On Collapsing Prefix Normal Words

Pamela Fleischmann[1(✉)], Mitja Kulczynski[1], Dirk Nowotka[1], and Danny Bøgsted Poulsen[2]

[1] Department of Computer Science, Kiel University, Kiel, Germany
{fpa,mku,dn}@informatik.uni-kiel.de
[2] Department of Computer Science, Aalborg University, Aalborg, Denmark
dannybpoulsen@cs.aau.dk

Abstract. Prefix normal words are binary words in which each prefix has at least the same number of 1s as any factor of the same length. Firstly introduced in 2011, the problem of determining the index (amount of equivalence classes for a given word length) of the prefix normal equivalence relation is still open. In this paper, we investigate two aspects of the problem, namely prefix normal palindromes and so-called collapsing words (extending the notion of critical words). We prove characterizations for both the palindromes and the collapsing words and show their connection. Based on this, we show that still open problems regarding prefix normal words can be split into certain subproblems.

1 Introduction

Two words are called abelian equivalent if the amount of each letter is identical in both words, e.g. *rotor* and *torro* are abelian equivalent albeit *banana* and *ananas* are not. Abelian equivalence has been studied with various generalisations and specifications such as abelian-complexity, k-abelian equivalence, avoidability of (k-)abelian powers and much more (cf. e.g., [6,10,11,13,17,22–24]). The number of occurrences of each letter is captured in the Parikh vector (also known as Parikh image or Parikh mapping) [21]: given a lexicographical order on the alphabet, the i^{th} component of this vector is the amount of the i^{th} letter of the alphabet in a given word. Parikh vectors have been studied in [12,16,19] and are generalised to Parikh matrices for saving more information about the word than just the amount of letters (cf. eg., [20,25]).

A recent generalisation of abelian equivalence, for words over the binary alphabet $\{0,1\}$, is prefix normal equivalence (pn-equivalence) [14]. Two binary words are pn-equivalent if their maximal numbers of 1s in any factor of length n are equal for all $n \in \mathbb{N}$. Burcsi et al. [5] showed that this relation is indeed an equivalence relation and moreover that each class contains exactly one uniquely determined representative - called a *prefix normal word*. A word w is said to be prefix normal if the prefix of w of any length has at least the number of 1s as any of w's factors of the same length. For instance, the word 110101 is prefix normal but 101101 is not, witnessed by the fact that 11 is a factor but not a prefix. Both words are pn-equivalent. In addition to being representatives of the

A. Leporati et al. (Eds.): LATA 2020, LNCS 12038, pp. 412–424, 2020.
https://doi.org/10.1007/978-3-030-40608-0_29

pne-classes, prefix normal words are also of interest since they are connected to Lyndon words, in the sense that every prefix normal word is a pre-necklace [14]. Furthermore, as shown in [14], the indexed jumbled pattern matching problem (see e.g. [2,4,18]) is connected to prefix normal forms: if the prefix normal forms are given, the indexed jumbled pattern matching problem can be solved in linear time $\mathcal{O}(n)$ of the word length n. The best known algorithm for this problem has a run-time of $\mathcal{O}(n^{1.864})$ (see [7]). Consequently there is also an interest in prefix normal forms from an algorithmic point of view. An algorithm for the computation of all prefix normal words of length n in run-time $\mathcal{O}(n)$ per word is given in [8]. Balister and Gerke [1] showed that the number of prefix normal words of length n is $2^{n-\Theta(\log^2(n))}$ and the class of a given prefix normal word contains at most $2^{n-O(\sqrt{n\log(n)})}$ elements. A closed formula for the number of prefix normal words is still unknown. In "OEIS" [15] the number of prefix normal words of length n (A194850), a list of binary prefix normal words (A238109), and the maximum size of a class of binary words of length n having the same prefix normal form (A238110), can be found. An extension to infinite words is presented in [9].

Our Contribution. In this work we investigate two conspicuities mentioned in [3,14]: palindromes and extension-critical words. Generalising the result of [3] we prove that prefix normal palindromes (pnPal) play a special role since they are not pn-equivalent to any other word. Since not all palindromes are prefix normal, as witnessed by 101101, determining the number of pnPals is an (unsolved) sub-problem. We show that solving this sub-problem brings us closer to determining the index, i.e. number of equivalence classes w.r.t. a given word length, of the pn-equivalence relation. Moreover we give a characterisation based on the maximum-ones function for pnPals. The notion of extension-critical words is based on an iterative approach: compute the prefix normal words of length $n+1$ based on the prefix normal words of length n. A prefix normal word w is called extension-critical if $w1$ is not prefix normal. For instance, the word 101 is prefix normal but 1011 is not and thus 101 is called extension-critical. This means that all non-extension-critical words contribute to the class of prefix normal words of the next word-length. We investigate the set of extension-critical words by introducing an equivalence relation *collapse*, grouping all extensional-critical words that are pn-equivalent w.r.t. length $n+1$. Finally we prove that (prefix normal) palindromes and the collapsing relation (extensional-critical words) are related. In contrast to [14] we work with suffix-normal words (least representatives) instead of prefix-normal words. It follows from Lemma 1 that both notions lead to the same results.

Structure of the Paper. In Sect. 2, the basic definitions and notions are presented. In Sect. 3, we present the results on pnPals. Finally, in Sect. 4, the iterative approach based on collapsing words is shown. This includes a lower bound and an upper bound for the number of prefix normal words, based on pnPals and the collapsing relation. Due to space restrictions all proofs are in the appendix.

