An Extension of the Lyndon and Schützenberger Result to Pseudoperiodic Words

Elena Czeizler, Eugen Czeizler, Lila Kari, and Shinnosuke Seki

January 13, 2009

Abstract

One of the particularities of information encoded as DNA strands is that a string u contains basically the same information as its Watson-Crick complement, denoted here as $\theta(u)$. Thus, any expression consisting of repetitions of u and $\theta(u)$ can be considered in some sense periodic. In this paper we give a generalization of Lyndon and Schützenberger's classical result about equations of the form $u^l = v^n w^m$, to cases where both sides involve repetitions of words as well as their complements. Our main results show that, for such extended equations, if $l \geq 5, n, m \geq 3$, then all three words involved can be expressed in terms of a common word t and its complement $\theta(t)$. Moreover, if $l \geq 5$, then $n = m = 3$ is an optimal bound. We also obtain a complete characterization of all possible overlaps between two expressions that involve only some word u and its complement $\theta(u)$.

1 Introduction

This paper is a theoretical study of *pseudoperiodic words*, notion motivated by the properties of information encoded as DNA strands for DNA computing purposes. Informally, a word is pseudoperiodic if it consists of repeated occurrences of another word and/or the image of that word under an antimorphic involution. The notion of antimorphic involution is the mathematical formalization of the Watson-Crick complementarity of DNA single strands, as detailed below.

DNA, in its single-stranded form, is a linear chain made up of four different types of units, called nucleotides, and can thus be viewed to a first approximation as a word over the four-letter alphabet $\{A, C, G, T\}$. A DNA single strand has an orientation, with one end known as the 5' end, and the other as the 3' end, based on their chemical properties. By convention, a word over the DNA alphabet represents the corresponding DNA single strand in the 5'-3' orientation. Another crucial feature is the Watson-Crick (WK) complementarity: A is complementary to T, and G is to C. Two complementary DNA single strands with opposite orientation will bind to each other by bonds between their individual bases to form a helical DNA double strand. The Watson-Crick complementarity operation is a fundamental bio-operation in DNA Computing experiments [1]. In this paper we investigate the consequences of Watson-Crick complementarity on the notion of periodicity of words.

Periodicity properties of words are among the main theoretical tools used in pattern-matching algorithms, see e.g. [2] and [3]. Recall that a word u is called *periodic* if there exists another word v, shorter than u, such that u is a prefix of v^i for some $i \geq 2$. Moreover, the way in which a word can be decomposed, and whether two words are powers of a common word are two questions which have been widely investigated in language theory, see, e.g., [4] and [5]. However, when dealing with DNA strands, note that a string u encodes the same information as its complement, $\theta(u)$, where θ denotes the WK complementarity function or its mathematical formalization as an antimorphic involution. In this context, e.g., the word $u^m\theta(u)^n$ can be considered periodic, since it consists of repetitions of the same information unit. (Other generalizations of the notion of periodicity include, e.g., the "weak periodicity" of [6] whereby a word is called *weakly periodic* if it consists of repetitions of words with the same Parikh vector. This type of period was called *abelian period* in [7].) In [8] the Fine and Wilf Theorem – one of the fundamental periodicity results on words, see e.g. [4] and [5] – was extended to deal with expressions involving both a word and its image under an antimorphic involution.

Here we extend another central periodicity result, due to Lyndon and Schützenberger, [9]. (See also [10] and Chapter 5 from [5] for some shorter proofs and [11] and [12] for some other generalizations.) The original result states that, if the concatenation of two periodic words v^n and w^m can be expressed in terms of a third period u, i.e., $u^l = v^n w^m$, for some $n, m, l \geq 2$, then all three words u, v, and w can be expressed in terms of a common word t, i.e., $u, v, w \in \{t\}^*$.

In our generalization, we consider repetitions involving both a word and its image under θ, i.e., the equation $\alpha(u, \theta(u)) = \beta(v, \theta(v)) \cdot \gamma(w, \theta(w))$ where $\alpha(u, \theta(u)) \in \{u, \theta(u)\}^l$, $\beta(v, \theta(v)) \in \{v, \theta(v)\}^n$, and $\gamma(w, \theta(w)) \in \{w, \theta(w)\}^m$ with $l, n, m \geq 2$. A conclusion of our main results is that, whenever $l \geq 5$, $n, m \geq 3$ we have $u, v, w \in \{t, \theta(t)\}^+$ for some word t, i.e., all three words can be expressed using a common word t and its image $\theta(t)$. Moreover, we provide examples showing that, for any $l \geq 5$, $n = m = 3$ is an optimal bound. In the case when $l = 3$ or $l = 4$, the problem of finding optimal bounds remains open. Our proofs are not generalizations of the methods used in the classical case, since one of the main properties used therein, i.e., the fact that the conjugate of a primitive word is still primitive, cannot be used here.

In our search for these bounds, we also obtain a characterization of all possible overlaps of two expressions $\alpha(v, \theta(v)), \beta(v, \theta(v)) \in \{v, \theta(v)\}^+$. In particular, we show that, contrary to the classical case (when the two expressions involve only a word v, but not its image under θ), the equality $\alpha(v, \theta(v)) \cdot x = y \cdot \beta(v, \theta(v))$ with x and y shorter than v, does not always force a decomposition of v of the form $v \in \{t, \theta(t)\}^+$ for some word t.

The paper is organized as follows. In Section 2, we fix our terminology and recall some known results. In Section 3, we provide the characterization of all possible overlaps of the form $\alpha(v, \theta(v)) \cdot x = y \cdot \beta(v, \theta(v))$ with

$\alpha(v,\theta(v)),\beta(v,\theta(v)) \in \{v,\theta(v)\}^+$ and x,y shorter than v. Finally, in Section 4 we provide our extension of Lyndon and Schützenberger's result.

2 Preliminaries

Let Σ be a finite alphabet. We denote by Σ^* the set of all finite words over Σ, by ϵ the empty word, and by Σ^+ the set of all nonempty finite words. The catenation of two words $u,v \in \Sigma^*$ is denoted by either uv or $u \cdot v$. The *length* of a word $w \in \Sigma^*$, denoted by $|w|$, is the number of letters occurring in it. We say that u is a *factor* (a *prefix*, a *suffix*, resp.) of v, if $v = t_1ut_2$ ($v = ut_2$, $v = t_1u$, resp.) for some $t_1,t_2 \in \Sigma^*$. We denote by $\mathrm{Pref}(v)$ (resp. $\mathrm{Suff}(v)$) the set of all prefixes (resp. suffixes) of the word v. We say that two words u and v overlap if $ux = yv$ for some $x,y \in \Sigma^*$ with $|x| < |v|$. An integer $p \geq 1$ is a *period* of a word $w = a_1 \ldots a_n$, with $a_i \in \Sigma$ for all $1 \leq i \leq n$, if $a_i = a_{i+p}$ for all $1 \leq i \leq n-p$.

A word $w \in \Sigma^+$ is called *primitive* if it cannot be written as a power of another word; that is, if $w = u^n$ then $n = 1$ and $w = u$. For a word $w \in \Sigma^+$, the shortest $u \in \Sigma^+$ such that $w = u^n$ for some $n \geq 1$ is called the *primitive root* of the word w and is denoted by $\rho(w)$. The following is a well-known property of primitive words, see, e.g., [4], [5].

Proposition 1. *Let $u \in \Sigma^+$ be a primitive word. If $u^2 = xuy$, then either $x = \epsilon$ or $y = \epsilon$.*

A mapping $\theta : \Sigma^* \to \Sigma^*$ is called an *antimorphism* if for any words $u,v \in \Sigma^*$, $\theta(uv) = \theta(v)\theta(u)$. Moreover, a mapping $\theta : \Sigma^* \to \Sigma^*$ is called an *involution* if, for all words $u \in \Sigma^*$, $\theta(\theta(u)) = u$. An antimorphic involution is a mathematical formalization of the WK complementarity of DNA single strands. Throughout this paper we will assume that θ is an antimorphic involution on a given alphabet Σ. A word $w \in \Sigma^*$ is called *θ-palindrome*, or *pseudopalindrome* if θ is not specified, if $w = \theta(w)$ (see [13] and [14]).

The notions of periodic and primitive words were extended in [8] in the following way. A word $w \in \Sigma^+$ is θ-periodic if $w = w_1 \ldots w_k$ for some $k \geq 2$ and words $t, w_1, \ldots, w_k \in \Sigma^+$ such that $w_i \in \{t, \theta(t)\}$ for all $1 \leq i \leq k$. Following [14], in less precise terms, a word which is θ-periodic with respect to a given but unspecified involutory morphism θ will be also called *pseudoperiodic*. The word t in the definition of a θ-periodic word w is called a θ-period of w. We call a word $w \in \Sigma^+$ *θ-primitive* if it is not θ-periodic. The set of θ-primitive words is strictly included in the set of primitive ones, see [8]; for instance, if we take $a \neq b$ and $\theta(a) = b$, $\theta(b) = a$, then the word ab is primitive, but not θ-primitive. We define *the θ-primitive root of* w, denoted by $\rho_\theta(w)$, as the shortest word t such that $w = w_1 \ldots w_k$ for some $k \geq 1$, $w_i \in \{t, \theta(t)\}$ for all $1 \leq i \leq k$, and $w_1 = t$. Note that if w is θ-primitive, then $\rho_\theta(w) = w$.

We say that two words u and v *commute* if $uv = vu$. We can characterize the commutation of two words in terms of primitive roots, see, e.g., [4], [5].

Theorem 2. *For $u, v \in \Sigma^*$, the following conditions are equivalent: i) u and v commute; ii) u and v satisfy a nontrivial relation, i.e., a nontrivial equation over two variables without constants; iii) u and v have the same primitive root.*

Two words u and v are said to be *conjugate* if there exist words x and y such that $u = xy$ and $v = yx$. In other words, v can be obtained via a cyclic permutation of u. The next known result, see, e.g., [4], [5], characterizes the conjugacy of two words.

Theorem 3. *Let $u, v \in \Sigma^+$. Then, the following conditions are equivalent: i) u and v are conjugate; ii) there exists a word z such that $uz = zv$; moreover, this holds if and only if $u = pq$, $v = qp$, and $z = (pq)^i p$, for some $p, q \in \Sigma^*$ and $i \geq 0$; iii) the primitive roots of u and v are conjugate.*

The following periodicity result is due to Lyndon and Schützenberger, [9].

Theorem 4. *If words u, v, w satisfy the relation $u^l = v^n w^m$ for some positive integers $l, n, m \geq 2$, then they are all powers of a common word, i.e., there exists a word t such that $u, v, w \in \{t\}^*$.*

The Fine and Wilf theorem, in its form for words, see [4], [5], illustrates another fundamental periodicity property. It states that if two words $u, v \in \Sigma^*$, with $n = |u|$, $m = |v|$, $d = \gcd(n, m)$, are such that if two powers u^i and v^j have a common prefix of length at least $n + m - d$, then u and v are powers of a common word, where $\gcd(n, m)$ denotes, as usual, the *greatest common divisor of n and m*. Moreover, the bound $n + m - d$ is optimal. The original result of Fine and Wilf, [15], was formulated for sequences of real numbers.

This theorem was extended in [8] for the case when instead of dealing with powers of two words u^i and v^j, we look at expressions over $\{u, \theta(u)\}$ and $\{v, \theta(v)\}$, respectively. Its weaker version, which will be very useful, is presented as well.

