

# On the Combinatorics of Palindromes and Antipalindromes

Chuan Guo<sup>1</sup>, Jeffrey Shallit<sup>1</sup>, and Arseny M. Shur<sup>2</sup>

<sup>1</sup> University of Waterloo, Ontario, Canada  
 {c3guo, shallit}@uwaterloo.ca

<sup>2</sup> Ural Federal University, Ekaterinburg, Russia  
 arseny.shur@urfu.ru

**Abstract.** We prove a number of results on the structure and enumeration of palindromes and antipalindromes. In particular, we study conjugates of palindromes, palindromic pairs, rich words, and the counterparts of these notions for antipalindromes.

## 1 Introduction

Combinatorial and algorithmic studies of palindromes can be traced back to the 1970's, when they were considered as a promising tool to construct a “hard” context-free language, which cannot be recognized by a linear-time random access machine. Nevertheless, palindrome-based languages were proved to be linear-time recognizable [5, 9, 10]. Recent topics of interest in the study of palindromes include, for example, *rich words* (containing the maximum possible number of distinct palindromes; see [1, 2, 6]) and *palstars* (products of even-length palindromes; see [9, 14, 15]). Also, there is a popular modification of the notion of palindrome, where the reversal of a word coincides not with the word itself, but with the image of the word under a certain involution of the alphabet; see, e.g., [7, 13]. In the binary case, there is a unique such modification, called an *antipalindrome*.

In this paper, we aim to fill certain gaps in the knowledge on combinatorics of palindromes and antipalindromes. The four subsequent sections are mostly independent. In Section 2, we study the distribution of palindromes among conjugacy classes and enumerate conjugates of palindromes. Section 3 is devoted to the words which are products of two palindromes; we prove some characterizations of this class of words and show that the number of  $k$ -ary words that are products of two odd-length palindromes is exactly  $k$  times the number of  $k$ -ary words of the same length that are products of two even-length palindromes. In Section 4, we analyze the growth function for the language of binary rich words. We give the first nontrivial lower bound, of order  $C\sqrt{n}$  for a constant  $C$ , and provide some empirical evidence that this growth function does indeed have subexponential growth. Finally, in Section 5 we focus on antipalindromes. We show that antipalindromes share many common properties with palindromes, with a notable exception: the notion of a rich word becomes trivial.

*Definitions and notation.* We study finite words over finite alphabets, using the array notation  $w = w[1..n]$  when appropriate. The notions of prefixes, suffixes, factors, periods and (integer) powers are as usual. We write  $|w|$  for the length of  $w$  and  $\varepsilon$  for the empty word. For two words  $v$  and  $w$  of length  $n$ , their *perfect shuffle*  $v \text{ III } w$  is the word  $v[1]w[1]v[2]w[2] \cdots v[n]w[n]$ . Thus, for example,  $\text{clip III aloe} = \text{calliope}$ . Given a word  $w$ , let  $w^R$  denote its reversal (e.g.,  $(\text{stressed})^R = \text{desserts}$ ). A word  $w$  is a *palindrome* if  $w = w^R$ . A word is *primitive* if it is not an integer power of a shorter word. Two words  $u$  and  $v$  are *conjugates* if  $u = xy$  and  $v = yx$  for some words  $x$  and  $y$ . Conjugacy is an equivalence relation. The following lemma is folklore.

**Lemma 1.** *Let  $u = z^i$  for a primitive word  $z$ . Then the conjugacy class of  $u$  contains exactly  $|z|$  words.*

We use two basic properties of periodic words due to Lyndon and Schützenberger [11].

**Lemma 2.** (i) *For any nonempty words  $u$  and  $v$ , the equality  $uv = vu$  holds if and only if  $u = z^i$  and  $v = z^j$  for some word  $z$  and positive integers  $i, j$ .*  
(ii) *For any nonempty words  $u, v$ , and  $w$ , the equality  $uw = vw$  holds if and only if  $u = xy$ ,  $v = yx$ ,  $w = (xy)^i x$  for some words  $x \neq \varepsilon$  and  $y$ , and nonnegative integer  $i$ .*

For a language  $L$  over an alphabet  $\Sigma$ , its *growth function* (also called combinatorial complexity or census function) is defined to be  $C_L(n) = |L \cap \Sigma^n|$ .

Below, we list some basic properties of palindromes.

**Proposition 1.** *For all integers  $m, n \geq 1$ , the word  $x^m$  is a palindrome if and only if  $x^n$  is a palindrome.*

**Proposition 2.** *For all nonempty palindromes  $u, v$ , the word  $uv$  is a palindrome if and only if both  $u, v$  are powers of some palindrome  $z$ .*

*Proof.* Suppose  $uv$  is a palindrome. Then  $uv = (uv)^R = v^R u^R = vu$ . By Lemma 2 (i),  $u = z^i$ ,  $v = z^j$  for integers  $i, j \geq 1$ . By Proposition 1, we have that  $z$  is a palindrome. Conversely, if  $u = z^i$  and  $v = z^j$ , we have  $uv = z^{i+j}$ , which is a palindrome by Proposition 1.  $\square$

**Proposition 3.** *The word  $x$  is an even-length palindrome iff there exists a word  $y$  such that  $x = y \text{ III } y^R$ .*

## 2 Conjugates of Palindromes

Here we study the distribution of palindromes in conjugacy classes and count conjugates of palindromes.

**Theorem 1.** *A conjugacy class contains at most two palindromes. A conjugacy class has two palindromes if and only if it contains a word of the form  $(xx^R)^i$ , where  $xx^R$  is a primitive word and  $i \geq 1$ .*

**Lemma 3.** *Suppose  $u \neq u^R$  and  $uu^R = z^i$  for a primitive word  $z$ . Then  $i$  is odd and  $z = xx^R$  for some  $x$ .*

*Proof.* If  $i$  is even, then  $uu^R = (z^{i/2})^2$ . Hence,  $u = u^R$ , contradicting the conditions of the lemma. So  $i$  is odd and then  $|z|$  is even. Let  $z = xx'$ , where  $|x| = |x'|$ . We see that  $x$  is a prefix of  $u$  and  $x'$  is a suffix of  $u^R$ . Hence,  $x' = x^R$ , as required.  $\square$

*Proof (of Theorem 1).* Let us prove that for any conjugacy class with two distinct palindromes, say  $uv$  and  $vu$ , there exists a word  $x$  and a number  $i$  such that  $xx^R$  is primitive,  $uv = (xx^R)^i$ , and  $vu = (x^Rx)^i$ . We use induction on  $n = |uv|$ . The base case is trivial, because such conjugacy classes do not exist for, say,  $n = 1$ .

