Similarity density of the Thue-Morse word with overlap-free infinite binary words

Chen Fei Du and Jeffrey Shallit
School of Computer Science, University of Waterloo, Waterloo, ON N2L 3G1, Canada
cfdu@uwaterloo.ca,
shallit@uwaterloo.ca

We consider a measure of similarity for infinite words that generalizes the notion of asymptotic or natural density of subsets of natural numbers from number theory. We show that every overlap-free infinite binary word, other than the Thue-Morse word $\mathbf{t}$ and its complement $\bar{\mathbf{t}}$, has this measure of similarity with $\mathbf{t}$ between $\frac{1}{4}$ and $\frac{3}{4}$. This is a partial generalization of a classical 1927 result of Mahler.

1 Introduction

The Thue-Morse word

$$\mathbf{t} = \mathtt{0110100110010110100101100110100}\cdots$$

is one of the most studied objects in combinatorics on words. It can be defined in a number of different ways, such as the fixed point of the morphism μ defined by $\mu(\mathtt{0}) := \mathtt{01}$ and $\mu(\mathtt{1}) := \mathtt{10}$ beginning with $\mathtt{0}$, or as the word whose n^{th} position is the number of $\mathtt{1}$s (modulo 2) in the binary representation of n.

The word $\mathbf{t}$ has a large number of interesting properties, many of which are covered in the survey [1]. For example, $\mathbf{t}$ is *overlap-free*: it contains no factor of the form $axaxa$, where x is a (possibly empty) word and a is a single letter. One that concerns us here is the following "fragility" property [4]: if the bits in any *finite* non-empty set of positions are "flipped" (i.e., changed to their binary complement) in the Thue-Morse word, the resulting word is no longer overlap-free.[1]

Of course, this is not true of arbitrary *infinite* sets of positions; for example, we can transform $\mathbf{t}$ to $\bar{\mathbf{t}}$ by flipping *all* the positions. Chao Hsien Lin (personal communication, October 2013) raised the following natural question.

Problem 1. Is it possible to flip an *infinite*, but density 0, set of positions in $\mathbf{t}$ and still get an overlap-free word?

Our main result (Theorem 18) solves Problem 1 in the negative. After making precise what we mean by "density", we use a certain automaton [10] encoding all the overlap-free infinite binary words to compare $\mathbf{t}$ to all other overlap-free infinite binary words and show that they differ from $\mathbf{t}$ in at least density $\frac{1}{4}$ of the positions. Furthermore, computational evidence suggests that the true lower bound is density $\frac{1}{3}$. However, we were unable to obtain a proof of this tighter bound. Finally, we consider the possibility of similar results holding for other words (in place of $\mathbf{t}$) or for larger classes of words (in place of overlap-free words).

[1] Note that the "fragility" property does not hold for an arbitrary overlap-free binary word; for example, both $\mathtt{0}\mathbf{t}$ and $\mathtt{1}\mathbf{t}$ are overlap-free. There are even overlap-free words in which blocks arbitrarily far from the beginning may be flipped and still remain overlap-free [10].

Z. Ésik and Z. Fülöp (Eds.): Automata and Formal Languages 2014 (AFL 2014)
EPTCS 151, 2014, pp. 231–245, doi:10.4204/EPTCS.151.16

2 Notation

We observe the following notational conventions throughout this paper. We let $\mathbb{N} := \{0,1,2,\dots\}$ denote the natural numbers. The upper-case Greek letters Σ, Δ, Γ represent finite alphabets. For each $n \in \mathbb{N}$, we let $\Sigma_n := \{0,1,2,\dots,n-1\}$.

As usual, Σ^ω denotes the set of all (right-)infinite words over Σ and $L^\omega := \{x_0x_1x_2\cdots \ : \ x_i \in L \setminus \{\varepsilon\}\}$ denote the set of all infinite words formed by concatenation from nonempty words of L. By x^ω we mean the infinite periodic word $xxx\cdots$.

We adopt the convention that, in the context of words, lower-case letters such as x, y, z refer to finite words (i.e., $x, y, z \in \Sigma^*$), while boldface letters $\mathbf{x}, \mathbf{y}, \mathbf{z}$ refer to infinite words (i.e., $\mathbf{x}, \mathbf{y}, \mathbf{z} \in \Sigma^\omega$).

To be consistent with $0 \in \mathbb{N}$, all words are zero-indexed, i.e., the first letter of the word is in position 0. For $x \in \Sigma^*$ and $m \le n \in \mathbb{N}$, $x[n]$ denotes the letter at the n^{th} position of x and $x[m..n]$ denotes the subword consisting of the letters from the m^{th} through n^{th} positions (inclusive) of x. For $x \in \Sigma_2^*$, $\overline{x}$ denotes the binary complement of x, i.e., the word obtained by changing all 0s to 1s and vice versa. We use the same notation just described for infinite words. In addition, for $\mathbf{x} \in \Sigma^\omega$ and $n \in \mathbb{N}$, $\mathbf{x}[n..\infty]$ denotes the (infinite) suffix of $\mathbf{x}$ starting from the n^{th} position of $\mathbf{x}$.

For a morphism $g : \Sigma^* \to \Sigma^*$ and $n \in \mathbb{N}$, we let g^n denote the n-fold composition of g, and $g^\omega : \Sigma^* \to \Sigma^\omega$ denote $\lim_{n\to\infty} g^n$ if the limit exists. The Thue-Morse morphism $\mu : \Sigma_2^* \to \Sigma_2^*$ is defined by $\mu(\mathtt{0}) := \mathtt{01}$ and $\mu(\mathtt{1}) := \mathtt{10}$. Iterates of the Thue-Morse morphism acting on 0 are denoted by $t_n := \mu^n(\mathtt{0})$. Note that $\mathbf{t} = \mu^\omega(\mathtt{0})$.

3 Similarity density of words

Let us express Problem 1 in another way: how similar can an arbitrary overlap-free word $\mathbf{w}$ be to $\mathbf{t}$? For $\mathbf{w}$ a shift of $\mathbf{t}$, this was essentially determined by the following result from a surprisingly little-known 1927 paper of Kurt Mahler on autocorrelation [7].

Theorem 2. *For all $k \in \mathbb{N}$, the limit*

$$\sigma(k) := \lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} (-1)^{\mathbf{t}[i]+\mathbf{t}[i+k]}$$

exists. Furthermore, we have $\sigma(0) = 1$, $\sigma(1) = -\frac{1}{3}$, and for all $n \in \mathbb{N}$, $\sigma(2n) = \sigma(n)$ and $\sigma(2n+1) = -\frac{1}{2}(\sigma(n) + \sigma(n+1))$.

(Also see [11, 12].) Then an easy induction on k gives

Corollary 3. *For all $k \in \mathbb{N} \setminus \{0\}$, $-\frac{1}{3} \le \sigma(k) \le \frac{1}{3}$.*

Mahler's result is not exactly what we want, but we can easily transform it. Rather than autocorrelation, we are more interested in a quantity we call "similarity density"; it measures how similar two words of the same length are, with a simple and intuitive definition for finite words that generalizes to infinite words by way of limits.

Definition 4. We interpret the Kronecker delta as a function of two variables $\delta : \Sigma^2 \to \Sigma_2$ as follows.

$$\delta(a,b) := \begin{cases} 0, & \text{if } a \ne b; \\ 1, & \text{if } a = b. \end{cases}$$

Definition 5. Let $n \in \mathbb{N} \setminus \{0\}$ and $x, y \in \Sigma^n$. The *similarity density* of x and y is

$$\mathrm{SD}(x,y) := \frac{1}{n} \sum_{i=0}^{n-1} \delta(x[i], y[i]).$$

Thus, two finite words of the same length have similarity density 1 if and only if they are equal.

Definition 6. Let $\mathbf{x}, \mathbf{y} \in \Sigma^\omega$. The *lower* and *upper similarity densities* of $\mathbf{x}$ and $\mathbf{y}$ are, respectively,

$$\begin{aligned} \mathrm{LSD}(\mathbf{x},\mathbf{y}) &:= \liminf_{n\to\infty} \mathrm{SD}(\mathbf{x}[0..n-1], \mathbf{y}[0..n-1]), \\ \mathrm{USD}(\mathbf{x},\mathbf{y}) &:= \limsup_{n\to\infty} \mathrm{SD}(\mathbf{x}[0..n-1], \mathbf{y}[0..n-1]). \end{aligned}$$

Remark 7. Our notion of similarity density is not a new idea. (Similar ideas can be found, e.g., in [8, 6].) It is inspired by the well-studied number-theoretic notion of *asymptotic* or *natural density* of subsets of natural numbers. The *lower* and *upper asymptotic densities* of $A \subseteq \mathbb{N}$ are, respectively,

$$\begin{aligned} \underline{d}(A) &:= \liminf_{n\to\infty} \frac{1}{n} \left| A \cap \{0, \ldots, n-1\} \right|, \\ \overline{d}(A) &:= \limsup_{n\to\infty} \frac{1}{n} \left| A \cap \{0, \ldots, n-1\} \right|. \end{aligned}$$

Similarity density generalizes asymptotic density in the following way. For $A \subseteq \mathbb{N}$, let $\chi_A \in \Sigma_2^\omega$ denote the characteristic sequence of A (i.e., $\chi_A[n] = \mathtt{1}$ iff $n \in A$). Then

$$\begin{aligned} \underline{d}(A) &= \mathrm{LSD}(\chi_A, \mathtt{1}^\omega), \\ \overline{d}(A) &= \mathrm{USD}(\chi_A, \mathtt{1}^\omega). \end{aligned}$$

Mahler's result can now be restated as follows.