Theorem 5 ([8]). *Let $u, v \in \Sigma^+$ be two distinct words with $|u| > |v|$. If there exist two expressions $\alpha(u, \theta(u)) \in u\{u, \theta(u)\}^*$ and $\beta(v, \theta(v)) \in v\{v, \theta(v)\}^*$ having a common prefix of length at least $2|u| + |v| - gcd(|u|, |v|)$, then $\rho_\theta(u) = \rho_\theta(v)$. Moreover, the bound $2|u| + |v| - gcd(|u|, |v|)$ is optimal.*

Theorem 6 ([8]). *Let $u, v \in \Sigma^+$, $\alpha(u, \theta(u)) \in u\{u, \theta(u)\}^*$, and $\beta(v, \theta(v)) \in v\{v, \theta(v)\}^*$ such that $\alpha(u, \theta(u)) = \beta(v, \theta(v))$. Then $\rho_\theta(u) = \rho_\theta(v)$.*

The next two results, also from [8], will be very useful in our considerations.

Lemma 7 ([8]). *For $u, v \in \Sigma^*$, if $uv = \theta(uv)$ and $vu = \theta(vu)$, then there exists a word $t \in \Sigma^+$ such that $u, v \in \{t, \theta(t)\}^*$.*

Lemma 8 ([8]). *Let $v \in \Sigma^+$ be a θ-primitive word. Then, $\theta(v)vx = yv\theta(v)$ for some words $x, y \in \Sigma^*$ with $|x|, |y| < |v|$, if and only if $v = \theta(v)$ and $x = y = \epsilon$. Similarly, $v\theta(v)v = xv^2y$ for some $x, y \in \Sigma^*$ if and only if $v = \theta(v)$ and either $x = \epsilon$ or $y = \epsilon$.*

The following result will prove very useful in our future considerations.

Lemma 9. *Let $u \in \Sigma^+$ such that $u = xz = zy$ for some $x, y, z \in \Sigma^+$ with $x = \theta(x)$ and $y = \theta(y)$. Then $x, y, z, u \in \{t, \theta(t)\}^*$ for some $t \in \Sigma^+$.*

Proof. The equation $u = xz = zy$ implies that $x = pq$, $y = qp$, and $z = (pq)^j p$ for some $p, q \in \Sigma^*$ and $j \geq 0$. Since $x = \theta(x)$ and $y = \theta(y)$, we have $pq = \theta(pq)$ and $qp = \theta(qp)$. Then, Lemma 7 implies that there exists a word $t \in \Sigma^+$ such that $p, q \in \{t, \theta(t)\}^*$. □

3 Overlaps between θ-primitive Words

It is well known that a primitive word v cannot occur nontrivially inside v^2, see Proposition 1. Thus, two expressions v^i and v^j, with $i, j \geq 1$, cannot overlap nontrivially on a sequence longer than $|v|$. A natural question is whether we can have some nontrivial overlaps between two expressions $\alpha(v, \theta(v)), \beta(v, \theta(v)) \in \{v, \theta(v)\}^+$, when $v \in \Sigma^+$ is a θ-primitive word. In this section we completely characterize all such nontrivial overlaps, and, moreover, in each case we also give the set of all solutions of the corresponding equation.

We begin our analysis by giving two intermediate results.

Theorem 10. *Let $v \in \Sigma^+$ be a θ-primitive word and $\alpha(v, \theta(v)), \beta(v, \theta(v)) \in \{v, \theta(v)\}^+$ such that $\alpha(v, \theta(v)) \cdot x = y \cdot \beta(v, \theta(v))$, with $x, y \in \Sigma^+$, $|x|, |y| < |v|$. Then, v^2 and $\theta(v)^2$ cannot occur simultaneously neither in $\alpha(v, \theta(v))$ nor in $\beta(v, \theta(v))$.*

Proof. Suppose that both v^2 and $\theta(v)^2$ occur in $\alpha(v, \theta(v))$; the case when they both occur in $\beta(v, \theta(v))$ is symmetric. Moreover, since θ is an involution, we can suppose without loss of generality that v^2 occurs before $\theta(v)^2$, thus implying that $v^2\theta(v)$ is a factor in $\alpha(v, \theta(v))$. Since v (resp. $\theta(v)$) is primitive, the border between any two consecutive v's (resp. $\theta(v)$'s) falls inside a $\theta(v)$ (resp. v), see Figure 1.

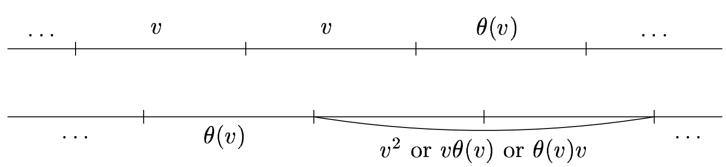

Figure 1: The case when $v^2\theta(v)$ is a factor in $\alpha(v, \theta(v))$

Thus, $v^2\theta(v)$ overlaps either with $\theta(v)v^2$ or with $\theta(v)v\theta(v)$ or with $\theta(v)^2 v$. In all three cases the nontrivial overlap between $v\theta(v)$ and $\theta(v)v$ contradicts the θ-primitivity of v, see Lemma 8. □

Theorem 11. *Let $v \in \Sigma^+$ be a θ-primitive word and $\alpha(v, \theta(v)), \beta(v, \theta(v)) \in \{v, \theta(v)\}^+$ such that $\alpha(v, \theta(v)) \cdot x = y \cdot \beta(v, \theta(v))$ for some $x, y \in \Sigma^+$ with*

$|x|, |y| < |v|$. *Then, neither $v\theta(v)v$ nor $\theta(v)v\theta(v)$ can occur either in $\alpha(v, \theta(v))$ or in $\beta(v, \theta(v))$.*

Proof. Suppose that $v\theta(v)v$ occurs in $\alpha(v, \theta(v))$. Since we assumed $x, y \in \Sigma^+$ and $|x|, |y| < |v|$, $v\theta(v)v$ contains as a proper factor an expression of length $2|v|$, $\gamma(v, \theta(v)) \in \{v, \theta(v)\}^2$, i.e., there exist some $p, q \in \Sigma^+$ such that $v\theta(v)v = p\gamma(v, \theta(v))q$. We already know, due to Lemma 8, that neither $v\theta(v)$ nor $\theta(v)v$ can be factors of $v\theta(v)v$. Thus, we have that either v^2 or $\theta(v)^2$ is a factor in $v\theta(v)v$. However, by Lemma 8, the first case is not possible. On the other hand, the second case contradicts the primitivity of $\theta(v)$, since $\theta(v)$ would occur as a factor of $\theta(v)^2$. Thus, $v\theta(v)v$ cannot occur in $\alpha(v, \theta(v))$. All the other cases can be proved similarly. □

As an immediate consequence of the previous two theorems, for a given θ-primitive word v, if $\alpha(v, \theta(v)) \cdot x = y \cdot \beta(v, \theta(v))$ with $x, y \in \Sigma^+$, $|x|, |y| < |v|$, then $\alpha(v, \theta(v))$ and $\beta(v, \theta(v))$ can be only of the following types v^i, $v^i\theta(v)$, $v\theta(v)^i$, $\theta(v)^i$, $\theta(v)^i v$, or $\theta(v)v^i$ for some $i \geq 1$. The next result refines this characterization further.

Table 1: Characterization of possible proper overlaps of the form $\alpha(v, \theta(v)) \cdot x = y \cdot \beta(v, \theta(v))$

Equation	Solution
$v^i x = y\theta(v)^i, i \geq 1$	$v = yp, x = \theta(y), p = \theta(p)$, and whenever $i \geq 2, y = \theta(y)$
$v^i x = y\theta(v)^{i-1}v, i \geq 1$	$v = (pq)^{j+1}p, x = qp, y = pq$, and whenever $i \geq 2, p = \theta(p), pq = \theta(q)p$
$v\theta(v)x = yv\theta(v)$,	$v = (pq)^{j+1}p, x = \theta(pq), y = pq$, with $j \geq 0, qp = \theta(qp)$;
$v\theta(v)^i x = yv^{i+1}, i \geq 1$	$v = r(tr)^n(rt)^{m+n}r, y = (rt)^n r(rt)^m, x = (rt)^{m+n}r$, where $r = \theta(r), t = \theta(t), n \geq 0, m \geq 1$
$v\theta(v)^i x = yv^i\theta(v), i \geq 2$	$v = (rt)^n r(rt)^{m+n}r, y = (rt)^n r(rt)^m, x = (tr)^m r(tr)^n$, where $r = \theta(r), t = \theta(t), n \geq 0, m \geq 1$

Theorem 12. *Let $v \in \Sigma^+$ be a θ-primitive word. Then, the only possible proper overlaps of the form $\alpha(v, \theta(v)) \cdot x = y \cdot \beta(v, \theta(v))$ with $\alpha(v, \theta(v)), \beta(v, \theta(v)) \in \{v, \theta(v)\}^+$, $x, y \in \Sigma^+$ and $|x|, |y| < |v|$ are given in Table 1 (modulo a substitution of v by $\theta(v)$) together with the characterization of their sets of solutions.*

Proof. Since θ is an involution, we can assume without loss of generality that $\alpha(v, \theta(v))$ starts with v. Then, due to the previous observation we know that $\alpha(v, \theta(v)) \in \{v^i, v^i\theta(v), v\theta(v)^i \mid i \geq 1\}$.
Case 1: Suppose first that $\alpha(v, \theta(v)) = v^i$ for some $i \geq 1$. Since v is θ-primitive, $v^i x = y\beta(v, \theta(v))$, and $|y|, |x| < |v|$, the border between any two consecutive v's falls inside a $\theta(v)$, see Figure 2; otherwise v would occur inside v^2 which would

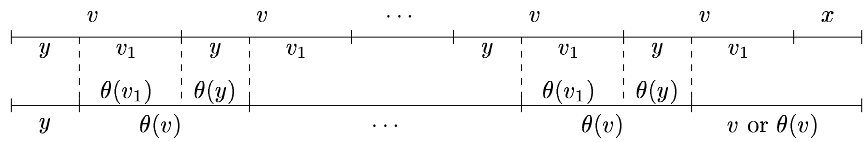

Figure 2: The case when $\alpha(v, \theta(v)) = v^i$

contradict its primitivity. Thus, $\beta(v, \theta(v)) \in \{\theta(v)^i, \theta(v)^{i-1}v\}$. Then, we can write $v = yv_1$ and $\theta(v) = \theta(v_1)\theta(y)$.

Suppose first that $\beta(v, \theta(v)) = \theta(v)^i$. Then, we immediately obtain $v_1 = \theta(v_1)$ and, if $i \geq 2$ also $y = \theta(y)$. Moreover, if we look at the end of the two sides of the equation $v^i x = y\theta(v)^i$, we also obtain that $x = \theta(y)$. Thus, a proper overlap of the form $v^i x = y\theta(v)^i$ with v θ-primitive is possible, and, moreover, the set of all solutions of this equation is characterized by the following formulas: $v = yv_1$ and $x = \theta(y)$, where $v_1 = \theta(v_1)$ and $y = \theta(y)$ whenever $i \geq 2$.