2 Preliminaries

Let $\mathbb{N}$ denote the set of natural numbers starting with 1, and let $\mathbb{N}_0 = \mathbb{N} \cup \{0\}$. Define $[n] = \{1, \ldots, n\}$, for $n \in \mathbb{N}$, and set $[n]_0 = [n] \cup \{0\}$.

An alphabet is a finite set Σ, the set of all finite words over Σ is denoted by Σ^*, and the empty word by ε. Let $\Sigma^+ = \Sigma^* \backslash \{\varepsilon\}$ be the free semigroup for the free monoid Σ^*. Let $w[i]$ denote the i^{th} letter of $w \in \Sigma^*$ that is $w = \varepsilon$ or $w = w[1] \ldots w[n]$. The *length* of a word $w = w[1] \ldots w[n]$ is denoted by $|w|$ and let $|\varepsilon| = 0$. Set $w[i..j] = w[i] \ldots w[j]$ for $1 \leq i \leq j \leq |w|$. Set $\Sigma^n = \{w \in \Sigma^* \,|\, |w| = n\}$ for all $n \in \mathbb{N}_0$. The number of occurrences of a letter $\mathsf{x} \in \Sigma$ in $w \in \Sigma^*$ is denoted by $|w|_{\mathsf{x}}$. For a given word $w \in \Sigma^n$ the *reversal* of w is defined by $w^R = w[n] \ldots w[1]$. A word $u \in \Sigma^*$ is a factor of $w \in \Sigma^*$ if $w = xuy$ holds for some words $x, y \in \Sigma^*$. If $x = \varepsilon$ then u is called a *prefix* of w and a *suffix* if $y = \varepsilon$. Let $\mathrm{Fact}(w), \mathrm{Pref}(w), \mathrm{Suff}(w)$ denote the sets of all factors, prefixes, and suffixes respectively. Define $\mathrm{Fact}_k(w) = \mathrm{Fact}(w) \cap \Sigma^k$ and $\mathrm{Pref}_k(w), \mathrm{Suff}_k(w)$ are defined accordingly. Notice that $|\mathrm{Pref}_k(w)| = |\mathrm{Suff}_k(w)| = 1$ for all $k \leq |w|$. The powers of $w \in \Sigma^*$ are recursively defined by $w^0 = \varepsilon$, $w^n = ww^{n-1}$ for $n \in \mathbb{N}$.

Following [14], we only consider binary alphabets, namely $\Sigma = \{\mathsf{0}, \mathsf{1}\}$ with the fixed lexicographic order induced by $\mathsf{0} < \mathsf{1}$ on Σ. In analogy to binary numbers we call a word $w \in \Sigma^n$ *odd* if $w[n] = \mathsf{1}$ and *even* otherwise.

For a function $f : [n] \to \Delta$ for $n \in \mathbb{N}_0$ and an arbitrary alphabet Δ the concatenation of the images defines a finite word $\mathsf{serialise}(f) = f(1)f(2)\ldots f(n) \in \Delta^*$. Since $\mathsf{serialise}$ is bijective, we will identify $\mathsf{serialise}(f)$ with f and use in both cases f (as long as it is clear from the context). This definition allows us to access f's *reversed function* $g : [n] \to \Delta; k \mapsto f(n-k+1)$ easily by f^R.

Definition 1. *The* maximum-ones functions *is defined for a word* $w \in \Sigma^*$ *by* $f_w : [|w|]_0 \to [|w|]_0; k \mapsto \max\{\, |v|_1 \mid v \in \mathrm{Fact}_k(w)\}$, *giving for each* $k \in [|w|]_0$ *the maximal number of* $\mathsf{1}$*s occuring in a factor of length* k. *Likewise the* prefix-ones and suffix-ones functions *are defined by* $p_w : [|w|]_0 \to [|w|]_0; k \mapsto |\mathrm{Pref}_k(w)|_1$ *and* $s_w : [|w|]_0 \to [|w|]_0; k \mapsto |\mathrm{Suff}_k(w)|_1$.

Definition 2. *Two words* $u, v \in \Sigma^n$ *are called* prefix-normal equivalent *(pn-equivalent,* $u \equiv_n v$*) if* $f_u = f_v$ *holds and* v*'s equivalence class is denoted by* $[v]_{\equiv} = \{u \in \Sigma^n \,|\, u \equiv_n v\}$. *A word* $w \in \Sigma^*$ *is called* prefix (suffix) normal *iff* $f_w = p_w$ *(*$f_w = s_w$ *resp.) holds. Let* $\sigma(w) = \sum_{i \in [n]} f_w(i)$ *denote the* maximal-one sum *of a* $w \in \Sigma^n$.

Remark 1. Notice that $s_w = p_{w^R}, f_w = f_{w^R}, p_w(i), s_w(i) \leq f_w(i)$ for all $i \in \mathbb{N}_0$. By $p_{w^R} = s_w$ and $f_w = f_{w^R}$ follows immediately that a word $w \in \Sigma^*$ is prefix normal iff its reversal is suffix normal.

Fici and Lipták [14] showed that for each word $w \in \Sigma^*$ there exists exactly one $w' \in [w]_{\equiv}$ that is prefix normal - the prefix normal form of w. We introduce the concept of *least representative*, which is the lexicographically smallest element of a class and thus also unique. As mentioned in [5] palindromes play a special role. Immediately by $w = w^R$ for $w \in \Sigma^*$, we have $p_w = s_w$, i.e. palindromes are the

Table 1. Prefix normal palindromes (pnPals).