Theorem 8. *For all $k \in \mathbb{N} \setminus \{0\}$, $\frac{1}{3} \le \mathrm{LSD}(\mathbf{t}, \mathbf{t}[k..\infty]) = \mathrm{USD}(\mathbf{t}, \mathbf{t}[k..\infty]) \le \frac{2}{3}$.*

Proof. Note that for all $i, k \in \mathbb{N}$, $(-1)^{\mathbf{t}[i]+\mathbf{t}[i+k]} = 2\delta(\mathbf{t}[i], \mathbf{t}[i+k]) - 1$. Hence, by Definition 6, Theorem 2, and Corollary 3, we obtain

$$\mathrm{LSD}(\mathbf{t}, \mathbf{t}[k..\infty]) = \mathrm{USD}(\mathbf{t}, \mathbf{t}[k..\infty]) = \frac{1}{2}(\sigma(k)+1) \in \frac{1}{2}\left(\left[-\frac{1}{3}, \frac{1}{3}\right] + 1\right) = \left[\frac{1}{3}, \frac{2}{3}\right]. \qquad \blacksquare$$

Remark 9. There exist overlap-free infinite binary words $\mathbf{w}$ with $\mathrm{LSD}(\mathbf{t},\mathbf{w}) < \mathrm{USD}(\mathbf{t},\mathbf{w})$. One example is the word $\mathbf{h} = \mathtt{00100110100101100110100110010110}\cdots$ whose n^{th} position is the number of $\mathtt{0}$s (modulo 2) in the binary representation of n. (Note that $\mathbf{h}[0] = \mathtt{0}$ as we take the binary representation of 0 to be ε.) We prove in Proposition 17 that $\mathrm{LSD}(\mathbf{t},\mathbf{h}) = \frac{1}{3}$ while $\mathrm{USD}(\mathbf{t},\mathbf{h}) = \frac{2}{3}$. See Figure 1, where this similarity density is graphed as a function of the length of the prefix.

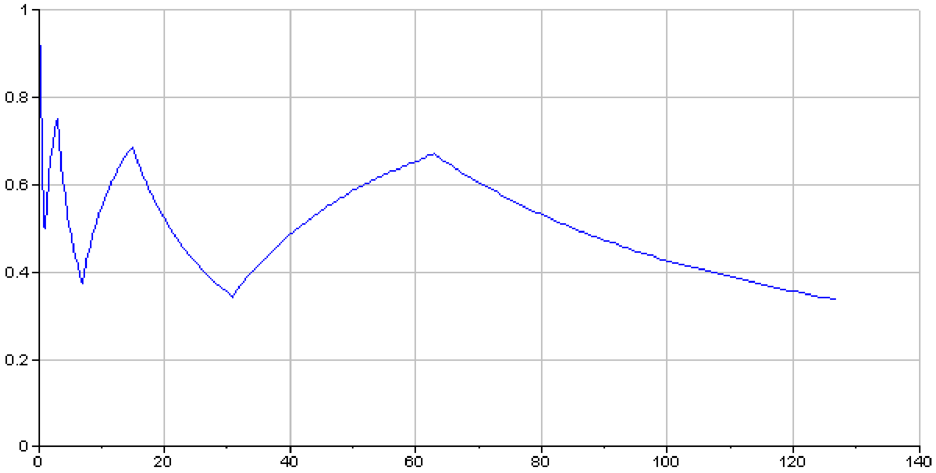

Figure 1: Similarity density of prefixes of **t** and **h**

Our main result (Theorem 18) is that the lower and upper similarity densities of **t** with *any* overlap-free infinite binary word other than **t** and $\overline{\mathbf{t}}$ are bounded below and above as in Theorem 8, but with the constants $\frac{1}{4}$ and $\frac{3}{4}$ instead of $\frac{1}{3}$ and $\frac{2}{3}$ respectively. However, computational evidence suggests that the tighest bounds are indeed $\frac{1}{3}$ and $\frac{2}{3}$, which, if true, would fully generalize Theorem 8 from nontrivial shifts of **t** to all overlap-free infinite binary words (other than **t** and $\overline{\mathbf{t}}$).

The following are basic properties of similarity density that we will use later. Their statements are all intuitive and their proofs are just basic exercises in algebra. Observation 10 states that similarity density can be computed using weighted averages. Observation 11 and Corollary 12 explain how complementation affects similarity density. Observation 13 states that the similarity densities of infinite words depends only on their tails, so we can ignore arbitrarily long prefixes. Observation 14 states that the similarity densities of infinite words can be obtained by considering similarity densities of prefixes where the length of the prefix grows by any constant instead of just by one in each iteration.

Observation 10. *Let $n,m \in \mathbb{N} \setminus \{0\}$, $u,v \in \Sigma^n$, and $x,y \in \Sigma^m$. Then*

$$\mathrm{SD}(ux,vy) = \frac{n}{n+m}\mathrm{SD}(u,v) + \frac{m}{n+m}\mathrm{SD}(x,y).$$

Proof.

$$\begin{aligned}
\mathrm{SD}(ux,vy) &= \frac{1}{n+m}\sum_{i=0}^{n+m-1}\delta((ux)[i],(vy)[i]) \\
&= \frac{1}{n+m}\left(\sum_{i=0}^{n-1}\delta(u[i],v[i]) + \sum_{i=0}^{m-1}\delta(x[i],y[i])\right) \\
&= \frac{n}{n+m}\cdot\frac{1}{n}\sum_{i=0}^{n-1}\delta(u[i],v[i]) + \frac{m}{n+m}\cdot\frac{1}{m}\sum_{i=0}^{m-1}\delta(u[i],v[i]) \\
&= \frac{n}{n+m}\mathrm{SD}(u,v) + \frac{m}{n+m}\mathrm{SD}(x,y).
\end{aligned}$$

■

Observation 11. *For all* $n \in \mathbb{N} \setminus \{0\}$ *and* $x, y \in \Sigma_2^n$,

(i) $\mathrm{SD}(\overline{x}, y) = 1 - \mathrm{SD}(x, y)$.

(ii) $\mathrm{SD}(\overline{x}, \overline{y}) = \mathrm{SD}(x, y)$.

Proof.

(i) $\mathrm{SD}(\overline{x}, y) = \frac{1}{n} \sum_{i=0}^{n-1} \delta(\overline{x}[i], y[i]) = \frac{1}{n} \sum_{i=0}^{n-1} (1 - \delta(x[i], y[i])) = 1 - \mathrm{SD}(x, y)$.

(ii) By (i) and symmetry of SD, we have $\mathrm{SD}(\overline{x}, \overline{y}) = 1 - \mathrm{SD}(x, \overline{y}) = 1 - (1 - \mathrm{SD}(x, y)) = \mathrm{SD}(x, y)$. ∎

Corollary 12. *For all* $\mathbf{x}, \mathbf{y} \in \Sigma_2^\omega$,

(i) $\mathrm{LSD}(\overline{\mathbf{x}}, \mathbf{y}) = 1 - \mathrm{USD}(\mathbf{x}, \mathbf{y})$ *and* $\mathrm{USD}(\overline{\mathbf{x}}, \mathbf{y}) = 1 - \mathrm{LSD}(\mathbf{x}, \mathbf{y})$.

(ii) $\mathrm{LSD}(\overline{\mathbf{x}}, \overline{\mathbf{y}}) = \mathrm{LSD}(\mathbf{x}, \mathbf{y})$ *and* $\mathrm{USD}(\overline{\mathbf{x}}, \overline{\mathbf{y}}) = \mathrm{USD}(\mathbf{x}, \mathbf{y})$.

Proof. Immediate by Definition 6, Observation 11, and basic properties of limits. ∎

Observation 13. *Let* $l \in \mathbb{N}$, $u, v \in \Sigma^l$ *and* $\mathbf{x}, \mathbf{y} \in \Sigma^\omega$. *Then* $\mathrm{LSD}(u\mathbf{x}, v\mathbf{y}) = \mathrm{LSD}(\mathbf{x}, \mathbf{y})$ *and* $\mathrm{USD}(u\mathbf{x}, v\mathbf{y}) = \mathrm{USD}(\mathbf{x}, \mathbf{y})$.

