Journal of Integer Sequences, Vol. 22 (2019),
Article 19.8.2

Finite Test Sets for Morphisms That Are Squarefree on Some of Thue's Squarefree Ternary Words

James D. Currie
Department of Mathematics and Statistics
University of Winnipeg
Winnipeg, Manitoba R3B 2E9
Canada
j.currie@uwinnipeg.ca

Abstract

Let S be one of $\{aba, cbc\}$ and $\{aba, aca\}$, and let w be an infinite squarefree word over $\Sigma = \{a, b, c\}$ with no factor in S. Suppose that $f : \Sigma \to T^*$ is a non-erasing morphism. We prove that the word $f(w)$ is squarefree if and only if f is squarefree on factors of w of length 8 or less.

1 Introduction

The papers of Thue on squarefree words [12, 13] are foundational to the area of combinatorics on words. A word w is *squarefree* if we cannot write $w = xyyz$, where y is a non-empty word. The longest squarefree words over the 2-letter alphabet $\{a, b\}$ are aba and bab, each of length 3, but Thue showed that arbitrarily long squarefree words exist over the 3-letter alphabet $\{a, b, c\}$. Infinite squarefree words over finite alphabets are routinely encountered in combinatorics on words, and are frequently used as building blocks in constructions. (See, for example, [9, 10].)

Let w be an infinite squarefree word over $\Sigma = \{a, b, c\}$. Thue showed that w must contain every squarefree word of length 2 over Σ . However, he showed that the same is not true for squarefree words of length 3 over Σ. For each of $S_1 = \{aba, cbc\}, S_2 = \{aba, aca\}$, and

$S_3 = \{aba, bab\}$, Thue constructed an infinite squarefree word over Σ with no factor in S_i. Constructions giving squarefree words equivalent to Thue's word with no factors in S_1 were independently discovered by Braunholtz [4] and Istrail [8]; Berstel [1] shows this equivalence. Their word is called **vtm** (for 'variation of Thue-Morse') by Blanchet-Sadri et al. [3], and has been used as the basis for various constructions [3, 6, 7]. These constructions require showing that $f(\mathbf{vtm})$ is squarefree for particular morphisms f. In this paper, we give a testable characterization of morphisms f such that $f(\mathbf{vtm})$ is squarefree; we do the same in the case where **vtm** is replaced by an infinite squarefree word over Σ with no factors in S_2. We leave as an open problem whether there is a characterization when we replace **vtm** by a word over Σ with no factor in S_3.

Theorem 1. *Let w be an infinite squarefree word over Σ such that either w has no factor in S_1, or w has no factor in S_2. Suppose that $f : \Sigma \to T^*$ is a non-erasing morphism. The word $f(w)$ is squarefree if and only if f is squarefree on factors of w of length 8 or less.*

Our theorem says that to establish squarefreeness of $f(w)$, the morphism f need only be checked for squarefreeness on a finite test set. Crochemore [5] proved a variety of similar theorems; in particular, a morphism f defined on Σ^* preserves squarefreeness exactly when f preserves squarefreeness on words of Σ^* of length at most 5. Note that while Crochemore's theorem requires testing the squarefreeness of $f(v)$ for every squarefree word $v \in \Sigma^*$ up to a certain length, we only test words v that are factors of w. Thus, while aba is squarefree, we do not require $f(aba)$ to be squarefree, for example. Finite test sets for morphisms preserving overlap-freeness have also been well-studied [11].

2 Preliminaries

Let $S = S_1$ or $S = S_2$. For the remainder of this paper, let w be an infinite squarefree word over Σ with no factor in S. Write $w = a_0a_1a_2a_3\cdots$ with $a_i \in \Sigma$. For the remainder of this section suppose that $f : \Sigma \to T^*$ is a non-erasing morphism that is squarefree on factors of w of length 8 or less.

Lemma 2. *Suppose $f(\xi)$ is a factor of $f(x)$, where $x \in \Sigma$ and ξ is a factor of w. Then $|\xi| \leq 3$.*

Proof. If x is a letter of ξ, but $f(\xi)$ is a factor of $f(x)$, we have

$$\begin{aligned} |f(x)| &\leq |f(\xi)| \\ &\leq |f(x)|. \end{aligned}$$

Since f is non-erasing, this forces $x = \xi$, giving $|\xi| = 1$.

