Overpals, Underlaps, and Underpals

Aayush Rajasekaran[1], Narad Rampersad[2], and Jeffrey Shallit[1(✉)]

[1] School of Computer Science, University of Waterloo, Waterloo, ON N2L 3G1, Canada
{arajasekaran,shallit}@uwaterloo.ca

[2] Department of Math/Stats, University of Winnipeg, 515 Portage Ave., Winnipeg, MB R3B 2E9, Canada
n.rampersad@uwinnipeg.ca

Abstract. An overlap in a word is a factor of the form $axaxa$, where x is a (possibly empty) word and a is a single letter; these have been well-studied since Thue's landmark paper of 1906. In this note we consider three new variations on this well-known definition and some consequences.

Keywords: Overlap · Automata · Avoidability in words

1 Introduction

An *overlap* is a word of the form $axaxa$, where x is a (possibly empty) word and a is a single letter. Examples include `alfalfa` in English and `entente` in French. Since Thue's work [2,15,16] in the early 20th century, overlaps and their avoidance have been well-studied in the literature (see, e.g., [12]).

Let μ be the morphism defined by $\mu(0) = 01$ and $\mu(1) = 10$. The Thue-Morse word $\mathbf{t}$ is defined to be the infinite fixed point of μ starting with 0. We have $\mathbf{t} = 0110100110010110 \cdots$. We recall two famous results about binary overlaps:

Theorem 1.

(a) The Thue-Morse word $\mathbf{t}$ *is overlap-free [2,16].*
(b) The number of binary overlap-free words of length n *is* $\Omega(n^{\alpha})$ *and* $O(n^{\beta})$ *for real numbers* $1 < \alpha < \beta$ *[3,4,7].*

In this paper we consider three variants of overlaps and study their properties.

2 Definitions and Notation

Throughout, we use the variables a, b, c to denote single letters, and the variables u, v, w, x, y, z to denote words. By $|x|$ we mean the length of a word x, and by x^R we mean its reversal. The empty word is written ε.

If a word w can be written in the form $w = xyz$ for (possibly empty) words x, y, z, then we say that y is a *factor* of w. We say that a word $x = x[1..n]$ has

S. Brlek et al. (Eds.): WORDS 2017, LNCS 10432, pp. 17–29, 2017.
DOI: 10.1007/978-3-319-66396-8_3

period p if $x[i] = x[i+p]$ for $1 \le i \le n-p$. We say that a word x is a (p/q)-power, for integers $p > q \ge 1$, if x has period q and length p. For example, the word `ionization` is a $(10/7)$-power. A 2-power is called a *square*. Finally, we say that a word z *contains* an α-power if z contains a factor x that is a (p/q)-power for some $p/q \ge \alpha$. Otherwise we say that z *avoids* α-powers or is α-power free. We say that a word z avoids $(\alpha+\epsilon)$-powers or is $(\alpha+\epsilon)$-power-free if, for all $p/q > \alpha$, z contains no factor that is a (p/q)-power. By x^ω we mean the infinite word $xxx\cdots$.

We recall the definition of three famous sequences. The *Rudin-Shapiro sequence* $\mathbf{r} = (r_n)_{n\ge0} = 0001001000011101000 10\cdots$ is defined by the relations $r_0 = 0$, and $r_{2n} = r_n$, $r_{4n+1} = r_n$, $r_{8n+3} = 1 - r_n$, and $r_{8n+7} = r_{2n+1}$ for $n \ge 0$. The *Fibonacci sequence* $\mathbf{f} = (f_n)_{n\ge0} = 0100101001001010 01010\cdots$ is the fixed point of the morphism $\varphi(0) = 01$, $\varphi(1) = 0$. The *Tribonacci sequence* $\mathbf{T} = (T_n)_{n\ge0} = 01020100102010102010\cdots$ is the fixed point of the morphism $\theta(0) = 01$, $\theta(1) = 02$, $\theta(2) = 0$.

3 Overpals

In our first variation, we replace the second occurrence of axa in an overlap with its reversal. Thus, an *overpal* is a word of the form $axax^Ra$, where x^R is the reverse of the (possibly empty) word x and a is a single letter. The English word `tartrate` contains an occurrence of an overpal corresponding to $a = $ `t` and $x = $ `ar`. The *order* of an overpal $axax^Ra$ is defined to be $|ax|$.

We start with some results about binary words.

3.1 The Binary Case

Lemma 2. *Every binary palindrome x of odd length $\ell \ge 7$ contains an occurrence of either aaa, ababa, or abbabba, for some distinct letters a, b.*

Proof. Let w be an odd-length palindrome of length ≥ 7. Then we can write w in the form $xabcdcbax^R$ for some letters a, b, c, d and x possibly empty. Then a check of all 16 possibilities for a, b, c, d gives the result.

