MONOCHROMATIC ARITHMETIC PROGRESSIONS IN AUTOMATIC SEQUENCES WITH GROUP STRUCTURE

IBAI AEDO, UWE GRIMM, NEIL MAÑIBO, YASUSHI NAGAI, AND PETRA STAYNOVA

ABSTRACT. We determine asymptotic growth rates for lengths of monochromatic arithmetic progressions in certain automatic sequences. In particular, we look at (one-sided) fixed points of aperiodic, primitive, bijective substitutions and spin substitutions, which are generalisations of the Thue–Morse and Rudin–Shapiro substitutions, respectively. For such infinite words, we show that there exists a subsequence $\{d_n\}$ of differences along which the maximum length $A(d_n)$ of a monochromatic arithmetic progression (with fixed difference d_n) grows at least polynomially in d_n. Explicit upper and lower bounds for the growth exponent can be derived from a finite group associated to the substitution. As an application, we obtain bounds for a van der Waerden-type number for a class of colourings parametrised by the size of the alphabet and the length of the substitution.

"Here the Maestro laid down his pen."
This paper is dedicated to our late friend and colleague, Uwe Grimm.

1. INTRODUCTION

The study of Ramsey-type properties of morphic words has a long history, spanning from the classic theorem of Graham and Rothschild [25] to more recent advances such as antipowers [17] and monochromatic factorisations [32, 42] in infinite words. A subset of Ramsey-type properties which has also gathered interest is a consideration of the arithmetic subsequences of automatic or morphic words. In [5], Avgustinovich and Frid show that any binary word occurs as an arithmetic subsequence of the Thue–Morse sequence (or more generally, a fixed point of any primitive bijective binary constant-length substitution), and go on to investigate properties of the arithmetic complexity of certain words over arbitrary finite alphabets. If we instead consider monochromatic arithmetic subsequences of a given substitutive word and fix the difference of the arithmetic progressions, we note that the length is bounded in the case of the Thue–Morse and more general Thue–Morse-like sequences, as shown, respectively, in [33] and, by the present authors, in [2]; see also [27, 28] for results regarding arithmetic progressions in model sets.

While van der Waerden's theorem ensures the existence of arbitrarily long arithmetic progressions within any finite colouring of $\mathbb{N}$, it does not immediately provide an estimate for the initial segment within which these can be found. The van der Waerden numbers were initially introduced to study this, and are defined as the minimal initial segment of the integers such that any colouring with n colours will give an arithmetic progression of length ℓ. Only a handful of van der Waerden numbers are known, and Gowers [24] gives hyper-exponential upper bounds for the rest.

Here, we pursue the line of enquiry that expands on results by Frid et al [5, 6, 21], Parshina [35–37], and the present authors [2, 34]. We focus on fixed points of constant-length substitutions over finite alphabets, and their images under codings.

Let w be a fixed point of a substitution ϱ over a finite alphabet $\mathcal{A}$. Fix a difference $d \geqslant 1$. We denote by $A_w(d)$ the maximum length of a monochromatic arithmetic progression of difference d which occurs in w. When the context is clear, we make w implicit and just refer to $A(d)$. We restrict to classes of substitutions which possess an explicit group structure, which provides direct access to bounding $A(d)$ for some specific values of d and allows some asymptotic estimates. This work expands the results in [2, 36, 37] to a more general setting, giving upper bounds of $A(d)$ for a larger class of automatic sequences over arbitrary finite alphabets. The following result establishes an asymptotic lower bound for $A(d)$ along a subsequence of differences for bijective substitutions.

Theorem 1. *Let ϱ be an aperiodic primitive bijective constant-length substitution on a finite alphabet $\mathcal{A}$ and let $w \in \mathcal{A}^{\mathbb{N}}$ be a one-sided fixed point of ϱ. There exists an increasing sequence $\{d_n\} \subset \mathbb{N}$ such that $A(d_n) \gtrsim d_n^{\alpha}$, for some $0 < \alpha \leqslant 1$.*

We introduce the notion of *g-palindromicity*, and show that this implies $\alpha = 1$ (i.e., there is a subsequence where $A(d)$ grows linearly); see Proposition 12. In particular, this holds for the family of cylic Thue–Morse substitutions on L letters; see Section 3.4.1 below. We also deal with non-bijective substitutions with a *supersubstitution* structure and show that they satisfy bounds similar to those for bijective substitutions. For Vandermonde substitutions, which can be constructed from Vandermonde matrices, one has the following result.

Theorem 2. *Let ϱ be a constant-length spin substitution arising from a Vandermonde matrix and let $w = \pi_G(v) \in \mathcal{A}^{\mathbb{N}}$ be the spin coding of a fixed point v of ϱ. There exists an increasing sequence $\{d_n\} \subset \mathbb{N}$ such that $A(d_n) \gtrsim d_n^{\alpha}/L$, for some $0 < \alpha \leqslant 1$.*

Under some mild assumptions on the substitution, one can also obtain upper bounds for $A(d)$. In particular, sufficient conditions for $A(d) < \infty$ to hold are given in Propositions 6 and 41. For a subclass of bijective substitutions, a d-dependent computable upper bounds are given in Corollary 26 and Proposotion 27.

The bounds used to prove Theorem 1 only depends on the size c of the alphabet and the length L of the substitution. This allows one to associate a *van der Waerden-type constant* $W(\mathcal{B}(c, L), M)$ to a family of bijective substitutions sharing these same attributes; see Proposition 21 for an general upper bound and Corollary 28 for lower bounds.

The paper is organised as follows. In Section 2, we provide some basic notions on combinatorics on words and substitutions. In Section 3, we focus on bijective substitutions, extending results on the lower bounds in [2, 36, 37] to this family. Note that this includes the class of symmetric morphisms in [22] and group substitutions in [23]. We prove Theorem 1 in Section 3.1. In Section 3.2, we develop the notion of a van der Waerden type-constant for bijective substitution and provide explicit upper bounds. We use recurrence properties for substitutive words and the results obtained in Section 3.1 to compute these bounds. Section 3.3 deals with upper bounds for $A(d)$ for bijective substitutions with additional properties. Poignant examples, including the family of cyclic Thue–Morse substitutions, are given in Section 3.4. We show that some of the results can be partially extended to the non-bijective case in Section 3.5 for substitutions admitting a supersubstitution structure. In Section 4, we deal with spin substitutions. We begin

with the case of the Rudin–Shapiro substitution in Sections 4.1 and 4.2 where we derive lower bounds using two combinatorial approaches, namely via the spin matrix and via the stagerred substitution approach. At this point, we would like to mention that the same bounds have been found for the Rudin–Shapiro sequence in [39] using a different method. The novelty of the approach we use in this work is that it extends to other automatic sequences derived from other spin matrices. We carry this out in Section 4.3 in the case where the relevant matrix is an $L \times L$-Vandermonde or discrete Fourier transform (DFT) matrix, where we prove Theorem 2. Finally, in Section 5, we end with some open questions and illustrate potential ways of extending our results to the non-bijective constant length case through a concrete example.

