A ternary square-free sequence avoiding factors equivalent to *abcacba*

James Currie*

Department of Mathematics & Statistics
University of Winnipeg
Winnipeg, MB
Canada R3B 2E9

j.currie@uwinnipeg.ca

Submitted: Mar 10, 2016; Accepted: May 7, 2016; Published: May 27, 2016
Mathematics Subject Classification: 68R15

Abstract

We solve a problem of Petrova, finalizing the classification of letter patterns avoidable by ternary square-free words; we show that there is a ternary square-free word avoiding letter pattern $xyzxzyx$. In fact, we

- characterize all the (two-way) infinite ternary square-free words avoiding letter pattern $xyzxzyx$
- characterize the lexicographically least (one-way) infinite ternary square-free word avoiding letter pattern $xyzxzyx$
- show that the number of ternary square-free words of length n avoiding letter pattern $xyzxzyx$ grows exponentially with n.

1 Introduction

A theme in combinatorics on words is **pattern avoidance**. A word w **encounters** word p if $f(p)$ is a factor of w for some non-erasing morphism f. Otherwise w **avoids** p. A standard question is whether there are infinitely many words over a given finite alphabet Σ, none of which encounters a given pattern p. Equivalently, one asks whether an ω-word over Σ avoids p.

The first problems of this sort were studied by Thue [11, 12] who showed that there are infinitely many words over $\{a, b, c\}$ which are **square-free** – i.e., do not encounter xx. He also showed that over $\{a, b\}$ there are infinitely many **overlap-free** words – which simultaneously avoid xxx and $xyxyx$. Thue also introduced a variation on pattern

*The author was supported by an NSERC Discovery Grant

avoidance by asking whether one could simultaneously avoid squares xx and factors from a finite set. For example, Thue showed that infinitely many words over $\{a, b, c\}$ avoid squares, and also have no factors aba or cbc.

In combinatorics, once an existence problem has been solved, it is natural to consider stronger questions: characterizations, enumeration problems and extremal problems. Since Thue, progressively stronger questions about pattern-avoiding sequences have been asked and answered:

- Gottschalk and Hedlund [3] characterized the doubly infinite binary words avoiding overlaps.
- How many square-free words of length n are there over $\{a, b, c\}$? The number of such words was shown to grow exponentially by Brandenburg [2].
- Let $\mathbf{w}$ be the lexicographically least square-free ω-word over $\{a, b, c\}$. As the author [1] has pointed out, the method of Shelton [8] allows one to test whether a given finite word over $\{a, b, c\}$ is a prefix of $\mathbf{w}$.

Interest in words avoiding patterns continues, and a recent paper by Petrova [7] studied **letter pattern avoidance** by ternary square-free words. A word w over $\{1, 2, 3\}$ avoids the **letter pattern** $P \in \{x, y, z\}^*$ if no factor of w is an image of P under a bijection from $\{x, y, z\}$ to $\{1, 2, 3\}$. For example, to avoid the letter pattern $xyzxzyx$, a word w cannot contain any of the factors $1231321, 1321231, 2132312, 2312132, 3123213$ and 3213123.

Petrova gives an almost complete classification of the letter patterns over $\{x, y, z\}$ which can be avoided by ternary square-free words. To do this, she uses the notion of 'codewalks', developed by Shur [9] as a generalization of the encodings introduced by Pansiot [6]. In addition to her classification, Petrova also gives upper and lower bounds on the critical exponents of ternary square-free words avoiding letter patterns $xyxzx$, $xyzxy$, and $xyxzyz$.

Regarding the particular letter pattern $xyzxzyx$, Petrova remarks at the end of her paper that '(p)roving its avoidance will finalize the classification of letter patterns avoidable by ternary square-free words.'

In this note, we show that there is a ternary square-free word avoiding letter pattern $xyzxzyx$. In fact, we

- characterize all the (two-way) infinite ternary square-free words avoiding letter pattern $xyzxzyx$ (Theorems 1 and 2)
- characterize the lexicographically least (one-way) infinite ternary square-free word avoiding letter pattern $xyzxzyx$ (Theorem 3)
- show that the number of ternary square-free words of length n avoiding letter pattern $xyzxzyx$ grows exponentially with n (Theorem 4).