For the inductive step, assume  $|u| \leq |v|$  without loss of generality. If  $|u| = |v|$ , then  $v = u^R$ . By Lemma 3 we get  $uv = (xx^R)^i$ ,  $vu = (x^Rx)^i$  for a primitive word  $xx^R$  and  $i \geq 1$ .

Now let  $|u| < |v|$ . Then  $v$  begins and ends with  $u^R$ . Applying Lemma 2 (ii), we obtain  $u^R = (st)^i s$ , where  $s \neq \varepsilon$  and  $i \geq 0$  and, respectively,  $v = (st)^{i+1} s$ . Looking at the central factor of the palindromes

$$uv = (s^R t^R)^i s^R \cdot st \cdot s(ts)^i, \quad (1)$$

$$vu = (st)^i s \cdot ts \cdot s^R (t^R s^R)^i, \quad (2)$$

we see that  $st$  and  $ts$  are also palindromes. If  $t = \varepsilon$ , then  $s$  is a palindrome, implying  $uv = vu$ , which is impossible. If  $st = ts$ , then by Lemma 2 (i) both  $s$  and  $t$  are powers of some  $z$ . By Proposition 1,  $s$ ,  $t$ , and  $z$  are palindromes, and then again  $uv = vu$ . Therefore, we obtain  $st \neq ts$ . So we can apply the inductive hypothesis to these two palindromes, getting  $st = (xx^R)^j$ ,  $ts = (x^Rx)^j$  for a primitive word  $xx^R$ . Then we can write

$$v = (xx^R)^{j(i+1)} s = s(x^Rx)^{j(i+1)}. \quad (3)$$

Since all conjugates of the word  $xx^R$  are distinct words by Lemma 1,  $x$ 's and their reversals occur in the same positions in both representations (3). Then  $s = (xx^R)^k x$  for some  $k$ ,  $0 \leq k < j$ . Then we can easily compute  $t$ ,  $s^R$ , and  $t^R$  to get  $uv = (xx^R)^{2j(i+1)}$ ,  $vu = (x^Rx)^{2j(i+1)}$ . Thus, we have finished the inductive step.

Note that a conjugacy class of a palindrome  $(xx^R)^i$ , where  $xx^R$  is primitive, clearly contains a different palindrome  $(x^Rx)^i$ . To finish the proof of the theorem it remains to show that such a class contains no other palindromes. Indeed, consider the class of  $w = (xx^R)^i$ . It consists of  $i$ th powers of conjugates of  $xx^R$  (see Lemma 1). Let  $u$  and  $v$  be such that  $uv = xx^R$  and  $(vu)^i$  is a palindrome. Then  $vu$  is a palindrome by Proposition 1. If  $|u| \neq |v|$ , say,  $|u| < |v|$ , then we apply the above argument to  $uv$  and  $vu$ , getting  $uv = (yy^R)^{2k}$  for some  $y$  and  $k$ . But this is impossible, because  $uv = xx^R$  is primitive. Hence,  $|u| = |v|$ , and thus  $vu = x^R x$ . Thus, the class of  $w$  contains exactly two palindromes.  $\square$

**Corollary 1.** *A conjugacy class of a word  $w = z^m$ , where  $z$  is primitive and  $m \geq 1$ , contains*

- (i) 0 or 1 palindrome, if  $|z|$  is odd;
- (ii) 0 or 2 palindromes, if  $|z|$  is even.

Let  $n = p_1^{i_1} \cdots p_k^{i_k}$ , where  $p_1, \dots, p_k$  are primes. Recall that the Möbius function  $\mu(n)$  equals  $(-1)^k$  if  $i_1 = \dots = i_k = 1$  (i.e., if  $n$  is *square-free*) and 0 otherwise. The following result apparently first appeared in [12].

**Lemma 4.** *The number  $\rho(k, n)$  of  $k$ -ary words of length  $n$  that are both (i) primitive and (ii) a palindrome satisfies the following formula*

$$\rho(k, n) = \sum_{d|n} \mu(d) k^{\lfloor (n/d+1)/2 \rfloor}. \quad (4)$$

We use (4) and Corollary 1 to count the conjugates of palindromes.

**Theorem 2.** *The number of conjugates of  $k$ -ary palindromes of length  $n$  is*

$$c(k, n) = \sum_{d|n} f(d) \cdot \rho(k, d), \text{ where } f(d) = \begin{cases} d, & \text{if } d \text{ is odd;} \\ d/2, & \text{if } d \text{ is even.} \end{cases} \quad (5)$$

*Proof.* If  $n$  is odd, it suffices to count the number of palindromes and multiply it by the number of distinct conjugates given in Lemma 1. Instead of non-primitive palindromes, we count their primitive roots, which are palindromes by Proposition 1. Thus we have

$$c(k, n) = \sum_{d|n} d \cdot \rho(k, d),$$

which is equivalent to (5) because  $d$  takes only odd values. If  $n$  is even, some classes contain two palindromes. Let  $w = z^i$ , where  $z$  is primitive. If  $|z|$  is odd, then the class of  $w$  contains 0 or 1 palindrome, while if  $|z|$  is even, the class of  $w$  contains 0 or 2 palindromes. In the latter case, we must divide the result of counting the conjugates of palindromes by 2. This gives precisely (5).  $\square$

### 3 Palindromic Pairs

In this section we consider palindromic pairs, which are words factorizable into two palindromes. First we give a few easy characterizations of palindromic pairs, and then prove Theorem 4 on the number of “even” and “odd” palindromic pairs.

Let  $\mathcal{P}$  be the set of palindromes over the alphabet  $\Sigma$ . Palindromic pairs are exactly the elements of  $\mathcal{P}^2$ . Recall that  $L_1/L_2 = \{x \in \Sigma^* \mid \exists y \in L_2 : xy \in L_1\}$ .