2. Preliminaries

2.1. **Combinatorics on words and substitutions.** Throughout this work, an *alphabet* $\mathcal{A}$ will be a finite collection of symbols called *letters*. We denote by $\mathcal{A}^*$ and $\mathcal{A}^+$ the sets of all *finite words* and *non-empty finite words* over $\mathcal{A}$, respectively. The sets of all *one-sided* and *two-sided* *infinite words* over $\mathcal{A}$ are denoted by $\mathcal{A}^\mathbb{N}$ and $\mathcal{A}^\mathbb{Z}$, respectively. These are also called *sequences* over $\mathcal{A}$. Here $\mathbb{Z}$ is the set of integers and $\mathbb{N} = \mathbb{Z}_{\geqslant 0}$.

The *length* of a word $w \in \mathcal{A}^+$ is denoted by $|w|$. For each $0 \leqslant i < |w|$, we write w_i to denote the ith letter of w. A *subword* of w is a word of the form $w_i w_{i+1} \cdots w_j$, for some $0 \leqslant i \leqslant j < |w|$. A *substitution* ϱ on $\mathcal{A}$ is a map $\varrho \colon \mathcal{A} \to \mathcal{A}^+$, which extends to a map on $\mathcal{A}^+$ by concatenation. This allows one to define ϱ^n inductively via $\varrho^n(a) := \varrho^{n-1}(\varrho(a))$ for $n \geqslant 2$. We call $\varrho^n(a)$ a *level-n superword* of type a. We say that ϱ is *primitive* if there exists an $n \in \mathbb{N}$ such that, for every $a \in \mathcal{A}$, the word $\varrho^n(a)$ contains all the letters in $\mathcal{A}$. A word $w \in \mathcal{A}^+$ is *legal* with respect to ϱ if there exists $n \in \mathbb{N}$ and a letter a such that w is a subword of $\varrho^n(a)$. We denote by $\mathcal{L}_n(\varrho)$ the set of all legal words of length n. The set $\mathcal{L}(\varrho) := \bigcup_{n \geqslant 1} \mathcal{L}_n(\varrho)$ of all legal words is called the *language* of ϱ.

If there exists an $L \in \mathbb{N}$ such that, for all $a \in \mathcal{A}$, the word $\varrho(a)$ has length L, we say that ϱ is a substitution of *constant length* L. Such maps are also called *uniform morphisms*. For each $0 \leqslant i \leqslant L-1$ and $a \in \mathcal{A}$, we define $\varrho_i(a)$ to be the ith letter of $\varrho(a)$. We call the map $\varrho_i \colon \mathcal{A} \to \mathcal{A}$ the ith *column* of ϱ. If ϱ_i is a bijection of $\mathcal{A}$, we call it a *bijective* column. If $\varrho_i(a) = b$ for some fixed $b \in \mathcal{A}$ for all $a \in \mathcal{A}$, we call ϱ_i a *coincidence*. If $1 < |\varrho_i(\mathcal{A})| < |\mathcal{A}|$, we call ϱ_i a *partial coincidence*.

One can find the columns of a power ϱ^n of ϱ via the following well-known result.

Fact 3. *Let $n \in \mathbb{N}$ and let ϱ be a substitution of constant length L. Let $0 \leqslant k \leqslant L^n - 1$. Then the kth column of the substitution ϱ^n is given by the functional composition*

$$(\varrho^n)_k = \varrho_{k_0} \circ \varrho_{k_1} \circ \cdots \circ \varrho_{k_{n-1}},$$

where $[k_{n-1}, \ldots, k_1, k_0]$ is the base-L expansion of k. $\qquad\qquad\square$

To continue, let $\mathcal{C}$ be another finite alphabet. A letter-to-letter map $\tau \colon \mathcal{A} \to \mathcal{C}$ is called a *coding*, which extends to a map $\tau \colon \mathcal{A}^\mathbb{N} \to \mathcal{C}^\mathbb{N}$. We say that $w \in \mathcal{A}^\mathbb{N}$ is a *fixed point* of ϱ if $\varrho(w) = w$. A fixed point w of ϱ is called *aperiodic* if there does not exist a finite word v such that $w = v^\infty$, i.e., w is not a concatenation of infinite copies of v. We call a substitution ϱ *aperiodic* if it does not admit any periodic fixed point.

Remark 4. Cobham's little theorem then implies that these infinite words are actually L-automatic (where L is the length of ϱ). An automatic sequence is one which can be retrieved

as an output of a deterministic finite state automaton with output (DFAO). We do not define what a DFAO here and refer the reader to [4] instead. ◇

2.2. **Monochromatic arithmetic progressions in substitution fixed points.** Consider $w \in \mathcal{A}^{\mathbb{N}}$. One can view w as an $|\mathcal{A}|$-colouring of $\mathbb{N}$, where the colours are in one-to-one correspondence with the elements of $\mathcal{A}$. Fix $d, M \geqslant 1$. We say that w contains a *monochromatic arithmetic progression* of difference d and length M if there exists a starting position $k \in \mathbb{N}$ such that $w_k = w_{k+dn}$, for $0 \leqslant n \leqslant M - 1$. The monochromatic arithmetic progression is infinite if $w_k = w_{k+dn}$, for all $n \in \mathbb{N}$. Monochromatic arithmetic progressions in two-sided infinite words are similarly defined in the obvious way.

Definition 5. Let $w \in \mathcal{A}^{\mathbb{N}}$ and $d \in \mathbb{N}$. We denote by $A_w(d)$ the maximum length of a monochromatic arithmetic progression of difference d that can be found in w.

Next, we define the *height* of the substitution ϱ. Note that this definition does not depend on the fixed point w; see [13]. Let ϱ be an aperiodic, primitive, constant-length substitution. Let w be a one-sided fixed point of (possibly some power of) ϱ. The height $h(\varrho)$ is given by

$$h(\varrho) := \max \{n \geqslant : \gcd(n, L) = 1, n \text{ divides } \gcd \{w_0 = w_a\}\}$$

We have the following sufficient condition for the finiteness of $A(d)$, for all $d \in \mathbb{N}$, for a fixed point w in terms of the columns of ϱ; compare [2, Prop. 8].

Proposition 6. *Let ϱ be an aperiodic, primitive, constant-length substitution with height 1. Let w be a fixed point of any power of ϱ. Then $A(d) < \infty$ for all $d \geqslant 1$ if and only if ϱ does not have a coincidence column.* □

This finiteness result carries over to codings of certain substitution fixed points; see Section 4 below for the treatment of the Rudin–Shapiro substitution and its generalisations.

