Diekert and D. Nowotka, editors, *Proc. DLT09*, volume LNCS 5583 of *Lecture Notes in Computer Science*, pages 183–194, Berlin, 2009. Springer-Verlag.

[9] E. Czeizler, L. Kari, and S. Seki. On a special class of primitive words. *Theoretical Computer Science*, 411(3):617–630, 2010.

[10] A. de Luca and A. De Luca. Pseudopalindrome closure operators in free monoids. *Theoretical Computer Science*, 362:282–300, 2006.

[11] A. de Luca and F. Mignosi. Some combinatorial properties of Sturmian words. *Theoretical Computer Science*, 136:361–385, 1994.

[12] N. J. Fine and H. S. Wilf. Uniqueness theorem for periodic functions. *Proceedings of the American Mathematical Society*, 16(1):109–114, February 1965.

[13] J. Justin. On a paper by Castelli, Mignosi, Restivo. *RAIRO - Theoretical Informatics and Applications*, 34:373–377, 2000.

[14] L. Kari, B. Masson, and S. Seki. Properties of pseudo-primitive words and their applications. Submitted, available at `http://hal.archives-ouvertes.fr/hal-00458695/fr/`, 2009.

[15] F. Mignosi, A. Restivo, and P. V. Silva. On Fine and Wilf's theorem for bidimensional words. *Theoretical Computer Science*, 292:245–262, 2003.