**Proposition 4.**  $\mathcal{P}/\mathcal{P} = \mathcal{P}^2$ .

*Proof.*  $\mathcal{P}/\mathcal{P} \subseteq \mathcal{P}^2$ : Suppose  $x \in \mathcal{P}/\mathcal{P}$ . Then there exists a palindrome  $y$  such that  $xy$  is a palindrome. If either  $x$  or  $y$  is empty, the result is clearly true. Otherwise  $xy = (xy)^R = y^R x^R = yx^R$ . Then by Lemma 2 (ii) there exist  $u \in \Sigma^+$ ,  $v \in \Sigma^*$  and an integer  $e \geq 0$  such that  $x = uv$ ,  $x^R = vu$ , and  $y = (uv)^i u$ . From  $x^R = vu$  we get  $x = u^R v^R$ . But  $x = uv$ , so  $u = u^R$  and  $v = v^R$ . So  $u, v \in \mathcal{P}$  and then  $x \in \mathcal{P}^2$ .

$\mathcal{P}^2 \subseteq \mathcal{P}/\mathcal{P}$ : If  $x = uv$  with  $u, v$  palindromes, then  $uvu$  is a palindrome. Thus, taking  $y = u$ , we have  $xy \in \mathcal{P}$  and  $y \in \mathcal{P}$ . Hence  $x \in \mathcal{P}/\mathcal{P}$ .  $\square$

Call a word  $x$  *credible* if it is a conjugate of its reverse  $x^R$ , like the English word **referee**.

**Proposition 5.** *The word  $x$  is in  $\mathcal{P}^2$  if and only if it is credible.*

*Proof.* Suppose  $x \in \mathcal{P}^2$ , that is, that  $x = uv$  where  $u, v$  are palindromes. Then  $x^R = v^R u^R = vu$ , so  $x$  is a conjugate of  $x^R$ .

Otherwise, assume  $x$  is a conjugate of  $x^R$ . Then there exist  $u, v$  such that  $x = uv$  and  $x^R = vu$ . But  $x^R = v^R u^R$ , so  $v = v^R$  and  $u = u^R$ , and  $x$  is the product of two palindromes.  $\square$

*Remark 1.* The growth function for  $\mathcal{P}^2$  was studied by Kemp [8], who computed a precise formula and described its asymptotics. For the binary alphabet, see sequence A007055 in the *On-Line Encyclopedia of Integer Sequences* [18].

A factorization of a palindromic pair  $w$  is a pair of palindromes  $u, v$  such that  $uv = w$  and  $v \neq \varepsilon$ . The number of factorizations of a palindromic pair is described in the following theorem, also due to Kemp [8].

**Theorem 3.** *A palindromic pair  $w$  has  $m$  factorizations if and only if  $w = z^m$  for a primitive word  $z$ .*

A palindromic pair  $w$  is even (resp., odd) if it can be factorized into two palindromes of even (resp., odd) length. Thus, an even-length palindromic pair is either even, or odd, or both, like the word

$$aabaab = aa \cdot baab = aabaa \cdot b. \quad (6)$$

Note that in view of Theorem 3, the last option applies to non-primitive words only. Let  $E(n, k)$  (resp.,  $O(n, k)$ ) denote the number of even (resp., odd)  $k$ -ary palindromic pairs of length  $n$ .

**Theorem 4.** *For all  $n$  and  $k$  we have  $O(n, k) = k \cdot E(n, k)$ .*

The proof is based on several lemmas. The following lemma is a well-known corollary of the Fine-Wilf property [4]:

**Lemma 5.** *Let  $p$  be the minimal period of a word  $w$ ,  $q$  be a period of  $w$ , and  $p, q \leq |w|/2$ . Then  $q$  is a multiple of  $p$ .*

We call an even-length word  $w$  *even-primitive* if it is not an integer power of a shorter even-length word. The difference between primitive and even-primitive words is clarified in the following lemma.

**Lemma 6.** *An even-length non-primitive word is even-primitive iff it is the square of an odd-length primitive word.*

*Proof.* Let  $w$  be an even-length non-primitive word. Then  $w = z^m$ , for a primitive word  $z$  and an integer  $m > 1$ . Such a pair  $(z, m)$  is unique. Indeed,  $|z|$  is a period of  $w$ ,  $|z| \leq |w|/2$ , and  $|z|$  is not a multiple of a shorter period of  $w$  due to primitivity. Then  $|z|$  is the minimal period of  $w$  by Lemma 5.

*Necessity.* Since  $w$  is even-primitive,  $|z|$  is odd. Then  $m$  is even, because  $|w| = m|z|$ . If  $m \geq 4$ , then  $w = (z^2)^{m/2}$  is not even-primitive. Hence,  $m = 2$ , as desired.

*Sufficiency* stems from the fact that  $|z|$  is the minimal period of  $w$ .  $\square$

A palindromic pair can have several factorizations according to Theorem 3. However, the even-primitive words have the following useful property (cf. (6)).

**Lemma 7.** *An even-length even-primitive word has at most one factorization into two even-length palindromes and at most one factorization into two odd-length palindromes.*

*Proof.* Let  $u_1, u_2, v_1, v_2$  be even-length palindromes such that  $u_1v_1 = u_2v_2 = w$  and  $|v_1| < |v_2|$  (recall that  $v_1 \neq \varepsilon$  by the definition of factorization). Then  $v_1$  is a suffix of  $v_2$  and, as both these words are palindromes, a prefix of  $v_2$  as well. By Lemma 2 (ii),  $v_1 = (xy)^s x$ ,  $v_2 = (xy)^{s+1} x$  for some words  $x \neq \lambda$  and  $y$ , and some integer  $s \geq 0$ ; in addition, it is clear that both  $x$  and  $y$  are palindromes. Hence,  $u_1 = u_2xy$ , and, since  $u_1, u_2, x, y$  are palindromes,  $u_1 = yxu_2$ . Applying Lemma 2 (ii) again, we finally get  $w = (yx)^t$  for some  $t \geq 2$ . Since  $|yx| = |v_2| - |v_1|$  is even,  $w$  is not even-primitive.