2.3. **Notation.** Here, we recall some standard notation concerning asymptotics of non-negative functions; compare [29, Ch. 1]. For functions $f, g \colon \mathbb{N} \to \mathbb{R}_{>0}$, we write

- $f(n) \sim g(n)$, if $\lim_{n\to\infty} |f(n)/g(n)| = 1$,
- $f(n) \gtrsim g(n)$, if there exists $h(n) \colon \mathbb{N} \to \mathbb{R}_{>0}$ with $f(n) \geqslant h(n)$ and $h(n) \sim g(n)$.

3. Bijective automata

A substitution ϱ of constant length L is called *bijective* if every column ϱ_i of ϱ is a bijection. We denote by $G = G^{(1)} = \langle \varrho_i \rangle_{0 \leqslant i \leqslant L-1}$ the group generated by the columns of ϱ, seen as a subgroup of the symmetric group $S_{|\mathcal{A}|}$. Throughout this section, ϱ will be a length-L substitution satisfying the following assumptions, which we denote by $(*)$ for brevity,

$$\text{aperiodic, primitive, bijective, } \varrho_0 = \text{id.} \tag{$*$}$$

Note that the condition of the zeroth column being the identity is natural for bijective substitutions and can be achieved by taking a suitable power. We refer the reader to [38, Ch. 9] for a comprehensive treatment of bijective substitutions; see also [18, 26].

Let w be a fixed point of (a power of) a constant-length substitution ϱ which satisfies $(*)$. The following proposition is a version of Proposition 6 for bijective substitutions that does not need the height-1 condition.

Proposition 7. *Let ϱ be an aperiodic, primitive and bijective substitution. Any fixed point w of a power of ϱ satisfies $A(d) < \infty$, for all $d \geqslant 1$.*

Proof. If w is not periodic, then by Lee-Moody-Solomyak's overlap algorithm [30, Theorem 4.7, Lemma A.9], the corresponding self-similar tiling is not pure point and does not admit infinite arithmetic progressions by [34, Theorem 5.1]. $\qquad \square$

In the next section, we provide lower bounds for $A(d)$ for specific values of d and prove Theorem 1.

3.1. **Lower bounds and polynomial growth of** $A(d)$**.** The following result shows that, for the fixed points of a primitive bijective substitution, we can find a sequence of differences d for which $A(d)$ grows polynomially in d.

Proposition 8. *Let ϱ be a length-L substitution satisfying $(*)$, and let G be the group generated by the columns of ϱ. Then any fixed point of ϱ satisfies, for every $k \geqslant 1$,*

$$L^k \leqslant A\left(\frac{L^{k|G|} - 1}{L^k - 1}\right) < \infty.$$

Proof. Let $k = 1$. Consider the substitution $\varrho^{|G|}$, which is of length $L^{|G|}$. Let $i_0, i_1, \ldots, i_{L-1}$ be the arithmetic progression of difference $d^{(1)} = \sum_{j=0}^{|G|-1} L^j$ with $i_0 = 0$. For every $0 \leqslant m \leqslant L - 1$, $i_m = \sum_{j=0}^{|G|-1} mL^j$, which has base-$L$ expansion $[m, m, \ldots, m]$. Then, for every $0 \leqslant m \leqslant L - 1$, the i_mth column of $\varrho^{|G|}$ is equal to the identity. Indeed, $\left(\varrho^{|G|}\right)_{i_m} = \varrho_m \circ \cdots \circ \varrho_m = (\varrho_m)^{|G|} = \mathrm{id}$, where the first equality holds by Fact 3, and the last equality holds because $g^{|G|} = \mathrm{id}$, for every group element $g \in G$. Since $\varrho^{|G|}$ has L columns equal to the identity substitution distributed in arithmetic progression of difference d, any fixed point of ϱ has a monochromatic arithmetic progression of difference d and length at least L. This completes the proof for $k = 1$. This proof extends to every positive integer k because, since ϱ has a column which is equal to the identity substitution (the leftmost column), the group generated by the columns of ϱ is equal to the group generated by the columns of ϱ^k [9]. This means for a fixed $k \geqslant 1$, one can take $\varrho^{k|G|}$ and construct $d^{(k)}$, this time with $0 \leqslant m \leqslant L^k - 1$. The finiteness of $A(d)$ follows from Proposition 7. $\qquad \square$

Remark 9. Note that one can replace $|G|$ with $\mathrm{lcm}\{\mathrm{ord}(g) \colon g \in G\}$ in Proposition 8 and get the same lower bound. $\qquad \diamondsuit$

The following is immediate from Proposition 8.

Corollary 10. *For all $d = \frac{L^{k|G|} - 1}{L^k - 1}$ with $k \geqslant 1$, $A(d) \gtrsim d^\alpha$, where $\alpha = (|G| - 1)^{-1}$.*

Theorem 1 follows directly from Corollary 10. Note that when $|G| = 2$, then there exists a subsequence of distances for which $A(d)$ grows linearly in d. This is exactly the subfamily treated in [2]. Below, we provide another sufficient condition for a bijective substitution (now on a possibly larger alphabet) to admit an infinite subsequence of differences along which $A(d)$ grows linearly in d. We begin with the following definition.

Definition 11. Let ϱ be a length-L substitution satisfying $(*)$, and let G be the group generated by the columns of ϱ. If there exists $g \in G$ such that $\varrho_i \cdot \varrho_{L-1-i} = g$, for all $0 \leqslant i \leqslant L - 1$, we say that ϱ is *g-palindromic*. If $g = \mathrm{id}$, we say that ϱ is *inverse palindromic*.

Proposition 12. *Let ϱ be a length-L substitution satisfying $(*)$. Suppose further that*

(1) *the group G generated by the columns of ϱ is Abelian,*
(2) *ϱ is g-palindromic, for $g \in G$.*