Suppose now that $\beta(v, \theta(v)) = \theta(v)^{i-1}v$. If we look at the end of the two sides of the equation, then we obtain $v = v_1 x$. Thus, $v = yv_1 = v_1 x$, implying that there exist some $p, q \in \Sigma^*$ and $j \geq 0$ such that $y = pq$, $x = qp$, $v_1 = (pq)^j p$, and $v = (pq)^{j+1}p$. If $q = \epsilon$, then $v = p^{j+2}$ which contradicts the primitivity of v. If $p = \epsilon$, then $v = q^{j+1}$ which either contradicts the primitivity of v or, when $j = 0$, implies that $v = y$ contradicting our assumption that $|y| < |v|$. Thus, we can suppose $p, q \in \Sigma^+$. Now, if $i \geq 2$, then $v_1 = \theta(v_1)$ and $y = \theta(y)$, i.e., $pq = \theta(pq)$, $p = \theta(p)$. If $j \geq 1$, then also $q = \theta(q)$, which contradicts the primitivity of v. Thus, if $i \geq 2$, then we must have $j = 0$. To conclude, a proper overlap of the form $v^i x = y\theta(v)^{i-1}v$ with v θ-primitive is possible, and, moreover the set of all solutions of this equation is characterized by the following formulas: $v = (pq)^{j+1}p$, $y = pq$, and $x = qp$; moreover, if $i \geq 2$ then $p = \theta(p)$ and $pq = \theta(q)p$.

Case 2: Suppose now that $\alpha(v, \theta(v)) = v^i\theta(v)$, for some $i \geq 1$. If $i \geq 2$, then $\beta(v, \theta(v))$ has to start with $\theta(v)^{i-1}$, since otherwise it would contradict the primitivity of v. If this $\theta(v)^{i-1}$ is followed by v, see Figure 3, then $v\theta(v)$ overlaps with $\theta(v)v$ with the overlap properly longer than v. Then Lemma 8

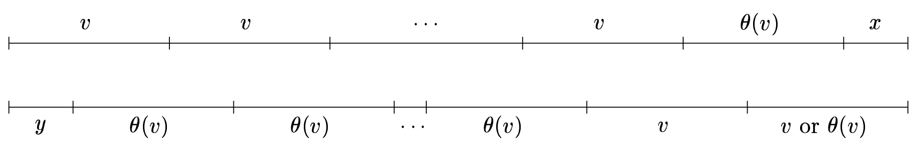

Figure 3: The case when $\alpha(v, \theta(v)) = v^i\theta(v)$

leads to a contradiction. Hence, $\beta(v, \theta(v))$ starts with $\theta(v)^i$, see Figure 4. But then, $\beta(v, \theta(v))$ can end neither with v due to Lemma 8, nor with $\theta(v)$ due to Proposition 1. Thus we must have $i = 1$. Then Proposition 1 and Lemma 8

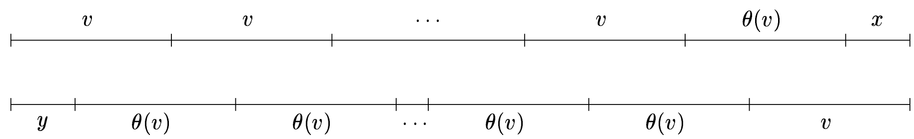

Figure 4: The case when $\alpha(v,\theta(v)) = v^i\theta(v)$

imply that $\beta(v,\theta(v))$ starts with v. If $\beta(v,\theta(v)) = v^2$, see Figure 5 a), then we can write $v = yv_1 = v_1v_2$, implying that $y = pq$, $v_2 = qp$, $v_1 = (pq)^jp$, and $v = (pq)^{j+1}p$ for some $j \geq 0$ and $p, q \in \Sigma^*$; moreover, just as before we can suppose again $p, q \in \Sigma^+$. Also, we immediately obtain $x = v_2 = \theta(v_2)$ and

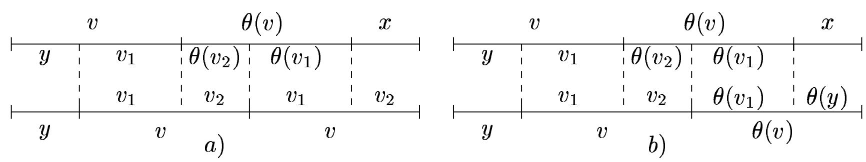

Figure 5: The equations: a) $v\theta(v)x = yv^2$ and b) $v\theta(v)x = yv\theta(v)$

$v_1 = \theta(v_1)$. If $j \geq 1$, then we have $p = \theta(p)$, $q = \theta(q)$, and since $qp = \theta(qp)$, this implies that $\rho(p) = \rho(q)$, contradicting the primitivity of v. If $j = 0$, then we have $p = \theta(p)$ and $qp = p\theta(q)$, which implies that $p = r(tr)^n$ and $q = (rt)^m$, for some $n \geq 0$, $m \geq 1$, $r = \theta(r)$, and $t = \theta(t)$, see [16]. Thus, a proper overlap of the form $v\theta(v)x = yv^2$ with v θ-primitive is possible, and, moreover, the set of all solutions of this equation is characterized by the following formula: $v = (rt)^nr(rt)^{m+n}r$, $x = (rt)^{m+n}r$, and $y = (rt)^nr(rt)^m$. The last case to consider is when $\beta(v,\theta(v)) = v\theta(v)$, see Figure 5 b). Then, we can write $v = yv_1 = v_1v_2$ and we obtain immediately $x = \theta(y)$ and $v_2 = \theta(v_2)$. Thus, a proper overlap of the form $v\theta(v)x = yv\theta(v)$, with v θ-primitive, is possible, and moreover the set of all solutions of this equation is characterized by the following formulas: $v = (pq)^{j+1}p$, $y = pq$, $x = \theta(pq)$, and $qp = \theta(qp)$.
Case 3: Suppose now that $\alpha(v,\theta(v)) = v\theta(v)^i$, for some $i \geq 2$; the case when $i = 1$ was already considered before. Since $\theta(v)$ is primitive, the border between any two $\theta(v)$'s falls inside v. If $\beta(v,\theta(v))$ starts with $\theta(v)$, then this $\theta(v)$ could not be followed by either v due to Lemma 8 or $\theta(v)$ due to Proposition 1. Therefore $\beta(v,\theta(v))$ has to begin with v, and moreover $\beta(v,\theta(v)) \in v^i\{v,\theta(v)\}$, see Figure 6. As $v\theta(v)x = yv^2$ in Case 2, we can write $v = yv_1 = v_1v_2$ with $y = (rt)^nr(rt)^m$, $v_1 = (rt)^nr$, and $v_2 = (rt)^{m+n}r$. Thus, proper overlaps of the forms both $v\theta(v)^ix = yv^{i+1}$ and $v\theta(v)^ix = yv^i\theta(v)$, with v θ-primitive, are possible. For the former equation, $x = v_2 = (rt)^{m+n}r$, while for the latter, $x = \theta(y) = (tr)^mr(tr)^n$. Note that the set of all solutions of $v\theta(v)x = yv^2$ and that of $v\theta(v)^ix = yv^{i+1}$ are characterized by the exactly same formulas. In contrast, the set of all solutions of $v\theta(v)x = yv\theta(v)$ and that of $v\theta(v)^ix = yv^i\theta(v)$

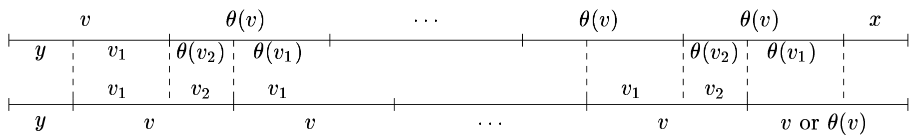

Figure 6: The case when $\alpha(v, \theta(v)) = v\theta(v)^i$ and $\beta(v, \theta(v))$ starts with v

are formulated in different ways. □

4 An Extension of Lyndon and Schützenberger's Result

For $u, v, w \in \Sigma^+$, let us consider some expressions $\alpha(u, \theta(u)) \in \{u, \theta(u)\}^+$, $\beta(v, \theta(v)) \in \{v, \theta(v)\}^+$, and $\gamma(w, \theta(w)) \in \{w, \theta(w)\}^+$ satisfying the equation $\alpha(u, \theta(u)) = \beta(v, \theta(v)) \cdot \gamma(w, \theta(w))$. Assume that for some positive integers $l, n, m \geq 0$, $|\alpha(u, \theta(u))| = l \times |u|$, $|\beta(v, \theta(v))| = n \times |v|$, and $|\gamma(w, \theta(w))| = m \times |w|$. We say that the triple (l, n, m) ***imposes θ-periodicity on*** u, v, w, (or shortly, ***imposes θ-periodicity***), if the equation $\alpha(u, \theta(u)) = \beta(v, \theta(v)) \cdot \gamma(w, \theta(w))$ admits only solutions of the form $u, v, w \in \{t, \theta(t)\}^*$ for some word $t \in \Sigma^+$. Note that, if (l, n, m) imposes θ-periodicity, then so does (l, m, n), and vice versa.

In the classical case of the equation $u^l = v^n w^m$, Lyndon and Schützenberger's result (Theorem 4) states that any triple (l, n, m) with $l, n, m \geq 2$ imposes classical periodicity on u, v, w, with the same period. In this section we extend this result by considering the more general equation $\alpha(u, \theta(u)) = \beta(v, \theta(v)) \cdot \gamma(w, \theta(w))$. Note that the fact that a certain triple (l, n, m) imposes θ-periodicity does not imply that (l', n', m') imposes θ-periodicity for $l' > l$ or $n' > n$ or $m' > m$.

The results of this section are summarized in Table 2. Overall, combining all the results from this section we obtain that $l \geq 5$, $n \geq 3$, $m \geq 3$ imposes θ-periodicity on u, v, and w. In contrast, for $l \geq 3$, once either $n = 2$ or $m = 2$, (l, n, m) does not always impose θ-periodicity, see Examples 1 and 2. Therefore, when $l \geq 5$, $(l, 3, 3)$ is the optimal bound. In the case when $l = 2$, $l = 3$, or $l = 4$, the problem of finding optimal bounds is still open.

l	n	m	θ-periodicity	
≥ 6	≥ 3	≥ 3	YES	Theorem 13
5	≥ 5	≥ 5	YES	Theorem 14
5	4	≥ 4	YES	Theorem 21
5	3	≥ 3	YES	Theorem 22
≥ 3	2	≥ 1	NO	Examples 1 and 2

Table 2: Result summary for the extended Lyndon-Schützenberger equation.

Example 1. *Let* $\Sigma = \{a, b\}$ *and* $\theta : \Sigma^* \to \Sigma^*$ *be the mirror image defined as* $\theta(a) = a$, $\theta(b) = b$, *and* $\theta(w_1 \dots w_n) = w_n \dots w_1$, *where* $w_i \in \{a, b\}$ *for all* $1 \le i \le n$. *Take now* $u = a^k b^2 a^{2k}$, $v = \theta(u)^l a^{2k} b^2 = (a^{2k} b^2 a^k)^l a^{2k} b^2$, *and* $w = a^2$, *for some* $k, l \ge 1$. *Then, although* $\theta(u)^{l+1} u^{l+1} = v^2 w^k$, *there is no word* $t \in \Sigma^+$ *with* $u, v, w \in \{t, \theta(t)\}^+$, *i.e., for any* $k, l \ge 1$, *the triple of numerical parameters* $(2l + 2, 2, k)$ *is not enough to impose* θ*-periodicity.*

Example 2. *Consider again* $\Sigma = \{a, b\}$ *and* $\theta : \Sigma^* \to \Sigma^*$ *be the mirror image defined in the previous example and take* $u = b^2(aba)^k$, $v = u^l b = (b^2(aba)^k)^l b$, *and* $w = aba$ *for some* $k, l \ge 1$. *Then, although* $u^{2l+1} = v\theta(v)w^k$, *there is no word* $t \in \Sigma^+$ *with* $u, v, w \in \{t, \theta(t)\}^+$, *i.e., for any* $k, l \ge 1$, $(2l + 1, 2, k)$ *is not enough to impose* θ*-periodicity.*

The next two results analyze the cases when we have triples (l, n, m) with $l \ge 6$ and $n, m \ge 3$ and respectively $(5, n, m)$ with $n, m \ge 5$.