Proof. If $l = 0$, then the proof is trivial. If $l > 0$, then we have

$$\begin{aligned}
\mathrm{LSD}(u\mathbf{x}, v\mathbf{y}) &= \liminf_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} \delta((u\mathbf{x})[i], (v\mathbf{y})[i]) \\
&= \liminf_{n\to\infty} \frac{1}{n+l} \sum_{i=0}^{n+l-1} \delta((u\mathbf{x})[i], (v\mathbf{y})[i]) \\
&= \liminf_{n\to\infty} \left(\underbrace{\frac{1}{n+l} \sum_{i=0}^{l-1} \delta(u[i], v[i])}_{\in [0, \frac{l}{n+l}] \xrightarrow{n\to\infty} 0} + \frac{1}{n+l} \sum_{i=0}^{n-1} \delta(\mathbf{x}[i], \mathbf{y}[i]) \right) \\
&= \liminf_{n\to\infty} \left(0 + \left(\frac{1}{n} - \frac{l}{n(n+l)} \right) \sum_{i=0}^{n-1} \delta(\mathbf{x}[i], \mathbf{y}[i]) \right) \\
&= \liminf_{n\to\infty} \left(\frac{1}{n} \sum_{i=0}^{n-1} \delta(\mathbf{x}[i], \mathbf{y}[i]) - \underbrace{\frac{l}{n(n+l)} \sum_{i=0}^{n-1} (1 - \delta(\mathbf{x}[i], \mathbf{y}[i]))}_{\in [0, \frac{l}{n+l}] \xrightarrow{n\to\infty} 0} \right) \\
&= \liminf_{n\to\infty} \left(\frac{1}{n} \sum_{i=0}^{n-1} (1 - \delta(\mathbf{x}[i], \mathbf{y}[i])) - 0 \right) \\
&= \mathrm{LSD}(\mathbf{x}, \mathbf{y}).
\end{aligned}$$

The proof is exactly the same for USD with lim inf replaced by lim sup. ∎

Observation 14. *Let* $M \in \mathbb{N} \setminus \{0\}$. *Then*

$$\begin{aligned}
\mathrm{LSD}(\mathbf{x}, \mathbf{y}) &= \liminf_{n\to\infty} \mathrm{SD}(\mathbf{x}[0..Mn-1], \mathbf{y}[0..Mn-1]), \\
\mathrm{USD}(\mathbf{x}, \mathbf{y}) &= \limsup_{n\to\infty} \mathrm{SD}(\mathbf{x}[0..Mn-1], \mathbf{y}[0..Mn-1]).
\end{aligned}$$

Proof. For any $n \in \mathbb{N} \setminus \{0\}$ and $k \in \{Mn, Mn+1, \ldots, M(n+1)-2\}$, by Observation 10, we have

$$\begin{aligned}
\mathrm{SD}(\mathbf{x}[0..k], \mathbf{y}[0..k]) &= \frac{Mn}{k+1}\,\mathrm{SD}(\mathbf{x}[0..Mn-1], \mathbf{y}[0..Mn-1]) \\
&\quad + \frac{k-Mn+1}{k+1}\,\mathrm{SD}(\mathbf{x}[Mn..k], \mathbf{y}[Mn..k]) \\
&\in \left[\frac{Mn}{M(n+1)-1}, \frac{Mn}{Mn+1}\right] \mathrm{SD}(\mathbf{x}[0..Mn-1], \mathbf{y}[0..Mn-1]) \\
&\quad + \left[\frac{1}{M(n+1)-1}, \frac{M-1}{Mn+1}\right] \mathrm{SD}(\mathbf{x}[Mn..k], \mathbf{y}[Mn..k]),
\end{aligned}$$

so since $\lim_{n\to\infty}[\frac{Mn}{M(n+1)-1}, \frac{Mn}{Mn+1}] = [1,1] = \{1\}$ and $\lim_{n\to\infty}[\frac{1}{M(n+1)-1}, \frac{M-1}{Mn+1}] = [0,0] = \{0\}$, all of the intermediate values $\mathrm{SD}(\mathbf{x}[0..k], \mathbf{y}[0..k])$ for $k \in \{Mn, Mn+1, .., M(n+1)-2\}$ get arbitrarily close to $\mathrm{SD}(\mathbf{x}[0..Mn-1], \mathbf{y}[0..Mn-1])$ as $n \to \infty$. Hence,

$$\begin{aligned}
\liminf_{n\to\infty} \mathrm{SD}(\mathbf{x}[0..n-1], \mathbf{y}[0..n-1]) &= \liminf_{n\to\infty} \mathrm{SD}(\mathbf{x}[0..Mn-1], \mathbf{y}[0..Mn-1]), \\
\limsup_{n\to\infty} \mathrm{SD}(\mathbf{x}[0..n-1], \mathbf{y}[0..n-1]) &= \limsup_{n\to\infty} \mathrm{SD}(\mathbf{x}[0..Mn-1], \mathbf{y}[0..Mn-1]). \qquad \blacksquare
\end{aligned}$$

4 Fife automaton for overlap-free infinite binary words

We recall the so-called “Fife automaton” for overlap-free infinite binary words from [10]. (Note that this automaton does not appear in the original paper of Fife [5].)

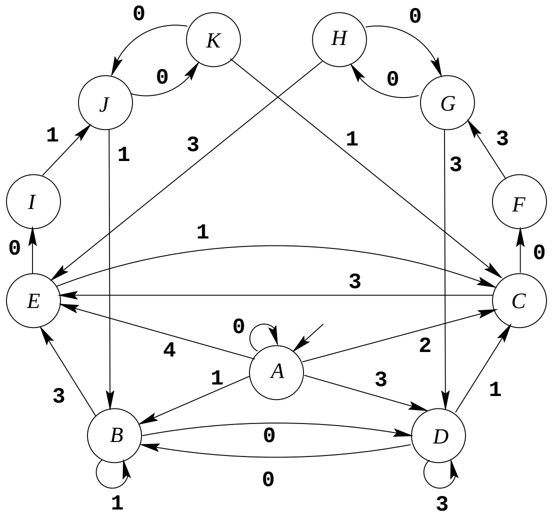

Figure 2: Automaton encoding all overlap-free infinite binary words

Here, infinite paths through the automaton encode all overlap-free infinite binary words, as follows.

Definition 15. First, each of the edge labels encodes a binary word, via $c : \Sigma_5 \to \Sigma_2^*$ defined by

$$\begin{aligned}
c(\mathtt{0}) &:= \varepsilon, \\
c(\mathtt{1}) &:= \mathtt{0}, \\
c(\mathtt{2}) &:= \mathtt{00}, \\
c(\mathtt{3}) &:= \mathtt{1}, \\
c(\mathtt{4}) &:= \mathtt{11}.
\end{aligned}$$

Then, the Fife-to-binary encoding $\mathrm{FBE} : (\Sigma_5^\omega \setminus \Sigma_5^* \mathtt{0}^\omega) \cup \left(\Sigma_5^* \mathtt{0}^\omega \times \Sigma_2\right) \to \Sigma_2^\omega$ is defined by

$$\begin{aligned}
\mathrm{FBE}(\mathbf{x}) &:= \prod_{n=0}^{\infty} \mu^n(c(\mathbf{x}[n])) && \text{for } \mathbf{x} \in \Sigma_5^\omega \setminus \Sigma_5^* \mathtt{0}^\omega; \\
\mathrm{FBE}(\mathbf{x}, a) &:= \left(\prod_{n=0}^{\infty} \mu^n(c(\mathbf{x}[n])) \right) \mu^\omega(a) && \text{for } (\mathbf{x}, a) \in \Sigma_5^* \mathtt{0}^\omega \times \Sigma_2.
\end{aligned}$$

Note that FBE is well-defined because c is only erasing for the letter $\mathtt{0}$ and μ is non-erasing, so for $\mathbf{x} \in \Sigma_5^\omega$, the concatenation $\prod_{n=0}^{\infty} \mu^n(c(\mathbf{x}[n]))$ is finite iff $\mathbf{x}$ ends in $\mathtt{0}^\omega$.

We now recall the basic property of the automaton from [10].

Theorem 16. *Let $\mathbf{w} \in \Sigma_2^\omega$. Then $\mathbf{w}$ is overlap-free iff there exists $\mathbf{x} \in \Sigma_5^\omega$ that encodes a valid path through the Fife automaton for overlap-free infinite binary words such that* $\mathrm{FBE}(\mathbf{x}) = \mathbf{w}$ *(if $\mathbf{x}$ does not end in $\mathtt{0}^\omega$) or* $\mathrm{FBE}(\mathbf{x}, a) = \mathbf{w}$ *(if $\mathbf{x}$ ends in $\mathtt{0}^\omega$) for some $a \in S$, where $S \subseteq \Sigma_2$ depends on the eventual cycle corresponding to the suffix $\mathtt{0}^\omega$ of the path encoded by $\mathbf{x}$: on state A and between states B and D ($S = \Sigma_2$), between states G and H ($S = \{\mathtt{1}\}$), or between states J and K ($S = \{\mathtt{0}\}$).*

Recall $\mathbf{h}$ as defined in Remark 9. Note that the definitions of $\mathbf{h}$ and $\mathbf{t}$ are very similar. This is related to the special path that encodes $\mathbf{h}$ in the Fife automaton for overlap-free infinite binary words [10]: $\mathbf{h} = \mathrm{FBE}(\mathtt{2}(\mathtt{31})^\omega)$. We will see later in our proof of our main result why this path is special. For now, we can use this path to compute the following result.

Proposition 17. $\mathrm{LSD}(\mathbf{h}, \mathbf{t}) = \mathrm{LSD}(\overline{\mathbf{h}}, \mathbf{t}) = \frac{1}{3}$ *and* $\mathrm{USD}(\mathbf{h}, \mathbf{t}) = \mathrm{USD}(\overline{\mathbf{h}}, \mathbf{t}) = \frac{2}{3}$.