Theorem 3. *A binary word contains an overpal if and only if it contains aaa, ababa, or abbabba for letters $a \ne b$.*

Proof. Suppose w contains an overpal $t = axax^Ra$. If $|x| = 0$, then $t = aaa$. If $|x| = 1$, then t is either $aaaaa$ or $ababa$. Otherwise $|x| \ge 2$, so $|t| \ge 7$, and the result follows by Lemma 2. On the other hand, if w contains any of aaa, $ababa$, or $abbabba$, then w contains an overpal.

Theorem 3 allows us to compute the generating function for the number of binary words avoiding overpals.

Corollary 4. *The generating function for the number of binary words avoiding overpals is*

$$\frac{2x^9 + 6x^8 + 8x^7 + 6x^6 + 6x^5 + 5x^4 + 4x^3 + 3x^2 + 2x + 1}{1 - x^2 - x^4}$$

Corollary 4 was apparently first noticed by Colin Barker, in a remark posted at the *On-Line Encyclopedia of Integer Sequences* about sequence A277277.

Proof. We use the `DAVID_IAN` Maple package [10,11], implementing the Goulden-Jackson cluster method [6], with the command

```
GJs(0,1,[0,0,0],[1,1,1],[0,1,0,1,0],[1,0,1,0,1],
[0,1,1,0,1,1,0],[1,0,0,1,0,0,1],x);
```

This gives us the above generating function counting the binary words avoiding the patterns *aaa*, *ababa*, and *abbabba*.

Corollary 5. *The number* $\mathrm{ovp}_2(n)$ *of binary words of length* n *avoiding overpals is, for* $n \geq 6$*, equal to* $a\alpha^n + b\beta^n + c\gamma^n + d\delta^n$ *where*

$$a \doteq 5.0968257035281799892230 10 \qquad b \doteq 0.0087471054719041322133 20$$

are the real zeroes of the polynomial $25Z^4 - 300Z^3 + 1240Z^2 - 1840Z + 16$*, and*

$$c = 3 + \frac{\sqrt{5}}{5} + (2\sqrt{5} - 2)^{1/2} i \qquad d = 3 + \frac{\sqrt{5}}{5} - (2\sqrt{5} - 2)^{1/2} i,$$

and $\alpha = ((1 + \sqrt{5})/2)^{1/2} \doteq 1.2720196495140689642524 2246$*,* $\beta = -\alpha$*,* $\gamma = i\alpha^{-1}$*,* $\delta = -i\alpha^{-1}$*.*

Proof. From the generating function, we know that $\mathrm{ovp}_2(n)$ satisfies the linear recurrence $\mathrm{ovp}_2(n) = \mathrm{ovp}_2(n-2) + \mathrm{ovp}_2(n-4)$ for $n \geq 10$. Now we use standard techniques to solve this linear recurrence.

Corollary 6. *There are* $\Theta(\alpha^n)$ *binary words of length* n *containing no overpals.*

We now turn to infinite words avoiding overpals. It is easy to construct a *periodic* binary word avoiding overpals: namely, $(0011)^\omega = 001100110011\cdots$. (To verify this, it suffices to enumerate its subwords of odd length ≤ 7 and check that none of them are of the form *aaa*, *ababa* or *abbabba*.)

Theorem 7. *The lexicographically least infinite binary word that avoids overpals is* $\mathbf{x} := 001(001011)^\omega$*.*

Proof. To verify that $\mathbf{x}$ avoids overpals, it suffices to enumerate its subwords of odd length ≤ 7. None are of the form *aaa*, *ababa* or *abbabba*.

Suppose there is an infinite binary word $\mathbf{w}$ that is lexicographically less than $\mathbf{x}$, but contains no overpals. Let v be the shortest prefix of $\mathbf{w}$ such that v is not a prefix of $\mathbf{x}$. Suppose $|v| = n$. At position n there must be a 0 in v and $\mathbf{w}$ and a 1 in $\mathbf{x}$. This means there are four possibilities: (i) $v = 000$; (ii) $v = 001(001011)^i000$ for some $i \geq 0$; (iii) $v = 001(001011)^i00100$ for some $i \geq 0$; (iv) $v = 001(001011)^i001010$ for some $i \geq 0$.

In cases (i) and (ii), v ends with the overpal 000, a contradiction. In case (iii), consider the letter at position $n+1$ of $\mathbf{w}$. If it is 0, then $v0$ is a prefix of $\mathbf{w}$ and ends with 000. If it is 1, then $v1$ is a prefix of $\mathbf{w}$ and ends with 1001001. Both cases give a contradiction. In case (iv), v ends with the overpal 01010, a contradiction.

