arXiv:1906.03689v1 [cs.DM] 9 Jun 2019

Borders, Palindrome Prefixes, and Square Prefixes

Daniel Gabric and Jeffrey Shallit
School of Computer Science
University of Waterloo
Waterloo, ON N2L 3G1
Canada
shallit@uwaterloo.ca

June 11, 2019

Abstract

We show that the number of length-n words over a k-letter alphabet having no even palindromic prefix is the same as the number of length-n unbordered words, by constructing an explicit bijection between the two sets. A similar result holds for those words having no odd palindromic prefix, again by constructing a certain bijection. Using known results on borders, it follows that the number of length-n words having no even (resp., odd) palindromic prefix is asymptotically $\gamma_k \cdot k^n$ for some positive constant γ_k. We obtain an analogous result for words having no nontrivial palindromic prefix. Finally, we obtain similar results for words having no square prefix, thus proving a 2013 conjecture of Chaffin, Linderman, Sloane, and Wilks.

1 Introduction

In this note, we work with finite words over a finite alphabet Σ. For reasons that will be clear later, we assume without loss of generality that $\Sigma = \Sigma_k = \{0, 1, \ldots, k-1\}$ for some integer $k \geq 1$.

We index words starting at position 1, so that $w[1]$ is the first symbol of w and $w[i..j]$ is the factor beginning at position i and ending at position j.

We let w^R denote the *reverse* of a word; thus, for example, $(\texttt{drawer})^R = \texttt{reward}$. A word w is a *palindrome* if $w = w^R$; an example in English is `radar`. A palindrome is *even* if it is of even length, and *odd* otherwise. If a palindrome is of length n, then its *order* is defined to be $\lfloor n/2 \rfloor$. A palindrome is *trivial* if it is of length ≤ 1.

A word w has an *even palindromic prefix* (resp., *odd palindromic prefix*) if there is some nonempty prefix p (possibly equal to w) that is a palindrome of even (resp., odd) length. Thus, for example, the English word `diffident` has the even palindromic prefix `diffid` of order 3, and the English word `selfless` has an odd palindromic prefix `selfles` of order 3.

A *border* of a word w is a word u that is both a prefix and suffix of u. If the only borders of w are the trivial ones (ϵ and w itself), we say it is *unbordered*. Otherwise it is *bordered*. For more about bordered words, see, for example, [10, 3].

Call a border u of a word w *long* if $|u| > |w|/2$ and *short* otherwise. If a word has a long border u, then by considering the overlap of the two occurrences of u, one as prefix and one as suffix, we see that w also has a short border. Given a word w, its set of short border lengths is $\{1 \leq i \leq |w|/2 \ : \ w[1..i] \text{ is a border of } w\}$.

By explicit counting for small n, one quickly arrives at the conjecture that $u_k(n)$, the *number* of length-n words over Σ_k that are unbordered, equals $v_k(n)$, the *number* of length-n words over Σ_k having no even palindromic prefix. This seems to be true, despite the fact that the individual words being counted differ in the two cases. As an example consider 0011, which has an even palindromic prefix 00 but is unbordered. Similarly, if $t_k(n)$ denotes the number of length-n words over Σ_k having no odd palindromic prefix, it is natural to conjecture that $t_k(n) = v_k(n)$ for k odd, and $t_k(n) = kt_k(n-1)$ for k even. The first few terms of the sequences $t_2(n)$ and $v_2(n)$ are given in the following table.

n	1	2	3	4	5	6	7	8	9	10	11	12
$t_2(n)$	2	4	4	8	12	24	40	80	148	296	568	1136
$v_2(n) = u_2(n)$	2	2	4	6	12	20	40	74	148	284	568	1116

The sequence $t_2(n)$ is sequence A308528 in the On-Line Encyclopedia of Integer Sequences (OEIS) [7], and the sequence $u_2(n)$ is sequence A003000.