Proof. Note that

$$\begin{aligned}
\mathbf{h} &= \mathrm{FBE}(\mathtt{2}(\mathtt{31})^\omega) = \mu^0(p(\mathtt{2})) \prod_{n=0}^{\infty} \left(\mu^{2n+1}(p(\mathtt{3})) \mu^{2n+2}(p(\mathtt{1})) \right) \\
&= \mu^0(\mathtt{00}) \prod_{n=0}^{\infty} \left(\mu^{2n+1}(\mathtt{1}) \mu^{2n+2}(\mathtt{0}) \right) = \mathtt{0} t_0 \prod_{n=0}^{\infty} \left(\overline{t_{2n+1}} t_{2n+2} \right) = \mathtt{0} \prod_{n=0}^{\infty} \left(t_{2n} \overline{t_{2n+1}} \right),
\end{aligned}$$

and since for each $n \in \mathbb{N}$, we have $|t_n| = 2^n$ and $1 + \sum_{i=0}^{n} 2^i = 2^{n+1}$, it follows that

$$\mathbf{h}[2^n \mathinner{..} 2^{n+1} - 1] = \begin{cases} t_n, & \text{if } n \equiv 0 \pmod{2}; \\ \overline{t_n}, & \text{if } n \equiv 1 \pmod{2}. \end{cases}$$

Note that for each $n \in \mathbb{N}$, we have $\mathbf{t}[2^n \mathinner{..} 2^{n+1} - 1] = t_{n+1}[2^n \mathinner{..} 2^{n+1} - 1] = \overline{t_n}$. Hence, for all $n \in \mathbb{N}$,

$$\mathbf{h}[2^n \mathinner{..} 2^{n+1} - 1] = \begin{cases} \overline{\mathbf{t}}[2^n \mathinner{..} 2^{n+1} - 1], & \text{if } n \equiv 0 \pmod{2}; \\ \mathbf{t}[2^n \mathinner{..} 2^{n+1} - 1], & \text{if } n \equiv 1 \pmod{2}, \end{cases}$$

whence

$$\mathrm{SD}(\mathbf{h}[2^n..2^{n+1}-1],\mathbf{t}[2^n..2^{n+1}-1]) = \begin{cases} 0, & \text{if } n \equiv 0 \pmod 2; \\ 1, & \text{if } n \equiv 1 \pmod 2. \end{cases}$$

If we consider two of these blocks at a time, we obtain, by Observation 10, that for all $n \in \mathbb{N}$,

$$\begin{aligned} \mathrm{SD}(\mathbf{h}[2^n..2^{n+2}-1],\mathbf{t}[2^n..2^{n+2}-1]) &= \frac{2^n}{2^n+2^{n+1}}\mathrm{SD}(\mathbf{h}[2^n..2^{n+1}-1],\mathbf{t}[2^n..2^{n+1}-1]) \\ &\quad + \frac{2^{n+1}}{2^n+2^{n+1}}\mathrm{SD}(\mathbf{h}[2^{n+1}..2^{n+2}-1],\mathbf{t}[2^{n+1}..2^{n+2}-1]) \\ &= \begin{cases} \frac{2}{3}, & \text{if } n \equiv 0 \pmod 2; \\ \frac{1}{3}, & \text{if } n \equiv 1 \pmod 2. \end{cases} \end{aligned}$$

Iterating Observation 10 finitely many times, we obtain that for all $n \in \mathbb{N}$,

$$\mathrm{SD}(\mathbf{h}[1..2^{2n}-1],\mathbf{t}[1..2^{2n}-1]) = \frac{2}{3},$$
$$\mathrm{SD}(\mathbf{h}[2..2^{2n+1}-1],\mathbf{t}[2..2^{2n+1}-1]) = \frac{1}{3}.$$

Furthermore, applying Observation 10 one letter at a time, we see that for $k \in [2^{2n}-1, 2^{2n+1}-1]$, $\mathrm{SD}(\mathbf{h}[1..k],\mathbf{t}[1..k])$ monotonically decreases (from $\frac{2}{3}$), and for $k \in [2^{2n+1}-1, 2^{2n+2}-1]$, $\mathrm{SD}(\mathbf{h}[1..k],\mathbf{t}[1..k])$ monotonically increases (back to $\frac{2}{3}$). Thus,

$$\mathrm{USD}(\mathbf{h}[1..\infty],\mathbf{t}[1..\infty]) = \limsup_{n\to\infty} \mathrm{SD}(\mathbf{h}[1..n],\mathbf{t}[1..n]) = \frac{2}{3}.$$

Similarly, for $k \in [2^{2n+1}-1, 2^{2n+2}-1]$, $\mathrm{SD}(\mathbf{h}[2..k],\mathbf{t}[2..k])$ monotonically increases (from $\frac{1}{3}$), and for $k \in [2^{2n+2}-1, 2^{2n+3}-1]$, $\mathrm{SD}(\mathbf{h}[2..k],\mathbf{t}[2..k])$ monotonically decreases (back to $\frac{1}{3}$), so

$$\mathrm{LSD}(\mathbf{h}[2..\infty],\mathbf{t}[2..\infty]) = \liminf_{n\to\infty} \mathrm{SD}(\mathbf{h}[2..n+1],\mathbf{t}[2..n+1]) = \frac{1}{3}.$$

Finally, by Observation 13, we conclude that $\mathrm{LSD}(\mathbf{h},\mathbf{t}) = \mathrm{LSD}(\mathbf{h}[2..\infty],\mathbf{t}[2..\infty]) = \frac{1}{3}$ and $\mathrm{USD}(\mathbf{h},\mathbf{t}) = \mathrm{USD}(\mathbf{h}[1..\infty],\mathbf{t}[1..\infty]) = \frac{2}{3}$, whence by Corollary 12(i), we obtain $\mathrm{LSD}(\overline{\mathbf{h}},\mathbf{t}) = 1-\mathrm{USD}(\mathbf{h},\mathbf{t}) = 1-\frac{2}{3} = \frac{1}{3}$ and $\mathrm{USD}(\overline{\mathbf{h}},\mathbf{t}) = 1-\mathrm{LSD}(\mathbf{h},\mathbf{t}) = 1-\frac{1}{3} = \frac{2}{3}$. ∎

5 Main result

We now state and prove our main result.

Theorem 18. *For all overlap-free* $\mathbf{w} \in \Sigma_2^\omega \setminus \{\mathbf{t},\overline{\mathbf{t}}\}$, $\frac{1}{4} \le \mathrm{LSD}(\mathbf{w},\mathbf{t}) \le \mathrm{USD}(\mathbf{w},\mathbf{t}) \le \frac{3}{4}$.

Our approach to proving Theorem 18 is to consider each overlap-free infinite binary word in terms of the path through the Fife automaton that encodes it. We divide the paths into four cases.

(1) ends in $\mathtt{0}^\omega$.

(2) does not end in $\mathtt{0}^\omega$, begins with $\mathtt{0}^n\mathtt{2}$ or $\mathtt{0}^n\mathtt{4}$ for some $n \in \mathbb{N}$, and contains exactly n $\mathtt{0}$s.

(3) does not end in $\mathtt{0}^\omega$, begins with $\mathtt{0}^n\mathtt{2}$ or $\mathtt{0}^n\mathtt{4}$ for some $n \in \mathbb{N}$, and contains more than n $\mathtt{0}$s.

(4) does not end in $\mathtt{0}^\omega$ and begins with $\mathtt{0}^n\mathtt{1}$ or $\mathtt{0}^n\mathtt{3}$ for some $n \in \mathbb{N}$.

Upon closer examination of the Fife automaton, case (2) can be subdivided into two cases: $\mathtt{0}^n\mathtt{2}(\mathtt{31})^\omega$ and their complements under FBE, $\mathtt{0}^n\mathtt{4}(\mathtt{13})^\omega$. It turns out that we can bootstrap Proposition 17 to obtain the same bounds for both of these cases. Case (1) follows from Mahler's theorem 8, but it will also follow from our own generalized version of it (albeit with weaker bounds). For cases (3) and (4), we observe that the infinite binary word corresponding to the path eventually "lags behind" the prefixes t_n of $\mathbf{t}$ in the sense that each successive n^{th} symbol in the path can only generate positions prior to 2^n, whence we can use a technical lemma that bounds the similarity density of t_n with nontrivial subwords of t_{n+1} to complete the proof.

Proposition 19. *For all $n \in \mathbb{N}$ we have* $\mathrm{LSD}(\mathrm{FBE}(\mathtt{0}^n\mathtt{2}(\mathtt{31})^\omega), \mathbf{t}) = \frac{1}{3}$ *and* $\mathrm{USD}(\mathrm{FBE}(\mathtt{0}^n\mathtt{2}(\mathtt{31})^\omega), \mathbf{t}) = \frac{2}{3}$.