We have shown there are ultimately periodic binary words avoiding overpals. We now turn to aperiodic binary words.

Theorem 8. *No* $(7/3)$*-power-free binary word contains an overpal.*

Proof. Suppose it did. From Lemma 2 any odd-length palindrome in such a word is of length $1, 3$, or 5. A palindrome of length 1 cannot be an overpal. The only overpals of length 3 are 000 and 111, each of which is a cube. Finally, the only overpals of length 5 are 00000 and 01010 and their complements, each of which contains a $(7/3)$-power.

Corollary 9. *The Thue-Morse word* $\mathbf{t}$ *contains no overpals.*

Theorem 10. *If* $\mu(x)$ *contains an overpal, then so does* x.

Proof. Suppose $\mu(x)$ contains an overpal. Then it contains an occurrence of aaa, $ababa$, or $abbabba$. However, it is easy to verify that neither aaa nor $abbabba$ can be the factor of a word that is an image under μ. For $ababa$ to be the factor of $\mu(x)$, it must be that x has the factor aaa, and hence an overpal.

Theorem 11. *The orders of overpals occurring in the Fibonacci word* $\mathbf{f}$ *are given, for* $n \geq 1$*, by the* n *whose Fibonacci representation is accepted by the following automaton.*

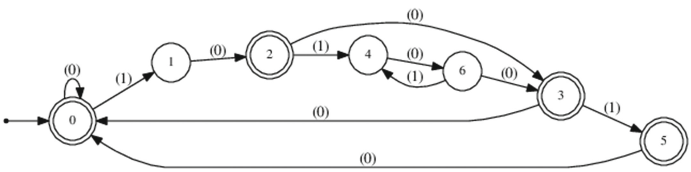

Fig. 1. Automaton accepting orders of overpals in the Fibonacci word

There are infinitely many orders for which there is no overpal factor of $\mathbf{f}$ *and infinitely many for which there are.*

Proof. We use the automatic theorem-proving software `Walnut` [9] with the predicate

```
def fiboverpal "?msd_fib (n=0) | Ei (n>=1) & (At (t<=2*n) =>
F[i+t] = F[i+2*n-t]) & F[i]=F[i+n]":
```

Corollary 12. *An overpal of order n exists in the Fibonacci word, for $n \geq 1$, if and only if there exists an integer m such that $n = \lfloor m\alpha + \frac{1}{2} \rfloor$, where $\alpha = (1+\sqrt{5})/2$.*

Proof. The proof is in six steps.

Step 1: Define the infinite binary word $\mathbf{p} = (p_i)_{i \geq 0}$, where $p_i = 1$ if the Fibonacci representation of i is accepted by the automaton in Fig. 1, and $p_i = 0$ otherwise. Using the usual extension of Cobham's theorem to Fibonacci numeration systems, p is given by the image under the coding τ of the fixed point $f^\omega(0)$, where

$$f(0) = 01 \qquad f(1) = 2 \qquad f(2) = 34$$
$$f(3) = 05 \qquad f(4) = 6 \qquad f(5) = 0 \qquad f(6) = 34$$

and $\tau(0123456) = 1011010$. This is obtained just by reading off the transitions of the automaton, where the image of a state is 1 if the state is accepting, and 0 otherwise.

Step 2: Let $h : \{0,1\}^* \to \{0,1\}^*$ be the morphism sending $1 \to 10110$, $0 \to 110$. A routine induction on n, which we omit, proves that

$$\tau(f^{3n}(0)) = \tau(f^{3n}(2)) = \tau(f^{3n}(3)) = \tau(f^{3n}(3)) = \tau(f^{3n}(4)) = \tau(f^{3n}(6)) = h^n(1)$$
$$\tau(f^{3n}(1)) = h^n(0)$$
$$\tau(f^{3n+3}(5)) = h^n(101)$$

for $n \geq 0$.

Step 3: We now use a result in a paper of Tan and Wen [14]. Define $\pi : \{0,1\}^* \to \{0,1\}^*$ to be the morphism sending $0 \to 1$, $1 \to 0$. Define $\lambda : \{0,1\}^* \to \{0,1\}^*$ to be the morphism corresponding to $\pi \circ h^2 \circ \pi$.

A cutting sequence $K_{q,\,r}$ is defined as the infinite binary sequence generated by the straight line $y = qx + r$ as it cuts a square grid. See [14] for more on cutting sequences. Let the fixed point of λ be generated by the cutting sequence $K_{\gamma,\beta}$. Tan and Wen give us the slope γ, and the intercept β of this line. We define the additional morphisms $\sigma, \rho : \{0,1\}^* \to \{0,1\}^*$, where σ sends $0 \to 01$, $1 \to 0$, and ρ sends $0 \to 10$, $1 \to 0$.