Then any fixed point of ϱ satisfies, for every $n \geqslant 1$ and even $\ell \geqslant 2$,

$$L^n \leqslant A\left(\frac{L^{n\ell}-1}{L^n+1}\right) < \infty.$$

Proof. To prove the lower bound for $n = 1$ and an even $\ell \geqslant 2$, we consider the substitution ϱ^ℓ, which has length L^ℓ. Let $i_1, i_2, \ldots, i_L$ be an arithmetic progression of difference $d = (L^\ell - 1)/(L+1) = \sum_{j=0}^{\ell-1}(-1)^{j+1}L^j$, where $i_k = k\,d$, for each $1 \leqslant k \leqslant L$. Using the identity

$$k\sum_{j=0}^{\ell-1}(-1)^{j+1}L^j \;=\; (k-1)\sum_{j=0}^{\frac{\ell-2}{2}}L^{2j+1} \;+\; (L-k)\sum_{j=0}^{\frac{\ell-2}{2}}L^{2j},$$

we see that the base-L representation of i_k is $[k-1, L-k, \ldots, k-1, L-k]$, with all the even digits equal to $L-k$, and all the odd digits equal to $k-1$. Then, the i_kth column of ϱ^ℓ is given by

$$
\begin{aligned}
(\varrho^\ell)_{i_k} \;&=\; \varrho_{L-k} \circ \varrho_{k-1} \circ \cdots \circ \varrho_{L-k} \circ \varrho_{k-1} &=\\
&=\; \underbrace{\varrho_{L-k} \circ \varrho_{L-1-(L-k)}}_{=g} \circ \cdots \circ \underbrace{\varrho_{L-k} \circ \varrho_{L-1-(L-k)}}_{=g} &=\; g^{\frac{\ell}{2}},
\end{aligned}
$$

where the first equality holds by Fact 3, and the last equality holds because ϱ is g-palindromic. This implies that $A(d) \geqslant L$, as required.

To prove the claim for an integer $n \geqslant 2$, it suffices to show that ϱ^n is g^n-palindromic when ϱ is g-palindromic. Notice that ϱ^n has length L^n. Let $[i_{n-1}, \ldots, i_1, i_0]$ be the base-L representation of an integer $0 \leqslant i \leqslant L^n - 1$. It is easy to check that the base-L representation of $L^n - 1 - i$ is $[L-1-i_{n-1}, \ldots, L-1-i_1, L-1-i_0]$. Then,

$$
\begin{aligned}
(\varrho^n)_i \circ (\varrho^n)_{L^n-1-i} \;&=\; \left(\varrho_{i_0} \circ \cdots \circ \varrho_{i_{n-1}}\right) \circ \left(\varrho_{L-1-i_0} \circ \cdots \circ \varrho_{L-1-i_{n-1}}\right) &=\\
&=\; \underbrace{\varrho_{i_0} \circ \varrho_{L-1-i_0}}_{=g} \circ \cdots \circ \underbrace{\varrho_{i_{n-1}} \circ \varrho_{L-1-i_{n-1}}}_{=g} &=\; g^n,
\end{aligned}
$$

where the first equality holds by Fact 3, the second equality holds because G is Abelian, and the last equality holds because ϱ is g-palindromic. So ϱ^n is g^n-palindromic. Similar to the $n = 1$ case, this implies that, for all integers $n \geqslant 2$, $A(d) \geqslant L^n$, as required. Finally, the finiteness of $A(d)$ follows by Proposition 7, for every positive integer n. $\qquad\square$

Notice that if we pick $\ell = 2|G|$ in Proposition 12, we get an analogue of Proposition 8 for another family of differences.

Remark 13. We observe also that, for inverse palindromic substitutions, the monochromatic arithmetic progression found in Proposition 12 can be extended by two. Indeed, from the base-L representations of $0d$ and $(L+1)d$, it is easy to see that $(\varrho^\ell)_{0d} = \mathrm{id}$ and $(\varrho^\ell)_{(L+1)d} = \varrho_{L-1}^\ell$. Since, for inverse palindromic substitutions $g = \varrho_{L-1} = \mathrm{id}$, this implies that $A(d) \geqslant L^n + 2$. $\;\Diamond$

The following is immediate from Proposition 12.

Corollary 14. *For all $d = \frac{L^{n\ell}-1}{L^n+1}$ with $n \geqslant 1$ and even $\ell \geqslant 2$, $A(d) \gtrsim d^\alpha$, where $\alpha = (\ell-1)^{-1}$. In particular, $A(d) \gtrsim d$ for differences $d = L^n - 1$.*

3.2. **Van der Waerden-type numbers.** Van der Waerden's theorem [40] states that, for every $c, M \geqslant 1$, there exists an $n \geqslant 1$ such that any colouring of $\{0, 1, \dots, n-1\}$ with c many colours contains a monochromatic arithmetic progression of length M. The smallest threshold of n, for given values of c and M, is the *van der Waerden number* $W(c, M)$. In this subsection, we define van der Waerden-type numbers for automatic sequences arising from substitutions ϱ satisfying condition ($*$), i.e., 'aperiodic, primitive, bijective, with $\varrho_0 = \mathrm{id}$', and provide explicit upper bounds.

A word $x \in \mathcal{A}^{\mathbb{N}}$ is called *linearly recurrent* if there exists a positive constant R_x, such that the distance between any two consecutive occurrences of a finite subword u of x is at most $R_x |u|$. We say that R_x is a *linear recurrence constant* for x. Since the fixed points of a primitive substitution ϱ are linearly recurrent (see [4, 14, 15]), and moreover, one can find an R_x that is independent of x and depends only on ϱ (see [14, Thm. 18]), we can associate a *linear recurrence constant* to ϱ, and denote it by $R = R_\varrho$. Let $u \in \mathcal{L}$. A *return word* v to u is a legal word such that (i) $vu \in \mathcal{L}$, (ii) u is a prefix of v, and (iii) u occurs exactly once in v. Below, we mention some well-known results on the linear recurrence constant for primitive substitutions; compare [14–16].

Proposition 15. *Let ϱ be a primitive constant-length substitution on a finite alphabet.*

   (1) *The substitution ϱ is linearly recurrent for the constant $R = L\zeta_2$, where ζ_2 is the maximum length of a return word for a legal word of length 2 in $\mathcal{L}$.*

   (2) *All legal words of length N appear in any legal word of length $(R+1)N$.*

Definition 16. Consider the class of length-L substitutions on c letters that satisfy the ($*$) condition. We denote by $\mathcal{B}(c, L)$ the set of all fixed points of substitutions from this class.

One can compute R that works for all $x \in \mathcal{B}(c, L)$ by computing an upper bound for ζ_2 that depends only on c and L. We obtain an upper bound for this gap using the level-2 induced substitution on $\mathcal{L}_2$, where $\mathcal{L}_2$ is the set of all length-2 legal words.

First, we identify $\mathcal{L}_2$ with the set of *right-collared words* of the form a_b, where $ab \in \mathcal{L}$. The level-2 induced substitution $\varrho^{(2)} \colon \mathcal{L}_2 \to (\mathcal{L}_2)^+$ is then the substitution arising from the original ϱ which respect the collaring. As an example, the level-2 induced substitution for Thue–Morse is given by

$$\varrho^{(2)} \colon 0_0 \mapsto 0_1 1_0 \qquad 1_0 \mapsto 1_0 0_0 \qquad 0_1 \mapsto 0_1 1_1 \qquad 1_1 \mapsto 1_0 0_1.$$

It is well known that $\varrho^{(2)}$ is also primitive whenever ϱ is primitive; see [7, Sec. 4.8.3]. Combining (1) in Proposition 15 with Wielandt's bound [41] for the index of primitivity yields the following result.