In fact, even more seems to be true: if S is any set of positive integers, then the number of length-n words w for which S gives the lengths of all the short borders of w is exactly the same as the number of length-n words having even palindromic prefixes with orders given by S. A similar claim (but slightly different) seems to hold for the odd palindromic prefixes. How can we explain this?

The obvious attempts at a bijection (e.g., map uvu to uu^Rv) don't work, because (for example) 00100 and 00010 both map to 00001. Nevertheless, there is a bijection, as we will see below, and this bijection provides even more information.

2 A bijection on Σ^n

The *perfect shuffle* of two words of length n, written $x \sqcup\!\sqcup\!\sqcup\, y$, is defined as follows: if $x = a_1a_2\cdots a_n$ and $y = b_1b_2\cdots b_n$, then

$$x \sqcup\!\sqcup\!\sqcup\, y = a_1b_1a_2b_2\cdots a_nb_n.$$

Thus, for example, `clip` $\sqcup\!\sqcup\!\sqcup$ `aloe` = `calliope`.

Clearly $(x \sqcup\!\sqcup\!\sqcup\, y)^R = y^R \sqcup\!\sqcup\!\sqcup\, x^R$, a fact we use below.

Lemma 1. *A word x is an even palindrome iff there is a word y such that $x = y \sqcup\!\sqcup\!\sqcup\, y^R$.*

Proof. Suppose x is an even palindrome. Then it is of the form $x = tt^R$ for some word t.

By "unshuffling", write t as $(t_1 \sqcup\!\sqcup\, t_2)a$, for words t_1 and t_2, where a is either empty or a single letter, depending on whether $|t|$ is even or odd. Then

$$x = tt^R = (t_1 \sqcup\!\sqcup\, t_2)aa(t_1 \sqcup\!\sqcup\, t_2)^R = (t_1 \sqcup\!\sqcup\, t_2)aa(t_2^R \sqcup\!\sqcup\, t_1^R) = (t_1at_2^R) \sqcup\!\sqcup\, (t_2at_1^R).$$

Letting $y = t_1at_2^R$, we see that $x = y \sqcup\!\sqcup\, y^R$.

Similarly, suppose $x = y \sqcup\!\sqcup\, y^R$. Write $y = y_1ay_2$, where y_1, y_2 are words of equal length, and a is either empty or a single letter, depending on whether $|y|$ is even or odd. Then

$$x = (y_1ay_2) \sqcup\!\sqcup\, (y_2^Ray_1^R) = (y_1 \sqcup\!\sqcup\, y_2^R)aa(y_2 \sqcup\!\sqcup\, y_1^R).$$

Letting $t = (y_1 \sqcup\!\sqcup\, y_2^R)a$, we see that $x = tt^R$. $\square$

For a related result, see [8].

We now define a certain map from Σ^n to Σ^n, as follows:

$$f(x) := (y \sqcup\!\sqcup\, z^R)a$$

if $x = yaz$ with $|y| = |z|$ and a empty or a single letter (depending on whether $|x|$ is even or odd). Thus, for example, $f(\texttt{preserve}) = \texttt{perverse}$ and $f(\texttt{cider}) = \texttt{cried}$. Clearly this map is a bijection.

Theorem 2. *Let w be a word and let $1 \leq i \leq |w|/2$. Then w has a border of length i iff $f(w)$ has an even palindromic prefix of order i.*

Roughly speaking, this theorem says that f "maps borders to orders".

Proof. Suppose w has a border of length i. Then $w = uvu$, where $|u| = i$. Write $v = v_1av_2$, where $|v_1| = |v_2|$ and a is either empty, or a single letter, depending on whether $|v|$ is even or odd. Then

$$f(w) = f(uv_1av_2u) = ((uv_1) \sqcup\!\sqcup\, (v_2u)^R)a = (u \sqcup\!\sqcup\, u^R)(v_1 \sqcup\!\sqcup\, v_2^R)a,$$

which by Lemma 1 has a palindromic prefix of length $2i$ and order i.