To get γ, we need to express λ as a composition of $\sigma \circ \pi$, $\rho \circ \pi$ and π. We hence write $\lambda = ((\sigma \circ \pi) \circ \pi \circ (\rho \circ \pi) \circ \pi \circ (\rho \circ \pi) \circ \pi)^2$. Tan and Wen gives the continued fraction expansion of the slope as $\gamma = [0; 1, \overline{1, 1, 1, 1, 1, 1}] = 1/\alpha$, where $\alpha = \frac{1+\sqrt{5}}{2}$ is the golden ratio.

To get β, we follow Tan and Wen to get the word $u = 010010$ that satifies $\lambda(01) = u01v, \lambda(10) = u10v, vu$ is a palindrome, for some word v. We also define $u_n = \lambda^{n-1}(u)\lambda^{n-2}(u) \cdots \lambda(u)u$. Let $|u_n|_0$ denote the number of zeroes in u_n.

The value of β is given by the unique number $x \in [-\gamma, 1+\gamma)$ that satisfies $e^{2\pi ix} = \lim_{n\to\infty} e^{-2\pi i(|u_n|_0+1)\gamma}$. We calculate this value as $\beta = 1 - \sqrt{5}/2$.

Finally, Tan and Wen assert that if the fixed point of λ is given by $K_{\gamma,\beta}$, then the fixed point of h is given by the cutting sequence $K_{1/\gamma,-\beta/\gamma}$. Thus, $h^\omega(1)$ is given by $K_{\alpha,1-\alpha/2}$.

Step 4: The Sturmian word $\mathbf{s}_{e,\,f} = (s_i)$ is the infinite binary word defined by $s_i = \lfloor e(i+1)+f \rfloor - \lfloor ei+f \rfloor - \lfloor e \rfloor$. We now use a classical result relating cutting sequences to Sturmian words (e.g., p. 56 of [8]) to conclude that $K_{\alpha,1-\alpha/2} = \mathbf{s}_{1/\alpha,(5-3\alpha)/2}$.

Step 5: We shift this Sturmian word right by 1 position, getting the equality $\mathbf{s}_{1/\alpha,(5-3\alpha)/2} = 1 \cdot \mathbf{s}_{1/\alpha,-1/(2\alpha)}$.

Step 6: Finally, we use the usual connection between Sturmian words and Beatty sequences (e.g., Lemma 9.1.3 of [1], generalized from characteristic words to the more general setting of Sturmian words) to conclude that $\mathbf{s}_{1/\alpha,-1/(2\alpha)} = b_1 b_2 \cdots$, where $b_n = 1$ if and only if there exists an integer $m \geq 1$ such that $n = \lfloor m\alpha + \frac{1}{2} \rfloor$.

Theorem 13. *There are exactly four overpals in the Rudin-Shapiro sequence, and they are given by* 000, 111, 0100010, 1011101.

Proof. We use `Walnut` [9] to find the orders of overpals in the Rudin-Shapiro sequence

```
eval RSOverpal "Ei (n>=1) & (At (t<=2*n) =>
(RS[i+t] = RS[i+2*n-t])) & (RS[i]=RS[i+n])":
```

The only accepted orders are 1 and 2. An exhaustive search yields the result.

3.2 Larger Alphabets

Understanding the words that avoid overpals over large alphabets is more challenging than the binary case. For one thing, there is no analogue of Lemma 2, as the following result shows:

Theorem 14. *Over a ternary alphabet, there are arbitrarily long odd-length palindromes containing no overpals.*

Proof. We know that $\mu^{2n}(0)$ is a palindrome for all $n \geq 0$, and furthermore, since it is a prefix of $\mathbf{t}$, it contains no overpals. Therefore, for all $n \geq 0$, the word $\mu^{2n}(0)2\mu^{2n}(0)$ is a palindrome containing no overpals, and it is of length $2^{2n+1}+1$.

Theorem 15.

(i) Every odd-length ternary palindrome of length ≥ 17 *contains a* $\frac{7}{4}$ *power.*
(ii) There are arbitrarily large odd-length ternary palindromes avoiding $(\frac{7}{4}+\epsilon)$*-powers.*

Proof.

(i) It suffices to examine all ternary palindromes of length 17.
(ii) Dejean's word [5] avoids $(\frac{7}{4}+\epsilon)$-powers and contains ternary palindromes of all odd lengths.

Theorem 16. *No infinite ternary word can avoid overpals and* $\frac{41}{22}$*-powers.*

Proof. We use the usual tree-traversal technique. The tree has 120844 internal nodes, and 241689 leaves. The longest such string is of length 228.

Conjecture 17. There is an infinite ternary word that avoids overpals and $(\frac{41}{22}+\epsilon)$-powers.