Word length	Prefix normal palindromes	# prefix normal words
1	$0, 1$	2
2	$0^2, 1^2$	3
3	$0^3, 101, 1^3$	5
4	$0^4, 1001, 1^4$	8
5	$0^5, 10001, 10101, 11011, 1^5$	14
6	$0^6, 100001, 110011, 1^6$	23
7	$0^7, 10^210^21, 10^51, 1010101, 1^20101^2, 1^20^31^2, 1^301^3, 1^7$	41
8	$0^8, 10^61, 1010^2101, 1^20^41^2, 1^201^201^2, 1^30^21^3, 1^8$	70

only words that can be prefix and suffix normal. Recall that not all palindromes are prefix normal witnessed by 101101.

Definition 3. *A palindrome is called* prefix normal palindrome *(pnPal) if it is prefix normal. Let* $\mathrm{NPal}(n)$ *denote the set of all prefix normal palindromes of length* $n \in \mathbb{N}$ *and set* $\mathrm{npal}(n) = |\mathrm{NPal}(n)|$. *Let* $\mathrm{Pal}(n)$ *be the set of all palindromes of length* $n \in \mathbb{N}$.

3 Properties of the Least-Representatives

Before we present specific properties of the least representatives (LR) for a given word length, we mention some useful properties of the maximum-ones, prefix-ones, and suffix-ones functions (for the basic properties we refer to [5,14] and the references therein). Since we are investigating only words of a specific length, we fix $n \in \mathbb{N}_0$. Beyond the relation $p_w = s_{w^R}$ the mappings p_w and s_w are determinable from each other. Counting the 1s in a suffix of length i and adding the 1s in the corresponding prefix of length $(n-i)$ of a word w, gives the overall amount of 1s of w, namely

$$p_w(n) = p_w(n-i) + s_w(i) \quad \text{and} \quad s_w(n) = p_w(i) + s_w(n-i).$$

For suffix (resp. prefix) normal words this leads to $p_w(i) = f_w(n) - f_w(n-i)$ resp. $s_w(i) = f_w(n) - f_w(n-i)$ witnessing the fact $p_w = s_w$ for palindromes (since both equation hold). Before we show that indeed pnPals form a singleton class w.r.t. $\equiv_n$, we need the relation between the lexicographical order and prefix and suffix normality.

Lemma 1. *The prefix normal form of a class is the lexicographically largest element in the class and the suffix-normal of a class is a LR.*

Lemma 1 implies that a word being prefix and suffix normal forms a singleton class w.r.t. $\equiv_n$. As mentioned $p_w = s_w$ only holds for palindromes.

Proposition 1. *For a word $w \in \Sigma^n$ it holds that $|[w]|_{\equiv} = 1$ iff $w \in \mathrm{NPal}(n)$.*

The general part of this section is concluded by a somewhat artificial equation which is nevertheless useful for pnPals : by $s_w(i) = p_w^R(i) - p_w^R(i+1) + s_w(i-1)$ with $p_w^R(n+1) = 0$ for $i \in [n]$ and $s_w = p_{w^R}$ we get

$$p_{w^R}(i) = p_w^R(i) - p_w^R(i+1) - p_{w^R}(i-1).$$

The rest of the section will cover properties of the LRs of a class.

Remark 2. For completeness, we mention that $\mathsf{0}^n$ is the only even LR w.r.t. $\equiv_n$ and the only pnPal starting with $\mathsf{0}$. Moreover, $\mathsf{1}^n$ is the largest LR. As we show later in the paper $\mathsf{0}^n$ and $\mathsf{1}^n$ are of minor interest in the recursive process due to their speciality.

The following lemma is an extension of [5, Lemma 1] for the suffix-one function by relating the prefix and the suffix of the word s_w for a least representative. Intuitively the suffix normality implies that the $\mathsf{1}$s are more at the end of the word w rather than at the beginning: consider for instance $s_w = 1123345$ for $w \in \Sigma^7$. The associated word w cannot be suffix normal since the suffix of length two has only one $\mathsf{1}$ ($s_w(2) = 1$) but by $s_w(5) = 3, s_w(6) = 4$, and $s_w(7) = 5$ we get that within two letters two $\mathsf{1}$s are present and consequently $f_w(2) \geq 2$. Thus, a word w is only least representative if the amount of $\mathsf{1}$s at the end of s_w does not exceed the amount of $\mathsf{1}$s at the beginning of s_w.

Lemma 2. *Let $w \in \Sigma^n$ be a LR. Then we have*

$$s_w(i) \geq \begin{cases} s_w(n) - s_w(n-i+1) & \text{if } s_w(n-i+1) = s_w(n-i), \\ s_w(n) - s_w(n-i+1) + 1 & \text{otherwise.} \end{cases}$$

The remaining part of this section presents results for prefix normal palindromes. Notice that for $w \in \mathrm{NPal}(n)$ with $w = \mathsf{x}v\mathsf{x}$ with $\mathsf{x} \in \Sigma, v$ is not necessarily a pnPal; consider for instance $w = \mathsf{10101}$ with $\mathsf{010} \in \mathrm{Pal}(3) \setminus \mathrm{NPal}(3)$. The following lemma shows a result for prefix normal palindromes which is folklore for palindromes substituting f_w by p_w or s_w.