If x is not a letter of ξ, then ξ is a squarefree word over a two-letter alphabet, so that $|\xi| \leq 3$. □

Lemma 3. *Suppose that $f(x)$ is a prefix or a suffix of $f(y)$ where $x, y \in \Sigma$. Then $x = y$.*

Proof. We give the proof where $f(x)$ is a prefix of $f(y)$. (The other case is similar.) Suppose $x \neq y$. Then xy must be a factor of w, and xy is squarefree. However, $f(xy)$ begins with the square $f(x)f(x)$, contradicting the squarefreeness of f on factors of w of length at most 8. □

Lemma 4. *There is no solution* $(\alpha, \beta, \gamma, \xi, p, s, t, x, y, z)$ *to*

$$\begin{aligned}
&\alpha\xi xyz\beta \text{ is a factor of } w;\\
&\alpha, \beta, \gamma, x, y, z \in \Sigma, \xi \in \Sigma^*;\\
&t \text{ is a suffix of } f(\gamma);\\
&s \text{ is a suffix of } f(\alpha);\\
&p \text{ is a non-empty prefix of } f(\beta);\\
&t = sf(\xi xyz)p.
\end{aligned}$$

Proof. Suppose, for the sake of getting a contradiction, that $(\alpha, \beta, \gamma, \xi, p, s, t, x, y, z)$ is a solution.

Here p is a prefix of $f(\beta)$, but also a suffix of t, that is a suffix of $f(\gamma)$. Thus we see that $f(\gamma\beta)$ contains square pp, so that $\gamma\beta$ is not a factor of w. This forces $\gamma = \beta$. On the other hand, since $z\beta$ is a factor of w, we conclude that $z \neq \gamma$. Again, $f(z)p$ is a prefix of $f(z\beta)$, but also a suffix of $f(\gamma)$, so that $f(\gamma z\beta)$ contains a square. Since $yz\beta$ is a factor of w, it follows that $y \neq \gamma$. Similarly, $f(\gamma yz\beta)$ contains a square, but $xyz\beta$ is a factor of w, so that that $x \neq \gamma$. Finally, we see that $f(\gamma xyz\beta)$ contains a square. Let δ be the last letter of $\alpha\xi$. We conclude that $\delta \neq \gamma$. However, now δxyz is a squarefree word of length 4 over the two-letter alphabet $\Sigma - \{\gamma\}$. This is impossible. □

The symmetrical lemma is proved analogously:

Lemma 5. *There is no solution* $(\alpha, \beta, \gamma, \xi, p, s, t, x, y, z)$ *to*

$$\begin{aligned}
&\alpha\xi xyz\beta \text{ is a factor of } w;\\
&\alpha, \beta, \gamma, x, y, z \in \Sigma, \xi \in \Sigma^*;\\
&t \text{ is a suffix of } f(\gamma);\\
&s \text{ is a non-empty suffix of } f(\alpha);\\
&p \text{ is a non-empty of } f(\beta);\\
&t = sf(xyz\xi)p.
\end{aligned}$$

Suppose that $f(w)$ contains a non-empty square xx, with $|x|$ as short as possible. Write $f(w) = uxxv$, such that

$$\begin{aligned}
u &= A_0 A_1 \cdots A_i'\\
ux &= A_0 A_1 \cdots A_j'\\
uxx &= A_0 A_1 \cdots A_k',
\end{aligned}$$

where $i \le j \le k$ are non-negative integers, and for each non-negative integer ℓ, $A_\ell = f(a_\ell)$, and A'_ℓ is a prefix of A_ℓ, but $A'_\ell \neq A_\ell$. This notation is not intended to exclude the possibilities that $i = 0$, $i = j$ and/or $j = k$.

Remark 6. Since f is squarefree on factors of w of length at most 8, but $f(a_i \cdots a_k)$ contains the square xx, we must have $k - i \ge 8$.

Remark 7. We cannot have $i = j$; otherwise suffix x of ux is a factor of $A_j = A_i$, and $A_{i+1} \cdots A_{k-1}$ is a factor of suffix x of uxx. Then, $f(a_{i+1}a_{i+2} \cdots a_{k-1})$ is a factor of $f(a_i)$, forcing $(k-1)-(i+1)+1 \le 3$ by Lemma 2, so that $k - i \le 4$, a contradiction. Reasoning, in the same way, we show that $i < j < k$.