4 Underpals

A word is said to be an *underpal* if it is of the form $axbx^Ra$ where x is a (possibly empty) word and a, b are letters with $a \neq b$. An example in English is the word `racecar`, with $a = \texttt{r}$, $x = \texttt{ac}$, and $b = \texttt{e}$.

Theorem 18. *A word contains an underpal if and only if it contains some word of the form* ab^ia *with* $a \neq b$ *and* i *odd.*

Proof. Let a word contain an underpal $z = axbx^Ra$. Either x ends in b, or it does not end in b.

Case 1: x ends in b. Then either $x = b^l$ for some $l \geq 1$, or $x = ycb^l$ for some word y and letter $c \neq b$.

Case 1a: $x = b^l$. Then $z = ab^ia$ for odd $i = 2l+1$.

Case 1b: $x = ycb^l$. Then $z = aycb^lbb^lcy^Ra$, which contains cb^ic with odd $i = 2l+1$.

Case 2: x does not end in b. If $x = \varepsilon$, then $z = ab^ia$, with odd $i = 1$. Otherwise $x = yc$, which gives $z = aycbcy^Ra$, which contains cb^ic with odd $i = 1$.

Thus, a word that contains an underpal must contain an ab^ia, with odd i, and so a word that avoids such factors must avoid underpals.

For the converse, suppose w contains $z = ab^ia$ with $a \neq b$ and i odd. Then $z = ab^lbb^la$ for some non-negative integer l. Since z is an underpal with $x = b^l$, the word w contains an underpal.

Theorem 19. *The number of length-n words avoiding underpals over a k-letter alphabet satisfies the recurrence* $f_k(0) = 1, f_k(1) = k, f_k(n) = (k-2)f_k(n-1) + kf_k(n-2) + k$.

Proof. Let Σ be an alphabet with $|\Sigma| = k$. For all p, define $L_{k,p} \subseteq \Sigma^*$ to be the language of words of length p that avoid underpals.

We now define two languages A_1 and A_2 as follows:

$$A_1 = \{ca^{n-1} : a \neq c \in \Sigma\}$$
$$A_2 = \{cx : x = a^l bz \in L_{k,\,n-1}, a \neq b, a \neq c, b \neq c, a,\ b,\ c \in \Sigma, l > 0, z \in \Sigma^*\}$$

Note that $|A_1| = k(k-1)$. Since we exclude unary words of the form a^{n-1}, the number of words of the form $x = a^l bz$ is $|L_{k,n-1}| - k = f_k(n-1) - k$. We thus get that $|A_2| = (k-2)(f_k(n-1) - k)$.

Define $A = A_1 \cup A_2$. All words in A are n-length words avoiding underpals since they must avoid $ab^i a$ with i odd, and so $A \subseteq L_{k,n}$.

Define $D \subseteq L_{k,n-2}$ as follows:

$$D = \{a^l bz \in L_{k,n-2}, a \neq b \in \Sigma, l > 0, l \text{odd}, z \in \Sigma^*\}.$$

Next, we define the languages B_1, B_2 and B_3 as follows:

$$B_1 = \{ccx : x = a^l bz \in D, c \in \Sigma, c \neq b\}$$
$$B_2 = \{bax : x = a^l bz \in D\}$$
$$B_3 = \{ccx : x \in L_{k,\,n-2}, x \notin D, c \in \Sigma\}$$

Clearly $|B_1| = (k-1)\,|D|$, and $|B_2| = |D|$, and $|B_3| = k(|L_{k,\,n-2}| - |D|)$.

Define $B = B_1 \cup B_2 \cup B_3$. All words in B are n-length words avoiding underpals since they must avoid $ab^i a$ with i odd, and so $B \subseteq L_{k,n}$.

Thus, we get

$$A \cup B \subseteq L_{k,n}. \tag{1}$$

Consider any word $z = d_1 d_2 \cdots d_n \in L_{k,\,n}$. Note that $d_2 d_3 \cdots d_n \in L_{k,n-1}$ and $d_3 d_4 \cdots d_n \in L_{k,n-2}$. We divide these words z into two cases:

Case 1. $d_1 = d_2$. If $z = d_1 d_1 a^l bx$, for some $a \neq b \in \Sigma$ and even $l \geq 0$, then $z \in B_3$. If $z = d_1 d_1 a^l bx$, for some $a \neq b \in \Sigma$ and odd l, then we consider d_1. If $d_1 \neq b$, then $z \in B_1$. If $d_1 = b$, then z contains $ba^l b$, with odd l, and thus $z \notin L_{k,\,n}$. If $z = d_1^n$, then $z \in B_3$.