2 Preliminaries

We will use several standard notations from combinatorics on words. An **alphabet** is a finite set whose elements are called **letters**. For an alphabet Σ, we denote by Σ^*, the set of all finite words over Σ; more formally, Σ^* is the free semigroup over Σ, written multiplicatively, with identity element ϵ. We refer to ϵ as the **empty word**. By a **morphism**, we mean a semigroup homomorphism.

If $w = uvz$, with $u, v, z \in \Sigma^*$, we refer to u, v and z as a **prefix**, **factor**, and **suffix** of w, respectively. A word w over Σ is **square-free** if it has no non-empty factor of the form xx.

By Σ^ω, we denote the ω-words over Σ, which are infinite to the right; more formally, an ω-word $\mathbf{w}$ over Σ is a function $\mathbf{w} : \mathbb{N} \to \Sigma$, where $\mathbb{N}$ denotes the set of positive integers. By $\Sigma^\mathbb{Z}$ we denote the $\mathbb{Z}$-words over Σ, which are doubly infinite. Depending on context, a 'word' over Σ may refer to a finite word, an ω-word or a $\mathbb{Z}$-word.

Let $S = \{1, 2, 3\}$, $T = \{a, b, c, d\}$ and $U = \{a, c, d\}$. We put natural orders on alphabets S, T and U:

$$1 < 2 < 3 \text{ and } a < b < c < d.$$

These induce lexicographic orders on words over these alphabets; the definition is recursive: if w is a word and x, y are letters, then $wx < wy$ if and only if $x < y$. For more background on combinatorics on words, see the books by Lothaire [4, 5].

Call a word over S **factor-good** if it has no factor of the form $xyzxzyx$ where $\{x, y, z\} = S$; i.e., the factors 1231321, 1321231, 2132312, 2312132, 3123213, 3213123 are forbidden. Call a word over S **good** if it is square-free and factor-good. Petrova's question is whether there are infinitely many good words.

3 Results on good words

Theorem 1 and Theorem 2 below characterize good $\mathbb{Z}$-words. These turn out to be in 2-to-1 correspondence with square-free $\mathbb{Z}$-words over U.

Let π be the morphism on S^* generated by

$$\pi(1) = 1, \pi(2) = 3, \pi(3) = 2;$$

thus, this morphism π relabels 2's as 3's and vice versa.

Let $f{:}T^* \to S^*$ be the morphism given by

$$f(a) = 1213, f(b) = 123, f(c) = 1323, f(d) = 1232.$$

Let $g{:}U^* \to T^*$ be the map where $g(u)$ is obtained from a word $u \in \{a, c, d\}^*$ by replacing each factor ac of u by abc, each factor da of u by dba and each factor dc of u by dbc.

Theorem 1. *There is a $\mathbb{Z}$-word over S which is good. In particular, if $\mathbf{u} \in U^\mathbb{Z}$ is square-free then $f(g(\mathbf{u}))$ is good.*

Theorem 2. *Let* $\mathbf{w} \in S^{\mathbb{Z}}$ *be good. Exactly one of the following is true:*

1. *There is a square-free word* $\mathbf{u} \in U^{\mathbb{Z}}$ *such that* $\mathbf{w} = f(g(\mathbf{u}))$.
2. *There is a square-free word* $\mathbf{u} \in U^{\mathbb{Z}}$ *such that* $\mathbf{w} = \pi(f(g(\mathbf{u})))$.

We can also characterize the lexicographically least good ω-word:

Theorem 3. *The lexicographically least good* ω*-word is* $f(g(\mathbf{u}))$*, where* $\mathbf{u}$ *is the lexicographically least square-free* ω*-word over* U.

There are 'many' finite good words, in the sense that the number of words grows exponentially with length. For each non-negative integer n, let $G(n)$ be the number of good words of length n.

Theorem 4. *The number of good words of length* n *grows exponentially with* n*. In particular, there are positive constants* A*,* B *and* $C > 1$ *such that*

$$\sum_{i=0}^{n} G(i) \geqslant A + B(C^n).$$

4 Proof of Theorem 2

The proof of Theorem 2 proceeds via a series of lemmas.