Proof. Note that

$$\begin{aligned}
\mathrm{FBE}(\mathtt{0}^n\mathtt{2}(\mathtt{31})^\omega) &= \prod_{k=0}^{n-1}\left(\mu^k(p(\mathtt{0}))\right)\mu^n(p(\mathtt{2}))\prod_{k=0}^{\infty}\left(\mu^{n+2k+1}(p(\mathtt{3}))\mu^{n+2k+2}(p(\mathtt{1}))\right)\\
&= \prod_{k=0}^{n-1}\left(\mu^k(\varepsilon)\right)\mu^n(\mathtt{00})\prod_{k=0}^{\infty}\left(\mu^{n+2k+1}(\mathtt{1})\mu^{n+2k+2}(\mathtt{0})\right)\\
&= t_n t_n \prod_{k=0}^{\infty}(\overline{t_{n+2k+1}}t_{n+2k+2})\\
&= t_n \prod_{k=0}^{\infty}(t_{n+2k}\overline{t_{n+2k+1}}).
\end{aligned}$$

From the proof of Proposition 17, we see that

$$\mathrm{FBE}(\mathtt{0}^n\mathtt{2}(\mathtt{31})^\omega)[2^n..\infty] = \begin{cases}\mathbf{h}[2^n..\infty], & \text{if } n \equiv 0 \pmod 2;\\ \overline{\mathbf{h}}[2^n..\infty], & \text{if } n \equiv 1 \pmod 2.\end{cases}$$

Hence, by Observation 13 and Proposition 17, we have

$$\begin{aligned}
(\mathrm{LSD},\mathrm{USD})(\mathrm{FBE}(\mathtt{0}^n\mathtt{2}(\mathtt{31})^\omega),\mathbf{t}) &= (\mathrm{LSD},\mathrm{USD})(\mathrm{FBE}(\mathtt{0}^n\mathtt{2}(\mathtt{31})^\omega)[2^n..\infty],\mathbf{t}[2^n..\infty])\\
&= \begin{cases}(\mathrm{LSD},\mathrm{USD})(\mathbf{h}[2^n..\infty],\mathbf{t}[2^n..\infty]), & \text{if } n \equiv 0 \pmod 2;\\ (\mathrm{LSD},\mathrm{USD})(\overline{\mathbf{h}}[2^n..\infty],\mathbf{t}[2^n..\infty]), & \text{if } n \equiv 1 \pmod 2,\end{cases}\\
&= \begin{cases}(\mathrm{LSD},\mathrm{USD})(\mathbf{h},\mathbf{t}), & \text{if } n \equiv 0 \pmod 2;\\ (\mathrm{LSD},\mathrm{USD})(\overline{\mathbf{h}},\mathbf{t}), & \text{if } n \equiv 1 \pmod 2,\end{cases}\\
&= \left(\frac{1}{3},\frac{2}{3}\right).
\end{aligned}$$

■

Lemma 20. *For all $n \in \mathbb{N}$ and $i \in [1, 2^n - 1]$,*

$$(a) \qquad \mathrm{SD}(t_n, t_{n+1}[i\,..\,2^n+i-1]) \in \begin{cases} \{\frac{1}{2}\}, & \textit{if } i = 2^{n-1}; \\ [\frac{1}{4}, \frac{3}{4}], & \textit{otherwise}. \end{cases}$$

$$\overline{(}a) \qquad \mathrm{SD}(\overline{t_n}, t_{n+1}[i\,..\,2^n+i-1]) \in \begin{cases} \{\frac{1}{2}\}, & \textit{if } i = 2^{n-1}; \\ [\frac{1}{4}, \frac{3}{4}], & \textit{otherwise}. \end{cases}$$

$$(a\overline{)} \qquad \mathrm{SD}(t_n, \overline{t_{n+1}}[i\,..\,2^n+i-1]) \in \begin{cases} \{\frac{1}{2}\}, & \textit{if } i = 2^{n-1}; \\ [\frac{1}{4}, \frac{3}{4}], & \textit{otherwise}. \end{cases}$$

$$\overline{(}a\overline{)} \qquad \mathrm{SD}(\overline{t_n}, \overline{t_{n+1}}[i\,..\,2^n+i-1]) \in \begin{cases} \{\frac{1}{2}\}, & \textit{if } i = 2^{n-1}; \\ [\frac{1}{4}, \frac{3}{4}], & \textit{otherwise}. \end{cases}$$

$$(b) \qquad \mathrm{SD}(t_n, t_n^2[i\,..\,2^n+i-1]) \in \begin{cases} \{0\}, & \textit{if } i = 2^{n-1}; \\ [\frac{1}{4}, \frac{3}{4}], & \textit{otherwise}. \end{cases}$$

$$\overline{(}b) \qquad \mathrm{SD}(\overline{t_n}, t_n^2[i\,..\,2^n+i-1]) \in \begin{cases} \{1\}, & \textit{if } i = 2^{n-1}; \\ [\frac{1}{4}, \frac{3}{4}], & \textit{otherwise}. \end{cases}$$

$$(b\overline{)} \qquad \mathrm{SD}(t_n, \overline{t_n}^2[i\,..\,2^n+i-1]) \in \begin{cases} \{1\}, & \textit{if } i = 2^{n-1}; \\ [\frac{1}{4}, \frac{3}{4}], & \textit{otherwise}. \end{cases}$$

$$\overline{(}b\overline{)} \qquad \mathrm{SD}(\overline{t_n}, \overline{t_n}^2[i\,..\,2^n+i-1]) \in \begin{cases} \{0\}, & \textit{if } i = 2^{n-1}; \\ [\frac{1}{4}, \frac{3}{4}], & \textit{otherwise}. \end{cases}$$

Proof. By induction on n.

- For $n = 0$, all eight cases are vacuously true due to $i \in \emptyset$.
- Suppose all eight cases hold for some $n \in \mathbb{N}$. For $i \in [1, 2^{n+1} - 1]$, using Observation 10 followed by the induction hypothesis, we calculate

$$\begin{aligned}
&\mathrm{SD}(t_{n+1}, t_{n+2}[i\,..\,2^{n+1}+i-1]) \\
&= \mathrm{SD}(t_n\overline{t_n}, (t_n\overline{t_n t_n}t_n)[i\,..\,2^{n+1}+i-1]) \\
&= \begin{cases} \mathrm{SD}(t_n\overline{t_n}, (t_n\overline{t_n t_n})[i\,..\,2^{n+1}+i-1]), & \text{if } i \in [1, 2^n-1]; \\ \mathrm{SD}(t_n\overline{t_n}, \overline{t_n t_n}), & \text{if } i = 2^n; \\ \mathrm{SD}(t_n\overline{t_n}, (\overline{t_n t_n}t_n)[i-2^n\,..\,2^n+i-1]), & \text{if } i \in [2^n+1, 2^{n+1}-1], \end{cases} \\
&= \begin{cases} \frac{2^n}{2^{n+1}}\,\mathrm{SD}(t_n, (t_n\overline{t_n})[i\,..\,2^n+i-1]) + \frac{2^n}{2^{n+1}}\,\mathrm{SD}(\overline{t_n}, (\overline{t_n t_n})[i\,..\,2^n+i-1]), & \text{if } i \in [1, 2^n-1]; \\ \frac{2^n}{2^{n+1}}\,\mathrm{SD}(t_n, \overline{t_n}) + \frac{2^n}{2^{n+1}}\,\mathrm{SD}(\overline{t_n}, \overline{t_n}), & \text{if } i = 2^n; \\ \frac{2^n}{2^{n+1}}\,\mathrm{SD}(t_n, (\overline{t_n t_n})[i-2^n\,..\,i-1]) + \frac{2^n}{2^{n+1}}\,\mathrm{SD}(\overline{t_n}, (\overline{t_n}t_n)[i-2^n\,..\,i-1]), & \text{if } i \in [2^n+1, 2^{n+1}-1], \end{cases} \\
&\in \begin{cases} \frac{1}{2}\{\frac{1}{2}\} + \frac{1}{2}\{0\}, & \text{if } i = 2^{n-1};\ (\text{by (a)}, \overline{(}\text{b}\overline{)}) \\ \frac{1}{2}[\frac{1}{4}, \frac{3}{4}] + \frac{1}{2}[\frac{1}{4}, \frac{3}{4}], & \text{if } i \in [1, 2^n-1] \setminus \{2^{n-1}\};\ (\text{by (a)}, \overline{(}\text{b}\overline{)}) \\ \frac{1}{2}\{0\} + \frac{1}{2}\{1\}, & \text{if } i = 2^n; \\ \frac{1}{2}\{1\} + \frac{1}{2}\{\frac{1}{2}\}, & \text{if } i = 2^n + 2^{n-1};\ (\text{by (b}\overline{)}, \overline{(}\text{a}\overline{)}) \\ \frac{1}{2}[\frac{1}{4}, \frac{3}{4}] + \frac{1}{2}[\frac{1}{4}, \frac{3}{4}], & \text{if } i \in [2^n+1, 2^{n+1}-1] \setminus \{2^n + 2^{n-1}\},\ (\text{by (b}\overline{)}, \overline{(}\text{a}\overline{)}) \end{cases}
\end{aligned}$$

$$= \begin{cases} \{\frac{1}{4}\}, & \text{if } i = 2^{n-1}; \\ [\frac{1}{4}, \frac{3}{4}], & \text{if } i \in [1, 2^n - 1] \setminus \{2^{n-1}\}; \\ \{\frac{1}{2}\}, & \text{if } i = 2^n; \\ \{\frac{3}{4}\}, & \text{if } i = 2^n + 2^{n-1}; \\ [\frac{1}{4}, \frac{3}{4}], & \text{if } i \in [2^n + 1, 2^{n+1} - 1] \setminus \{2^n + 2^{n-1}\}, \end{cases}$$