Suppose $f(w)$ has an even palindromic prefix of order i. Write $w = yaz$, so that $f(w) = (y \sqcup\!\sqcup\, z^R)a$. Write $y = y_1y_2$ and $z = z_1z_2$ such that $|y_1| = |z_2| = i$. Now

$$f(w) = ((y_1y_2) \sqcup\!\sqcup\, (z_2^Rz_1^R))a = (y_1 \sqcup\!\sqcup\, z_2^R)(y_2 \sqcup\!\sqcup\, z_1^R)a.$$

It follows that $y_1 \sqcup\!\sqcup\, z_2^R$ is a palindrome and hence $z_2 = y_1$. Hence w has a length-i border, namely y_1. $\square$

Corollary 3. *Let $S \subseteq \{1, \ldots, \lfloor n/2 \rfloor\}$. Then the number of length-n words whose short borders are exactly those in S equals the number of length-n words whose even palindromic prefixes are of orders exactly those in S.*

Let $\mathrm{epp}_{k,S}(n)$ denote the number of length-n words over a k-letter alphabet having even palindrome prefixes of order i for each $i \in S$, and no other orders.

Proposition 4. *We have* $\mathrm{epp}_{k,S}(n+1) = k \cdot \mathrm{epp}_{k,S}(n)$ *for n even.*

Proof. Let n be even. Let w be a word over a k-letter alphabet with even palindrome prefix lengths given by S, and let a be a single letter. Then clearly wa has exactly the same palindromic prefixes as w. Since a is arbitrary, the result follows. □

3 Odd palindrome prefixes

Let S be any subset of $\{1, 2, \ldots, \lfloor n/2 \rfloor\}$. Let $\mathrm{opp}_{k,S}(n)$ denote the number of length-n words over a k-letter alphabet having odd palindrome prefixes of order i for each $i \in S$, and no others.

Proposition 5. *We have* $\mathrm{opp}_{k,S}(n+1) = k \cdot \mathrm{opp}_{k,S}(n)$ *for n odd.*

Proof. Exactly like the proof of Proposition 4. □

Theorem 6. *We have*

(a) $\mathrm{opp}_{k,S}(n) = \mathrm{epp}_{k,S}(n)$ *for n odd; and*

(b) $\mathrm{opp}_{k,S}(n) = k \cdot \mathrm{epp}_{k,S}(n-1)$ *for n even.*

Proof. We begin by proving $\mathrm{opp}_{k,S}(n) = k \cdot \mathrm{epp}_{k,S}(n-1)$ for n odd. We do this by creating a k to 1 map from the length-n words with odd palindrome prefix orders given by S to the length-$(n-1)$ words with even palindrome prefix orders given by S.

Here is the map. Let $w = a_1 a_2 \cdots a_n$ be a word of odd length, and define $g(w) = (a_1 + a_2)(a_2 + a_3) \cdots (a_{n-1} + a_n)$, where the addition is performed modulo k. We claim that this is a k to 1 map, and furthermore, it maps words with odd palindrome prefix orders given by S to words with even palindrome prefix orders also given by S.

To see the first claim, observe that if both $g(w)$ and a_1 are given, then we can uniquely reconstruct w. Since a_1 is arbitrary, this gives a k to 1 map.

To see the second claim, suppose $w = a_1 a_2 \cdots a_n$ has an odd palindrome prefix of order i. Then $a_{2i+2-j} = a_j$ for $1 \leq j \leq i + 1$. Hence, applying the map g to w gives

$$(a_1 + a_2)(a_2 + a_3) \cdots (a_{i-1} + a_i)(a_i + a_{i+1})(a_{i+1} + a_i)(a_i + a_{i-1}) \cdots (a_3 + a_2)(a_2 + a_1),$$

which is clearly an even palindrome of order i.