Lemma 3. *For $w \in \mathrm{NPal}(n) \setminus \{\mathsf{0}^n\}, v \in \mathrm{Pal}(n)$ with $w = \mathsf{1}v\mathsf{1}$ we have*

$$f_w(k) = \begin{cases} \mathsf{1} & \text{if } k = 1, \\ f_v(k-1) + \mathsf{1} & \text{if } 1 < k \leq |w| - 1, \\ f_w(|v|+1) + \mathsf{1} & \text{if } k = |w|. \end{cases}$$

In the following we give a characterisation of when a palindrome w is prefix normal depending on its maximum-ones function f_w and a derived function $\overline{f_w}$. In particular we observe that $f_w = \overline{f_w}^R$ if and only if w is a prefix normal palindrome. Intuitively $\overline{f_w}$ captures the progress of f_w in reverse order. This is an intriguing result because it shows that properties regarding prefix and suffix normality can be observed when f_w, s_w, p_w are considered in their serialised representation.

Table 2. Number of pnPals. [15] (A308465)

i	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
#	2	2	3	3	5	4	8	7	12	11	21	18	36	31	57
i	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
#	55	104	91	182	166	308	292	562	512	1009	928	1755	1697	3247	2972

Definition 4. *For $w \in \Sigma^n$ define $\overline{f}_w : [n] \to [n]$ by $\overline{f}_w(k) = \overline{f}_w(k-1) - (f_w(k-1) - f_w(k-2))$ with the extension $f_w(-1) = f_w(0) = 0$ of f and $\overline{f}_w(0) = f_w(n)$. Define $\overline{p}_w$ and $\overline{s}_w$ analogously.*

Example 1. Consider the pnPal $w = \mathsf{11011}$ with $f_w = 12234$. Then $\overline{f}_w$ is 43221 and we have $f_w = \overline{f}_w^R$. On the other hand for $v = \mathsf{101101} \in \mathrm{Pal}(6) \setminus \mathrm{NPal}(6)$ we have $p_v = 112334$ and $f_v = 122334$ and $\overline{f}_v = 432211$ and thus $\overline{f}_v^R \neq f_v$.

The following lemma shows a connection between the reversed prefix-ones function and the suffix-ones function that holds for all palindromes.

Lemma 4. *For $w \in \mathrm{Pal}(n)$ we have $s_w \equiv \overline{p}_w^R$.*

By Lemma 4 we get $p_w \equiv \overline{p}_w^R$ since $p_w \equiv s_w$ for a palindrome w. As advocated earlier, our main theorem of this part (Theorem 1) gives a characterisation of pnPals. The theorem allows us to decide if a word is a pnPal by only looking at the maximum-ones-function, thus a comparison of all factors is not required.

Theorem 1. *Let $w \in \Sigma^n \setminus \{ 0^n \}$. Then w is a pnPal if and only if $f_w = \overline{f}_w^{\mathsf{R}}$.*

Table 2 presents the amount of pnPals up to length 30 These results support the conjecture in [5] that there is a different behaviour for even and odd length of the word.

4 Recursive Construction of Prefix Normal Classes

In this section we investigate how to generate LRs of length $n+1$ using the LRs of length n. This is similar to the work of Fici and Lipták [14] except they investigated appending a letter to prefix normal words while we explore the behaviour on prepending letters to LRs. Consider the words $v = \mathsf{1001}$ and $w = \mathsf{0011}$, both being (different) LRs of length 4. Prepending a $\mathsf{1}$ to them leads to $\mathsf{11001}$ and $\mathsf{10011}$ which are pn-equivalent. We say that v and w *collapse* and denote it by $v \leftrightarrow w$. Hence for determining the index of $\equiv_n$ based on the least representatives of length $n-1$, only the least representative of one class matters.

Definition 5. *Two words $w, v \in \Sigma^n$* collapse *if $1w \equiv_{n+1} 1v$ holds. This is denoted by $w \leftrightarrow v$.*

Prepending a 1 to a non LR will never lead to a LR. Therefore It is sufficient to only look at LRs. Since collapsing is an equivalence relation, denote the equivalence class w.r.t. $\leftrightarrow$ of a word $w \in \Sigma^*$ by $[w]_{\leftrightarrow}$. Next, we present some general results regarding the connections between the LRs of lengths n and $n+1$. As mentioned in Remark 2, 0^n and 1^n are for all $n \in \mathbb{N}$ LRs. This implies that they do not have to be considered in the recursive process.

Remark 3. By [14] a word $w0 \in \Sigma^{n+1}$ is prefix-normal if w is prefix-normal. Consequently we know that if a word $w \in \Sigma^n$ is suffix normal, $0w$ is suffix normal as well. This leads in accordance to the naïve upper bound of $2^n + 1$ to a naïve lower bound of $|\Sigma^n / \equiv_n|$ for $|\Sigma^{n+1} / \equiv_{n+1}|$.

Remark 4. The maximum-ones functions for $w \in \Sigma^*$ and $0w$ are equal on all $i \in [|w|]$ and $f_{0w}(|w|+1) = f_w(|w|)$ since the factor determining the maximal number of 1's is independent of the leading 0. Prepending 1 to a word w may result in a difference between f_w and f_{1w}, but notice that since only one 1 is prepended, we always have $f_{1w}(i) \in \{f_w(i), f_w(i)+1\}$ for all $i \in [n]$. In both cases we have $s_w(i) = s_{\mathsf{x}w}(i)$ for $\mathsf{x} \in \{0,1\}$ and $i \in [|w|]$ and $s_{0w}(n+1) = s_w(n)$ as well as $s_{1w}(n+1) = s_w(n) + 1$.