For $\ell \in \{i, j, k\}$, let A''_ℓ be the suffix of A_ℓ such that $A_\ell = A'_\ell A''_\ell$. By our choice of A'_ℓ, $A''_\ell \neq \epsilon$. Then

$$x = A''_i A_{i+1} \cdots A_{j-1} A'_j = A''_j A_{j+1} \cdots A_{k-1} A'_k. \tag{1}$$

Remark 8. Since $k - i \ge 8$, we must have $k - j - 1 \ge 3$ and/or $j - i - 1 \ge 3$.

Lemma 9. *We must have* $|A''_i| + |A'_k| \le |x|$ *and* $|A''_j| + |A'_j| \le |x|$.

Proof. We give the proof that $|A''_j| + |A'_j| \le |x|$. (The proof of the other assertion is similar.) Suppose for the sake of getting a contradiction that $|A''_j| + |A'_j| > |x|$. Then $|A''_j| > |x| - |A'_j| = |A''_i A_{i+1} \cdots A_{j-1}|$. It follows that $A''_j = A''_i A_{i+1} \cdots A_{j-1} A'''_j$ for some non-empty prefix A'''_j of A_j. Similarly, one shows that $A'_j = A''''_j A_{j+1} \cdots A_{k-1} A'_k$ for some non-empty suffix A''''_j of A_j. Now $|a_{i+1} \cdots a_{j-1}| = (j-1)-(i+1)+1 = j-i-1$, and $|a_{j+1} \cdots a_{k-1}| = (k-1)-(j+1)+1 = k-j-1$. However, either $j - i - 1 \ge 3$, or $k - j - 1 \ge 3$. If $j - i - 1 \ge 3$, then $A''_j = A''_i A_{i+1} \cdots A_{j-1} A'''_j$ for some non-empty prefix A'''_j of A_j contradicts Lemma 4, letting $s = A''_i$, $\alpha = a_i$, $\xi = a_{i+1} \cdots a_{j-4}$, $xyz = a_{j-3}a_{j-2}a_{j-1}$, $p = A'''_j$, $\beta = a_j$.

In the case where $k - j - 1 \ge 3$, we get the analogous contradiction using Lemma 5. □

Lemma 10. *We have* $j - i = k - j$, $A''_i = A''_j$, $A'_j = A'_k$, *and* $A_{i+\ell} = A_{j+\ell}$, $1 \le \ell \le j - i - 1$.

Proof. To begin with we show that $A'_j = A'_k$. Since both words are suffixes of x, it suffices to show that $|A'_j| = |A'_k|$. Suppose not. Suppose that $|A'_k| < |A'_j|$.

By the previous lemma, $|A'_j| \le |A_{j+1} \cdots A_{k-2} A_{k-1} A'_k|$. Let m be greatest such that $|A_m \cdots A_{k-1} A'_k| \ge |A'_j|$. Thus $j + 1 \le m \le k - 1$, and

$$|A_{m+1} \cdots A_{k-1} A'_k| < |A'_j| \le |A_m A_{m+1} \cdots A_{k-1} A'_k|.$$

If $A'_j = A_m A_{m+1} \cdots A_{k-1} A'_k$, then A_m is a non-empty prefix of A'_j, forcing $a_m = a_j$ so that $A_m = A'_j = A_j$, by Lemma 3. This is contrary to our choice of A'_j.

Similarly, suppose $|A'_k| < |A'_j|$. Suppose that $|A'_k| < |A'_j|$.

Again by the previous lemma, $|A'_k| \le |A_{i+1} \cdots A_{j-2} A_{j-1} A'_j|$. Let m be greatest such that $|A_m \cdots A_{j-1} A'_j| \ge |A'_k|$. Thus $i + 1 \le m \le j - 1$, and

$$|A_{m+1} \cdots A_{j-1} A'_j| < |A'_k| \le |A_m A_{m+1} \cdots A_{j-1} A'_j|.$$

If $A'_k = A_m A_{m+1} \cdots A_{j-1} A'_j$, then A_m is a non-empty prefix of A'_k, forcing $a_m = a_k$ so that $A_m = A'_k = A_k$, by Lemma 3, contrary to our choice of A'_k.