Hence for n odd we get

$$\mathrm{opp}_{k,S}(n) = k \cdot \mathrm{epp}_{k,S}(n-1) = \mathrm{epp}_{k,S}(n),$$

where we have used Proposition 4.

And for n even we get

$$\mathrm{opp}_{k,S}(n) = k \cdot \mathrm{opp}_{k,S}(n-1) = k^2 \cdot \mathrm{epp}_{k,S}(n-2) = k \cdot \mathrm{epp}_{k,S}(n-1).$$

□

Remark 7. It is seductive, but wrong, to think that the map g also maps even-length palindrome prefixes in a k to 1 manner to odd-length palindrome prefixes, but this is not true (consider what happens to the center letter).

4 Applications

As an application of our results we can (for example) determine the asymptotic fraction of length-n words having no nontrivial even palindrome prefix (resp., having no nontrivial odd palindrome prefix).

Corollary 8. *For all $k \geq 2$ there is a constant $\gamma_k > 0$ such that the number of length-n words having no nontrivial even palindrome prefix (resp., having no nontrivial odd palindrome prefix) is asymptotically equal to $\gamma_k \cdot k^n$.*

Proof. Follows immediately from the same result for unbordered words; see [6, 1, 4]. For related results, see [9]. ☐

5 Interlude: the permutation defined by f

The map f defined in Section 2 can be considered as a permutation on $a_1 a_2 \cdots a_n$. In this case, we write it as f_n. For example, if $n = 7$, the resulting permutation f_n is

$$\begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 \\ 1 & 7 & 2 & 6 & 3 & 5 & 4 \end{pmatrix}.$$

This is an interesting permutation that has been well-studied in the context of card-shuffling, where it is called the *milk shuffle*. A classic result about the milk shuffle is the following [5]:

Theorem 9. *The order of the permutation f_n is the least m such that $2^m = \pm 1 \pmod{2n-1}$.*

This is sequence A003558 in the OEIS.

6 No palindrome prefix

In this section we consider the words having no nontrivial palindrome prefix. (Recall that a palindrome is trivial if it is of length ≤ 1.) This is only of interest for alphabet size $k \geq 3$, for if $k = 2$, the only such words are of the form 01^i and 10^i.

Let $a(n) = a_k(n)$ denote the number of such words over a k-letter alphabet. We use the technique of [1, 4] to show that $a_k(n) \sim \rho_k k^n$ for a constant ρ_k and large n.

Proposition 10. *For $n \geq 1$ we have*

$$a(2n) = ka(2n-1) - a(n) \tag{1}$$
$$a(2n+1) = ka(2n) - a(n+1). \tag{2}$$

Proof. Consider the words of length $2n-1$ having no nontrivial palindrome prefix. By appending a new letter, we get $ka(2n-1)$ words. However, some of these words will be palindromes of length $2n$. But since no proper prefix is a palindrome, the first half of these palindromes will be counted by $a(n)$. This gives (1). A similar argument works to prove (2). □

For $k = 3$ the corresponding sequence is given below and is sequence A252696 in the OEIS:

n	1	2	3	4	5	6	7	8	9	10	11	12
$a_3(n)$	3	6	12	30	78	222	636	1878	5556	16590	49548	148422

Now define $t(n) = t_k(n)$ by $t(n) = a(n)k^{-n}$. From (1) and (2) we get

$$t(2n) = t(2n-1) - t(n)k^{-n}$$
$$t(2n+1) = t(2n) - t(n+1)k^{-n}.$$

It now follows that

$$t(2n) = t(2n-2) - (k+1)t(n)k^{-n}. \tag{3}$$

By telescoping cancellation applied to (3), we now get

$$t(2n) = \frac{k-1}{k} - (k+1)\sum_{i=2}^{n} t(i)k^{-i},$$

or

$$t(2n) = 2 - (k+1)\sum_{i=1}^{n} t(i)k^{-i}.$$

If we now set $f(X) = \sum_{n\geq 1} t(n)X^n$, we get

$$\rho_k := \lim_{n\to\infty} t(n) = 2 - (k+1)f(1/k).$$

Note that ρ_k is the limiting frequency of words having no nontrivial palindrome prefix.