Lemma 17. *For given $c, L \geqslant 2$ and $x \in \mathcal{B}(c, L)$, a linear recurrence constant for x is*

$$R = 2L^{c^4 - 2c^2 + 3} - L.$$

Proof. The index of primitivity of a $c \times c$ primitive matrix M is bounded from above by $c^2 - 2c + 2$, i.e., $M^{c^2 - 2c + 2} > 0$ (seen entry-wise). Let ϱ be a substitution satisfying the conditions in Definition 16, $\varrho^{(2)}$ the level-2 induced substitution, and $M^{(2)}$ be the substitution matrix of $\varrho^{(2)}$. Note that there are at most $|\mathcal{A}|^2 = c^2$ length-2 legal words for ϱ. This means the size of $M^{(2)}$ is at most $c^2 \times c^2$. Applying Wielandt's bound [41], we get that $\left(M^{(2)}\right)^{c^4 - 2c^2 + 2}$ is a strictly positive matrix. It follows that $\left(\varrho^{(2)}\right)^{c^4 - 2c^2 + 2}(a_b)$ contains all collared words in $\mathcal{L}_2$.

Fix $a_b \in \mathcal{L}_2$. We know that x can be written as a concatenation of level-$(c^4 - 2c^2 + 2)$ superwords of ϱ, all of which admitting at least one occurrence of ab (possibly at the border) by the argument above. It follows that, for any $a_b \in \mathcal{L}_2$, the longest return word to a_b has length at most $2L^{c^4-2c^2+2} - 1$. A direct application of (1) in Proposition 15 proves the claim. $\quad\square$

Remark 18. We comment on the generality of the proof of the previous lemma. First, note that it only depends on the size c of the alphabet and the length L of the alphabet, and hence it gives a linear recurrence constant for all substitutions in the class considered in this section, parametrised by c and L. Second, since bijectivity is invoked nowhere in the proof, such a bound can be used for extensions to more general classes. $\quad\Diamond$

Note that the set $\mathcal{B}(c, L)$ is a non-empty proper subset of $\mathcal{A}^{\mathbb{N}}$. As a direct consequence of van der Waerden's theorem, we have the following.

Proposition 19. *Given c, $L \geqslant 2$ and $M \geqslant 1$, there exists a positive integer n such that every length-n subword of every element of $\mathcal{B}(c, L)$ contains a length-M monochromatic arithmetic progression.*

Definition 20. Given c, $L \geqslant 2$ and M, we call the smallest threshold of the number n predicted by Proposition 19 a *van der Waerden-type number for $\mathcal{B}(c, L)$*, and we denote it by $W(\mathcal{B}(c, L), M)$.

It is clear that $W(\mathcal{B}(c, L), M) \leqslant W(c, M)$.

Proposition 21. *For $c, L \geqslant 2$ and $M \geqslant 1$, one has*

$$W(\mathcal{B}(c, L), M) \leqslant (R+1)L^{kc!},$$

where $k = \lceil \log_L M \rceil$ and $R = 2L^{c^4-2c^2+3} - L$.

Proof. By Proposition 8, we know that, for any $n \geqslant 1$, the maximum length of monochromatic arithmetic progressions in any element in $\mathcal{B}(c, L)$ satisfies,

$$A(d_n) \geqslant L^n , \quad \text{where } d_n = \frac{L^{n|G|} - 1}{L^n - 1} .$$

Let k be the least non-negative integer such that $L^k \geqslant M$, which one can write as $k = \lceil \log_L M \rceil$. Then, $A(d_k) \geqslant L^k \geqslant M$. Since the arithmetic progression from Proposition 8 starts at 0, every fixed point x of ϱ has a prefix y of length $1 + (L^k - 1) \cdot d_k = L^{k|G|}$ containing a monochromatic arithmetic progression of difference d_k and length M.

From the discussion above, x is linearly recurrent for some constant $R_x > 0$. By Property 2 in Proposition 15, all subwords of x of length $L^{k|G|}$ (in particular the subword y) appear in every subword of length $(R_x + 1)L^{k|G|}$. So, every subword of x of length $(R_x + 1)L^{k|G|}$ contains a monochromatic arithmetic progression of difference d_k and length M. From Lemma 17, one can choose R_x to be $2L^{c^4-2c^2+3} - L$. To complete the proof, notice that $|G| \leqslant |S_c| = c!$. $\quad\square$

Proposition 21 can be reformulated as follows: if ϱ is a length-L substitution on c letters which satisfies $(*)$ and $M \geqslant 1$, every legal word of ϱ of length at least $(R+1)L^{kc!}$ contains a monochromatic arithmetic progression of length M.

Example 22. Consider the case $c = L = 2$, which is generated by the Thue–Morse substitution

$$\varrho: \begin{array}{ccc} a & \mapsto & ab \\ b & \mapsto & ba \end{array}.$$

The substitution ϱ has two fixed points. The methods in [15] yield $R_\varrho \leqslant 16$, which can be further improved to $R_\varrho = 9$ using the software Walnut; see [16, Ex. 3.13]. Using this result we obtain $W(\mathcal{B}(2,2), M) \leqslant 10 \cdot 4^k$, where $k = \lceil \log_2 M \rceil$. Thus the bounds for the first few van der Waerden-type numbers are

$$W(\mathcal{B}(2,2), M) \leqslant 640 \qquad \text{for } M = 6, 7, 8,$$
$$W(\mathcal{B}(2,2), M) \leqslant 2560 \qquad \text{for } 8 < M \leqslant 16,$$
$$W(\mathcal{B}(2,2), M) \leqslant 10240 \qquad \text{for } 16 < M \leqslant 32,$$
$$W(\mathcal{B}(2,2), M) \leqslant 40960 \qquad \text{for } 32 < M \leqslant 64,$$

which are significantly lower than the respective bounds for the general van der Waerden numbers. $\Diamond$

The bound for R established in Proposition 21 is far from optimal. For example, for the family $\mathcal{B}(2,2)$ studied in Example 22, we obtain $R \leqslant 2^{12} - 2 = 4094$ and consequently,

$$\text{where} \qquad W(\mathcal{B}(2,2), M) \leqslant 4095 \cdot 4^k, k = \lceil \log_2 M \rceil,$$

which is a weaker bound than the bound obtained in Example 22 using the optimal value of the recurrence constant ($R = 9$). It would be interesting to obtain a better method to compute the constant R in Proposition 21 and hence improve the upper bound of $W(\mathcal{B}(c, L), M)$.