We thus conclude that $A'_j = A'_k$, as desired.

Next, we show that $j - 1 = k - j$ and $A_{i+\ell} = A_{j+\ell}$, $1 \le \ell \le j - i - 1$. Suppose that $j - i \le k - j$. (The other case is similar.) Suppose now that we have shown that for some ℓ, $0 \le \ell < j - i - 1$ that

$$A_{j-\ell} \cdots A_{j-1} A'_j = A_{k-\ell} \cdots A_{k-1} A'_k. \tag{2}$$

This is true when $\ell = 0$; i.e., $A'_j = A'_k$.

From (1), one of $A_{j-\ell-1} A_{j-\ell} \cdots A_{j-1} A'_j$ and $A_{k-\ell-1} A_{k-\ell} \cdots A_{k-1} A'_k$ is a suffix of the other. Together with (2), this implies that one of $A_{j-\ell-1}$ and $A_{k-\ell-1}$ is a suffix of the other. By Lemma 3, this implies that $a_{j-\ell-1} = a_{k-\ell-1}$, and by combining this with (2),

$$A_{j-\ell-1} \cdots A_{j-1} A'_j = A_{k-\ell-1} \cdots A_{k-1} A'_k. \tag{3}$$

By induction we conclude that

$$A_{j-\ell} \cdots A_{j-1} A'_j = A_{k-\ell} \cdots A_{k-1} A'_k, 0 \le \ell \le j - i - 1,$$

which implies $A_{j-\ell} = A_{k-\ell}$, $1 \le \ell \le j - i - 1$. In particular, we note that

$$A_{i+1} \cdots A_{j-1} = A_{k-j+i+1} \cdots A_{k-1}. \tag{4}$$

If we now have $k - j > j - i$, then $k - j + i > j$, and (1) and (4) imply that $A''_i = A''_j A_{j+1} \cdots A_{k-j+i}$. Then A_{k-j+i} is a suffix of A''_i and Lemma 3 forces $A_{k-j+i} = A_i$. Then (1) and (4) force $A_i = A'_i$, contrary to our choice of A'_i. We conclude that $k - j = j - i$. From (1) and (4) we conclude that $A''_i = A''_j$, as desired. □

Corollary 11. *The word w contains a factor $\alpha z \beta z \gamma$, $\alpha, \beta, \gamma \in \Sigma$, $\alpha, \gamma \ne \beta$, $|z| \ge 3$, such that $\alpha\beta\gamma$ is not a factor of w.*

Proof. By Lemma 10, w contains a factor $\alpha z \beta z \gamma$, where $\alpha = a_i$, $\beta = a_j$, $\gamma = a_k$, $z = a_{i+1} \cdots a_{j-1} = a_{j+1} \cdots a_{k-1}$. This gives $|z| = j - i - 1 = k - j - 1 \ge 3$.

Since w is squarefree, we cannot have $a_i = a_j$; otherwise w contains the square $(a_0 z)^2$; similarly, $a_j \ne a_k$. To see that $\alpha\beta\gamma$ is not a factor of w, we note that f is squarefree on factors of w of length at most 8, but $f(\alpha\beta\gamma) = A_i A_j A_k$ contains the square $(A''_i A'_j)(A''_j A'_k)$; this is a square since $A''_i = A''_j$ and $A'_j = A'_k$. Since $|A_i A_j A_k| = 3 \le 8$, we conclude that $A_i A_j A_k$ is not a factor of w. □

3 Results

Lemma 12. *If $S = \{aba, cbc\}$, the only squarefree words of length 3 that are not factors of w are aba and cbc. In addition, w contains no factor of the form azbza or czbzc.*

Proof. Thue [13] showed that a squarefree word over Σ not containing aba or cbc as a factor contains every other length 3 squarefree word as a factor.

Suppose w contains a factor $azbza$. Since aba is not a factor of w, $z \neq \epsilon$. Since az and bz are factors of w, and hence squarefree, the first letter of z cannot be a or b, and must be c. Similarly, the last letter of z must be c. But w contains zbz, and thus cbc. This is a contradiction. Therefore w contains no factor $azbza$.