$$\begin{aligned} &\mathrm{SD}(t_{n+1}, t_{n+1}^2[i\,..\,2^{n+1} + i - 1]) \\ &= \mathrm{SD}(t_n\overline{t_n}, (t_n\overline{t_n}t_n\overline{t_n})[i\,..\,2^{n+1} + i - 1]) \\ &= \begin{cases} \mathrm{SD}(t_n\overline{t_n}, (t_n\overline{t_n}t_n)[i\,..\,2^{n+1} + i - 1]), & \text{if } i \in [1, 2^n - 1]; \\ \mathrm{SD}(t_n\overline{t_n}, \overline{t_n}t_n), & \text{if } i = 2^n; \\ \mathrm{SD}(t_n\overline{t_n}, (\overline{t_n}t_n\overline{t_n})[i - 2^n\,..\,2^n + i - 1]), & \text{if } i \in [2^n + 1, 2^{n+1} - 1], \end{cases} \\ &= \begin{cases} \frac{2^n}{2^{n+1}}\,\mathrm{SD}(t_n, (t_n\overline{t_n})[i\,..\,2^n + i - 1]) + \frac{2^n}{2^{n+1}}\,\mathrm{SD}(\overline{t_n}, (\overline{t_n}t_n)[i\,..\,2^n + i - 1]), & \text{if } i \in [1, 2^n - 1]; \\ \frac{2^n}{2^{n+1}}\,\mathrm{SD}(t_n, \overline{t_n}) + \frac{2^n}{2^{n+1}}\,\mathrm{SD}(\overline{t_n}, t_n), & \text{if } i = 2^n; \\ \frac{2^n}{2^{n+1}}\,\mathrm{SD}(t_n, (\overline{t_n}t_n)[i - 2^n\,..\,i - 1]) + \frac{2^n}{2^{n+1}}\,\mathrm{SD}(\overline{t_n}, (t_n\overline{t_n})[i - 2^n\,..\,i - 1]), & \text{if } i \in [2^n + 1, 2^{n+1} - 1], \end{cases} \\ &\in \begin{cases} \frac{1}{2}\{\frac{1}{2}\} + \frac{1}{2}\{\frac{1}{2}\}, & \text{if } i = 2^{n-1};\ (\text{by (a)}, \overline{\text{(a)}}) \\ \frac{1}{2}[\frac{1}{4}, \frac{3}{4}] + \frac{1}{2}[\frac{1}{4}, \frac{3}{4}], & \text{if } i \in [1, 2^n - 1] \setminus \{2^{n-1}\};\ (\text{by (a)}, \overline{\text{(a)}}) \\ \frac{1}{2}\{0\} + \frac{1}{2}\{0\}, & \text{if } i = 2^n; \\ \frac{1}{2}\{\frac{1}{2}\} + \frac{1}{2}\{\frac{1}{2}\}, & \text{if } i = 2^n + 2^{n-1};\ (\text{by } \overline{\text{(a)}}, \text{(a)}) \\ \frac{1}{2}[\frac{1}{4}, \frac{3}{4}] + \frac{1}{2}[\frac{1}{4}, \frac{3}{4}], & \text{if } i \in [2^n + 1, 2^{n+1} - 1] \setminus \{2^n + 2^{n-1}\},\ (\text{by } \overline{\text{(a)}}, \text{(a)}) \end{cases} \\ &= \begin{cases} \{\frac{1}{2}\}, & \text{if } i = 2^{n-1}; \\ [\frac{1}{4}, \frac{3}{4}], & \text{if } i \in [1, 2^n - 1] \setminus \{2^{n-1}\}; \\ 0, & \text{if } i = 2^n; \\ \{\frac{1}{2}\}, & \text{if } i = 2^n + 2^{n-1}; \\ [\frac{1}{4}, \frac{3}{4}], & \text{if } i \in [2^n + 1, 2^{n+1} - 1] \setminus \{2^n + 2^{n-1}\}, \end{cases} \end{aligned}$$

hence proving (a) and (b) also hold for $n+1$. By Observation 11, the remaining six cases also hold for $n+1$. ∎

Corollary 21. *For all* $n \in \mathbb{N}$*,* $i \in [0, 2^n - 1]$ *with* $\gcd(i, 2^n) \le 2^{n-2}$*, and* $x, y_0, y_1 \in \{t_n, \overline{t_n}\}$*,*

$$\mathrm{SD}(x, (y_0y_1)[i\,..\,i + 2^n - 1]) \in [\tfrac{1}{4}, \tfrac{3}{4}].$$

Proof. Follows immediately from Lemma 20. ∎

Corollary 22. *For all* $n, i \in \mathbb{N}$ *with* $\gcd(i, 2^n) \le 2^{n-2}$ *and* $\mathbf{x}, \mathbf{y} \in \{t_n, \overline{t_n}\}^\omega$*,*

$$\frac{1}{4} \le \mathrm{LSD}(\mathbf{x}, \mathbf{y}[i\,..\,\infty]) \le \mathrm{USD}(\mathbf{x}, \mathbf{y}[i\,..\,\infty]) \le \frac{3}{4}.$$

Proof. Note that for any $j \in \mathbb{N}$, $\gcd(i + j \cdot 2^n, 2^n) = \gcd(i, 2^n) \le 2^{n-2}$. Also for any $j \in \mathbb{N}$, since $\mathbf{x}, \mathbf{y} \in \{t_n, \overline{t_n}\}^\omega$ and $|t_n| = |\overline{t_n}| = 2^n$, we have $\mathbf{x}[2^n j\,..\,2^n(j+1) - 1] \in \{t_n, \overline{t_n}\}$ and $\mathbf{y}[i + 2^n j\,..\,i + 2^n(j+1) - 1] = (y_0y_1)[(i \bmod 2^n)\,..\,(i \bmod 2^n) + 2^n - 1]$ for some $y_0, y_1 \in \{t_n, \overline{t_n}\}$. Hence, for any $j \in \mathbb{N}$, by Corollary 21,

$$\mathrm{SD}(\mathbf{x}[2^n j\,..\,2^n(j+1) - 1], \mathbf{y}[i + 2^n j\,..\,i + 2^n(j+1) - 1]) \in \left[\frac{1}{4}, \frac{3}{4}\right],$$

whence by Observation 10,

$$\mathrm{SD}(\mathbf{x}[0\,.\,.\,2^n(j+1)-1],\mathbf{y}[i\,.\,.\,i+2^n(j+1)-1]) \in \left[\frac{1}{4},\frac{3}{4}\right],$$

whence by Observation 14,

$$\begin{aligned}(\mathrm{LSD},\mathrm{USD})(\mathbf{x},\mathbf{y}[i\,.\,.\,\infty]) &= \left(\liminf_{j\to\infty},\limsup_{j\to\infty}\right)\mathrm{SD}(\mathbf{x}[0\,.\,.\,2^n j-1],\mathbf{y}[i\,.\,.\,i+2^n j-1]) \\ &\in \left(\left[\frac{1}{4},\frac{3}{4}\right],\left[\frac{1}{4},\frac{3}{4}\right]\right). \qquad \blacksquare\end{aligned}$$

Corollary 23. *For all $i \in \mathbb{N}\setminus\{0\}$, $\frac{1}{4} \le \mathrm{LSD}(\mathbf{t},\mathbf{t}[i\,.\,.\,\infty]) \le \mathrm{USD}(\mathbf{t},\mathbf{t}[i\,.\,.\,\infty]) \le \frac{3}{4}$.*

Proof. Since $i > 0$, we have $4\max_{m\in\mathbb{N}}\gcd(i,2^m) = 2^n$ for some $n \in \mathbb{N}$. Note that $\gcd(i,2^n) = 2^{n-2}$. Also note that $\mathbf{t} = \mu^n(\mathbf{t}) \in \{t_n,\overline{t_n}\}^\omega$. Hence, by Corollary 22,

$$\frac{1}{4} \le \mathrm{LSD}(\mathbf{t},\mathbf{t}[i\,.\,.\,\infty]) \le \mathrm{USD}(\mathbf{t},\mathbf{t}[i\,.\,.\,\infty]) \le \frac{3}{4}. \qquad \blacksquare$$

We now have all the tools needed to prove Theorem 18.

Proof of Theorem 18. Let $\mathbf{w} \in \Sigma_2^\omega \setminus \{\mathbf{t},\overline{\mathbf{t}}\}$. By Theorem 16, there exists $\mathbf{x} \in \Sigma_5^\omega$ that encodes a valid path through the Fife automaton for overlap-free infinite binary words such that $\mathrm{FBE}(\mathbf{x}) = \mathbf{w}$ or $\mathrm{FBE}(\mathbf{x},a) = \mathbf{w}$ for some $a \in \Sigma_2$. From inspection of the Fife automaton for overlap-free infinite binary words, we see that $\mathbf{x}$ must fall into one of the following four cases.

(1) $\mathbf{x}$ ends in $\mathtt{0}^\omega$.

(2) $\mathbf{x}$ does not end in $\mathtt{0}^\omega$, begins with $\mathtt{0}^n\mathtt{2}$ or $\mathtt{0}^n\mathtt{4}$ for some $n \in \mathbb{N}$, and contains exactly n $\mathtt{0}$s.

(3) $\mathbf{x}$ does not end in $\mathtt{0}^\omega$, begins with $\mathtt{0}^n\mathtt{2}$ or $\mathtt{0}^n\mathtt{4}$ for some $n \in \mathbb{N}$, and contains more than n $\mathtt{0}$s.

(4) $\mathbf{x}$ does not end in $\mathtt{0}^\omega$ and begins with $\mathtt{0}^n\mathtt{1}$ or $\mathtt{0}^n\mathtt{3}$ for some $n \in \mathbb{N}$.