Lemma 5. *Suppose* $u \in U^*$*. Then* $f(g(u))$ *is factor-good.*

Lemma 6. *The map* $f \circ g : U^* \to S^*$ *is square-free: Suppose* $u \in U^*$ *is square-free. Then so is* $f(g(u))$.

Suppose that $\mathbf{w} \in \Sigma^{\mathbb{Z}}$ is good. Since $\mathbf{w}$ is square-free,

$$\mathbf{w} \in \{12, 123, 1232, 13, 132, 1323\}^{\mathbb{Z}}.$$

These are just the square-free words over $\{1, 2, 3\}$ which begin with 1 and contain exactly a single 1; evidently we can partition w into such blocks.

Proof. □

Lemma 7. *Let* w *be a good word. Then either* $|w|_{1231} = 0$ *or* $|w|_{1321} = 0$.

Proof. If the lemma is false, then either

- w contains a finite factor with prefix 1231 and suffix 1321 or
- w contains a factor with prefix 1321 and suffix 1231.

Without loss of generality up to relabeling, suppose that w contains a factor with prefix 1231 and suffix 1321. Since it is good, w cannot have 1231321 as a factor. Consider then a shortest factor $1231v1321$ of w; thus $|1231v1321|_{1231} = 1$.

Exhaustively listing good words $1231u$ with $|1231u|_{1231} = 1$, we find that there are only finitely many, and exactly three which are maximal with respect to right extension: 12312131232123, 123132312131232123, 12313231232123. It follows that one of these is a right extension of $1231v1321$; however, none of the three has 1321 as a factor. This is a contradiction. ☐

Interchanging 2's and 3's if necessary, suppose that $\mathbf{w}_{1321} = 0$. Thus

$$\mathbf{w} \in \{12, 123, 1232, 13, 1323\}^{\mathbb{Z}}.$$

Lemma 8. *Suppose* $\mathbf{t} \in 1213\{12, 123, 1232, 13, 1323\}^{\omega}$ *is good. Then*

$$\mathbf{t} \in \{1213, 123, 1232, 1323\}^{\omega}.$$

Proof. We prove this via a series of claims:

Claim 9. *Neither of 132313 and 21232 is a factor of* $\mathbf{t}$.

Proof of Claim. Since $\mathbf{t} \in 1213\{12, 123, 1232, 13, 1323\}^{\omega}$, if 132313 is a factor of $\mathbf{t}$, then so is one of 1323131 and 13231323, both of which end in squares. This is impossible, since $\mathbf{t}$ is good. Similarly, if 21232 is a factor of $\mathbf{t}$, so is one of 121232 and 12321232, both of which begin with squares. ☐

Claim 10. *Suppose that* $t12uv$ *is a factor of* $\mathbf{t}$*, where* t*,* u*,* $v \in \{12, 123, 1232, 13, 1323\}$*. Then* $u = 13$.

Proof of Claim. Word u must be 13 or 1323; otherwise, $12u$ begins with the square 1212. Suppose $u = 1323$. By the previous claim, v must have prefix 12. But then $2uv$ has prefix $2132312 = xyzxzyx$, where $x = 2$, $y = 1$, $z = 3$; this is impossible. Thus $u = 13$. ☐

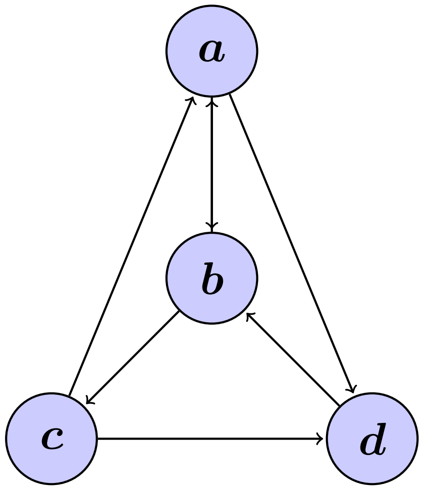
Claim 11. *Suppose that* $tu13v$ *is a factor of* $\mathbf{t}$*, where* t*,* u*,* $v \in \{12, 123, 1232, 13, 1323\}$*. Then* $u = 12$.