The same argument works for odd-length palindromes  $u_1, u_2, v_1, v_2$ .  $\square$

**Lemma 8.** *Let  $z$  be an even-primitive word,  $m > 1$  be an integer. Then  $z^m$  is an even (resp., odd) palindromic pair iff  $z$  is.*

*Proof. Necessity.* If  $z^m = uv$  for palindromes  $u$  and  $v$ , then  $u = (xy)^t x$ ,  $v = y(xy)^{m-t-1}$  for some words  $x, y$  and some integer  $t$  such that  $xy = z$  and  $0 \leq t \leq m-1$ . Since  $u$  and  $v$  are palindromes,  $x$  and  $y$  are also palindromes, implying that  $z$  is a palindromic pair. Since  $|z| = |xy|$  is even by definition, the numbers  $|x|$  and  $|u|$  (resp.,  $|y|$  and  $|v|$ ) have the same parity, whence the result.

*Sufficiency.* If  $z = xy$  for palindromes  $x$  and  $y$ , then  $z^m = x \cdot y(xy)^{m-1}$ . The word  $y(xy)^{m-1}$  is a palindrome and its length has the same parity as  $y$ .  $\square$

In what follows we suppose that the size  $k > 1$  of the alphabet is fixed (for the unary alphabet the theorem is straightforward), and write  $E(n)$  (resp.,  $O(n)$ ) for  $E(n, k)$  (resp.,  $O(n, k)$ ). We also assume that  $n$  is even, because  $E(n) = O(n) = 0$  for odd  $n$ . Let  $E'(n)$  (resp.,  $O'(n)$ ) be the number of even-primitive even (resp., odd) palindromic pairs of length  $n$ .

**Lemma 9.** *If  $O'(n) = k \cdot E'(n)$  for all  $n$ , then  $O(n) = k \cdot E(n)$  for all  $n$ .*

*Proof.* Any word  $w$  of length  $n$  can be uniquely represented as  $w = z^{n/2d}$ , where  $2d$  is a divisor of  $n$  and the word  $z$  of length  $2d$  is even-primitive. Then by Lemma 8 the number of even (resp., odd) palindromic pairs of this form is equal to the number of even (resp., odd) palindromic pairs  $z$ . The latter number is exactly  $E'(2d)$  (resp.,  $O'(2d)$ ). Hence,

$$O(n) = \sum_{2d|n} O'(2d) = k \cdot \sum_{2d|n} E'(2d) = k \cdot E(n). \quad \square$$

*Proof (of Theorem 4).* In view of Lemma 9, it suffices to prove the equality  $O'(n) = k \cdot E'(n)$  for all even  $n$ .

Let  $P_i(n)$  be the set of all palindromic pairs  $w = uv$  such that  $|w| = n$ ,  $|u| = i$ . Then the sets  $P_e(n)$  of all even palindromic pairs and  $P_o(n)$  of all odd palindromic pairs of length  $n$  can be written as

$$P_e(n) = P_0(n) \cup P_2(n) \cup \dots \cup P_{n-2}(n) \quad (7a)$$

$$P_o(n) = P_1(n) \cup P_3(n) \cup \dots \cup P_{n-1}(n) \quad (7b)$$

In general, the sets  $P_i(n)$  may intersect, but Lemma 7 tells us that an even-primitive word belongs to at most one set of (7a) and at most one set (7b). Furthermore, let  $2d$  be a divisor of  $n$ , and consider an even palindromic pair of the form  $w = z^{n/2d}$ , where  $z$  is even-primitive. Then  $z$  is an even palindromic pair by Lemma 8. By Lemma 7,  $z$  has a unique factorization into two even-length palindromes, say,  $z = xy$ . Hence,  $w$  has exactly  $n/2d$  factorizations into two even-length palindromes (see the proof of Lemma 8):

$$w = x \cdot y(xy)^{n/2d-1} = xyx \cdot y(xy)^{n/2d-2} = (xy)^{n/2d-1}x \cdot y.$$

Thus,  $w$  belongs to exactly  $n/2d$  sets (7a). In the same way, an odd palindromic pair of the form  $w = z^{n/2d}$  belongs to exactly  $n/2d$  sets (7b). To get the value of  $E'(n)$  (resp.,  $O'(n)$ ), we must sum up the cardinalities of all sets (7a) (resp., (7b)) and subtract the total contribution of the words that are not even-primitive. Since the number of even (resp., odd) palindromic pairs of the form  $w = z^{n/2d}$  is the same as the number of even-primitive even (resp., odd) palindromic pairs  $z$  (of length  $2d$ ), we have

$$E'(n) = \sum_{i=0}^{n/2-1} \#P_{2i}(n) - \sum_{\substack{2d|n, \\ 2d < n}} \frac{n}{2d} \cdot E'(2d), \quad O'(n) = \sum_{i=0}^{n/2-1} \#P_{2i+1}(n) - \sum_{\substack{2d|n, \\ 2d < n}} \frac{n}{2d} \cdot O'(2d). \quad (8)$$

Now let us compute  $\#P_i(n)$ . If  $i = 2j$ , then to determine a word from  $P_i(n)$  we can arbitrarily choose the first  $j$  letters to fix the first palindrome and the last  $(n/2 - j)$  letters to fix the second one. In total, we have  $n/2$  letters, concluding that  $\#P_i(n) = k^{n/2}$ , independent of  $i$ . On the other hand, if  $i = 2j+1$ , then to determine a word from  $P_i(n)$  we must choose the first  $j+1$  letters to fix the first palindrome and the last  $(n/2 - j)$  letters to fix the second one (the central letters of both odd-length palindromes must be chosen). Thus, in this case we have  $\#P_i(n) = k^{n/2+1}$ , also independent of  $i$ . Therefore, we rewrite (8) as

$$E'(n) = \frac{n}{2} \cdot k^{n/2} - \sum_{\substack{2d|n, \\ 2d < n}} \frac{n}{2d} \cdot E'(2d), \quad O'(n) = \frac{n}{2} \cdot k^{n/2+1} - \sum_{\substack{2d|n, \\ 2d < n}} \frac{n}{2d} \cdot O'(2d). \quad (9)$$

Finally, we prove the equality  $O'(n) = k \cdot E'(n)$  by induction. The base case is  $n = 2$ . All words of length 2 are even-primitive, and all of them are odd palindromic pairs. In contrast, the only even palindromic pairs of length 2 are palindromes (of the form  $aa$ ). Thus,  $O'(2) = k^2$ ,  $E'(2) = k$ . For the inductive step, note that the ratio of the first terms for  $O'(n)$  and  $E'(n)$  in (9) is  $k$ , and the inductive assumption implies that the ratio of the second terms (i.e., sums) is also  $k$ . The theorem is proved.  $\square$

## 4 Rich Words

Recall that a word of length  $n$  is (palindromic) rich if it contains  $n$  distinct nonempty palindromic factors. Basic properties of rich words were proved in [2,6]. Some of them are collected in the following

**Proposition 6.** (i) A factor of a rich word is rich.

(ii) A reversal of a rich word is rich.