Case 2: $d_1 \neq d_2$. If $z = d_1 d_2^{n-1}$, then $z \in A_1$. If $z = d_1 d_2^l bx$, for some $b \neq d_2 \in \Sigma$ and even $l > 0$, then we consider d_1. If $d_1 \neq b$, then $z \in A_2$. If $d_1 = b$, then $z = d_1 d_2 d_2^{l-1} d_1 x$, where $l-1$ is odd. In this case, $z \in B_2$. If $z = d_1 d_2^l bx$, for some $b \neq d_2 \in \Sigma$ and odd l, then we consider the value of d_1. If $d_1 \neq b$, then $z \in A_2$. If $d_1 = b$, then z contains $bd_2^l b$, with odd l, and thus $z \notin L_{k,\,n}$. We thus see that for all $z \in L_{k,\,n}, z \in A \cup B$, and hence

$$L_{k,\,n} \subseteq A \cup B. \tag{2}$$

Combining Eqs. (1) and (2) gives us $L_{k,\,n} = A \cup B$, which gives

$$f_k(n) = |L_{k,\,n}| = |A \cup B|\,. \tag{3}$$

Since the words in A_1 have exactly two different letters, while those in A_2 have at least 3 different letters, the sets A_1 and A_2 are disjoint.

The sets B_1 and B_2 are disjoint since they disagree on the first two letters. The sets B_1 and B_3 are disjoint since they disagree on the last $n-2$ letters. The sets B_2 and B_3 are disjoint since they disagree on the last $n-2$ letters.

Note that for all words $x = a_1 a_2 \cdots a_n \in A$ we have $a_1 \neq a_2$. The only words $y = b_1 b_2 \cdots b_n \in B$ for which $b_1 \neq b_2$ are in B_2, and are thus of the form $y = baa^l bz$, for some $a \neq b \in \Sigma$ and $z \in \Sigma^*$. Such words y cannot be in A, because A excludes all words with prefix $ba^p b$ for all $a \neq b \in \Sigma$, $p > 0$. This shows that the sets A and B are disjoint.

We have

$$|A| = |A_1| + |A_2| = k(k-1) + (k-2)(f_k(n-1) - k) = (k-2)f_k(n-1) + k. \quad (4)$$

We also have

$$|B| = |B_1| + |B_2| + |B_3| = (k-1)\,|D| + |D| + k(|L_{k,\,n-2}| - |D|) = kf_k(n-2). \quad (5)$$

Since A and B are disjoint, $|A \cup B| = |A| + |B| = (k-2)f_k(n-1) + kf_k(n-2) + k$. Combining this with Eq. (3) gives us $f_k(n) = (k-2)f_k(n-1) + kf_k(n-2) + k$. Finally, $f_k(0) = 1$, since the empty string avoids underpals, and $f_k(1) = k$, since all strings of length 1 avoid underpals.

Corollary 20. *The number $f_k(n)$ of length-n words avoiding underpals over a k-letter alphabet, for $k \geq 2$ and $n \geq 0$, is given by $f_k(n) = a\alpha^n + b\beta^n + c$, where*

$$\alpha = \frac{k-2+\sqrt{k^2+4}}{2} \qquad a = \frac{(k-1)(3(k^2+4)+(k+6)\sqrt{k^2+4})}{2(2k-3)(k^2+4)}$$

$$\beta = \frac{k-2-\sqrt{k^2+4}}{2} \qquad b = \frac{(k-1)(3(k^2+4)-(k+6)\sqrt{k^2+4})}{2(2k-3)(k^2+4)}$$

$$c = \frac{k}{3-2k}.$$

Proof. By the usual techniques for handling linear recurrences, we know that $f_k(n) = (k-1)f_k(n-1) + 2f_k(n-2) - kf_k(n-3)$. This means that $f_k(n)$ is expressible as a linear combination of the n'th powers of the zeroes of the polynomial $X^3 + (1-k)X^2 - 2X + k$. Solving the resulting linear system, using Maple as an assistant, gives the result.

Remark 21. For $k = 2$ this simplifies to $f_2(n) = 2^{(n+3)/2} - 2$ for n odd and $f_2(n) = 3 \cdot 2^{n/2} - 2$ for n even.

The *run length encoding* of a binary word is the integer sequence giving the lengths of maximal blocks of 0s and 1s. For example, the run length encoding of 0011101011 is 2, 3, 1, 1, 1, 2.