Theorem 13. *Let* $u, v, w \in \Sigma^+$, $n, m \ge 3$, $l \ge 6$, $u_i \in \{u, \theta(u)\}$ *for* $1 \le i \le l$, $v_j \in \{v, \theta(v)\}$ *for* $1 \le j \le n$, *and* $w_k \in \{w, \theta(w)\}$ *for* $1 \le k \le m$. *If* $u_1 \dots u_l = v_1 \dots v_n\ w_1 \dots w_m$, *then there exists a word* $t \in \Sigma^+$ *such that* $u, v, w \in \{t, \theta(t)\}^+$.

Proof. Let us suppose that $|v_1 \dots v_n| \ge |w_1 \dots w_m|$; the other case is symmetric and can be solved similarly. Then, $|v_1 \dots v_n| \ge \frac{1}{2}|u_1 \dots u_l| \ge 3|u|$, since $l \ge 6$. Since $n \ge 3$, this means that $u_1 \dots u_l$ and $v_1 \dots v_n$ share a common prefix of length larger than both $3|u|$ and $3|v|$. Thus, we can apply Theorem 5, to obtain that $u, v \in \{t, \theta(t)\}^+$ for some θ-primitive word $t \in \Sigma^+$. Moreover, since $u_1 \dots u_l = v_1 \dots v_n\ w_1 \dots w_m$, this implies $w_1 \dots w_m \in \{t, \theta(t)\}^*$. Since t is θ-primitive, Theorem 6 implies that also $w \in \{t, \theta(t)\}^+$. □

Theorem 14. *Let* $u, v, w \in \Sigma^+$, $n, m \ge 5$, $u_i \in \{u, \theta(u)\}$ *for* $1 \le i \le 5$, $v_j \in \{v, \theta(v)\}$ *for* $1 \le j \le n$, *and* $w_k \in \{w, \theta(w)\}$ *for* $1 \le k \le m$. *If* $u_1 u_2 u_3 u_4 u_5 = v_1 \dots v_n\ w_1 \dots w_m$, *then there exists a word* $t \in \Sigma^+$ *such that* $u, v, w \in \{t, \theta(t)\}^+$.

Proof. Since $u_1 u_2 u_3 u_4 u_5 = v_1 \dots v_n\ w_1 \dots w_m$ and $n, m \ge 5$, we immediately obtain that $|u| > |v|$ and $|u| > |w|$. Assume now that $n|v| \ge m|w|$; the other case is symmetric. Thus, $n|v| \ge 2|u| + \frac{1}{2}|u|$ and we take $n|v| = 2|u| + l$ for some $l \ge \frac{1}{2}|u|$.

We claim now that $l \ge |v|$. If $l \ge |u|$, then we are done since we already know that $|u| > |v|$. So, let $\frac{1}{2}|u| \le l < |u|$. If $n \ge 6$, then $n|v| = 2|u| + l < 3|u|$ and thus $|v| < \frac{1}{2}|u| \le l$. Thus, the only case remaining now is when $n = 5$. Then, $5|v| = 2|u| + l \ge 2|u| + \frac{1}{2}|u|$, which implies $|v| \ge \frac{1}{2}|u|$. But then we have that $4|v| \ge 2|u|$ while $5|v| = 2|u| + l$. Hence, also in this case we obtain $|v| \le l$.

Thus, $u_1 u_2 u_3 u_4 u_5$ and $v_1 \dots v_n$ have a common prefix of length $n|v| = 2|u| + l \ge 2|u| + |v|$. This means, due to Theorem 5, that there exists a θ-primitive word $t \in \Sigma^+$ such that $u, v \in \{t, \theta(t)\}^+$. But then, since $u_1 u_2 u_3 u_4 u_5 = v_1 \dots v_n\ w_1 \dots w_m$, we immediately obtain that also $w \in \{t, \theta(t)\}^+$. □

The triple $(5, n, m)$ also turns out to impose θ-periodicity for any $n \ge 4$ and $m \ge 7$.

Theorem 15. *Let $u, v, w \in \Sigma^+$, $n \geq 4$, $m \geq 7$, $u_i \in \{u, \theta(u)\}$ for $1 \leq i \leq 5$, $v_j \in \{v, \theta(v)\}$ for $1 \leq j \leq n$, and $w_k \in \{w, \theta(w)\}$ for $1 \leq k \leq m$. If $u_1u_2u_3u_4u_5 = v_1 \ldots v_n \; w_1 \ldots w_m$, then there exists a word $t \in \Sigma^+$ such that $u, v, w \in \{t, \theta(t)\}^+$.*

Proof. Unless the border between v_n and w_1 falls inside u_3, Theorem 5 concludes the existence of such t. So, assume that the border falls inside u_3. If the border between u_2 and u_3 falls inside some v_i except v_n, then, due to Theorem 5, we obtain $u, v, w \in \{t, \theta(t)\}^+$ for some $t \in \Sigma^+$. Otherwise, we have that $(n-1)|v| < 2|u|$, which means $|v| < \frac{2}{n-1}|u| \leq \frac{2}{3}|u|$. Similarly, if the border between u_3 and u_4 does not fall inside w_1, we reach the existence of such t; otherwise $|w| < \frac{2}{m-1}|u| \leq \frac{1}{3}|u|$. Under the condition that v_n and w_1 straddle these respective borders, the equation cannot hold because v and w are too short. □

We already know from Example 2 that for any $m \geq 1$, the triple $(5, 2, m)$ is not enough to impose θ-periodicity. So, we investigate next what would be the optimal bound for the extension of the Lyndon and Schützenberger result when the first parameter is 5. Note that, without loss of generality, we can assume $n \leq m$. Then, due to Theorem 14, all we have to investigate are the cases $(5, 3, m)$ for $m \geq 3$ and $(5, 4, m)$ for $m \geq 4$. The next intermediate lemma will be useful in the analysis of these cases.

Lemma 16. *Let $u \in \Sigma^+$ such that $u = xy$ and $y \in \mathrm{Pref}(u)$ for some θ-palindrome words $x, y \in \Sigma^+$. If $|y| \geq |x|$, then $\rho(x) = \rho(y) = \rho(u)$.*

Proof. We have $u = xy = yz$ for some $z \in \Sigma^+$ of the same length as x. The length condition implies that $x \in \mathrm{Pref}(y)$. Since $x = \theta(x)$ and $y = \theta(y)$, this means that $x \in \mathrm{Suff}(y)$ and hence $z = x$. So we have $u = xy = yx$, and hence x, y, and u share their primitive root. □

Unlike in the case of the original Lyndon-Schützenberger equation, the investigation of our extension entails the consideration of an enormous amount of cases since for each variable u_i, v_j, w_k we have two possible values. However, in almost all cases, it is enough to consider the common prefix between $u_1 \ldots u_l$ and $v_1 \ldots v_n$ or the common suffix between $u_1 \ldots u_l$ and $w_1 \ldots w_m$ to prove that either the equation imposes θ-periodicity or the equation cannot hold.

Note that for the $(5, 3, m)$ or $(5, 4, m)$ extensions of the Lyndon-Schützenberger equation, we only have to consider the case when the border between v_n and w_1 is inside u_3 because otherwise Theorem 5 imposes θ-periodicity. Also even if the border is inside u_3, if $m|w| \geq 2|u| + |w|$, then we reach the same conclusion. Moreover, we can assume that w is θ-primitive since otherwise we would just increase the value of the parameter m. These observations justify the assumptions which will be made in the following propositions.

Proposition 17. *Let $u, v \in \Sigma^+$ such that v is a θ-primitive word, $u_1, u_2, u_3 \in \{u, \theta(u)\}$, and $v_1, \cdots, v_{2m+1} \in \{v, \theta(v)\}$ for some $m \geq 1$. If $v_1 \cdots v_{2m+1}$ is a proper prefix of $u_1u_2u_3$ and $2m|v| < 2|u| < (2m+1)|v|$, then $u_2 \neq u_1$ and $v_1 = \cdots = v_{2m+1}$. Moreover, $v_1 = yp$ and $u_1u_2 = (yp)^{2m}y$ for some $y, p \in \Sigma^*$ such that $y = \theta(y)$ and $p = \theta(p)$.*

Proof. Since θ is an involution, we may assume without loss of generality that $u_1 = u$ and $v_1 = v$. Note that $|v| < |u|$ and, due to the length condition, the border between u_1 and u_2 falls inside v_{m+1} while the one between u_2 and u_3 falls inside v_{2m+1}. Now, we have two cases depending on whether u_2 is equal to u_1.

Case 1: Suppose first that $u_2 \neq u_1$, i.e., $u_2 = \theta(u)$. Since $u_1 u_2 = u\theta(u)$ is θ-palindrome, $v_1 \cdots v_{2m} \in \mathrm{Pref}(u_1 u_2)$ implies $\theta(v_{2m}) \cdots \theta(v_1) \in \mathrm{Suff}(u_1 u_2)$. Applying Theorem 12 to the overlap between $v_1 \cdots v_{2m}$ and $\theta(v_{2m}) \cdots \theta(v_1)$ gives the following subcases: a) $v_1 = \cdots = v_{2m} = v$, and b) $v_1 = v, v_2 = \cdots = v_{2m} = \theta(v)$. For case b), because of the θ-primitivity of v, applying Theorem 12 to the overlap between $v_{2m}v_{2m+1}$ and $\theta(v_2)\theta(v_1)$ implies that v_{2m+1} can be neither v nor $\theta(v)$. Thus, this subcase is not possible.

Next, we consider the subcase a), and prove that v_{2m+1} must be v. Let us suppose otherwise, i.e., $v_{2m+1} = \theta(v)$, and we analyze two cases depending on whether u_3 is u or $\theta(u)$. If $u_3 = u$, then $v_{2m}v_{2m+1} = v\theta(v)$ overlaps with $\theta(v_1)v_1 = \theta(v)v$ because $v_1 \in \mathrm{Pref}(u)$, which contradicts Theorem 12. Otherwise, i.e., $u_3 = \theta(u)$, we look at the overlap between $v_{m+1} = v$ and $\theta(v_{m+1}) = \theta(v)$. Note that this overlap is θ-palindrome and, moreover, since the border between u_1 and u_2 cuts this overlap exactly in half, see Figure 7, we can say it is of the form $z\theta(z)$ for some $z \in \Sigma^+$. Then $v = z\theta(z)y$ for some θ-palindrome word y. Note that, due to length constraints, $z\theta(z) \in \mathrm{Pref}(v_{2m+1})$ and $\theta(v) = yz\theta(z)$. If $|z\theta(z)| \geq |y|$, then Lemma 16 implies that $\rho(z\theta(z)) = \rho(y)$, which contradicts the θ-primitivity of v. Otherwise, since $|z\theta(z)| < |y|$ we have $z \in \mathrm{Pref}(y)$, and hence $\theta(z) \in \mathrm{Suff}(y)$. So, if we look at the border between u_2 and u_3, then $yz\theta(z)^2 = z\theta(z)^2 y$. Thus $\rho(y) = \rho(z\theta(z)^2)$, and hence $y, z \in \{t, \theta(t)\}^+$ for some $t \in \Sigma^+$, again a contradiction with the θ-primitivity of v.