(iii) A word  $w$  is rich if and only if the longest palindromic suffix of any its prefix  $w[1..i]$  has no other occurrences in  $w[1..i]$ .

Let  $\mathcal{R}$  be the language of binary rich words. Here we analyze its growth function (sequence A216264 in the OEIS [18]), starting with a lower bound. This is, to our knowledge, the first nontrivial (not polynomial) lower bound obtained for the number of binary rich words.

Every word has a unique *run-length encoding* of the form  $a_1^{s_1} a_2^{s_2} \cdots a_k^{s_k}$ , for  $s_1, \dots, s_k \geq 1$ , where the letters  $a_i$  and  $a_{i+1}$  are distinct for any  $i$ . We call each term  $a_i^{s_i}$  a *block*. Consider the language  $I$  of binary words whose run-length encoding satisfies  $s_i \leq s_{i+2}$  for all  $i = 1, \dots, k-2$ . The language  $I$  is close to the languages of intermediate growth studied in [16].

**Theorem 5.**  $I \subseteq \mathcal{R}$ .

*Proof.* Let  $w = 0^{s_1} 1^{s_2} 0^{s_3} \cdots a_k^{s_k} \in I$  for  $s_1, \dots, s_k \geq 1$ . We prove the richness of all the prefixes of  $w$  (including  $w$  itself) by induction on the length of  $w$ .

The base case is trivial. Now assume that  $w[1..i]$  is rich and  $w[i] = 0$  (the case  $w[i] = 1$  is analysed in the same way). We add the letter  $w[i+1]$  and search for the “new” palindromic suffix described in Proposition 6 (iii). If  $w[1..i] = 0^i$ , there is nothing to prove: the new suffix is  $0^{i+1}$  if  $w[i+1] = 0$  and  $1$  if  $w[i+1] = 1$ . So let  $w[1..i] = 0^{s_1} 1^{s_2} \cdots 0^{s'_l}$ , where  $s'_l \leq s_l$  and  $1 < l \leq k$ .

If  $s'_l \geq s_{l-2}$  and  $w[i+1] = 0$ , then the new palindromic suffix is  $0^{s'_l+1}$ ; if  $s'_l > s_{l-2}$  and  $w[i+1] = 1$ , then the new suffix is  $10^{s'_l}1$ . Thus, two nontrivial cases remain.

*Case 1:*  $s'_l < s_{l-2}$ ; clearly then  $w[i+1] = 0$ . The word  $w[1..i]$  ends with a palindrome  $0^{s'_l}$ , but it has more occurrences in  $w[1..i]$ . Hence, it is not the longest palindromic suffix. Then this suffix, which we denote by  $v$ , begins with  $0^{s'_l}$  and ends with  $1^{s_{l-1}} 0^{s'_l}$ . If  $v = 0^{s'_l} 1^{s_{l-1}} 0^{s'_l}$ , then  $0v$  is a suffix of  $w[1..i]$  because  $s'_l < s_{l-2}$ . Hence, the palindrome  $0v0$  is a suffix of  $w[1..i+1]$ , and it has no earlier occurrences because  $v$  occurs in  $w[1..i]$  only once.

If  $v$  intersects more than three blocks, then  $v = 0^{s'_l} (1^{s_{l-1}} 0^{s_{l-2}})^k 1^{s_{l-1}} 0^{s'_l}$  for some  $k \geq 1$ . If  $v$  is preceded by  $0$  in  $w$ , then the suffix  $0v0$  of  $w[1..i+1]$  is a new palindrome, as before. Finally, if  $v$  is preceded by  $1$ , consider the palindrome  $v' = 0^{s'_l} (1^{s_{l-1}} 0^{s_{l-2}})^{k-1} 1^{s_{l-1}} 0^{s'_l}$ . The word  $w[1..i+1]$  ends with  $0v'0$ , and this is a new palindrome, because  $0v'$  occurs in  $w[1..i]$  only as a suffix. (Indeed,  $0v'$  begins with  $0^{s'_l+1}$ , and there is no large enough block of zeroes to the left of  $v$ .)

*Case 2:*  $s'_l = s_{l-2}$  and  $w[i+1] = 1$ . Since  $0^{s'_l} 1^{s_{l-1}} 0^{s'_l}$  is a palindromic suffix of  $w[1..i]$ , the longest such suffix is  $v = (0^{s'_l} 1^{s_{l-1}})^k 0^{s'_l}$  for some  $k \geq 1$ . Clearly,  $1v$  is a suffix of  $w[1..n]$ , implying that  $w[1..i+1]$  ends with the new palindrome  $1v1$ .