3.3. **Upper bounds of $A(d)$ for Abelian bijective substitutions.** Throughout the whole Section 3, we consider substitutions ϱ satisfying condition $(*)$. In this subsection, we add the additional assumption that the group G generated by the columns of ϱ is Abelian. From Lemma 17, there exists a positive integer N such that for any $a \in \mathcal{A}$, $\varrho^N(a)$ contains all legal words of length 2, i.e., $N := \min\{n : \mathcal{L}_2 \subseteq \mathcal{L}(\varrho^n(a)), a \in \mathcal{A}\}$, where $\mathcal{L}(\varrho^n(a))$ is the set of all words appearing in $\varrho^n(a)$. Let $v \in \mathcal{A}^\mathbb{N}$ be a fixed point of ϱ. The goal of this section is to provide an upper bound on $A(d)$ for v.

We begin with the following results regarding certain columns of ϱ^{N+M} under the existence of certain progressions in v, where $M \geqslant 1$.

Lemma 23. *Let $d, M \geqslant 1$, with $d < L^{N+M}$. Let $\ell := \gcd(d, L^{N+M})$ and assume that there exists a non-negative integer n such that $v_n = v_{n+jd}$, for $j = 0, 1, \ldots, L^{N+M}/\ell$. Then for all $k \in \mathbb{Z}$ such that $0 \leqslant n + k\ell < n + k\ell + d < L^{N+M}$, one has*

$$\left(\varrho^{N+M}\right)_{n+k\ell} = \left(\varrho^{N+M}\right)_{n+k\ell+d}.$$

Proof. We have a trivial inclusion

$$\{[n + id]_{L^{N+M}} \mid i = 0, 1, \ldots, L^{N+M}/\ell - 1\} \subset \{[m]_{L^{N+M}} \mid m \in \mathbb{Z}, m \equiv n \mod \ell\},$$

where $[\cdot]_{L^{N+M}}$ denotes the equivalence class of natural numbers mod L^{N+M}. For $0 \leqslant i < j < \frac{L^{N+M}}{\ell}$, we have $n + id \not\equiv n + jd \mod L^{N+M}$ by the definition of ℓ. This means the two sets above have the same cardinality, and hence are the same set.

We see for each k satisfying the condition above, there exists $i \in \{0, 1, \ldots, (L^{N+M}/\ell) - 1\}$ such that $n + k\ell \equiv n + id \pmod{L^{N+M}}$. This implies there is a positive integer s for which

$$n + k\ell + sL^{N+M} = n + id \quad \text{and} \quad n + k\ell + d + sL^{N+M} = n + (i+1)d.$$

By assumption, $v_{n+id} = v_{n+(i+1)d}$. Note that v_{n+id} is the $(n+k\ell)$th letter in $\varrho^{N+M}(v_s)$ and $v_{n+(i+1)d}$ is the $(n+k\ell+d)$th letter in $\varrho^{N+M}(v_s)$. This means in the $(n+k\ell)$th and $(n+k\ell+d)$th

9

columns of ϱ^{N+M}, the images of one letter v_s are the same. Since the column group G is Abelian and acts on $\mathcal{A}$ transitively, we see that the columns $\left(\varrho^{N+M}\right)_{n+k\ell}$ and $\left(\varrho^{N+M}\right)_{n+k\ell+d}$ (seen as permutations of $\mathcal{A}$) must coincide, thus proving the claim. $\qquad\square$

We now relate the column equality result in Lemma 23 to existence of infinitely long progressions in v.

Lemma 24. *Let $M \geqslant 1, d \leqslant L^M$, and $\ell = \gcd(L^M, d)$. Let n be a non-negative integer such that $\left(\varrho^{N+M}\right)_{n+k\ell} = \left(\varrho^{N+M}\right)_{n+k\ell+d}$ for each integer k satisfying $0 \leqslant n + k\ell < n + k\ell + d < L^{N+M}$. Then, we have $v_n = v_{n+jd}$ for all $j \in \mathbb{N}$. That is, there exists an arithmetic progression of infinite length and difference d starting at n.*

Proof. For each $j \in \mathbb{N}$, v_{n+jd} and $v_{n+(j+1)d}$ are included either in a supertile of length L^M or two consecutive such supertiles. This means there exists a $t \geqslant 0$ such that v_{n+jd} is the $(n + jd - L^M t)$th letter in $\varrho^M(v_t v_{t+1})$ and $v_{n+(j+1)d}$ is the $(n + (j + 1)d - L^M t)$th letter in $\varrho^M(v_t v_{t+1})$. By assumption, the word $v_t v_{t+1}$ appears in $\varrho^N(a)$, for each $a \in \mathcal{A}$. There is an s such that the sth letter in $\varrho^N(a)$ is v_t and the $(s + 1)$th letter is v_{t+1}. This means that v_{n+jd} is the $(n+jd-L^M s-L^M t)$th letter in $\varrho^{N+M}(a)$ and $v_{n+(j+1)d}$ is the $(n+(j+1)d-L^M s-L^M t)$th letter in $\varrho^{N+M}(a)$. By the definition of ℓ and the assumption on the columns for ϱ^{N+M}, the $(n + jd - L^M s - L^M t)$th column and the $(n + (j + 1)d - L^M s - L^M t)$th column are the same, and we have $v_{n+jd} = v_{n+(j+1)d}$. Since j is arbitrary, the claim follows. $\qquad\square$

Proposition 25. *Let ϱ be a length-L substitution which satisfies $(*)$ and whose column group G is Abelian. Let v be a fixed point of ϱ. Fix a difference d and let M be a positive integer such that $d \leqslant L^M$. Suppose $\gcd(d, L^M) = \gcd(d, L^{N+M}) =: \ell$. We then have $A(d) \leqslant \frac{L^{N+M}}{\ell}$.*

Proof. Suppose there exists $n \geqslant 1$ such that $v_n = v_{n+jd}$, for $j = 0, 1, \ldots, L^{N+M}/\ell$. It follows from Lemmas 23 and 24 that $v_n = v_{n+jd}$, for all $j \in \mathbb{N}$. This contradicts Proposition 7 stating that v does not admit infinitely long monochromatic progressions, which immediately implies the claim on $A(d)$. $\qquad\square$

Combining Proposition 25 with Proposition 8, we get the following.

Corollary 26. *Let v be a fixed point of an Abelian, length-L substitution ϱ which satisfies $(*)$. Then, for all $k \geqslant 1$,*

$$L^k \leqslant A\left(\frac{L^{k|G|} - 1}{L^k - 1}\right) \leqslant L^{k|G|+N},$$

where G is the group generated by the columns of ϱ and $N = \min\{n : \mathcal{L}_2 \subseteq \mathcal{L}(\varrho^n(a))\}$.

Note that from Lemma 17, N is bounded from above by $c^4 - 2c^2 + 3$, where c is the size of the alphabet (this is one more than the bound for the index of primitivity for $M^{(2)}$ to include the case when a length-2 legal word appears at the boundary).