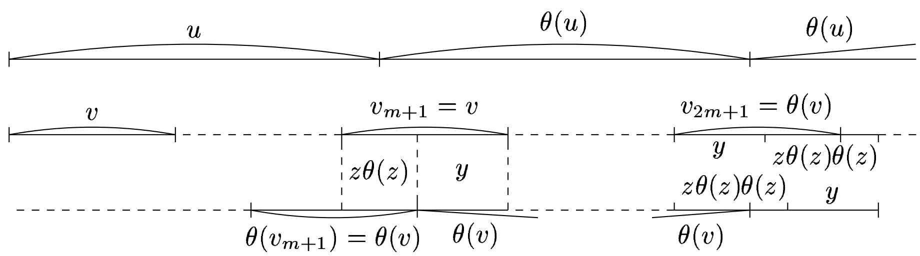

Figure 7: v_{m+1} overlaps with $\theta(v_{m+1})$ and the overlap is split exactly in half by the border between u_1 and u_2.

In conclusion, if $u_1 \neq u_2$, then we must have $v_1 = \cdots = v_{2m+1}$. Using Theorem 12, we can get the expressions of v and $u\theta(u)$ based on two θ-palindromes y and p.

Case 2: Let us suppose next that $u_2 = u_1 = u$. Now, if we look at the overlap between $v_1 \cdots v_m$ and $v_{m+1} \cdots v_{2m}$, then we see that all five cases from Theorem 12 are possible.

Firstly we consider the subcase a) when $v_1 = \cdots = v_m = v$ and $v_{m+1} = \cdots = v_{2m} = \theta(v)$, which is illustrated in Figure 8. As mentioned before, the border between u_1 and u_2 falls inside v_{m+1}, and hence in this case $u_1 = v^m z$

for some $z \in \mathrm{Pref}(v_{m+1})$; moreover $|z| < \frac{1}{2}|v|$ since $2|u| < (2m+1)|v|$. Then, we can write $v_{m+1} = \theta(v) = zy$ for some $y \in \Sigma^+$ with $y = \theta(y)$, see Figure 8. Moreover, using length arguments, we have that the right end of u_2 falls inside v_{2m+1} after exactly $2|z|$ characters. Since $u = v^m z$ and $\theta(z) \in \mathrm{Suff}(v)$, we obtain $\theta(z)z \in \mathrm{Pref}(v_{2m+1})$. Also, since $\theta(v) = zy$ and $|z| < \frac{1}{2}|v|$, we have $|y| > |z|$.

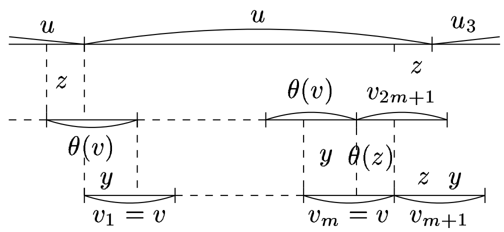

Figure 8: $v_1 \dots v_m$ and $v_{m+1} \dots v_{2m}$ overlap. Note that unless $u_3 = u$, we cannot assume that v_{m+1} overlaps with v_{2m+1}.

If $u_3 = u$, then $v_m v_{m+1} = v\theta(v)$ and $v_{2m}v_{2m+1} = \theta(v)v_{2m+1}$ overlap. So, due to Theorem 12, v_{2m+1} must be $\theta(v)$. So $z = \theta(z)$, and hence $\theta(v) = zy$ and $y \in \mathrm{Pref}(v_{m+1})$, i.e., $y \in \mathrm{Pref}(\theta(v))$. Then, since $|y| > |z|$, Lemma 16 implies $\rho(y) = \rho(z)$, which contradicts the θ-primitivity of v.

If $u_3 = \theta(u)$, then we consider two cases depending on the value of v_{2m+1}. First, suppose that $v_{2m+1} = v$. Since $\theta(z)z \in \mathrm{Suff}(u)$, we have $\theta(z)z \in \mathrm{Pref}(u_3)$ and we have two cases depending on $|v|$ and $2|\theta(z)z| = 4|z|$. If $|v| \leq 4|z|$, then $v_{2m+1} = v = \theta(z)zx$ for some $x \in \mathrm{Pref}(\theta(z)z)$. Since $|y| = |x| + |z|$ and $y, \theta(z)z \in \mathrm{Pref}(v)$, we have $x \in \mathrm{Pref}(y)$ and $z \in \mathrm{Suff}(y)$, which means $y = xz$. Thus, we have $v = \theta(z)zx = xz\theta(z)$. But we already know from [8] that this equation implies $x, z \in \{t, \theta(t)\}^+$ for some $t \in \Sigma^+$, which contradicts the θ-primitivity of v. Otherwise, i.e., $4|z| < |v|$, since u_2u_3 is θ-palindrome, $v_{2m}v_{2m+1}$ and $\theta(v_{2m+1})\theta(v_{2m})$ overlap with the overlap of length at least $|v|$. Since $|v| > 4|z|$, we can let $v = (\theta(z)z)^2 v_p$ for some $v_p \in \mathrm{Pref}(v) \cap \mathrm{Suff}(v)$, see

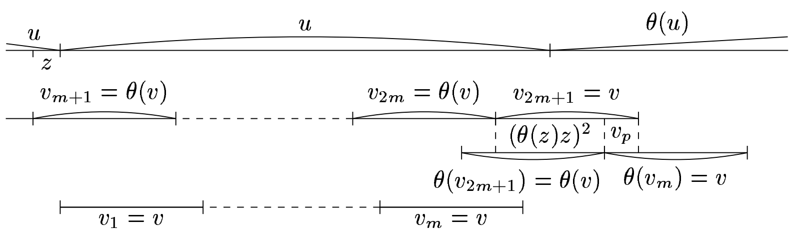

Figure 9: $v_{2m}v_{2m+1} = \theta(v)v$ overlaps with its image under θ. In addition, $v_{m+1} = \theta(v)$ overlaps with $v_1 = v$.

Figure 9. Then when we look at the overlap between $v_{m+1} = \theta(v)$ and $v_1 = v$, we can say that $\theta(v) = (z\theta(z))^2\theta(v_p)$. Hence $v = v_p(z\theta(z))^2 = (\theta(z)z)^2 v_p$. Since $(z\theta(z))^2$ and $(\theta(z)z)^2$ are θ-palindromes, Lemma 9 leads to a contradiction with

the θ-primitivity of v.

Next, suppose $v_{2m+1} = \theta(v)$. Then $z = \theta(z)$. If $2|z| < |v| \leq 4|z|$, then $v_{2m+1} = \theta(v) = z^k z_p$, where $k \in \{2, 3\}$ and $z_p \in \text{Pref}(z)$. This means that $z^3 \in \text{Suff}(u)$ since $z^2 \in \text{Suff}(v_m)$. It follows that $v_{2m} = \theta(v)$ has z as its suffix, which leads to a contradiction with the θ-primitivity of v since $\theta(v) = z^k z_p$. Otherwise, i.e., when $4|z| < |v|$, we can let $v_{2m+1} = \theta(v) = z^4 v_p$ for some $v_p \in \text{Pref}(v)$ with $v_p = \theta(v_p)$ (refer to Figure 9, but keeping in mind that now $v_{2m+1} = \theta(v)$). Since $v_1 = v$ overlaps with v_{m+1}, we have $z^3 \in \text{Pref}(v)$. Also the overlap between v_m and $v_{2m}v_{2m+1}$ implies that $v_p z \in \text{Suff}(v)$ (note that $v_p = \theta(v_p)$ in this case). Thus, $v = v_p z^4 = z^3 v_p z$, which contradicts the θ-primitivity of v since v_p is nonempty.

Secondly, we consider the subcase b). If $m \geq 2$, then $v_1 = \cdots = v_{m-1} = v$ and $v_m = \cdots = v_{2m} = \theta(v)$. This means that $v_{m-1}v_m = v\theta(v)$ and $v_{2m}v_{2m+1} = \theta(v)v_{2m+1}$ overlap, so Theorem 12 implies that v_{2m+1} cannot be either v or $\theta(v)$. For $m = 1$, we have $v_1 = v_2 = v$. If $v_3 = v$, the Fine and Wilf's theorem implies that $\rho(u) = \rho(v)$. Then, however, the length conditions $|v| < |u| < 2|v|$ implies that v is not primitive, a contradiction. Thus, $v_3 = \theta(v)$. Since u starts with v, we can write $v = xy = yz$, for some $x, y, z \in \Sigma^+$ with $z = \theta(z)$ and $2|u| - 2|v| = 2|z|$, as illustrated in Figure 10. Thus, $x = pq$, $z = qp$, and $y = (pq)^i p$ for some $p, q \in \Sigma^*$ and $i \geq 0$. Moreover, since $2|u| < 3|v| < 3|u|$, we have $|v| > 2|z|$, which means that $z^2 \in \text{Suff}(v)$, i.e., $z^2 \in \text{Pref}(\theta(v))$. Hence $z \in \text{Suff}(u)$. But, we already had that $x \in \text{Suff}(u)$, which implies that $pq = qp$. Thus, $\rho(p) = \rho(q) = \rho(x) = \rho(y)$, which contradicts the θ-primitivity of v. Now

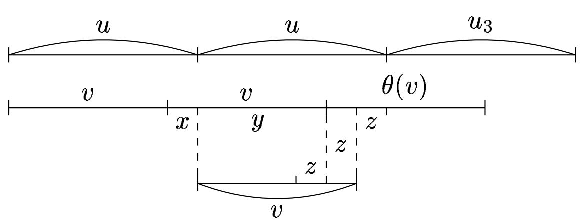

Figure 10: How $v_2v_3 = v\theta(v)$ and $v_1 = v$ overlap in the subcase b) for $m = 1$

we consider the other subcases. Note that in these subcases $m \geq 2$. The subcase c) when $v_1 = \cdots = v_m = v_{m+1} = v$ and $v_{m+2} = \cdots = v_{2m} = \theta(v)$ is illustrated in Figure 11. Then $v = yz$ for some $y, z \in \Sigma^+$ with $y = \theta(y)$, $z = \theta(z)$, and $y \in \text{Suff}(v)$. Since $|y| \geq |z|$, Lemma 16 leads to a contradiction with the primitivity of v. The remaining subcases d) is when $v_1 = \cdots = v_{m-1} = v$, $v_m = \theta(v)$, $v_{m+1} = v$, and $v_{m+2} = \cdots = v_{2m} = \theta(v)$. In this subcase, $v_{m-1}v_m = v\theta(v)$ overlaps with $v_{2m}v_{2m+1} = \theta(v)v_{2m+1}$ with an overlapped part of length at least $|v|$. Thus, Theorem 12 implies that v_{2m+1} can be neither v nor $\theta(v)$.