Theorem 22. *A finite binary word avoids underpals if and only if its run length encoding is of the form $i_1, i_2, \ldots, i_t$ where $i_2, i_3, \ldots, i_{t-1}$ are all even. An infinite binary word that does not end in a^ω for $a \in \{0,1\}$ avoids underpals if and only if its run length encoding is of the form $i_1, i_2, i_3, \ldots$ where $i_2, i_3, \ldots$ are all even.*

Theorem 23. *The Fibonacci word has underpals of order n for exactly those n accepted by the automaton below (Fig. 2).*

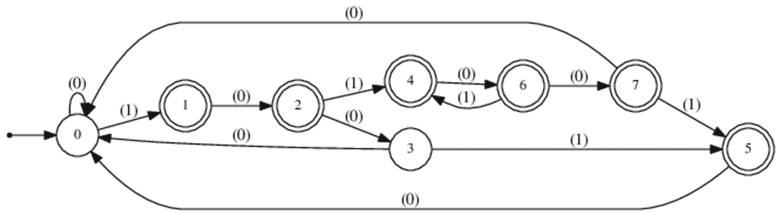

Fig. 2. Automaton accepting orders of underpals in the Fibonacci word

Theorem 24. *The Fibonacci word has both underpals and overpals of order n for exactly those n accepted by the automaton below (Fig. 3).*

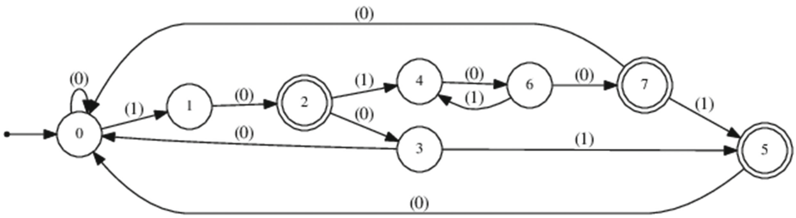

Fig. 3. Automaton accepting orders for which there are both overpals and underpals in the Fibonacci word

Theorem 25. *Every binary word of length* ≥ 17 *avoiding underpals contains a* 4*th power.*

Proof. By explicit enumeration of the 1022 binary words of length 17 avoiding underpals.

Theorem 26. *There is an infinite binary word avoiding underpals and avoiding* $(4+\epsilon)$*-powers.*

Proof. Let h be the doubling morphism $0 \to 00$ and $1 \to 11$. Applying h to the Thue-Morse word $\mathbf{t}$ gives a binary word $h(\mathbf{t})$ that contains no underpals and avoids $(4+\epsilon)$-powers.

Theorem 27. *Every ternary word of length* ≥ 6 *avoiding underpals has a square.*

Proof. By enumerating all ternary words of length 6 avoiding underpals.

Theorem 28. *There is an infinite ternary word avoiding underpals and* $(2+\epsilon)$*-powers.*

Proof. Take any infinite squarefree word w over a ternary alphabet $\{0,1,2\}$ and apply the morphism $h : 0 \to 01, 1 \to 10, 2 \to 22$. Then $h(w)$ has no overlaps, overpals, or underpals.

5 Underlaps

In analogy with overlaps, we can define *underlaps*. An *underlap* is a word of the form $axbxa$ with x a (possibly empty) word, and a, b letters with $a \neq b$. Note that x is a bispecial factor of the underlap. An example in English is the word `ginning`, with $a = \texttt{g}$, $x = \texttt{in}$, and $b = \texttt{n}$.

Theorem 29.

(a) The only underlaps in the Thue-Morse sequence $\mathbf{t}$ *are* $\{010, 101, 0011010, 0101100, 1010011, 1100101\}$.
(b) The only underlaps in the Fibonacci sequence are $\{010, 101, 00100\}$.
(c) The only underlaps in the Rudin-Shapiro sequence are $\{010, 101, 00100, 01110, 10001, 11011, 0001000, 1110111\}$.
(d) The only underlaps in the Tribonacci sequence are $\{010, 020, 101, 10201, 20102, 001020100\}$.

We now prove a theorem giving the relationship between underlaps and underpals. These concepts actually coincide for binary words.

Theorem 30.

(a) If z contains an underpal, then it contains an underlap.
(b) If z is over a binary alphabet and contains an underlap, then it contains an underpal.

Proof. (a) Suppose z contains an underpal. Then it can be written in the form $z = uaxbx^Rav$ where $a \neq b$.

Case 1: If x contains some letter $c \neq b$, write $x = ycb^i$ for some $i \geq 0$. Then z contains the word $xbx^R = ycb^ibb^icy^R$, which contains the word cb^ibb^ic, which is an underlap.

Case 2: Otherwise $x = b^i$ for some $i \geq 0$. Then z contains the word $axbx^Ra = ab^ibb^ia$, which is an underlap.

(b) Now suppose z contains an underlap and is over the alphabet $\{0,1\}$. Then it can be written in the form $z = uaxbxav$ where $a \neq b$.

Case 1: x has no a's. Since x is over a binary alphabet, it must be the case that $x = b^i$ for some $i \geq 0$. Then $axbxa = ab^ibb^ia$, which is an underpal.