Firstly we improve the naïve upper bound to $2|\Sigma^n / \equiv_n|$ by proving that only LRs in Σ^n can become LRs in Σ^{n+1} by prepending 1 or 0.

Proposition 2. *Let $w \in \Sigma^n$ not be LR. Neither $0w$ nor $1w$ are LRs in Σ^{n+1}.*

By Proposition 1 prefix (and thus suffix) normal palindromes form a singleton class. This implies immediately that a word $w \in \Sigma^n$ such that $1w$ is a prefix normal palindrome, does not collapse with any other $v \in \Sigma^n \backslash \{w\}$. The next lemma shows that even prepending once a 1 and once a 0 to different words leads only to equivalent words in one case.

Lemma 5. *Let $w, v \in \Sigma^n$ be different LRs. Then $0w \equiv_n 1v$ if and only if $v = 0^n$ and $w = 0^{n-1}1$.*

By Lemma 5 and Remark 3 it suffices to investigate the collapsing relation on prepanding 1s. The following proposition characterises the LR $1w$ among the elements $1v \in [1w]_{\equiv}$ for all LRs $v \in \Sigma^n$ with $w \leftrightarrow v$ for $w \in \Sigma^n$.

Proposition 3. *Let $w \in \Sigma^n$ be a LR. Then $1w \in \Sigma^{n+1}$ is a LR if and only if $f_{1w}(i) = f_w(i)$ holds for $i \in [n]$ and $f_{1w}(n+1) = f_w(n) + 1$.*

Corollary 1. *Let $w \in \mathrm{NPal}(n)$. Then $f_{w1}(i) = f_w(i)$ for $i \in [n]$ and $f_{w1}(n+1) = f_w(n) + 1$. Moreover $s_{w1}(i) = s_w(i)$ for $i \in [n]$ and $s_{w1}(n+1) = s_w(n) + 1$.*

This characterization is unfortunately not convenient for determining either the number of LRs of length $n+1$ from the ones from length n or the collapsing LRs of length n. For a given word w, the maximum-ones function f_w has to be determined, f_w to be extended by $f_w(n) + 1$, and finally the associated word - under the assumption $f_{1w} \equiv s_{1w}$ has to be checked for being suffix normal.

For instance, given $w = 100101$ leads to $f_w = 11223$, and is extended to $f_{1w} = 112234$. This would correspond to 110101 which is not suffix normal and thus w is not extendable to a new LR. The following two lemmata reduce the amount of LRs that needs to be checked for extensibility.

Lemma 6. *Let $w \in \Sigma^n$ be a LR such that $1w$ is a LR as well. Then for all LRs $v \in \Sigma^n \backslash \{w\}$ collapsing with w, $f_v(i) \leq f_w(i)$ holds for all $i \in [n]$, i.e. all other LRs have a smaller maximal-one sum.*

Corollary 2. *If $w, v \in \Sigma^n$ and $1w \in \Sigma^{n+1}$ are LRs with $w \leftrightarrow v$ and $v \neq w$ then $w \leq v$.*

Remark 5. By Corollary 2 the lexicographically smallest LR w among the collapsing leads to the LR of $[1w]$. Thus if w is a LR not collapsing with any lexicographically smaller word then $1w$ is LR.

Before we present the theorem characterizing exactly the collapsing words for a given word w, we show a symmetry-property of the LRs which are not extendable to LRs, i.e. a property of words which collapse.

Lemma 7. *Let $w \in \Sigma^n$ be a LR. Then $f_{1w}(i) \neq f_w(i)$ for some $i \in [n]$ iff $f_{1w}(n-i+1) \neq f_w(n-i+1)$.*

By [5, Lemma 10] a word $w1$ is prefix normal if and only if $|\operatorname{Suff}_k(w)|_1 < |\operatorname{Pref}_{k+1}(w)|_1$ for all $k \in \mathbb{N}$. The following theorem extends this result for determining the collapsing words w' for a given word w.

Theorem 2. *Let $w \in \Sigma^n$ be a LR and $w' \in \Sigma^n \backslash \{w\}$ with $|w|_1 = |w'|_1 = s \in \mathbb{N}$. Let moreover $v \not\leftrightarrow w$ for all $v \in \Sigma^*$ with $v \leq w$. Then $w \leftrightarrow w'$ iff*

1. *$f_{w'}(i) \in \{f_w(i), f_w(i) - 1\}$ for all $i \in [n]$,*
2. *$f_{w'}(i) = f_w(i)$ implies $f_{1w'}(i) = f_w(i)$,*
3. $$f_{w'}(i) \geq \begin{cases} f_{w'}(n) - f_{w'}(n-i+1) & \text{if } f_{w'}(n-i+1) = f_{w'}(n-i), \\ f_{w'}(n) - f_{w'}(n-i+1) + 1 & \text{otherwise.} \end{cases}$$

Theorem 2 allows us to construct the equivalence classes w.r.t. the least representatives of the previous length but more tests than necessary have to be performed: Consider, for instance $w = 11101100111011111$ which is a smallest LR of length 17 not collapsing with any lexicographically smaller LR. For w we have $f_w = 1 \cdot 2 \cdot 3 \cdot 4 \cdot 5 \cdot 5 \cdot 6 \cdot 7 \cdot 8 \cdot 8 \cdot 8 \cdot 9 \cdot 10 \cdot 10 \cdot 11 \cdot 12 \cdot 13$ where the dots just act as separators between letters. Thus we know for any w' collapsing with w, that $f_{w'}(1) = 1$ and $f_{w'}(17) = 13$. The constraints $f_{w'}(2) \in \{f_{w'}(2), f_{w'}(2) + 1\}$ and $f_{w'}(2) \leq f_w(2)$ implies $f_{w'}(2) \in \{1, 2\}$. First the check that $f_{w'}(10) = 4$ is impossible excludes $f_{w'}(2) = 1$. Since no collapsing word can have a factor of length 2 with only one 1, a band in which the possible values range can be defined by the unique greatest collapsing word w'. It is not surprising that this word is connected with the prefix normal form. The following two lemmata define the band in which the possible collapsing words f_w are.