Replacing a by c and vice versa in the preceding argument shows that w contains no factor $czbzc$. □

Combining the last two lemmas gives this corollary:

Corollary 13. *If $S = \{aba, cbc\}$ and $f : \Sigma^* \to T^*$ is squarefree on factors of w of length at most 8, then $f(w)$ is squarefree.*

Lemma 14. *If $S = \{aba, aca\}$, the only squarefree words of length 3 that are not factors of w are aba and aca. In addition, w contains no factor of the form $azbza$ or $azcza$, $|z| \geq 3$.*

Proof. Thue [13] showed that a squarefree word over Σ not containing aba or aca as a factor contains every other length 3 squarefree word as a factor.

Suppose w contains a factor $azbza$, $|z| \geq 3$. Since aba is not a factor of w, $z \neq \epsilon$. Since az and bz are factors of w, and hence squarefree, the first letter of z cannot be a or b, and must be c. Similarly, the last letter of z must be c. However, since az is a factor of w, but aca is not, the second letter of z cannot be a and must be b. Write $z = cbz'c$. Then $azbza = acbz'cbcbz'ca$ contains the square $cbcb$, that is impossible. We conclude that w contains no factor $azbza$, $|z| \geq 3$.

Replacing c by b and vice versa in the preceding argument shows that w contains no factor $azcza$, $|z| \geq 3$. □

Corollary 15. *If $S = \{aba, aca\}$ and $f : \Sigma^* \to T^*$ is squarefree on factors of w of length at most 8, then $f(w)$ is squarefree.*

Lemma 16. *The squarefree word $azbza$ where $z = cabcbacabcacbacabcbac$ has no factors aba or bab. It follows that any analogous theorem for S_3, with an analogous proof, would require us to replace 8 by a value of at least $|azbza| = 45$.*

Proof. This is established by a finite check. □

4 Acknowledgments

The author's research was supported by the Natural Sciences and Engineering Research Council of Canada (NSERC), [funding reference number 2017-03901]. The author thanks the careful referee and editor for their comments.

References

[1] J. Berstel, Sur la construction de mots sans carré, *Sém. Théor. Nombres Bordeaux* (1978–1979), 18.01–18.15.

[2] J. Berstel, *Axel Thue's Papers on Repetitions in Words: a Translation*, Publications du Laboratoire de Combinatoire et d'Informatique Mathématique **20** Université du Québec à Montréal, 1995.

[3] F. Blanchet-Sadri, J. Currie, N. Fox, and N. Rampersad, Abelian complexity of fixed point of morphism $0 \to 012, 1 \to 02, 2 \to 1$, *INTEGERS* **14** (2014), A11.

[4] C. Braunholtz, An infinite sequence of three symbols with no adjacent repeats, *Amer. Math. Monthly* **70** (1963), 675–676.

[5] M. Crochemore, Sharp characterizations of squarefree morphisms, *Theoret. Comput. Sci.* **18** (1982), 221–226.

[6] J. D. Currie, Which graphs allow infinite nonrepetitive walks?, *Discrete Math.* **87** (1991), 249–260.

[7] James Currie, Tero Harju, Pascal Ochem, and Narad Rampersad, Some further results on squarefree arithmetic progressions in infinite words, *Theoret. Comput. Sci.* **799** (2019), 140–148.

[8] S. Istrail, On irreductible languages and nonrational numbers, *Bull. Math. Soc. Sci. Math. R. S. Roumanie* **21** (1977), 301–308.

[9] M. Lothaire, *Combinatorics on Words*, Cambridge University Press, 1997.

[10] M. Lothaire, *Algebraic Combinatorics on Words*, Cambridge University Press, 2002.

[11] G. Richomme and F. Wlazinski, Overlap-free morphisms and finite test-sets, *Discrete Appl. Math.* **143** (2004), 92–109.

[12] Axel Thue, Über unendliche Zeichenreihen, *Norske Vid. Selsk. Skr. Mat. Nat. Kl.* **7** (1906), 1–22.

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

2010 *Mathematics Subject Classification*: Primary 68R15.
Keywords: Thue-Morse sequence, squarefree word, nonrepetitive word, factor.

(Concerned with sequences A010060 and A036577.)

Received October 3 2019; revised versions received November 30 2019; December 7 2019. Published in *Journal of Integer Sequences*, December 9 2019.

Return to Journal of Integer Sequences home page.