Proof of Claim. Word u must end with 2; otherwise, $u13v$ contains the square 3131. Thus u must be 12 or 1232. Suppose $u = 1232$. By the first claim, t must have suffix 3. But then $tu13$ has suffix $3123213 = xyzxzyx$, where $x = 3$, $y = 1$, $z = 2$; this is impossible. Thus $u = 12$. ☐

We have proved that 12 and 13 only appear in $\mathbf{t}$ in the context 1213. It follows that $\mathbf{t} \in \{1213, 123, 1232, 1323\}$. ☐

Corollary 12. *Word* $\mathbf{w} \in \{1213, 123, 1232, 1323\}^{\mathbb{Z}}$.

Figure 1: Directed graph $\mathcal{D}$



Proof. We know that $\mathbf{w} \in \{12, 123, 1232, 13, 1323\}^{\mathbb{Z}}$. If neither of 121 and 131 is a factor of $\mathbf{w}$, then $\mathbf{w}$ is concatenated from copies of $A = 1323$, $B = 1232$ and $C = 123$. However, CB and $AC1$ contain squares, while $BA12$ contains 2132312, which cannot be a factor of a good word. This implies that A, B and C always occur in $\mathbf{w}$ in the cyclical order $A \to B \to C \to A$, and $\mathbf{w}$ contains the square $ABCABC$, which is impossible. We conclude that one of 121 and 131 is a factor of $\mathbf{w}$. However, as in the proof of Claims 10 and 11, factors 12 and 13 can only occur in $\mathbf{w}$ in the context 1213, so the result follows. □

By Corollary 12, $f^{-1}(\mathbf{w})$ exists. Let $\mathbf{v} \in f^{-1}(\mathbf{w})$.

Lemma 13. *None of ac, aba, bd, cb, da and dc is a factor of* $\mathbf{v}$.

Proof. One checks that $f(ac)$, $f(aba)$, $f(bd)$, $f(cb)1$, $f(da)$ contain squares, and thus cannot be factors of $\mathbf{w}$. It follows that ac, aba, cb, da and dc are not factors of $\mathbf{v}$. On the other hand, as in the proof of the previous lemma, $f(d) = 1232$ only appears in $\mathbf{w}$ in the context 123213. It follows that if cd is a factor of $\mathbf{v}$, then $f(cd)13 = 1323123213$ is a factor of $\mathbf{w}$. However, this has the suffix $3123213 = xyzxzyx$ where $x = 3$, $y = 1$, $z = 2$. This is impossible. □

Remark 14. It follows that $\mathbf{v}$ can be walked on the directed graph $\mathcal{D}$ of Figure 1.

Let $h : \{a, b, c, d\}^* \to \{a, c, d\}^*$ be the morphism generated by $h(a) = a$, $h(b) = \epsilon$, $h(c) = c$, $h(d) = d$. Thus $h(w)$ is obtained by deleting all occurrences of b in a word w. Suppose that w is a factor of $\mathbf{v}$. If w does not begin or end with b, then

$$w = g(h(w)).$$

Let $\mathbf{u} = h(\mathbf{v}) \in U^{\mathbb{Z}}$. It follows that $\mathbf{v} = g(\mathbf{u})$, so that

$$\mathbf{w} = f(g(\mathbf{u})).$$

Word $\mathbf{u}$ must be square-free; otherwise its image $\mathbf{w}$ contains a square. Thus the first alternative in Theorem 2 holds.

The other situation occurs if we decide, after Lemma 7, that $\mathbf{w}_{1231} = 0$. As we remarked at that point in our argument, this amounts to interchanging 2's and 3's, i.e., applying π. In such a case, we find that

$$\mathbf{w} = \pi(f(g(\mathbf{u}))).$$

This completes the proof of Theorem 2.