Case 1: $\mathbf{w}$ ends in either $\mathbf{t}$ or $\overline{\mathbf{t}}$, so since $\mathbf{w} \notin \{\mathbf{t},\overline{\mathbf{t}}\}$, it follows that $\mathbf{w} \in \{z\mathbf{t},z\overline{\mathbf{t}}\}$ for some $z \in \Sigma_2^+$. By Observation 13, we have

$$(\mathrm{LSD},\mathrm{USD})(\mathbf{w},\mathbf{t}) \in \{(\mathrm{LSD},\mathrm{USD})(\mathbf{t},\mathbf{t}[|z|\,.\,.\,\infty]),(\mathrm{LSD},\mathrm{USD})(\overline{\mathbf{t}},\mathbf{t}[|z|\,.\,.\,\infty])\},$$

whence by Corollary 23 and Corollary 12, we obtain $(\mathrm{LSD},\mathrm{USD})(\mathbf{w},\mathbf{t}) \in (\{[\frac{1}{4},\frac{3}{4}],[1-\frac{3}{4},1-\frac{1}{4}]\},\{[\frac{1}{4},\frac{3}{4}],[1-\frac{3}{4},1-\frac{1}{4}]\}) = ([\frac{1}{4},\frac{3}{4}],[\frac{1}{4},\frac{3}{4}])$, as desired.

Case 2: From inspection of the Fife automaton for overlap-free infinite binary words, we see that $\mathbf{x} \in \{\mathtt{0}^n\{\mathtt{2}(\mathtt{31})^\omega,\mathtt{4}(\mathtt{13})^\omega\} : n \in \mathbb{N}\}$. Note that $\mathrm{FBE}(\mathtt{0}^n\mathtt{4}(\mathtt{13})^\omega) = \overline{\mathrm{FBE}(\mathtt{0}^n\mathtt{2}(\mathtt{31})^\omega)}$. Hence, by Proposition 19 and Corollary 12, we obtain $(\mathrm{LSD},\mathrm{USD})(\mathbf{w},\mathbf{t}) \in \{(\frac{1}{3},\frac{2}{3}),(1-\frac{2}{3},1-\frac{1}{3})\} = \{(\frac{1}{3},\frac{2}{3})\} \subset ([\frac{1}{4},\frac{3}{4}],[\frac{1}{4},\frac{3}{4}])$, as desired.

Case 3: From inspection of the Fife automaton for overlap-free infinite binary words, we see that $\mathbf{x} \in \{\mathtt{0}^n\{\mathtt{2}(\mathtt{31})^{\frac{m}{2}},\mathtt{4}(\mathtt{13})^{\frac{m}{2}}\}\mathtt{0}\{\mathtt{1},\mathtt{3}\}\mathbf{y} : n,m \in \mathbb{N},\mathbf{y} \in \{\mathtt{0},\mathtt{1},\mathtt{3}\}^\omega\}$, whence

$$\begin{aligned}\mathbf{w} &\in \Sigma_2^\omega \cap \left(\bigcup_{n,m\in\mathbb{N}} \Sigma_2^{2^{n+m+1}}\{t_{n+m+2},\overline{t_{n+m+2}}\}\prod_{k=n+m+3}^{\infty}\{\varepsilon,t_k,\overline{t_k}\}\right) \\ &\subseteq \bigcup_{n,m\in\mathbb{N}} \Sigma_2^{2^{n+m+1}}\{t_{n+m+2},\overline{t_{n+m+2}}\}\{t_{n+m+3},\overline{t_{n+m+3}}\}^\omega,\end{aligned}$$

so there is a $k \in \mathbb{N}$ such that $\mathbf{w}[2^k \,..\, \infty] \in \{t_{k+1}, \overline{t_{k+1}}\}\{t_{k+2}, \overline{t_{k+2}}\}^\omega$. By Observation 13 and Corollary 22, we obtain

$$\begin{aligned}
(\mathrm{LSD}, \mathrm{USD})(\mathbf{t}, \mathbf{w}) &= (\mathrm{LSD}, \mathrm{USD})(\mathbf{t}[2^{k+2} \,..\, \infty], \mathbf{w}[2^{k+2} \,..\, \infty]) \\
&= (\mathrm{LSD}, \mathrm{USD})(\underbrace{\mathbf{t}[2^{k+2} \,..\, \infty]}_{\in \{t_{k+2}, \overline{t_{k+2}}\}^\omega}, \underbrace{(\mathbf{w}[3 \cdot 2^k \,..\, \infty]}_{\in \{t_{k+2}, \overline{t_{k+2}}\}^\omega})[2^k \,..\, \infty]) \\
&\in \left(\left[\frac{1}{4}, \frac{3}{4} \right], \left[\frac{1}{4}, \frac{3}{4} \right] \right),
\end{aligned}$$

as desired.

Case 4: From inspection of the Fife automaton for overlap-free infinite binary words, we see that $\mathbf{x} \in \{\mathtt{0}^n\{\mathtt{1},\mathtt{3}\}\mathtt{0}^m\{\mathtt{1},\mathtt{3}\}\mathbf{y} \,:\, n, m \in \mathbb{N}, \mathbf{y} \in \{\mathtt{0},\mathtt{1},\mathtt{3}\}^\omega\}$, whence

$$\begin{aligned}
\mathbf{w} &\in \Sigma_2^\omega \cap \left(\bigcup_{n,m \in \mathbb{N}} \{t_n, \overline{t_n}\}\{t_{n+m+1}, \overline{t_{n+m+1}}\} \prod_{k=n+m+2}^{\infty} \{\varepsilon, t_k, \overline{t_k}\} \right) \\
&\subseteq \bigcup_{n,m \in \mathbb{N}} \{t_n, \overline{t_n}\}\{t_{n+m+1}, \overline{t_{n+m+1}}\}\{t_{n+m+2}, \overline{t_{n+m+2}}\}^\omega,
\end{aligned}$$

so there are $k, l \in \mathbb{N}$ such that $\mathbf{w} \in \{t_k, \overline{t_k}\}\{t_{k+l+1}, \overline{t_{k+l+1}}\}\{t_{k+l+2}, \overline{t_{k+l+2}}\}^\omega$. By Observation 13 and Corollary 22, we obtain

$$\begin{aligned}
(\mathrm{LSD}, \mathrm{USD})(\mathbf{t}, \mathbf{w}) &= (\mathrm{LSD}, \mathrm{USD})(\mathbf{t}[2^{k+l+2} \,..\, \infty], \mathbf{w}[2^{k+l+2} \,..\, \infty]) \\
&= (\mathrm{LSD}, \mathrm{USD})(\underbrace{\mathbf{t}[2^{k+l+2} \,..\, \infty]}_{\in \{t_{k+l+2}, \overline{t_{k+l+2}}\}^\omega}, \underbrace{(\mathbf{w}[2^k + 2^{k+l+1} \,..\, \infty])}_{\in \{t_{k+l+2}, \overline{t_{k+l+2}}\}^\omega}[2^{k+l+1} - 2^k \,..\, \infty]) \\
&\in \left(\left[\frac{1}{4}, \frac{3}{4} \right], \left[\frac{1}{4}, \frac{3}{4} \right] \right),
\end{aligned}$$

as desired. ■

6 Future work

Using the Fife automaton for overlap-free infinite binary words, we computed similarity densities of long prefixes of all overlap-free infinite binary words (up to a certain length) with prefixes of $\mathbf{t}$. Inspection of the compuation results immediately suggests the following improvement to Theorem 18.

Conjecture 24. For all overlap-free $\mathbf{w} \in \Sigma_2^\omega \setminus \{\mathbf{t}, \bar{\mathbf{t}}\}$, we have $\frac{1}{3} \leq \mathrm{LSD}(\mathbf{w}, \mathbf{t}) \leq \mathrm{USD}(\mathbf{w}, \mathbf{t}) \leq \frac{2}{3}$.

Note that the bounds in Conjecture 24 are tight due to Proposition 17. Computational evidence also suggests that these bounds are also tight for many other overlap-free infinite binary words.

However, Conjecture 24 cannot be proved just by using the technique we used to prove Theorem 18. This is because the bounds in Lemma 20 (and, more transparently, Corollary 21) are tight. For example, $\mathrm{SD}(t_2, t_3[1 \,..\, 4]) = \mathrm{SD}(\mathtt{0110}, \mathtt{1101}) = \frac{1}{4}$. More generally, for any $n \in \mathbb{N}$, we have $\mathrm{SD}(t_{n+2}, t_{n+3}[2^n \,..\, 2^{n+2} + 2^n - 1]) = \frac{1}{4}$.

On the other hand, our proof of Theorem 18 never used the overlap-free property directly; we merely used it indirectly via the Fife automaton. As such, our proof of Theorem 18 works for all images of FBE provided the argument to FBE is of the form required for one of the four cases presented in the proof, regardless of whether the resulting word is overlap-free. Namely, we have the following more general, but much more cumbersome, theorem.