Lemma 8. *Let $w \in \Sigma^n \backslash \{0^n\}$ be a LR with $v \not\leftrightarrow w$ for all $v \in \Sigma^n$ with $v \leq w$. Set $u := (1w[1..n-1])^R$. Then $w \leftrightarrow u$ and for all LRs $v \in \Sigma^n \backslash \{u\}$ with $v \leftrightarrow w$ and all $i \in [n]$ $f_v(i) \geq f_u(i)$, thus $\sigma(u) = \sum_{i\in[n]} f_u(i) \leq \sum_{i\in[n]} f_v(i) = \sigma(v)$.*

Notice that $w' = (1w[1..n-1])^R$ is not necessarily a LR in $\Sigma^n/\equiv_n$ witnessed by the word of the last example. For w we get $u = 1110111001101111$ with $f_u(8) = f_w(8)$ and $f_u(10) = 7 \neq 8 = f_w(10)$ violating the symmetry property given in Lemma 7. The following lemma alters w' into a LR which represents still the lower limit of the band.

Lemma 9. *Let $w \in \Sigma^n$ be a LR such that $1w$ is also a LR. Let $w' \in \Sigma^n$ with $w \leftrightarrow w'$, and I the set of all $i \in [\lfloor \frac{n}{2} \rfloor]$ with*

$$(f_{w'}(i) = f_w(i) \wedge f_{w'}(n-i+1) \neq f_w(n-i+1)) \text{ or}$$
$$(f_{w'}(i) \neq f_w(i) \wedge f_{w'}(n-i+1) = f_w(n-i+1))$$

and $f_w(j) = f_{w'}(j)$ for all $j \in [n]\backslash I$. Then $\hat{w}$ defined such that $f_{\hat{w}}(j) = f_{w'}(j)$ for all $j \in [n]\backslash I$ and $f_{\hat{w}}(n-i+1) = f_{w'}(n-i+1)+1$ ($f_{\hat{w}}(i) = f_{\hat{w}}(i)+1$ resp.) for all $i \in I$ holds, collapses with w.

Remark 6. Lemma 9 applied to $(1w[1..n-1])^R$ gives the lower limit of the band. Let $\hat{w}$ denote the output of this application for a given $w \in \Sigma^n$ according to Lemma 9.

Continuing with the example, we firstly determine $\hat{w}$ for $w = 1111011\,1001101111$. We get with $u = w[n-1..1]1$ Since for all collapsing $w' \in \Sigma^n$ we have $f_{\hat{w}}(i) \leq f_{w'}(i) \leq f_w(i), w'$ is determined for $i \in [17]\backslash\{5, 9, 13\}$. Since the value for 5 determines the one for 13 there are only two possibilities, namely $f_{w'}(5) = 5$ and $f_{w'}(9) = 7$ and $f_{w'}(5) = 4$ and $f_{w'}(9) = 8$. Notice that the words w' corresponding to the generated words $f_{w'}$ are not necessarily LRs of the shorter length as witnessed by the one with $f_{w'}(5) = 5$ and $f_{w'}(9) = 7$. In this example this leads to at most three words being not only in the class but also in the list of former representatives. Thus we are able to produce an upper bound for the cardinality of the class. Notice that in any case we only have to test the first half of w''s positions by Lemma 7. This leads to the following definition.

Table 3. f for $w = 11110111001101111$.

i	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
f_w	1	2	3	4	5	5	6	7	8	8	8	9	10	10	11	12	13
f_u	1	2	3	4	4	5	6	7	7	7	8	9	9	10	11	12	13
$f_{\hat{w}}$	1	2	3	4	4	5	6	7	7	8	8	9	9	10	11	12	13

Definition 6. *Let* $h_d : \Sigma^* \times \Sigma^* \to \mathbb{N}_0$ *be the Hamming-distance. The* palindromic distance $p_d : \Sigma^* \to \mathbb{N}_0$ *is defined by* $p_d(w) = h_d(w[1..\lfloor \frac{n}{2} \rfloor], (w[\lceil \frac{n}{2} \rceil + 1..|w|])^R)$. *Define the* palindromic prefix length $p_\ell : \Sigma^* \to \mathbb{N}_0$ *by* $p_\ell(w) = \max\{ k \in [|w|] \mid \exists u \in \mathrm{Pref}_k(w) : p_d(u) = 0 \}$.

The palindromic distance gives the minimal number of positions in which a bit has to be flipped for obtaining a palindrome. Thus, $p_d(w) = 0$ for all palindromes w, and, for instance, $p_d(\mathtt{110011001}) = 2$ since the first half of w and the reverse of the second half mismatch in two positions. The palindromic prefix length determines the length of w's longest prefix being a palindrome. For instance $p_\ell(\mathtt{1101}) = 2$ and $p_\ell(\mathtt{01101}) = 4$. Since a LR w determines the upper limit of the band and $w[n-1..1]\mathtt{1}$ the lower limit, the palindromic distance of $ww[n-1..1]\mathtt{1}$ is in relation to the positions of f_w in which collapsing words may differ from w.