Case 2: x has one a. Write $x = b^i ab^j$ for some $i, j \geq 0$. Then $axbxa = ab^i ab^j bb^i ab^j a = ab^i ab^{i+j+1} ab^j a$. If either i (resp., j) is odd, this contains $ab^i a$ (resp., $ab^j a$), which is an underpal. Otherwise i and j are both even, so $i+j+1$ is odd and $ab^{i+j+1}a$ is an underpal.

Case 3: x has two or more a's. By identifying the first and last occurrences of a, write $x = b^i ayab^j$. Then $axbxa = ab^i ayab^j bb^i ayab^j a = ab^i ayab^{i+j+1} ayab^j a$. If i (resp., j) is odd, this contains $ab^i a$ (resp., $ab^j a$), which is an underpal. Otherwise i and j are both even, so $i + j + 1$ is odd and $ab^{i+j+1}a$ is an underpal.

As an example of a word over the ternary alphabet that contains an underlap but no underpal, consider 001120110.

Theorem 31. *Every binary word of length ≥ 9 has either an overlap or an underlap.*

Proof. It suffices to examine all 512 binary words of length 9.

Theorem 32. *There are exponentially many ternary words avoiding overlaps, underlaps, overpals, and underpals.*

Proof. Take any squarefree word w over a ternary alphabet $\{0, 1, 2\}$ and apply the morphism $h : 0 \to 01, 1 \to 10, 2 \to 22$. Then $h(w)$ has no overlaps, underlaps, overpals, or underpals. Since there are exponentially many squarefree ternary words (the best lower bound known is $\Omega(952^{n/53})$ [13]), the result follows.

References

1. Allouche, J.P., Shallit, J.: Automatic Sequences: Theory, Applications, Generalizations. Cambridge University Press, Cambridge (2003)
2. Berstel, J.: Axel Thue's papers on repetitions in words: a translation, No. 20. In: Publications du Laboratoire de Combinatoire et d'Informatique Mathématique, Université du Québec à Montréal, February 1995
3. Blondel, V.D., Cassaigne, J., Jungers, R.M.: On the number of α-power-free binary words for $2<\alpha \leq 7/3$. Theoret. Comput. Sci. **410**, 2823–2833 (2009)
4. Cassaigne, J.: Counting overlap-free binary words. In: Enjalbert, P., Finkel, A., Wagner, K.W. (eds.) STACS 1993. LNCS, vol. 665, pp. 216–225. Springer, Heidelberg (1993). doi:10.1007/3-540-56503-5_24
5. Dejean, F.: Sur un théorème de Thue. J. Combin. Theor. Ser. A **13**, 90–99 (1972)
6. Goulden, I., Jackson, D.: An inversion theorem for cluster decompositions of sequences with distinguished subsequences. J. London Math. Soc. **20**, 567–576 (1979)
7. Jungers, R.M., Protasov, V.Y., Blondel, V.D.: Overlap-free words and spectra of matrices. Theor. Comput. Sci. **410**, 3670–3684 (2009)
8. Lothaire, M.: Algebraic Combinatorics on Words, Encyclopedia of Mathematics and its Applications. Cambridge University Press, Cambridge (2011)
9. Mousavi, H.: Automatic theorem proving in Walnut (2016). https://arxiv.org/abs/1603.06017
10. Noonan, J., Zeilberger, D.: DAVID_IAN Maple package (1999). http://www.math.rutgers.edu/~zeilberg/gj.html

11. Noonan, J., Zeilberger, D.: The Goulden-Jackson cluster method: extensions, applications, and implementations. J. Differ. Equ. Appl. **5**, 355–377 (1999)
12. Rampersad, N.: Overlap-free words and generalizations. Ph.D. thesis, University of Waterloo (2008)
13. Sollami, M., Douglas, C.C., Liebmann, M.: An improved lower bound on the number of ternary squarefree words. J. Integer Sequences **19**, 3 (2016). Article 16.6.7
14. Tan, B., Wen, Z.Y.: Invertible substitutions and Sturmian sequences. Eur. J. Comb. **24**, 983–1002 (2003)
15. Thue, A.: Über unendliche Zeichenreihen. Norske vid. Selsk. Skr. Mat. Nat. Kl. **7**, 1–22 (1906). Reprinted in Selected Mathematical Papers of Axel Thue, T. Nagell, editor, Universitetsforlaget, Oslo, 1977, pp. 139–158
16. Thue, A.: Über die gegenseitige Lage gleicher Teile gewisser Zeichenreihen. Norske vid. Selsk. Skr. Mat. Nat. Kl. **1**, 1–67 (1912). Reprinted in Selected Mathematical Papers of Axel Thue, T. Nagell, editor, Universitetsforlaget, Oslo, 1977, pp. 413–478