To conclude, we showed that $u_2 \neq u_1$ and $v_1 = \cdots = v_{2m+1}$. □

Proposition 18. *Let $u, v \in \Sigma^+$ such that v is θ-primitive, $u_1, u_2, u_3 \in \{u, \theta(u)\}$, and $v_1, \cdots, v_{2m} \in \{v, \theta(v)\}$ for some $m \geq 2$. If $v_1 \cdots v_{2m} \in \text{Pref}(u_1u_2u_3)$ and $(2m-1)|v| < 2|u| < 2m|v|$, then we have one of the following two cases:*

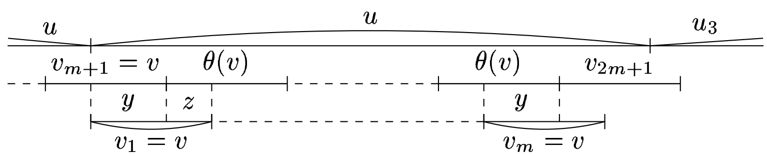

Figure 11: When $v_1 = \cdots = v_m = v_{m+1} = v$ and $v_{m+2} = \cdots = v_{2m} = \theta(v)$

1. *$u_1 \neq u_2$ and $v_1 = \cdots = v_{2m}$. In this case, $v_1 = yp$ and $u_1u_2 = (yp)^{2m-1}y$ for some $y, p \in \Sigma^*$ such that $y = \theta(y)$ and $p = \theta(p)$, or*

2. *$u_1 = u_2$, $v_1 = \cdots = v_m$, and $v_{m+1} = \cdots = v_{2m} = \theta(v_1)$. Moreover, $u_1 = \{r(tr)^i(rt)^{i+j}r\}^{m-1}r(tr)^i(rt)^j$ and $v_1 = r(tr)^i(rt)^{i+j}r$ for some $r, t \in \Sigma^*$ such that $r = \theta(r)$, $t = \theta(t)$, $i \geq 0$, and $j \geq 1$.*

Proof. Just as in the proof of Proposition 17, we can assume without loss of generality that $u_1 = u$ and $v_1 = v$. Then, we analyze two cases depending on whether $u_2 = u_1$.

Case 1: Let us look first at the case when $u_2 \neq u_1$, i.e., $u_2 = \theta(u)$, which differs only slightly from Case 1 from the proof of Proposition 17. Indeed, it is enough to consider only the case when $u_3 = \theta(u)$, $v_1 = \cdots = v_{2m-1} = v$ and prove that $v_{2m} = v$. Let us suppose for now that $v_{2m} = \theta(v)$ and let $v = yx$ and $x = z\theta(z)$ for some $x, y, z \in \Sigma^+$ with $x = \theta(x)$ and $y = \theta(y)$, as illustrated in Figure 12.

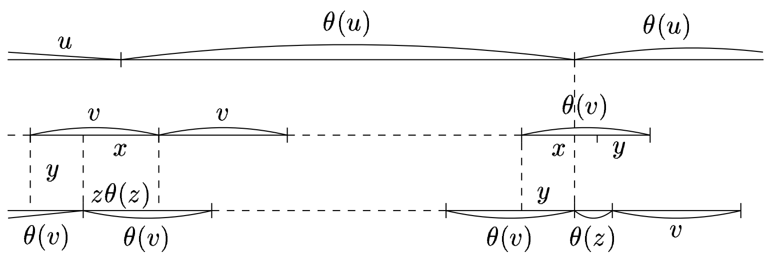

Figure 12: When $u_3 = \theta(u)$, $v_{2m} = \theta(v) = xy$ overlaps with $y\theta(z)v$ because $\theta(z)v \in \text{Pref}(\theta(u))$.

Note that $\theta(v) = xy = z\theta(z)y$ and $y \in \text{Pref}(\theta(v))$. If $|y| \geq |x|$, then Lemma 16 implies that $\rho(x) = \rho(y)$, which is a contradiction with θ-primitivity of v. If $|z| \leq |y| < |x|$, then $z \in \text{Pref}(y)$ and $z\theta(z)y \in \text{Pref}(y\theta(z)y)$, as illustrated in Figure 12. Thus, $z\theta(z)y = y\theta(z)z$, which implies that $y, z \in \{t, \theta(t)\}^+$, see [8], contradicting the θ-primitivity of v. If $\frac{1}{2}|z| \leq |y| < |z|$, then we have $y \in \text{Pref}(z)$ and $y\theta(z)y \in \text{Pref}(z\theta(z)y)$, see Figure 13 *i)*. Then, let $\theta(z) = z_1y = yz_2$ for some $z_1, z_2 \in \Sigma^+$ with $z_1 = \theta(z_1)$ and $z_2 = \theta(z_2)$ since $y \in \text{Pref}(z)$. Then, since $zy = y\theta(z)$ we have $z_2y^2 = y^2z_2$, and hence $\rho(y) = \rho(z_2)$, which contradicts the θ-primitivity of v as $v = yx = yz_2y^2z_2$. If $|y| < \frac{1}{2}|z|$, then we have $\theta(z) = z_3y = y^2z_4$ for some $z_3, z_4 \in \Sigma^+$ with $z_3 = \theta(z_3)$ and $z_4 = \theta(z_4)$, see Figure 13 *ii)*. Now, since $zy = y\theta(z)$, we have $z_4y^3 = y^3z_4$. This leads us to the same contradiction as above because $v = yz_4y^4z_4$.

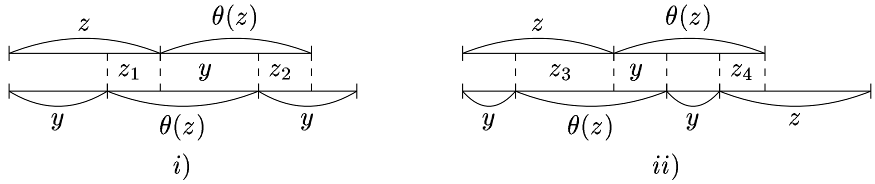

Figure 13: How $v_{2m} = z\theta(z)y$ overlaps with $y\theta(z)v$ when *i)* $\frac{1}{2}|z| \le |y| \le |z|$, or *ii)* $|y| \le \frac{1}{2}|z|$ in Case 1 of Proposition 18

Thus, if $u_1 \neq u_2$, then we must have $v_1 = \cdots = v_{2m} = v$. The representations of v_1 and u_1u_2 can be obtained using Theorem 12.

Case 2: Let us look next at the case when $u_2 = u_1 = u$, illustrated in Figure 14 and let $v = xy$ with $x \in \text{Suff}(v_m)$ and $y \in \text{Pref}(v_{m+1})$. Moreover, note that $|x| < |y|$ since $|x| = m|v| - |u|$ and $(2m-1)|v| < 2|u|$. Now, if we look at the overlap between $v_1 \cdots v_m$ and $v_m \cdots v_{2m-1}$, then due to Theorem 12, we get the following subcases: a) $v_1 = \cdots = v_{m-1} = v$ and $v_m = v_{m+1} = \cdots = v_{2m-1} = \theta(v)$; b) $v_1 = \cdots = v_m = v$, $v_{m+1} = \cdots = v_{2m-1} = \theta(v)$.

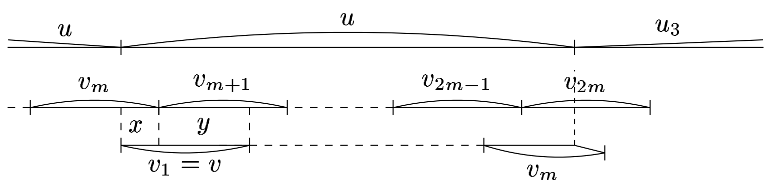

Figure 14: If $u_2 = u$, we can regard that $v_1 \ldots v_m$ overlaps with $v_m \ldots v_{2m-1}$ not depending on the value of u_3.

First, let us consider the subcase a). If $u_3 = u$, then $v_{m-1}v_m = v\theta(v)$ overlaps with $v_{2m-1}v_{2m} = \theta(v)v_{2m}$ and thus, due to Theorem 12, v_{2m} cannot be either v or $\theta(v)$. Otherwise, $u_3 = \theta(u)$ and note that $x = \theta(x)$ and $y = \theta(y)$ since $v_m = v_{m+1} = \theta(v)$. Then, since the overlapped part between v_{2m-1} and v_m is x, we obtain $x \in \text{Pref}(\theta(v))$. Since $\theta(v) = yx$ and $|x| < |y|$, we have $x \in \text{Pref}(y)$, i.e., $x \in \text{Suff}(y)$. Thus $x \in \text{Suff}(u)$, that is, $x \in \text{Pref}(\theta(u))$. Since $u_3 = \theta(u)$ and $v_m = \theta(v) = yx$, we can say that $v_{m-1}v_m$ overlaps with $v_{2m-1}v_{2m}$, which results in the same conclusion as above. Thus, the subcase a) is not possible.

For the subcase b), we prove that $v_{2m} = \theta(v)$. Let us start our analysis by supposing that $v_{2m} = v$. First, since $v_m = v$ ends with x, let $v = zwx$ for some $z, w \in \Sigma^+$ with $|w| = |x|$. If $u_3 = \theta(u)$, since $v_{2m} = v = zwx$, we obtain that $w \in \text{Pref}(u_3)$, i.e., $\theta(w) \in \text{Suff}(u)$. But this means that $w = \theta(w)$, since the right end of the first u cuts $v_m = v = zwx$ after exactly $|zw|$ characters. Since the overlap between v_{2m-1} and v_m is x, we have $xz = zw$ with $x = \theta(x)$ and $w = \theta(w)$. Then Lemma 9 implies that $x, z, w \in \{t, \theta(t)\}^+$ for some $t \in \Sigma^+$, a contradiction with the θ-primitivity of $v = zwx$. If $u_3 = u$, we immediately obtain $v = xy$ with $y \in \text{Pref}(v)$. Since $|x| < |y|$, the same contradiction derives from these relations due to Lemma 16.

In conclusion, for this case, i.e., when $u_2 = u_1$, we obtain that $v_1 = \cdots = v_m = v$ and $v_{m+1} = \cdots = v_{2m} = \theta(v)$. By applying Theorem 12 to the overlap between $v_1 \dots v_m$ and $v_m \dots v_{2m-1}$, we get the decompositions of u and v using two θ-palindromes r and t. □

These propositions show that if we suppose v to be θ-primitive, then the values of u_1, u_2, u_4, and u_5 determine the values of $v_1, \dots, v_n$ and $w_1, \dots, w_m$ uniquely, modulo a substitution of v by $\theta(v)$, or of w by $\theta(w)$. Thus, they decrease significantly the number of cases to be considered. Furthermore, the value of u_3 may put an additional useful restriction on v or w as shown in the following lemma.

Lemma 19. *Let $u, v \in \Sigma^+$ such that v is a θ-primitive word, $u_1, u_2, u_3 \in \{u, \theta(u)\}$, and $v_1, \cdots, v_n \in \{v, \theta(v)\}$ for some $n \geq 3$. If $v_1 \cdots v_n \in \text{Pref}(u_1u_2u_3)$, $u_1 \neq u_2$, $u_1 = u_3$, and $(n-1)|v| < 2|u| < n|v|$, then $|v| < \frac{4}{2n-1}|u|$.*

Proof. Since θ is an involution, we may assume without loss of generality that $u_1 = u$ and $v_1 = v$. Propositions 17 and 18 imply that $v_1 = \cdots = v_n = v$. Let now $v = xy$ for some $x, y \in \Sigma^+$ such that $x = \theta(x)$ and $y = \theta(y)$, as illustrated in Figure 15. Since $v \in \text{Pref}(u)$, we obtain that $y \in \text{Pref}(v)$. If

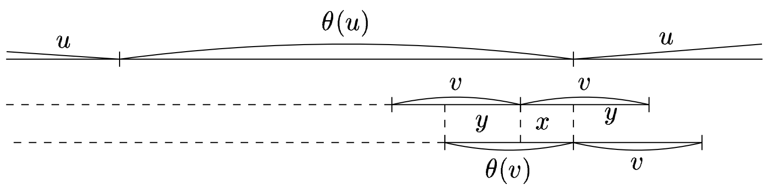

Figure 15: Since u begins with v, y is a prefix of v.