Theorem 3. *If* $w \in \Sigma^n$ *and* $\mathtt{1}w$ *are both LRs then* $|[w]_\leftrightarrow| \leq 2^{\lceil \frac{p_d(ww[n-1..1]\mathtt{1}}{2} \rceil}$.

For an algorithmic approach to determine the LRs of length n, we want to point out that the search for collapsing words can also be reduced using the palindromic prefix length. Let $w_1, \ldots, w_m$ be the LRs of length $n-1$. For each w we keep track of $|w| - p_\ell(w)$. For each w_i we check firstly if $|w_i| - p_\ell(w_i) = 1$ since in this case the prepended $\mathtt{1}$ leads to a palindrome. Only if this is not the case, $[w_i]_\leftrightarrow$ needs to be determined. All collapsing words computed within the band of w_i and $\hat{w}_i$ are deleted in $\{w_{i+1}, \ldots, w_m\}$.

In the remaining part of the section we investigate the set $\mathrm{NPal}(n)$ w.r.t. $\mathrm{NPal}(\ell)$ for $\ell < n$. This leads to a second calculation for an upper bound and a refinement for determining the LRs of $\Sigma^n / \equiv_n$ faster.

Lemma 10. *If* $w \in \mathrm{NPal}(n) \backslash \{\mathtt{1}^n\}$ *then* $\mathtt{1}w$ *is not a LR but* $w\mathtt{1}$ *is a LR.*

Remark 7. By Lemma 10 follows that all words $w \in \mathrm{NPal}(n)$ collapse with a smaller LR. Thus, for all $n \in \mathbb{N}$, an upper bound for $|\Sigma^{n+1} / \equiv_{n+1}|$ is given by $2|\Sigma^n / \equiv_n| - \mathrm{npal}(n)$.

For a closed recursive calculation of the upper bound in Remark 7, the exact number $\mathrm{npal}(n)$ is needed. Unfortunately we are not able to determine $\mathrm{npal}(n)$ for arbitrary $n \in \mathbb{N}$. The following results show relations between prefix normal palindromes of different lengths. For instance, if $w \in \mathrm{NPal}(n)$ then $\mathtt{1}w\mathtt{1}$ is a prefix normal palindrome as well. The importance of the pnPals is witnessed by the following estimation.

Theorem 4. *For all* $n \in \mathbb{N}_{\geq 2}$ *and* $\ell = |\Sigma^n / \equiv_n|$ *we have*

$$\ell + \mathrm{npal}(n-1) \leq |\Sigma^{n+1} / \equiv_{n+1}| \leq \ell + \mathrm{npal}(n+1) + \frac{\ell - \mathrm{npal}(n+1)}{2}.$$

The following results only consider pnPals that are different from $\mathtt{0}^n$ and $\mathtt{1}^n$. Notice for these special palindromes that $\mathtt{0}^n\mathtt{0}^n, \mathtt{1}^n\mathtt{1}^n, \mathtt{1}^n\mathtt{1}\mathtt{1}^n, \mathtt{0}^n\mathtt{0}\mathtt{0}^n$, $\mathtt{1}\mathtt{1}^n\mathtt{1}^n\mathtt{1}, \mathtt{1}\mathtt{0}^n\mathtt{0}^n\mathtt{1} \in \mathrm{NPal}(k)$ for an appropriate $k \in \mathbb{N}$ but $\mathtt{0}^n\mathtt{1}\mathtt{0}^n \notin \mathrm{NPal}(2n+1)$.

Lemma 11. *If* $w \in \mathrm{NPal}(n) \backslash \{1^n, 0^n\}$ *then neither* ww *nor* $w1w$ *are prefix normal palindromes.*

Lemma 12. *Let* $w \in \mathrm{NPal}(n) \backslash \{0^n\}$ *with* $n \in \mathbb{N}_{\geq 3}$. *If* $w0w$ *is also a prefix normal palindrome then* $w = 1^k$ *or* $w = 1^k 01 u 101^k$ *for some* $u \in \Sigma^*$ *and* $k \in \mathbb{N}$.

A characterisation for $w1w$ being a pnPal is more complicated. By $w \in \mathrm{NPal}(n)$ follows that a block of 1s contains at most the number of 1s of the previous block. But if such a block contains strictly less 1s the number of 0s in between can increase by the same amount the number of 1s decreased.

Lemma 13. *Let* $w \in \mathrm{NPal}(n) \backslash \{1^n, 0^n\}$. *If* $1ww1$ *is also a prefix normal palindrome then* $10 \in \mathrm{Pref}(w)$.

Lemmas 11, 12, and 13 indicate that a characterization of prefix normal palindromes based on smaller ones is hard to determine.

5 Conclusion

Based on the work in [14], we investigated prefix normal palindromes in Sect. 3 and gave a characterisation based on the maximum-ones function. At the end of Sect. 4 results for a recursive approach to determine prefix normal palindromes are given. These results show that easy connections between prefix normal palindromes of different lengths cannot be expected. By introducing the collapsing relation we were able to partition the set of extension-critical words introduced in [14]. This leads to a characterization of collapsing words which can be extended to an algorithm determining the corresponding equivalence classes. Moreover we have shown that palindromes and the collapsing classes are related.