One of the restrictions in Proposition 25 is that, for an arbitrary L, one is only able to give upper bounds for $A(d)$ for differences which satisfy the gcd-condition. In what follows, we give a subclass of lengths for which it is possible to give an upper bound for $A(d)$ for all d.

Proposition 27. *Let ϱ, v, and L be as in Proposition 25. Let $L = p_1^{n_1} p_2^{n_2} \cdots p_t^{n_t}$ be the prime factorisation of L with $p_1 < p_2 < \cdots < p_t$ and assume $n_1 \leqslant n_2 \leqslant \cdots \leqslant n_t$. Then, for each $d \geqslant 1$ there exist $M \geqslant 1$ with $d \leqslant L^M$ such that $\gcd(d, L^{N+M}) = \gcd(d, L^M)$. Moreover, we*

have $A(d) \leqslant L^{N+1} d^B$, where $B = \frac{\log L}{n_1 \log p_1}$. In particular, if $t = 1$, (that is, L is a power of a prime), $A(d) \lesssim L^{N+1} d$.

Proof. Choose $M \in \mathbb{N}$ such that $p_1^{n_1(M-1)} \leqslant d < p_1^{n_1 M}$. The equality of the greatest common divisors follows from $p_i^{n_i M} \nmid d$, for all $1 \leqslant i \leqslant t$. Set $B = \frac{\log L}{n_1 \log p_1}$. By Proposition 25, we have

$$A(d) \leqslant \frac{L^{N+M}}{\ell} = \frac{L^{N+1}}{\ell}(p_1^{n_1 B})^{M-1} = \frac{L^{N+1}}{\ell}(p_1^{n_1(M-1)})^B \leqslant \frac{L^{N+1}}{\ell} d^B \leqslant L^{N+1} d^B.$$

The last claim follows since $L = p_1^{n_1}$ implies $B = 1$. □

One can leverage the previous proposition to obtain lower bounds for van der Waerden-type numbers $W(\mathcal{B}(c, L), M)$, for certain values of L and M.

Corollary 28. *Let $c, m > 1$ and assume L admits the same form as in Proposition 27, i.e., it has prime factorisation $L = p_1^{n_1} p_2^{n_2} \cdots p_t^{n_t}$, where $p_1 < p_2 < \cdots < p_t$ are such that $n_1 \leqq n_2 \leqq \cdots \leqq n_t$. Then we have*

$$W(\mathcal{B}(c, L), L^{N_0+1} m^{\lceil B \rceil} + 1) > L^{N_0+1} m^{\lceil B \rceil + 1} + 1,$$

where $N_0 = c^4 - 2c^2 + 3$ and $B = \frac{\log L}{n_1 \log p_1}$.

Proof. We will prove a stronger statement that there exists an $x \in \mathcal{B}(c, L)$ such that any of its subwords of length $L^{N_0+1} m^{\lceil B \rceil + 1} + 1$ does not contain monochromatic arithmetic progressions of length $L^{N_0+1} m^{\lceil B \rceil} + 1$.

Take a ϱ, which is a primitive, aperiodic, bijective substitution of length L such that its column group is abelian and $\varrho_0 = \mathrm{id}$. Such a substitution always exists. Fix a length L and the size of the alphabet c. Without loss of generality, one can force the column group to be the cyclic group $G = C_c$ of order c, which is Abelian and acts transitively on $\mathcal{A} = \{0, \ldots, c-1\}$. Transitivity is already sufficient to ensure primitivity; see [9, Prop. 2.3].

It remains to construct an aperiodic substitution with that group profile, for any given length. Here, we use a criterion for aperiodicity provided in [26, Prop. 4.1], which states that a sufficient condition for aperiodicity for primitive and bijective substitutions is the existence of two length-two legal words which share either the same starting letter or the same ending letter; see also [9, Prop. 2.5].

We first handle the case when $L \geqslant 3$. For such lengths, we choose $\varrho_0 = \varrho_1 = \mathrm{id}$ and $\varrho_2 = (12 \cdots 0)$, where ϱ_2 generates G, and we fill the other positions with permutations from C_c. From construction, we immediatelty see that 00 and 01 are both legal, and hence implies aperiodicity.

For the case $L = 2$, we pick $\varrho_0 = \mathrm{id}$ and $\varrho_1 = (12 \cdots 0)$ and show that this substitution is aperiodic. Note that, under ϱ, $0 \mapsto 01$ and $(c-1) \mapsto (c-1)0$. Applying ϱ^2 to $(c-1)$ yields $\varrho^2(c-1) = (c-1)001$, which means 00 and 01 are both legal with respect to ϱ. By the same argument for the previous cases, we obtain aperiodicity for all such substitutions.

Let N be a natural number such that for any alphabet a, $\varrho^N(a)$ contains all of the two-letter legal words for ϱ. By the argument of Lemma 17, we have $N \leqslant N_0$. Let x be a fixed point for ϱ.

If $d' \leqslant m$, by Proposition 27, the maximal length of monochromatic arithmetic progression is less than or equal to $L^{N_0+1}(d')^{\lceil B \rceil}$, which is less than or equal to $L^{N_0+1} m^{\lceil B \rceil}$. There are no monochromatic arithmetic progressions of difference d' and length $L^{N_0+1} m^{\lceil B \rceil} + 1$ anywhere in x, and so anywhere in its subwords.

If $d' > m$, the progressions of difference d' and length $L^{N_0+1}m^{\lceil B \rceil} + 1$ span as long as $L^{N_0+1}m^{\lceil B \rceil}d' + 1$ and cannot be contained in a subword of length $L^{N_0+1}m^{\lceil B \rceil+1} + 1$. In either case, the subwords of x with length $L^{N_0+1}m^{\lceil B \rceil+1} + 1$ do not contain monochromatic arithmetic progressions of length $L^{N_0+1}m^{\lceil B \rceil} + 1$. $\hfill\square$

3.4. Examples.

3.4.1. *Thue–Morse sequence over L letters.* The Thue–Morse sequence over the alphabet $\mathcal{A}_L = \{0, 1, \ldots, L-1\}$ is the infinite word $v = v_0 v_1 v_2 \cdots$, where v_i is given by the sum (modulo L) of the digits in the base-L representation of i; see [11]. This sequence can also be defined to be the fixed point, with first letter 0, of the primitive, length-L, bijective substitution φ defined as [8]

$$\varphi(a) = \varphi_0(a)\, \varphi_1(a)\, \cdots\, \varphi_{L-1}(a), \qquad \text{where} \qquad \varphi_i(a) = a + i \pmod{L}. \tag{1}$$

The exact values of $A(L^n - 1)$ for $L = 2$ and $L = 3$ where obtained (in [35] and [36], respectively), and it was shown that the same arguments can be used for any prime number L (see [37]). The result for $L = 2$ was reestablished in [2] using a different approach. The key argument of this approach can be easily generalised for all L, giving Proposition 30 below as a result.