Thus, in all cases,  $w[1..i+1]$  ends with a new palindrome. The inductive step is finished.  $\square$

**Theorem 6.** The growth function of the language of binary rich words satisfies

$$\ln(C_{\mathcal{R}}(n)) \geq \frac{2\pi}{\sqrt{3}} \cdot \sqrt{n} - O(\ln n). \quad (10)$$

*Proof.* Let  $p(n)$  [ $p(n, k)$ ] denote the number of integer partitions (resp., partitions with exactly  $k$  parts) of  $n$ . There is a natural injection of partitions of  $n$  into words of length  $n$ : a partition  $s_1 + \dots + s_k = n$ , where the parts are written in increasing order, defines the word  $w = 0^{s_1} 1^{s_2} 0^{s_3} \dots a_k^{s_k}$ . Note that  $w \in I$ , implying  $C_I(n) \geq p(n)$ . By the famous Hardy-Ramanujan-Uspensky formula,

$$p(n) \sim \frac{e^{\pi\sqrt{2n/3}}}{4n\sqrt{3}} \text{ as } n \rightarrow \infty. \quad (11)$$

So we already have the bound similar to (10), but with a smaller constant in the leading term. To get the desired constant, we assume  $n$  to be even (since  $C_I(n)$  is an increasing function, substituting the bound obtained for  $n = 2m$  for  $n = 2m+1$  gives the general bound of the same order of growth). Note that any pair of partitions  $s_1 + \dots + s_k = t_1 + \dots + t_k = n/2$  can be mapped to the word  $0^{s_1} 1^{t_1} \dots 0^{s_k} 1^{t_k} \in I$ , and this map is injective. Then we have

$$C_I(n) > \sum_{k=1}^{n/2-1} (p(n/2, k))^2 > (\max_k p(n/2, k))^2 > \frac{4(p(n/2)^2)}{n^2}. \quad (12)$$

Substituting (11) and taking logarithms, we obtain (10).  $\square$

*Remark 2.* More precise estimates of  $C_I(n)$  cannot give better bounds for  $C_{\mathcal{R}}(n)$ , because all inequalities in (12) affect only the  $O$ -term in (10). Using Proposition 6 (i,ii), one can extend  $I$ , closing it under taking factors and reversals; but the effect of such extensions is swallowed by the  $O$ -term as well.

How good is this lower bound, which is roughly  $37\sqrt{n}$  divided by a polynomial? It is unclear, because no good upper bound for  $C_{\mathcal{R}}(n)$  has yet been obtained. The function  $C_{\mathcal{R}}(n)$  is submultiplicative due to Proposition 6 (i), so its growth rate  $\lim_{n \rightarrow \infty} (C_{\mathcal{R}}(n))^{1/n}$  is majorized by any value  $(C_{\mathcal{R}}(n))^{1/n}$ , according to Fekete's lemma [3]. Also, the ratio  $\frac{C_{\mathcal{R}}(n)}{C_{\mathcal{R}}(n-1)}$  gives us a clue about the growth of this function. Until recently, the number of known values of  $C_{\mathcal{R}}(n)$  was quite small ( $n \leq 25$  on OEIS, posted by the second author). But short rich words constitute a substantial share of all short words, giving us  $(C_{\mathcal{R}}(25))^{1/25} \approx 1.818$  and  $\frac{C_{\mathcal{R}}(25)}{C_{\mathcal{R}}(24)} \approx 1.599$ . Recently Mikhail Rubinchik (personal communication) invented a new technique and computed this sequence up to  $n = 60$ . From his calculations we get  $(C_{\mathcal{R}}(60))^{1/60} \approx 1.605$  and  $\frac{C_{\mathcal{R}}(60)}{C_{\mathcal{R}}(59)} \approx 1.394$ . Such a fast drop is unusual for exponential growth functions of languages closed under factors (see [17] and the references therein). So this is an argument in favor of a subexponential growth of  $C_{\mathcal{R}}(n)$ .

A much stronger argument is provided in Table 1 below:  $C_{\mathcal{R}}(n) \leq n\sqrt{n}$  for  $4 \leq n \leq 60$ , and, moreover, the function  $n\sqrt{n}$  seems to grow faster. So we propose the following conjecture, which implies that the bound of Theorem 6 is quite reasonable.

*Conjecture 1.* One has  $C_{\mathcal{R}}(n) = O\left(\frac{n}{g(n)}\right)^{\sqrt{n}}$ , or, equivalently,  $\ln(C_{\mathcal{R}}(n)) = O(\sqrt{n}(\ln n - g(n)))$ , for some infinitely growing function  $g(n)$ .

**Table 1.** Number of rich words compared to the  $n^{\sqrt{n}}$  function.

$n$	$C_{\mathcal{R}}(n)$	$n^{\sqrt{n}}$	Ratio
4	16	16	1
5	32	$\approx 36.55$	0.875
...			
25	3 089 518	$\approx 9.766 \cdot 10^6$	0.335
26	4 903 164	$\approx 1.641 \cdot 10^7$	0.316
...			
59	1 530 103 385 844	$\approx 4.001 \cdot 10^{13}$	0.038
60	2 132 734 033 216	$\approx 5.936 \cdot 10^{13}$	0.036

## 5 Antipalindromes

In this section, the alphabet is  $\{0, 1\}$ . For a word  $x \in \{0, 1\}^*$ , its *negation*  $\bar{x}$  is obtained by changing each 0 in  $x$  to 1 and vice versa. A word  $x$  is an *antipalindrome* if  $x = \bar{x}^R$ . Thus, for example, 001011 is an antipalindrome; note that all antipalindromes have even length. Let  $\mathcal{A}$  denote the set of all antipalindromes. There are definite similarities between the properties of palindromes and antipalindromes. The following analogs of Propositions 1-3 are straightforward.

**Proposition 7.** *For all integers  $m, n \geq 1$ ,  $x^m$  is an antipalindrome if and only if  $x^n$  is an antipalindrome.*

**Proposition 8.** *For all nonempty antipalindromes  $u, v$ , the word  $uv$  is an antipalindrome iff both  $u$  and  $v$  are powers of an antipalindrome  $z$ .*

**Proposition 9.** *The word  $x$  is an antipalindrome if and only if there exists a word  $z$  such that  $x = \bar{z} \text{III} z^R$ .*

A binary word  $x$  is called an *antipalstar* if it is the concatenation of 1 or more antipalindromes. A nonempty antipalstar is called *prime* if it is not the product of two or more even-length antipalindromes. An antipalstar is always an antipalindrome, but the converse is false, because for any antipalindromes  $x, y$  the word  $xyx$  will be an antipalindrome. The following theorem is a counterpart of the decomposition property for *palstars* [9].