Theorem 25. *For all* $\mathbf{x} \in \{1,2,3,4\}\Sigma_5^* 0^\omega \cup 0^*\{2(31),4(13)\}^\omega$

$$\cup\, 0^*\{2(31)^*\{\varepsilon,3\},4(13)^*\{\varepsilon,1\}\}0\{1,3\}\{0,1,3\}^\omega \cup (0^*\{1,3\})^2\{0,1,3\}^\omega$$

and

$$\mathbf{w} \in \begin{cases} \{\mathrm{FBE}(\mathbf{x},0),\mathrm{FBE}(\mathbf{x},1)\}, & \textit{if } \mathbf{x} \textit{ ends in } 0^\omega; \\ \{\mathrm{FBE}(\mathbf{x})\}, & \textit{otherwise}, \end{cases}$$

we have

$$\frac{1}{4} \le \mathrm{LSD}(\mathbf{w},\mathbf{t}) \le \mathrm{USD}(\mathbf{w},\mathbf{t}) \le \frac{3}{4}.$$

Note that Theorem 25 is indeed more general than Theorem 18, since, for example, 13^ω is not a valid path in the Fife automaton for overlap-free infinite binary words (indeed, $\mathrm{FBE}(13^\omega)$ begins with the overlap 01010) and $\mathrm{FBE}(13^\omega)$ also is not just a shift of $\mathbf{t}$ or $\bar{\mathbf{t}}$, but Theorem 25 nevertheless implies that $\frac{1}{4} \le \mathrm{LSD}(\mathrm{FBE}(13^\omega),\mathbf{t}) \le \mathrm{USD}(\mathrm{FBE}(13^\omega),\mathbf{t}) \le \frac{3}{4}$.

Together, Conjecture 24 and Theorem 25 suggest the following more general question.

Question 26. For each $n \in \mathbb{N} \setminus \{0,1\}$, $r,s \in [0,1]$, and $\mathbf{x} \in \Sigma_n^\omega$, let

$$S_{n,r,s}(\mathbf{x}) := \{\mathbf{y} \in \Sigma_n^\omega \,:\, r \le \mathrm{LSD}(\mathbf{x},\mathbf{y}) \le \mathrm{USD}(\mathbf{x},\mathbf{y}) \le s\}.$$

What are $S_{2,\frac{1}{4},\frac{3}{4}}(\mathbf{t})$ and $S_{2,\frac{1}{3},\frac{2}{3}}(\mathbf{t})$?

Another avenue of investigation is to consider what makes $\mathbf{t}$ so special in the sense of Theorem 18. As mentioned in the introduction, Theorem 18 is false if we replace $\mathbf{t}$ with an arbitrary overlap-free infinite binary word. However, perhaps there are specific words other than $\mathbf{t}$ and $\bar{\mathbf{t}}$ that do share similar properties. In other words, we raise the following question.

Question 27. Let $\mathscr{O}$ denote the set of all overlap-free infinite binary words.
What is $\{\mathbf{x} \in \Sigma_2^\omega \,:\, \mathscr{O} \subseteq S_{2,\frac{1}{4},\frac{3}{4}}(\mathbf{x})\}$? What if we replace $\frac{1}{4}, \frac{3}{4}$ with $\frac{1}{3}, \frac{2}{3}$?

A third avenue of investigation is to consider what occurs in words that avoid higher powers in place of being overlap-free (which are essentially $(2+\varepsilon)$- or 2^+-powers). In fact, there is a Fife automaton characterizing $\frac{7}{3}$-power-free infinite binary words having the same encoding mechanism as the Fife automaton for overlap-free infinite binary words but with more states and different transitions [3, 9]. However, initial inspection of the automaton for $\frac{7}{3}$-power-free infinite binary words suggests that our proof of Theorem 18 cannot be extended to account for all $\frac{7}{3}$-power-free infinite binary words because there are many more edges labeled 2 and 4 in the Fife automaton for $\frac{7}{3}$-power-free infinite binary words, resulting in valid paths that contain infinitely many 2s and 4s, but our proof of Theorem 18 heavily relied on there being at most one occurrence of 2 or 4 (which must be preceeded by a string of 0s if it occurs) in the path taken through the automaton so that the infinite binary word corresponding to the path eventually "lags behind" the prefixes t_n of $\mathbf{t}$ in the sense that each successive n^{th} symbol in the path can only generate positions prior to 2^n. Nevertheless, computational evidence suggests that Theorem 18 and even Conjecture 24 can be generalized even further.

Conjecture 28. For all $\frac{7}{3}$-power-free $\mathbf{w} \in \Sigma_2^\omega \setminus \{\mathbf{t},\bar{\mathbf{t}}\}$, $\frac{1}{3} \le \mathrm{LSD}(\mathbf{w},\mathbf{t}) \le \mathrm{USD}(\mathbf{w},\mathbf{t}) \le \frac{2}{3}$.

Finally, we revisit the notion, already mentioned in Remark 7, that LSD and USD are not new ideas, and not just in number theory. In fact, $1-\mathrm{LSD}$ is a pseudometric on $\Sigma^{\mathbb{N}}$, called the Besicovitch pseudometric, which has already been studied from the perspective of discrete dynamical systems such as

[2]. Also studied in [2] is the Weyl pseudometric, which suggests the following slightly different notion of similarity density, considering all blocks of a given size instead of just blocks from the beginning.

$$\begin{aligned}\mathrm{LSD}_{\mathrm{Weyl}}(\mathbf{x},\mathbf{y}) &= \liminf_{n\to\infty}\ \inf_{k\in\mathbb{N}} \mathrm{SD}(\mathbf{x}[k..k+n-1],\mathbf{y}[k..k+n-1]),\\ \mathrm{USD}_{\mathrm{Weyl}}(\mathbf{x},\mathbf{y}) &= \limsup_{n\to\infty}\ \sup_{k\in\mathbb{N}} \mathrm{SD}(\mathbf{x}[k..k+n-1],\mathbf{y}[k..k+n-1]).\end{aligned}$$

With this notion of Weyl similarity density, analogous to the Besicovitch case, we have that $1-\mathrm{LSD}_{\mathrm{Weyl}}$ is the Weyl pseudometric. The Besicovitch and Weyl pseudometrics share some topological properties, but the Besicovitch pseudometric is complete while the Weyl pseudometric is not [2]. This fact suggests one might be able to shed further light on some of the questions above by also considering the Weyl similarity density; perhaps several different notions of similarity density, when taken together, can characterize the overlap-free infinite binary words.

Acknowledgments. We are grateful to Chao Hsien Lin for having suggested the question we study here. We thank the referees for a careful reading, and Joel Ouaknine and Stefan Kiefer for having pointed out the paper [2].

References

[1] J.-P. Allouche & J. Shallit (1999): *The ubiquitous Prouhet-Thue-Morse sequence*. In C. Ding, T. Helleseth & H. Niederreiter, editors: *Sequences and Their Applications, Proceedings of SETA '98*, Springer-Verlag, pp. 1–16, doi:10.1007/978-1-4471-0551-0_1.

[2] F. Blanchard, E. Formenti & P. Kůrka (1997): *Cellular automata in the Cantor, Besicovitch, and Weyl topological spaces*. *Complex Systems* 11, pp. 107–123.

[3] V. D. Blondel, J. Cassaigne & R. M. Jungers (2009): *On the number of α-power-free binary words for $2 < \alpha \leq 7/3$*. *Theoret. Comput. Sci.* 410, pp. 2823–2833, doi:10.1016/j.tcs.2009.01.031.

[4] S. Brown, N. Rampersad, J. Shallit & T. Vasiga (2006): *Squares and overlaps in the Thue-Morse sequence and some variants*. *RAIRO Inform. Théor. App.* 40, pp. 473–484, doi:10.1051/ita:2006030.

[5] E. D. Fife (1980): *Binary sequences which contain no BBb*. *Trans. Amer. Math. Soc.* 261, pp. 115–136, doi:10.1090/S0002-9947-1980-0576867-5.

[6] E. Grant, J. Shallit & T. Stoll (2009): *Bounds for the discrete correlation of infinite sequences on k symbols and generalized Rudin-Shapiro sequences*. *Acta Arith.* 140, pp. 345–368, doi:10.4064/aa140-4-5.

[7] K. Mahler (1927): *On the translation properties of a simple class of arithmetical functions*. *J. Math. and Phys.* 6, pp. 158–163.

[8] P. Ochem, N. Rampersad & J. Shallit (2008): *Avoiding approximate squares*. *Internat. J. Found. Comp. Sci.* 19, pp. 633–648, doi:10.1142/S0129054108005863.

[9] N. Rampersad, J. Shallit & A. Shur (2011): *Fife's theorem for (7/3)-powers*. In P. Ambroz, S. Holub & Z. Masakova, editors: *WORDS 2011, 8th International Conference*, pp. 189–198, doi:10.4204/EPTCS.63.25.

[10] J. Shallit (2011): *Fife's theorem revisited*. In G. Mauri & A. Leporati, editors: *Developments in Language Theory*, *Lecture Notes in Computer Science* 6795, Springer-Verlag, pp. 397–405, doi:10.1007/978-3-642-22321-1_34.

[11] R. Yarlagadda & J. E. Hershey (1984): *Spectral properties of the Thue-Morse sequence*. *IEEE Trans. Commun.* 32, pp. 974–977, doi:10.1109/TCOM.1984.1096162.

[12] R. Yarlagadda & J. E. Hershey (1990): *Autocorrelation properties of the Thue-Morse sequence and their use in synchronization*. *IEEE Trans. Commun.* 38, pp. 2099–2102, doi:10.1109/26.64649.