The concrete values for prefix normal palindromes and the index of the collapsing relation remain an open problem as well as the cardinality of the equivalence classes w.r.t. the collapsing relation. Further investigations of the prefix normal palindromes and the collapsing classes lead directly to the index of the prefix equivalence.

Acknowledgments. We would like to thank Florin Manea for helpful discussions and advice.

References

1. Balister, P., Gerke, S.: The asymptotic number of prefix normal words. J. Comb. Theory **784**, 75–80 (2019)
2. Burcsi, P., Cicalese, F., Fici, G., Lipták, Z.: Algorithms for jumbled pattern matching in strings. Int. J. Found. CS **23**(2), 357–374 (2012)
3. Burcsi, P., Fici, G., Lipták, Z., Ruskey, F., Sawada, J.: On combinatorial generation of prefix normal words. In: Kulikov, A.S., Kuznetsov, S.O., Pevzner, P. (eds.) CPM 2014. LNCS, vol. 8486, pp. 60–69. Springer, Cham (2014). https://doi.org/10.1007/978-3-319-07566-2_7

4. Burcsi, P., Fici, G., Lipták, Z., Ruskey, F., Sawada, J.: Normal, abby normal, prefix normal. In: Ferro, A., Luccio, F., Widmayer, P. (eds.) FUN 2014. LNCS, vol. 8496, pp. 74–88. Springer, Cham (2014). https://doi.org/10.1007/978-3-319-07890-8_7
5. Burcsi, P., Fici, G., Lipták, Z., Ruskey, F., Sawada, J.: On prefix normal words and prefix normal forms. TCS **659**, 1–13 (2017)
6. Cassaigne, J., Richomme, G., Saari, K., Zamboni, L.Q.: Avoiding Abelian powers in binary words with bounded Abelian complexity. Int. J. Found. CS **22**(04), 905–920 (2011)
7. Chan, T.M., Lewenstein, M.: Clustered integer 3SUM via additive combinatorics. In: 47th ACM Symposium on TOC, pp. 31–40. ACM (2015)
8. Cicalese, F., Lipták, Z., Rossi, M.: Bubble-flip—a new generation algorithm for prefix normal words. In: Klein, S.T., Martín-Vide, C., Shapira, D. (eds.) LATA 2018. LNCS, vol. 10792, pp. 207–219. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-77313-1_16
9. Cicalese, F., Lipták, Z., Rossi, M.: On infinite prefix normal words. In: Proceedings of the SOFSEM, pp. 122–135 (2019)
10. Coven, E.M., Hedlund, G.A.: Sequences with minimal block growth. TCS **7**(2), 138–153 (1973)
11. Currie, J., Rampersad, N.: Recurrent words with constant Abelian complexity. Adv. Appl. Math. **47**(1), 116–124 (2011)
12. Dassow, J.: Parikh mapping and iteration. In: Calude, C.S., PĂun, G., Rozenberg, G., Salomaa, A. (eds.) WMC 2000. LNCS, vol. 2235, pp. 85–101. Springer, Heidelberg (2001). https://doi.org/10.1007/3-540-45523-X_5
13. Ehlers, T., Manea, F., Mercas, R., Nowotka, D.: k-Abelian pattern matching. J. Discrete Algorithms **34**, 37–48 (2015)
14. Fici, G., Lipták, Z.: On prefix normal words. In: Mauri, G., Leporati, A. (eds.) DLT 2011. LNCS, vol. 6795, pp. 228–238. Springer, Heidelberg (2011). https://doi.org/10.1007/978-3-642-22321-1_20
15. OEIS Foundation Inc.: The on-line encyclopedia of integer sequencess (2019). http://oeis.org/
16. Karhumäki, J.: Generalized Parikh mappings and homomorphisms. Inf. Control **47**(3), 155–165 (1980)
17. Keränen, V.: Abelian squares are avoidable on 4 letters. In: Kuich, W. (ed.) ICALP 1992. LNCS, vol. 623, pp. 41–52. Springer, Heidelberg (1992). https://doi.org/10.1007/3-540-55719-9_62
18. Lee, L.-K., Lewenstein, M., Zhang, Q.: Parikh matching in the streaming model. In: Calderón-Benavides, L., González-Caro, C., Chávez, E., Ziviani, N. (eds.) SPIRE 2012. LNCS, vol. 7608, pp. 336–341. Springer, Heidelberg (2012). https://doi.org/10.1007/978-3-642-34109-0_35
19. Mateescu, A., Salomaa, A., Salomaa, K., Yu, S.: On an extension of the Parikh mapping, 06 September 2000. http://citeseer.ist.psu.edu/440186.html
20. Mateescu, A., Salomaa, A., Yu, S.: Subword histories and Parikh matrices. J. Comput. Syst. Sci. **68**(1), 1–21 (2004)
21. Parikh, R.J.: On context-free languages. J. ACM **13**, 570–581 (1966)
22. Puzynina, S., Zamboni, L.Q.: Abelian returns in Sturmian words. J. Comb. Theory **120**(2), 390–408 (2013)

23. Richomme, G., Saari, K., Zamboni, L.Q.: Abelian complexity of minimal subshifts. J. Lond. Math. Soc. **83**(1), 79–95 (2010)
24. Richomme, G., Saari, K., Zamboni, L.Q.: Balance and Abelian complexity of the Tribonacci word. Adv. Appl. Math. **45**(2), 212–231 (2010)
25. Salomaa, A.: Connections between subwords and certain matrix mappings. TCS **340**(2), 188–203 (2005)