$|x| \leq |y|$, then Lemma 16 leads to a contradiction with the θ-primitivity of v. Thus $|y| < |x|$, which implies that $|y| < \frac{1}{2}|v|$. This means that $|v| < \frac{4}{2n-1}|u|$ because $|y| = n|v| - 2|u|$. □

All we did so far in studying the extended Lyndon-Schützenberger equation $u_1 \dots u_5 = v_1 \dots v_n \; w_1 \dots w_m$ was to consider either the common prefix of $v_1 \dots v_n$ and $u_1 \dots u_5$, or the common suffix of $w_1 \dots w_m$ and $u_1 \dots u_5$. Next, we combine them together and consider the whole equation. The following lemma proves to be useful for our considerations.

Lemma 20. *Let $u, v \in \Sigma^+$ such that v is a θ-primitive word, $u_1, u_2, u_3 \in \{u, \theta(u)\}$ and $v_1, \cdots, v_n \in \{v, \theta(v)\}$ for some $n \geq 3$. If $v_1 \cdots v_n = u_1u_2z$ for some $z \in \text{Pref}(u_3)$, $u_1 = u_2$, and $(n-1)|v| < 2|u|$, then $v_1 = xyx$ and $z = x^2$ for some $x, y \in \Sigma^+$ such that $x = \theta(x)$ and $yx = \theta(yx)$.*

Proof. Just as before, we can assume that $u_1 = u_2 = u$ and $v_1 = v$. Propositions 17 and 18 imply that n must be even, so let $n = 2m$ for some $m \geq 2$, and $u = \{r(tr)^i(rt)^{i+j}r\}^{m-1}r(tr)^i(rt)^{i+j}$ and $v = r(tr)^i(rt)^{i+j}r$ for some $r, t \in \Sigma^*$ such that $r = \theta(r)$, $t = \theta(t)$, $i \geq 0$, and $j \geq 1$. By taking $x = r(tr)^i$ and $y = (rt)^j$, we complete the proof. □

Next, we prove that the triple $(5, 4, m)$ imposes θ-periodicity for any $m \geq 4$.

Theorem 21. *Let $u, v, w \in \Sigma^+$, $u_1, u_2, u_3, u_4, u_5 \in \{u, \theta(u)\}$, $v_1, v_2, v_3, v_4 \in \{v, \theta(v)\}$, and $w_1, \cdots, w_m \in \{w, \theta(w)\}$ for some $m \geq 4$. If these words satisfy $u_1u_2u_3u_4u_5 = v_1v_2v_3v_4\ w_1 \cdots w_m$, then u is not θ-primitive and $u, v, w \in \{t, \theta(t)\}^+$ for some $t \in \Sigma^+$.*

Proof. First note that we can assume that w is θ-primitive, since otherwise we would just increase the numerical parameter m. If u is not θ-primitive, that is, $u \in \{p, \theta(p)\}^k$ for some θ-primitive word $p \in \Sigma^+$ and $k \geq 2$, then the equation can be rewritten as $p_1p_2 \cdots p_{5k} = v_1v_2v_3v_4w_1 \ldots w_m$, where $p_i \in \{p, \theta(p)\}$ for $1 \leq i \leq 5k$. But then, due to Theorem 13, we obtain that $v, w \in \{p, \theta(p)\}^+$. Furthermore, we can assume that also v is θ-primitive. Indeed, if it is not, then $v \in \{q, \theta(q)\}^j$ for some θ-primitive word q and $j \geq 2$. Then, the equation becomes $u_1 \ldots u_5 = q_1 \ldots q_{4j}w_1w_2 \ldots w_m$, where $q_i \in \{q, \theta(q)\}$ for $1 \leq i \leq 4j$. But this implies that $u, w \in \{q, \theta(q)\}^+$ due to Theorem 15. Since u and w are assumed to be θ-primitive, $u, w \in \{q, \theta(q)\}$ and we have $5|q| < 4j|q| + m|q|$, which contradicts the fact that u, v, and w satisfy the equation $u_1 \ldots u_5 = q_1 \ldots q_{4j}w_1w_2 \ldots w_m$. Even when v is θ-primitive, if $m \geq 7$ then the same argument leads to the same contradiction.

Now we will show that if u, v, and w are θ-primitive, then the equation cannot hold for $m \leq 6$. Since θ is involution, we can assume that $u_1 = u$, $v_1 = v$, and $w_1 = w$. Let us start by supposing that u, v, and w satisfy $u_1u_2u_3u_4u_5 = v_1v_2v_3v_4\ w_1 \cdots w_m$. Now, we have several cases depending on where the border between v_4 and w_1 is located. If it is left to or on the border between u_2 and u_3, then Theorem 5 implies that $u, w \in \{t, \theta(t)\}^+$ for some θ-primitive word $t \in \Sigma^+$, which further implies that also $v \in \{t, \theta(t)\}^+$. In fact, $u, v, w \in \{t, \theta(t)\}$ because they are θ-primitive. Then $|u_1 \ldots u_5| = 5|t|$, while $|v_1v_2v_3v_4w_1 \ldots w_m| = (4 + m)|t|$ with $m \geq 4$, which is a contradiction. The case when the border between v_4 and w_1 is right to or on the border between u_3 and u_4 will lead the contradiction along the same argument.

So let us suppose that $|u_1u_2| < |v_1v_2v_3v_4| < |u_1u_2u_3|$. Note that under this supposition, $|v|, |w| < |u|$. If $m|w| \geq 2|u| + |w| - 1$, then $u_3u_4u_5$ and $w_1 \ldots w_m$ share a suffix long enough to impose the θ-periodicity onto u and w due to Theorem 5. However, as explained before, this leads to a contradiction. This argument also applies to $u_1u_2u_3$ and $v_1v_2v_3v_4$. As a result, it is enough to consider the case when $3|v| < 2|u| < 4|v|$ and $(m-1)|w| < 2|u| < m|w|$.

There are 16 cases to be considered depending on the values of u_2, u_3, u_4, and u_5. Note that once these values are determined, the values of v_1, v_2, v_3, v_4 and $w_1, \cdots, w_m$ are set uniquely due to Propositions 17 and 18. We number these cases from 0 to 15 by regarding $u_2u_3u_4u_5$ as the 4-bit number based on the conversion $u \to 0$ and $\theta(u) \to 1$. For example, case 6 is $u_2u_3u_4u_5 = u\theta(u)\theta(u)u$.

First, we consider the case 2, that is, $uuu\theta(u)u = v_1 \cdots v_4\ w_1 \cdots w_m$. Since $3|v| < 2|u| < 4|v|$, $|v| < \frac{2}{3}|u|$. Moreover, Lemma 19 implies that $|w| < \frac{4}{2m-1}|u|$. Then $5|u| - (4|v| + m|w|) > 0$ which contradicts the fact that u, v, and w satisfy the given equation. The same arguments work for the cases when either $u_1u_2u_3 = u\theta(u)u$ (i.e., cases 8, 9, 10, 11), or $u_3u_4u_5 = u\theta(u)u$ (i.e., cases 2, 10), or $u_3u_4u_5 = \theta(u)u\theta(u)$ (i.e., cases 5, 13).

Secondly we consider the case 1, that is, $uuuu\theta(u) = v_1 \cdots v_4w_1 \cdots w_m$. Let $uux = v_1 \cdots v_4$, $yu\theta(u) = w_1 \cdots w_m$ for some $x, y \in \Sigma^+$ such that $u = xy$. We immediately obtain now, due to Lemma 20, that $x = \theta(x)$. Since $x \in \mathrm{Pref}(u_3)$, this means that $x \in \mathrm{Suff}(u_5)$, which implies that $w_m \in \mathrm{Suff}(x)$ or $x \in \mathrm{Suff}(w_m)$. In both cases, we obtain that $u_3u_4u_5$ and $w_mw_1w_2 \ldots w_m$ share a common suffix of length at least $2|u| + |w| - 1$. Then we employ Theorem 5 to lead a contradiction. Among the cases left to be investigated, the only one where we cannot apply this technique is case 0.

Now, case 0 is $u_1 = u_2 = u_3 = u_4 = u_5 = u$. Applying Propositions 17 and 18, we have that $m = 2k$ for some $k \geq 2$, $w_1 = \cdots = w_k = w$, $w_{k+1} = \cdots = w_{2k} = \theta(w)$, $v_1 = v_2 = v$, and $v_3 = v_4 = \theta(v)$. Note that $k \in \{2, 3\}$ since $4 \leq m \leq 6$. Then, Lemma 20 implies that $u = xyxxy = (y'x'x')^{k-1}y'x' = x^2x'^2$, $v = xyx$, and $\theta(w) = x'y'x'$ for some $x, y, x', y' \in \Sigma^+$ with $x = \theta(x)$, $yx = \theta(yx)$, $x' = \theta(x')$, and $x'y' = \theta(x'y')$.

When $k = 2$, i.e., $xyxxy = y'x'x'y'x'$, we have three subcases depending on the lengths of xy and $y'x'$. If $|xy| < |y'x'|$, then by looking at the two sides of the equality $xyxxy = y'x'x'y'x'$, we obtain $y'x' = xyz = \theta(z)xy$ and $x = zx'\theta(z)$ for some $z \in \Sigma^+$. Substituting $x = zx'\theta(z)$ into $xyz = \theta(z)xy$ we get $z = \theta(z)$, and hence $y'x' = xyz = zxy$. Thus, $y'x', xy, z \in \{p\}^+$ for some primitive word p. Let $z = p^i$ and $y'x' = p^j$ for some $i, j \geq 1$. Then $y'x' = zxy$ and $x = zx'z$ imply that $p^j = p^{2i}x'p^iy$. Since p is primitive, we obtain that $\rho(x') = p$, which contradicts the θ-primitivity of $\theta(w) = x'y'x'$. For the case when $|xy| > |y'x'|$ we can use similar arguments to reach a contradiction. Finally, if $|xy| = |y'x'|$, then $x = x'$, which is a contradiction with the θ-primitivity of u since $u = xxx'x'$.

When $k = 3$, i.e., $u = xyxxy = (y'x'x')^2y'x'$, we first note that $|xy| > |y'x'|$ and $|xyx| > |y'x'x'|$. If $|xy| \geq |y'x'x'|$, then, by the Fine and Wilf's theorem, $\rho(xyx) = \rho(y'x'x')$. Since xyx is strictly longer than $y'x'x'$, this means that $v = xyx$ is not primitive, which is a contradiction. Otherwise, i.e., $|y'x'| < |xy| < |y'x'x'|$, let $xy = y'x'z$ for some $z \in \mathrm{Pref}(x')$. Since $x' = \theta(x')$, the equation $xyxxy = (y'x'x')^2y'x'$ also implies that $xy = \theta(z)y'x'$. Moreover, since $xy = y'x'z = \theta(z)y'x'$ and $\theta(z) \in \mathrm{Suff}(x')$, we obtain $z = \theta(z)$. Thus $xy, y'x', z \in \{q\}^+$ for some primitive word $q \in \Sigma^+$, which, just as above, contradicts the θ-primitivity of $\theta(w)$. □

Next, we prove that the triple $(5, 3, m)$ imposes θ-periodicity for any $m \geq 3$.