**Theorem 7.** *Every antipalstar can be factored uniquely as the concatenation of prime antipalstars.*

*Proof.* If some antipalstar has multiple factorizations into prime antipalstars, then some prime antipalindrome  $u$  has another prime antipalindrome  $v$  as a prefix. Let  $v$  be the shortest antipalindrome in such pairs. If  $|v| \leq |u|/2$ , then  $u = vzv$ , where  $z$  is either empty or an antipalindrome; this contradicts the primality of  $u$ . Let  $|v| > |u|/2$ ,  $u = x\bar{x}^R$ ,  $v = y\bar{y}^R$ . Then  $v = xy$ ,  $\bar{x}^R = yz$  for some nonempty  $y, z$ . Since  $v$  is an antipalindrome, we have  $v = \overline{xy}^R = \bar{y}^R \bar{x}^R = \bar{y}^R yz$ . Since  $\bar{y}^R y$  is an antipalindrome, this contradicts the minimality of  $v$ . Thus, no prime antipalindrome has another prime antipalindrome as a prefix.  $\square$

Next we look at the number of antipalindromic factors in a word.

**Theorem 8.** *A word  $w$  of length  $n \geq 1$  has at most  $n - 1$  distinct nonempty factors that are antipalindromes.*

*Proof.* For any nonempty antipalindromic factor of  $w$ , consider its leftmost occurrence in  $w$ . We show that no two such occurrences end in the same position in  $w$ . If they did, say  $x$  and  $y$ , with  $|x| < |y|$ , then  $x$  is a suffix of  $y$ . But then  $x^R = \bar{x}$  is a prefix of  $y^R = \bar{y}$ . So  $x$  is a prefix of  $y$ . Since  $|x| < |y|$ , we have found an occurrence of  $x$  to the left of its leftmost occurrence, a contradiction.

Since no nonempty antipalindromic factor of  $w$  ends at position 1, the number of possible end positions of such factors is at most  $n - 1$ , whence the result.  $\square$

One can introduce the notion of an *a-rich word* as a word with maximum possible number of antipalindromic factors. But the following theorem shows that these words are trivial, in contrast with the rich words.

**Theorem 9.** *For all  $n \geq 1$ , there are exactly two a-rich words of length  $n$ .*

*Proof.* Let us prove that any word having the factor 00 or 11 is not a-rich. Consider such a word  $w$  and the position in which its leftmost factor of the form  $aa$  ends. An antipalindrome ending with  $aa$  must begin with  $\bar{a}\bar{a}$ ; hence,  $w$  has no antipalindrome ending in the chosen position. From the proof of Theorem 8 we know that an a-rich word contains nonempty antipalindromes ending at every position, except position 1, so  $w$  is not a-rich.

Thus, only two words of each length  $n$  remain. These words are  $(10)^k$  and  $(01)^k$  if  $n = 2k$  is even, and  $(10)^k 1$  and  $(01)^k 0$  if  $n = 2k + 1$  is odd.

To see that these words have  $n - 1$  distinct antipalindromes, consider a word  $w = 1010 \dots$ ; the other word admits the same proof. Note that  $w[1..2k] = (10)^k$  and  $w[2..2k+1] = (01)^k$  for  $k \geq 1$  are the leftmost occurrences of antipalindromes. This gives an antipalindrome ending at every position except position 1, for the total of  $n - 1$ .  $\square$

We call a word  $x$  *creaky* if it is a conjugate of its reversed complement  $\bar{x}^R$ .

**Proposition 10.** *The word  $x$  is in  $\mathcal{A}^2$  if and only if it is creaky.*

*Proof.* Suppose  $x \in \mathcal{A}^2$ . Then  $x = uv$  where  $u, v$  are antipalindromes. So  $\bar{x}^R = \overline{v^R u^R} = vu$ , a conjugate of  $x$ .

For the other direction, suppose  $x$  is a conjugate of  $\bar{x}^R$ . Then  $x = uv$  and  $\bar{x}^R = vu$ . From  $x = uv$  we get  $x^R = v^R u^R$  and then  $\bar{x}^R = \overline{v^R u^R}$ . It follows that  $v = \overline{v^R}$  and  $u = \overline{u^R}$ .  $\square$

**Proposition 11.** *For all integers  $m, n \geq 1$ , the word  $x^m$  is creaky iff  $x^n$  is creaky.*

*Proof.* It suffices to prove the result for  $m = 1$ . Suppose  $x$  is creaky. Then there exist  $u, v$  such that  $x = uv$  and  $\bar{x}^R = vu$ . Then  $x^n = (uv)^n$ , which is clearly a conjugate of  $(vu)^n = (\bar{x}^R)^n = \overline{(x^n)^R}$ .

Suppose  $x^n$  is creaky. Then there exist  $u, v$  such that  $x^n = uv$  and  $\overline{(x^n)^R} = vu$ . Then there exist an integer  $m$  and words  $x', x''$  such that  $u = x^m x'$ ,  $v = x'' x^{n-m-1}$ , where  $x' x'' = x$ . So  $vu = (x'' x')^n$ , and it follows that  $\bar{x}^R = x'' x'$ , a conjugate of  $x$ .  $\square$

In analogy with palindromic pairs, we define factorization of a creaky word  $w$  as a pair of antipalindromes  $u, v$ , where  $v \neq \varepsilon$  and  $uv = w$ . The following analog of Theorem 3 holds.

**Theorem 10.** *A creaky word  $w$  has  $m$  factorizations if and only if  $w = z^m$  for a primitive word  $z$ .*

*Proof.* A useful observation, used several times in this proof, is that if  $v \in \mathcal{A}$ , then for any word  $u$  any two of the following conditions (i)  $u \in \mathcal{A}$ , (ii)  $u$  is a prefix of  $v$ , (iii)  $u$  is a suffix of  $v$ , imply the third one.

First we take a creaky word with two different factorizations,  $w = u_1v_1 = u_2v_2$ , and prove that it is not primitive. We can assume  $|u_1| < |u_2|$ . If  $u_1 = \varepsilon$ , then  $w \in \mathcal{A}$ . By Proposition 8,  $w$  is a nontrivial power of an antipalindrome. Now let  $u_1 \neq \varepsilon$ . We prove the following fact by induction on  $|w|$ : for some antipalindromes  $x$  and  $y$ , one has  $w = (xy)^m$ ,  $u_1 = (xy)^jx$ ,  $u_2 = (xy)^kx$ , where  $0 \leq j < k < m$ . The base case is trivial (the shortest creaky words have a unique factorization), so we proceed with the inductive step.