Using (3), we now get

$$\begin{aligned}
f(X)(1-X) &= t(1)X + \sum_{n\ge 2}(t(n)-t(n-1))X^n \\
&= t(1)X + \left(\sum_{n\ge 1}(t(2n)-t(2n-1))X^{2n}\right) + \left(\sum_{n\ge 1}(t(2n+1)-t(2n))X^{2n+1}\right) \\
&= X - \left(\sum_{n\ge 1} t(n)k^{-n}X^{2n}\right) + \left(\sum_{n\ge 1} t(n+1)k^{-n}X^{2n+1}\right) \\
&= X - f(X^2/k) - \frac{k}{X}\left(f(X^2/k) - \frac{X^2}{k}\right) \\
&= 2X - \left(1+\frac{k}{X}\right) f(X^2/k),
\end{aligned}$$

and so we get a functional equation for $f(X)$:

$$f(X) = \frac{2X}{1-X} + \frac{X+k}{X(X-1)} f(X^2/k).$$

By iterating this functional equation, and using the fact that $f(\epsilon) \sim \epsilon$ for small real ϵ, we get an expression for $f(1/k)$:

$$\left(\lim_{n\to\infty} \frac{\prod_{i=1}^{n}(k^{2^i}+1)}{k^{n+1}\prod_{i=1}^{n}(k^{2^i-1}-1)}\right) - 2\sum_{n\ge 1} k^{2^{2n-1}-2n}(k^{2^{2n-1}-1}+1)\frac{\prod_{i=1}^{2n-2}(k^{2^i}+1)}{\prod_{i=1}^{2n}(k^{2^i-1}-1)}.$$

This is very rapidly converging; for $k=3$ only 6 terms are enough to get 60 decimal places of $f(1/k)$:

$$\begin{aligned}
f(1/3) &= 0.430377520029471213293382335121830467895548542549528870740458\cdots \\
\rho_3 &= 0.278489919882115146826470659512678128417805829801884517038l6\cdots
\end{aligned}$$

7 Square prefixes

It is natural to conjecture that our bijections connecting words with no border and no even palindrome prefix might also apply to words having no square prefix. However, this is not the case. Let $s_k(n)$ denote the number of length-n words over Σ_k having no square prefix. When $k=2$, for example, the two sequences $s_k(n)$ and $v_k(n)$ differ for the first time at $n=10$, as the following table indicates.

n	1	2	3	4	5	6	7	8	9	10	11	12
$v_2(n)$	2	2	4	6	12	20	40	74	148	284	568	1116
$s_2(n)$	2	2	4	6	12	20	40	74	148	286	572	1124

The sequence $s_2(n)$ is sequence A122536 in the OEIS.

Chaffin, Linderman, Sloane, and Wilks [2, §3.7] conjectured that $s_2(n) \sim \alpha_2 \cdot 2^n$ for a constant $\alpha_2 \doteq 0.27$. In this section we prove this conjecture in more generality.

Theorem 11. *The limit* $\lim_{n\to\infty} s_k(n)/k^n$ *exists and equals a constant* α_k *with* $\alpha_k > 1 - 1/(k-1)$.

Proof. Let $d_k(n) = k^n - s_k(n)$ be the number of length-n words over Σ_k having a nonempty square prefix, and let $c_k(n)$ be the number of squares of length $2n$ over Σ_k having no nonempty proper square prefix. Hence $c_k(1) = k$ and $c_k(2) = k(k-1)$.

The first few values of $c_2(n)$ and $d_2(n)$ are given in the following table.

n	1	2	3	4	5	6	7	8	9	10	11	12
$c_2(n)$	2	2	4	6	10	20	36	72	142	280	560	1114
$d_2(n)$	0	2	4	10	20	44	88	182	364	738	1476	2972

The sequence $c_2(n)$ is sequence A216958 in the OEIS, and the sequence $d_2(n)$ is sequence A121880.