Theorem 22. *Let $u, v, w \in \Sigma^+$, $u_1, u_2, u_3, u_4, u_5 \in \{u, \theta(u)\}$, $v_1, v_2, v_3 \in \{v, \theta(v)\}$, and $w_1, \cdots, w_m \in \{w, \theta(w)\}$ with $m \geq 3$. If these words verify* *the equation* $u_1u_2u_3u_4u_5 = v_1v_2v_3\ w_1\cdots w_m$*, then* u *is not* θ*-primitive and* $u, v, w \in \{t, \theta(t)\}^+$ *for some* $t \in \Sigma^+$*.*

Proof. As in the proof of Theorem 21, we can assume that w is θ-primitive. Also if u is not θ-primitive, then, just as before, Theorem 13 results in $u, v, w \in \{t, \theta(t)\}^+$ for some $t \in \Sigma^+$. So let us assume that u is θ-primitive. Moreover, we can assume that v is θ-primitive. Indeed, if it is not, then $v \in \{p, \theta(p)\}^j$ for some θ-primitive word p and $j \geq 2$. Then the equation becomes $u_1u_2u_3u_4u_5 = p_1 \dots p_{3j}w_1w_2 \dots w_m$, where $p_i \in \{p, \theta(p)\}$ for $1 \leq i \leq 3j$. For the case $m \geq 5$ and the case $m = 4$, Theorems 14 and 21 lead us to the contradiction, respectively. If $m = 3$, we can change the roles of v and w, and reduce it to the case when v is θ-primitive. In the following, we assume that u, v, and w are θ-primitive and prove that the equation cannot hold.

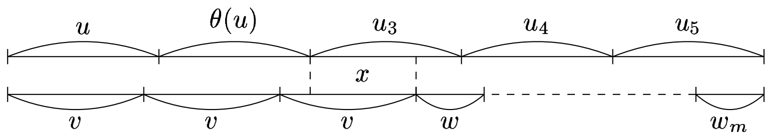

Figure 16: $u_1u_2u_3u_4u_5 = v_1v_2v_3w_1\cdots w_m$ for Theorem 22

Now, since θ is an involution, we can assume that $u_1 = u$, $v_1 = v$, and $w_1 = w$. As in the proof of Theorem 21, in all cases except when the border between v_3 and w_1 falls inside u_3, we get a contradiction. Furthermore, using the same arguments as in the previous proof, we can assume that $2|v| < 2|u| < 3|v|$ and $(m-1)|w| < 2|u| < m|w|$. Moreover, due to Proposition 17, $u_2 = \theta(u)$ and $v_1 = v_2 = v_3 = v$, see Figure 16. Then $u\theta(u)x = v^3$ for some $x \in \Sigma^+$, which satisfies $x = \theta(x)$ due to the same proposition. Since $x \in \text{Pref}(u_3)$, if $u_3 \neq u_5$, then $x \in \text{Suff}(u_5)$ which implies that either $w_m \in \text{Suff}(x)$ or $x \in \text{Suff}(w_m)$. In both cases, we obtain that $u_3u_4u_5$ and $w_mw_1w_2\cdots w_m$ share a common suffix of length at least $2|u|+|w|-1$. Hence, Theorem 5 implies that $u, w \in \{t, \theta(t)\}^+$ for some $t \in \Sigma^+$ and thus also $v \in \{t, \theta(t)\}^+$ which leads to the same contradiction as above. Otherwise, $u_3 = u_5$ and we have the following four cases left:

1. $u\theta(u)u\theta(u)u = vvvw_1\cdots w_m$,
2. $u\theta(u)\theta(u)u\theta(u) = vvvw_1\cdots w_m$,
3. $u\theta(u)uuu = vvvw_1\cdots w_m$,
4. $u\theta(u)\theta(u)\theta(u)\theta(u) = vvvw_1\cdots w_m$.

Let us start by considering the first equation. Since v is θ-primitive, using Lemma 19, we have $|v| < \frac{4}{5}|u|$ and $|w| < \frac{4}{2m-1}|u|$. However, then $5|u| - (3|v| + m|w|) > 5|u| - \frac{12}{5}|u| - \frac{4m}{2m-1}|u| = \frac{6m-13}{5(2m-1)}|u| > 0$ because $m \geq 3$. Hence, $5|u| > 3|v| + m|w|$ contradicting our supposition that the words u, v, and w satisfy the equation $u\theta(u)u\theta(u)u = vvvw_1\cdots w_m$.

For the second equation, Propositions 17 and 18 imply that $w_1 = w_2 = \cdots = w_m = w$. Since $u\theta(u) = v^2v_p$ for some $v_p \in \text{Pref}(v)$ and $u\theta(u)$ is θ-palindrome, we have $u\theta(u) = \theta(v_p)\theta(v)^2$. Note that $\theta(v_p) \in \mathrm{Suff}(\theta(v))$. Also $u\theta(u) = w_s w^{m-1}$ for some $w_s \in \mathrm{Suff}(w)$. Since $m \geq 3$, the Fine and Wilf's theorem implies that $\rho(\theta(v)) = \rho(w)$ and thus we obtain again the same contradiction as above.

Next we consider the third equation. Since $u_4 = u_5$, Propositions 17 and 18 imply that $m = 2k$ for some $k \geq 2$ and $w_1 = \cdots = w_k = w$ and $w_{k+1} = \cdots = w_{2k} = \theta(w)$. Let $w^k\theta(w)^k = z_1 z_2 u^2$ for some $z_1, z_2 \in \Sigma^+$ with $|z_1| = |z_2| = k|w| - |u|$, as illustrated in Figure 17. Then, $z_1 z_2 \in \mathrm{Suff}(u)$, which due to length

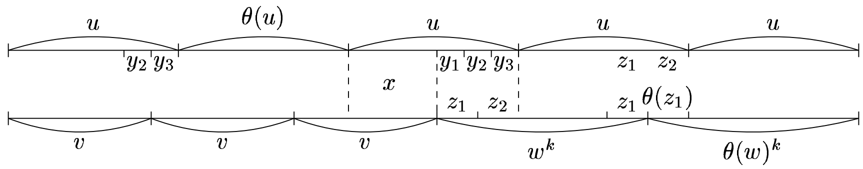

Figure 17: The suffix of u_3 can be written in two ways as $y_1y_2y_3$ and z_1z_2.

conditions means that $z_1 \in \mathrm{Suff}(w^k)$. Thus, $\theta(z_1) \in \mathrm{Pref}(\theta(w)^k)$ which implies immediately that $z_2 = \theta(z_1)$. Similarly, we can let $u\theta(u)u = v^3 y_1y_2y_3$ for some $y_1, y_2, y_3 \in \Sigma^+$ with $|y_1| = |y_2| = |y_3| = |u| - |v|$. Then $y_1y_2y_3 = z_1\theta(z_1)$, which implies $y_3 = \theta(y_1)$ and $y_2 = \theta(y_2)$. Note that since $|w^k\theta(w)^k| \geq 4|w|$ and $(2k-1)|w| < 2|u| < 2k|w|$ we obtain $3|y_1y_2y_3| = 3(2k|w| - 2|u|) < (2k-1)|w| < 2|u|$ which implies $|y_1y_2y_3| < \frac{2}{3}|u|$. This further implies that $|x| = |u| - |y_1y_2y_3| > |y_1|$. If we look at the second v, since $y_3 \in \mathrm{Suff}(u)$, using length arguments, we obtain that $y_3 \in \mathrm{Pref}(v)$, and hence $y_3 \in \mathrm{Pref}(u)$. Since $|y_3| < |x|$, this means that $y_3 \in \mathrm{Pref}(x)$ and hence $\theta(y_3) \in \mathrm{Suff}(x)$, which further implies $\theta(y_3) \in \mathrm{Suff}(v)$. Thus $y_2 = \theta(y_3)$ because $y_2 \in \mathrm{Suff}(v)$, which results in $y_1 = y_2 = y_3$ and, moreover, they are all θ-palindromes. Hence $y_1y_2 = \theta(y_2)\theta(y_1) = \theta(y_1y_2)$, which is a prefix of $\theta(v)$. This means that $u\theta(u)u$ and $v^3\theta(v)$ share a prefix of length at least $2|u| + |v| - 1$. Consequently $\rho_\theta(u) = \rho_\theta(v)$ which leads to the same contradiction as before.

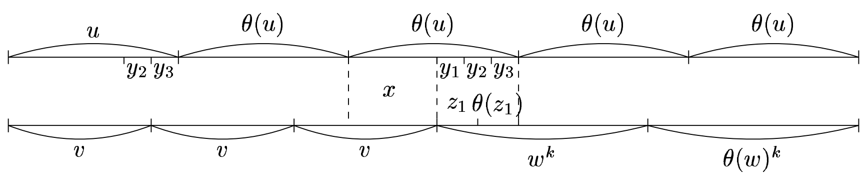

Figure 18: The suffix of u_3 can be written in two ways as $y_1y_2y_3$ and $z_1\theta(z_1)$.

Lastly, we consider the fourth equation, illustrated in Figure 18. Just as in the case of the third equation, $y_3 = \theta(y_1)$ and $y_2 = \theta(y_2)$. Moreover, using length arguments, we have $v = y_2y_3x$ for some $x \in \Sigma^+$ and hence $u\theta(u) = (y_2y_3x)^2y_2y_3$. This implies that $\theta(y_3) = y_2$ so that $y_1 = y_2 = y_3$. The rest is as same as for the third equation.

In conclusion, if u is θ-primitive, then, using length arguments, we always reach a contradiction. On the other hand, if u is not θ-primitive, then we proved that there exists a word $t \in \Sigma^+$ such that $u, v, w \in \{t, \theta(t)\}^+$. $\square$

References

[1] Adleman, L.M.: Molecular computation of solutions to combinatorial problems. Science **266**(5187) (November 1994) 1021–1024

[2] Crochemore, M., Hancart, C., Lecroq, T.: Algorithms on Strings. Cambridge University Press (2007)

[3] Crochemore, M., Rytter, W.: Jewels of Stringology. World Scientific (2002)

[4] Choffrut, C., Karhumäki, J.: Combinatorics of words. In Rozenberg, G., Salomaa, A., eds.: Handbook of Formal Languages. Volume 1. Springer-Verlag, Berlin-Heidelberg-New York (1997) 329–438

[5] Lothaire, M.: Combinatorics on Words. Volume 17 of Encyclopedia of Mathematics and its Applications. Addison-Wesley (1983)

[6] Cummings, L.J., Smyth, W.F.: Weak repetitions in strings. Journal of Combinatorial Mathematics and Combinatorial Computing **24** (1997) 33–48

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

[8] Czeizler, E., Kari, L., Seki, S.: On a special class of primitive words. In: Proc. of MFCS 2008. Volume 5162 of Lecture Notes in Computer Science., Berlin-Heidelberg, Springer (2008) 265–277

[9] Lyndon, R.C., Schützenberger, M.P.: The equation $a^m = b^n c^p$ in a free group. Michigan Mathematical Journal **9** (1962) 289–298

[10] Harju, T., Nowotka, D.: The equation $x^i = y^j z^k$ in a free semigroup. Semigroup Forum **68** (2004) 488–490

[11] Harju, T., Nowotka, D.: On the equation $x^k = z_1^{k_1} z_2^{k_2} \dots z_n^{k_n}$ in a free semigroup. Theoretical Computer Science **330**(1) (2005) 117–121

[12] Lentin, A.: Sur l'équation $a^m = b^n c^p d^q$ dans un monoïde libre. Comptes Rendus de l'Académie des Sciences Paris **260** (1965) 3242–3244

[13] Kari, L., Mahalingam, K.: Watson-Crick palindromes to appear in Natural Computing.

[14] de Luca, A., de Luca, A.: Pseudopalindrome closure operators in free monoids. Theoretical Computer Science **362** (2006) 282–300

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

[16] Kari, L., Mahalingam, K.: Watson-Crick conjugate and commutative words. In: Proc. of DNA 13. Volume 4848 of Lecture Notes in Computer Science. (2008) 273–283