5 Proof of Theorem 1

Proof of Lemma 5. Let $w = f(g(u))$. Each length 7 factor of w is a factor of $f(g(u'))$, some factor $u' \in U^3$. A finite check establishes that $f(u')$ is factor-good for each $u \in U^3$. □

Proof of Lemma 6. Suppose for the sake of getting a contradiction, that XX is a non-empty square in $w = f(v)$. If $|X| \leqslant 2$, then XX is a factor of $f(v')$, some factor v' of v with $|v'| = 2$. However, only need to consider

$$v' \in \{ab, ad, ba, bc, ca, cd, db\}.$$

(As per Remark 14, we can walk v' on $\mathcal{D}$.) In each case, we check that $f(v')$ is square-free. From now on, then, suppose that $|X| \geqslant 3$; in this case we can write

$$XX = qf(v_1 v_2 \cdots v_{n-1})p = qf(v_{n+1} v_{n+2} \cdots v_{2n-1})p,$$

where $v_0 v_1 \cdots v_{n-1} v_n v_{n+1} v_{n+2} \cdots v_{2n-1} v_{2n}$ is a factor of v, q is a suffix of $f(v_0)$, p is a prefix of $f(v_{2n})$, $f(v_n) = pq$, and the $v_i \in T$. It follows that $v_i = v_{n+i}$, $1 \leqslant i \leqslant n-1$.

If $v_0 = v_n$, then v contains the square $(v_0 v_1 v_2 \cdots v_{n-1})^2$; similarly, if $v_n = v_{2n}$, then v contains the square $(v_1 v_2 v_3 \cdots v_n)^2$. Since v is square-free, we deduce that $v_n \neq v_0, v_{2n}$. From the condition that $f(v_n)$ is concatenated from a prefix of v_{2n} and a suffix of v_0, where $v_n \neq v_0, v_{2n}$, we deduce that $v_n = b$.

From the definition of g and the fact that $v_n = b$, we have $v_{n-1} v_n v_1 \in \{abc, dba, dbc\}$. If $v_{n-1} = d$, the definition of g would force $v_n = v_{2n} = b$, contradicting $v_{2n} \neq v_n$. We conclude that $v_{n-1} v_n v_1 = abc$. However, if $v_1 = c$, the definition of g forces $v_n = v_0 = b$, contradicting $v_0 \neq v_n$. □

6 Proof of Theorem 3

Let $\mathbf{u}$ be the lexicographically least square-free ω-word over $U = \{a, c, d\}$, and let $\mathbf{t} = f(g(\mathbf{u}))$. It follows that $\mathbf{u}$ has prefix ac, so that $\mathbf{t}$ has prefix $p = f(g(ac)) = f(abc) = 12131231323$. A finite search shows that p is the lexicographically least good word of length 11. It will therefore suffice to show that $\mathbf{t}$ is the lexicographically least good ω-word with prefix p.

Suppose that $\mathbf{t}_1$ is a good ω-word with prefix p. By Lemma 7, it follows that $|\mathbf{t}_1|_{1321} = 0$, and from the proof of Theorem 2, we conclude that $\mathbf{t}_1 = f(g(\mathbf{u_1}))$, for some square-free word $\mathbf{u_1}$. It remains to show that $\mathbf{u_1}$ is lexicographically greater than or equal to $\mathbf{u}$. Suppose not.

Since $\mathbf{t_1}$ has prefix p, word ac must be a prefix of $\mathbf{u_1}$, and $\mathbf{u}$, $\mathbf{u_1}$ agree on a prefix of length at least 2. Let qrs and qrt be prefixes of $\mathbf{u_1}$ and $\mathbf{u}$, respectively, where $r, s, t \in \{a, c, d\}$, and s is lexicographically less than t.

- If $r = a$, then we cannot have $s = a$, since $\mathbf{u_1}$ is square-free. We therefore must have $s = c$ and $t = d$. It follows that $\mathbf{t_1}$ has prefix $f(g(qa)bc) = f(g(qa))1231323$, and $\mathbf{t}$ has prefix $f(g(qa)d) = f(g(qa))1232$, and we see that $\mathbf{t_1}$ is lexicographically less than $\mathbf{t}$. This contradicts the minimality of $\mathbf{t}$.