Since  $u_1 \in \mathcal{A}$ , it is both a prefix and a suffix of  $u_2$ . By Lemma 2 (ii) we have  $u_1 = (xy)^i x$ ,  $u_2 = (xy)^{i+1} x$  for some words  $x, y$  and some  $i \geq 0$ . Similarly we have  $v_2 = (x'y')^{i'} x'$ ,  $v_1 = (x'y')^{i'+1} x'$ . Clearly,  $x, y, x', y' \in \mathcal{A}$ . Furthermore, the suffix  $yx$  of  $u_2$  coincides with the prefix  $x'y'$  of  $v_1$ . So we have  $w = (xy)^{i+i'+1} xx'$ . If  $x' = y$ , then we are done, so assume that  $|x'| < |y|$  (the case  $|x'| > |y|$  is similar). Since  $yx = x'y'$  are different factorizations of a creaky word which is shorter than  $w$ , we apply the inductive hypothesis to get  $yx = (zt)^m$ ,  $x' = (zt)^j z$ ,  $y = (zt)^k z$ . Then we have  $u_1 = (tz)^{mi+m-k-1} t$ ,  $u_2 = (tz)^{m(i+1)+m-k-1} t$ ,  $w = (tz)^{m(i+i'+1)+m-k+j}$ . The inductive step is finished.

We proved that a primitive creaky word has a unique factorization. Now let us take  $w = z^m$  for a primitive  $z$  and let  $z = z'z''$  be the factorization of  $z$ . Then clearly all words of the form  $z^i z' = (z'z'')^i z'$  and  $z'' z^i = z''(z'z'')^i$  are antipalindromes. Thus,  $w$  has  $m$  factorizations of the form  $z^i z' \cdot z'' z^{m-i-1}$ . Conversely, suppose that  $w$  has a factorization  $z^i \hat{z} \cdot \tilde{z} z^{m-i-1}$ . Then  $\hat{z}, \tilde{z} \in \mathcal{A}$ , implying  $\hat{z} = z'$  and  $\tilde{z} = z''$ . Thus, the number of factorizations is exactly  $m$ .  $\square$

Using Theorem 10, it is easy to relate the growth function of  $\mathcal{A}^2$  (sequence A045655 in the OEIS [18]) to palindromic pairs.

**Theorem 11.**  $C_{\mathcal{A}^2}(n) = E(n, 2)$  for all  $n \geq 0$ .

*Proof.* Antipalindromes can be mapped to even-length palindromes with the bijection negating the right half of a word. We can extend this idea to creaky words and even palindromic pairs. For a creaky word  $w = z^m$  with  $z$  primitive, we know that  $z$  is creaky (Proposition 11) and has a unique factorization  $z = uv$  (Theorem 10). Negating the right halves of both  $u$  and  $v$ , we get an even palindromic pair  $\hat{z} = \hat{u}\hat{v}$ . The even palindromic pair  $\hat{w} = \hat{z}^m$  will be the image of  $w$ . Theorem 3 allows one to invert this mapping and thus obtain a bijection between creaky words and even palindromic pairs of a fixed length.  $\square$

## References

1. Bucci, M., de Luca, A., De Luca, A.: Rich and periodic-like words. In: Proc. 13th Int. Conf. on Developments in Language Theory. DLT 2009. LNCS, vol. 5583, pp. 145–155. Springer (2009)
2. Droubay, X., Justin, J., Pirillo, G.: Episturmian words and some constructions of de Luca and Rauzy. *Theoret. Comput. Sci.* 255, 539–553 (2001)
3. Fekete, M.: Über der verteilung der wurzeln bei gewissen algebraischen gleichungen mit ganzzahligen koeffizienten. *Math. Zeitschrift* 17, 228–249 (1923)
4. Fine, N.J., Wilf, H.S.: Uniqueness theorems for periodic functions. *Proc. Amer. Math. Soc.* 16, 109–114 (1965)
5. Galil, Z., Seiferas, J.: A linear-time on-line recognition algorithm for “Palstar”. *J. ACM* 25, 102–111 (1978)
6. Glen, A., Justin, J., Widmer, S., Zamboni, L.: Palindromic richness. *European J. Combinatorics* 30(2), 510–531 (2009)
7. Kari, L., Mahalingam, K.: Watson-Crick palindromes in DNA computing. *Natural Computing* 9, 297–316 (2010)
8. Kemp, R.: On the number of words in the language  $\{w \in \Sigma^* \mid w = w^R\}^2$ . *Discrete Math.* 40, 225–234 (1982)
9. Knuth, D.E., Morris, J., Pratt, V.: Fast pattern matching in strings. *SIAM J. Comput.* 6, 323–350 (1977)
10. Kosolobov, D., Rubinchik, M., Shur, A.M.:  $\text{Pal}^k$  is linear recognizable online. In: Proc. 41th Int. Conf. on Theory and Practice of Computer Science (SOFSEM 2015). LNCS, vol. 8939, pp. 289–301. Springer (2015)
11. Lyndon, R.C., Schützenberger, M.P.: The equation  $a^M = b^N c^P$  in a free group. *Michigan Math. J.* 9, 289–298 (1962)
12. Nester, M.R.: Mathematical investigations of some plant interaction designs. Ph.D. thesis, University of Queensland, Brisbane, Australia (1999)
13. Pelantova, E., Starosta, S.: Almost rich words as morphic images of rich words. *Internat. J. Found. Comp. Sci.* 23, 1067–1083 (2012)
14. Rampersad, N., Shallit, J., Wang, M.W.: Inverse star, borders, and palstars. *Inform. Process. Lett.* 111, 420–422 (2011)
15. Richmond, L.B., Shallit, J.: Counting the palstars. *Electronic J. Combinatorics* 21(3), # P3.25 (2014)
16. Shur, A.M.: On intermediate factorial languages. *Discrete Appl. Math.* 157, 1669–1675 (2009)
17. Shur, A.M.: Growth properties of power-free languages. *Computer Science Review* 6, 187–208 (2012)
18. Sloane, N.J.A.: The on-line encyclopedia of integer sequences. Available at <http://oeis.org>