Let w be a word of length n. Either its shortest square prefix is of length 2 (and there are $c_k(1)k^{n-2}$ such words), or of length 4 (and there are $c_k(2)k^{n-4}$ such words), or ...

So $d_k(n)$, the number of words of length n having a nonempty square prefix, is exactly $\sum_{2i\le n} c_k(i)k^{n-2i}$. Hence $d_k(n)/k^n = \sum_{2i\le n} c_k(i)k^{-2i}$. Thus $\lim_{n\to\infty} d_k(n)/k^n$ exists iff the infinite sum $\sum_{i=1}^{\infty} c_k(i)k^{-2i}$ converges. But, since $c_k(i) \le k^i$, this sum converges to some constant $\beta_k < 1/(k-1)$, by comparison with the sum $\sum_{i=1}^{\infty} k^{-i} = 1/(k-1)$. It follows that $s_k(n) = k^n - d_k(n) \sim (1-\beta_k)k^n$. Letting $\alpha_k = 1 - \beta_k$, the result follows. □

To estimate the value of β_k (and hence α_k) we use the inequalities

$$\sum_{i=1}^{n} c_k(i)k^{-2i} \le \beta_k \le \left(\sum_{i=1}^{n} c_k(i)k^{-2i}\right) + \sum_{i=n+1}^{\infty} k^{-i} = \left(\sum_{i=1}^{n} c_k(i)k^{-2i}\right) + (k^{-n})/(k-1).$$

Taking, for example, $n = 20$, we get $\beta_2 \in (0.7299563, 0.7299574)$ and hence $\alpha_2 \in (0.2700426, 0.2700437)$. This can be compared to the analogous constant $\gamma_2 \doteq 0.2677868$ for even palindromes.

References

[1] G. Blom. Overlapping binary sequences (problem 94-20). *SIAM Review* **37** (1995), 619–620.

[2] B. Chaffin, J. P. Linderman, N. J. A. Sloane, and A. R. Wilks. On curling numbers of integer sequences. *J. Integer Sequences* **16** (2013), 13.4.3 (electronic), `https://cs.uwaterloo.ca/journals/JIS/VOL16/Sloane/sloane3.html`

[3] A. Ehrenfeucht and D. M. Silberger. Periodicity and unbordered segments of words. *Discrete Math.* **26** (1979), 101–109.

[4] Š. Holub and J. Shallit. Periods and borders of random words. In N. Ollinger and H. Vollmer, editors, *33rd Symposium on Theoretical Aspects of Computer Science (STACS 2016)*, Vol. 47 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pp. 44:1–44:10. Schloss Dagstuhl–Leibniz-Zentrum für Informatik, Dagstuhl, Germany, 2016. Available at http://drops.dagstuhl.de/opus/volltexte/2016/5745.

[5] P. Lévy. Sur quelques classes de permutations. *Compositio Math.* **8** (1951), 1–48.

[6] P. Tolstrup Nielsen. A note on bifix-free sequences. *IEEE Trans. Inform. Theory* **IT-19** (1973), 704–706.

[7] N. J. A. Sloane et al. The on-line encyclopedia of integer sequences. Available at https://oeis.org, 2018.

[8] N. Rampersad, J. Shallit, and M.-w. Wang. Inverse star, borders, and palstars. *Inform. Process. Lett.* **111** (2011), 420–422.

[9] L. B. Richmond and J. Shallit. Counting the palstars. *Electronic J. Combinatorics* **21** (2014), P3.25 (electronic), https://www.combinatorics.org/ojs/index.php/eljc/article/view/v21i3p25

[10] D. M. Silberger. Borders and roots of a word. *Portugal. Math.* **30** (1971), 191–199.