- If $r = c$, then we must have $s = a$ and $t = d$. It follows that $\mathbf{t_1}$ has prefix $f(g(qca)) = f(g(qc))1213$, and $\mathbf{t}$ has prefix $f(g(qcd)) = f(g(qc))1232$, and again $\mathbf{t_1}$ is lexicographically less than $\mathbf{t}$, giving a contradiction.

- If $r = d$, then we must have $s = a$ and $t = c$. It follows that $\mathbf{t_1}$ has prefix $f(g(qd)ba) = f(g(qd))1231213$, and $\mathbf{t}$ has prefix $f(g(qd)bc) = f(g(qc))1231323$, and again $\mathbf{t_1}$ is lexicographically less than $\mathbf{t}$.

We conclude that $\mathbf{u_1}$ is lexicographically greater than or equal to $\mathbf{u}$, and $\mathbf{u}$ is the lexicographically least square-free ω-word over U, as claimed.

7 Proof of Theorem 4

Let $C(n)$ be the number of length n square-free words over U. As shown by Brandenburg [2], for $n > 2$, $C(n) \geqslant 6\left(2^{\frac{n}{21}}\right)$. The map $f \circ g$ is injective. Since g simply adds b's between some pairs of letters, $|u| \leqslant |g(u)| < 2|u|$; also, $3|u| \leqslant |f(u)| \leqslant 4|u|$. Let $u \in U^*$ be square-free. By the Lemmas 5 and 6, $f(g(u))$ is good. Also, $3|u| \leqslant |f(g(u))| < 8|u|$. We deduce that distinct square-free words over U of lengths between 3 and $(n+1)/8$ correspond to distinct good words of lengths between 9 and n. It follows that

$$\sum_{i=3}^{\lfloor (n+1)/8 \rfloor} 6\left(2^{\frac{n}{22}}\right) \leqslant \sum_{i=9}^{n} G(i),$$

and the theorem follows with $A = \sum_{i=0}^{8} G(i)$, $B = 6$ and $C = 2^{\frac{1}{22}}$.

Remark 15. The growth rate of ternary square-free words is now very well understood, because of the sharp analysis by Shur [10]. One could definitely tighten the bounds of the above proof; perhaps sharp bounds could be given building on Shur's work.

References

[1] J.-P. Allouche, J. Currie, J. Shallit, Extremal infinite overlap-free words, *Electron. J. Combin.*, **5** (1998), #R27.

[2] F.-J. Brandenburg, Uniformly growing k-th power-free homomorphisms. *Theoret. Comput. Sci.* **23** (1983), 69–82.

[3] W.H. Gottschalk, G.A. Hedlund, A characterization of the Morse minimal set *Proc. Amer. Math. Soc.*, **15** (1964), 70–74.

[4] M. Lothaire, *Combinatorics on Words*. Cambridge University Press, Cambridge (1997).

[5] M. Lothaire, *Algebraic Combinatorics on Words*, Cambridge University Press, Cambridge (2002).

[6] J. J. Pansiot, A propos d'une conjecture de F. Dejean sur les répétitions dans les mots, *Disc. Appl. Math.* **7** (1984), 297–311.

[7] E. A. Petrova, Avoiding Letter Patterns in Ternary Square-Free Words, *Electron. J. Combin.* **23** (1) (2016) #P1.18

[8] R. O. Shelton, Aperiodic words on three symbols, *J. Reine Angew. Math.*, **321** (1981), 195–209.

[9] A. M. Shur, On ternary square-free circular words. *Electron. J. Combin.* **17** 2010, #R140.

[10] A.M. Shur, Two-sided bounds for the growth rates of power-free languages, *Developments in Language Theory 2009*, *Lect. Notes Comp. Sci.* **5583** (2009), 466–477.

[11] A. Thue, Über unendliche Zeichenreihen, *Norske Vid. Selsk. Skr. I. Mat. Nat. Kl.* Christiana (1906), 1–22.

[12] A. Thue, Über die gegenseitige Lage gleicher Teile gewisser Zeichentreihen, *Norske Vid. Selske Skr.*, *I Mat. Nat. Kl. Christiania* **1** (1912), 1–67.