The group G generated by the columns of φ is the cyclic group C_L of order L. We write $C_L = \langle g \rangle$ multiplicatively, where g corresponds to adding 1 (mod L). We can easily see that φ_ℓ, the ℓ-th column of φ, is given by g^ℓ. We next show that φ is actually g^{L-1}-palindromic and so, Proposition 12 can also be directly applied to get a subsequence along which $A(d)$ grows faster.

Proposition 29. *Let φ be the generalised Thue–Morse substitution over L letters from Eq. (1). Then, φ is g^{L-1}-palindromic. Consequently, for any fixed point of φ one has*

$$A(L^n - 1) \geqslant L^n, \tag{2}$$

for all $n \geqslant 1$.

Proof. It suffices to show that, for each $0 \leqslant i \leqslant L - 1$, one has $\varphi_i \cdot \varphi_{L-(i+1)} = g^{L-1}$. This follows immediately from $\varphi_i = g^i$. The lower bound for $A(L^n - 1)$ follows from Proposition 12 by choosing $\ell = 2$. $\hfill\square$

Note that we can improve the lower bounds given in Eq. (2) when $n \equiv 0 \bmod L$ by looking at the concatenation of three level-n superwords, which we carry out below. As mentioned earlier, this result generalises that in [2] for $L = 2$ to any arbitrary L.

Proposition 30. *Let φ be the generalised Thue–Morse substitution over L letters from Eq. (1). For any fixed point of φ, one has*

$$A(L^n - 1) \geqslant \begin{cases} L^n + 2L, & \text{if } n \equiv 0 \bmod L, \\ L^n, & \text{otherwise.} \end{cases}$$

Proof. The case when $n \not\equiv 0 \bmod L$ is already covered in Proposition 29 so we assume from hereon that $n \equiv 0 \bmod L$. From the proof of Proposition 12, we have that, for φ^n, one has $(\varphi^n)_i \cdot (\varphi^n)_{L^n-(i+1)} = \left(g^{L-1}\right)^n = \text{id}$, for all $0 \leqslant i \leqslant L^n - 1$. This means, if we now look at φ^{2n}, we get $\left(\varphi^{2n}\right)_{i_m} = \text{id}$ with $i_m = m(L^n - 1)$ and $1 \leqslant m \leqslant L^n$. Note further that $\left(\varphi^{2n}\right)_{i_0} = \left(\varphi^{2n}\right)_{i_{L^n+1}} = \text{id}$. Altogether, this yields a monochromatic arithmetic progression of

as of length $L^n + 2$ within the superword $\varphi^{2n}(a)$. The goal is now to look at progressions of as in $\varphi^{2n}(a-1)$ and $\varphi^{2n}(a+1)$ of the same difference. We then extend the progression from $\varphi^{2n}(a)$ to a longer progression in $\varphi^{2n}((a-1)(a)(a+1))$. Note that the word $(a-1)(a)(a+1) \in \mathcal{L}_3(\varphi)$, for any $a \in \mathcal{A}$.

We first look at the supertile $\varphi^{2n}(a-1)$. We show that for $0 \leqslant m \leqslant L - 2$, at positions $i_m = L^{2n} - (m+1)L^n + (m+1)$, one has $\left(\varphi^{2n}\right)_{i_m} = g$. These are the positions which correspond to the continuation of the progression from $\varphi^{2n}(a)$ with difference $d = L^n - 1$; see Figure 1. One can check that the L-ary expansion of i_m reads

$$[\overbrace{L-1, L-1, \ldots, L-(m+1)}^{n \text{ digits}}, \overbrace{0, 0, \ldots, m+1}^{n \text{ digits}}]$$

From Fact 3, we get that

$$\left(\varphi^{2n}\right)_{i_m} = (\varphi_{L-1} \circ \varphi_0)^{n-1} \circ (\varphi_{L-(m+1)} \circ \varphi_{m+1}) = (g^{L-1})^{n-1}(g^{L-1} \cdot g)$$
$$= (g^{L-1})^n \cdot g = g,$$

where the second equality holds since φ is g-palindromic and the last equality holds since $n \equiv 0 \bmod L$. Note that this is only true for $0 \leqslant m \leqslant L - 2$, since for $m = L - 1$, one gets $i_m = [L-1, L-1, \ldots, L-1, 0, 0, \ldots, 1, 0]$. Carrying out the same calculation, we get $\left(\varphi^{2n}\right)_{i_m} = (g^{L-1})^{n-2} \cdot (g \cdot g^{L-1}) = g \cdot (g^{L-1})^{n-1} \neq g$, since $n - 1$ is coprime with L. This means the extension of the arithmetic progression in $\varphi^{2n}(a-1)$ has length at most $L - 1$.

One can do an analogous analysis for the supertile on the right, which is $\varphi^{2n}(a+1)$. Here the relevant positions are of the form $j_m = (m+1)L^n - (m+2)$, and one needs to show that $\left(\varphi^{2n}\right)_{j_m} = g^{L-1}$. Since the proof uses the same arguments above, we leave it to the reader. Note that here, one can also show that, for $m = L - 1$, $\left(\varphi^{2n}\right)_{j_m} \neq g^{L-1}$, which implies that the extension to the right also has length at most $L - 1$. Considering the progression of as which straddles across these three supertiles verifies the claim. $\square$

FIGURE 1. A monochromatic arithmetic progression of as of length $L^n + 2L$ accommodated within the supertiles $\varphi^{2n}(a-1)$, $\varphi^{2n}(a)$ and $\varphi^{2n}(a+1)$. Here level-$2n$ supertiles are written as $L^n \times L^n$ blocks, which are read from left to right, and then top to bottom. The bottom-most shaded square in $\varphi^{2n}(a-1)$ corresponds to $i_0 = L^{2n} - L^n + 1$ while the top-most shaded square in $\varphi^{2n}(a+1)$ is at $j_0 = L^n - 2$.

We conjecture that the lower bounds given in Proposition 30 are actually exact values. This has been settled when n is prime in [37]. We now look at other differences d. From Proposition 8,

we directly obtain the lower bound

$$A\left(\frac{L^{Ln} - 1}{L^n - 1}\right) \geqslant L^n,$$

for all $n \geqslant 1$. This result can also be geometrically visualised as in the previous proposition.

Example 31 (Ternary Thue–Morse). We fix $L = 3$ and consider the ternary Thue–Morse sequence, which is the fixed point $v = 012\cdots$ of the substitution

$$\varphi\colon \begin{matrix} 0 \mapsto 012 \\ 1 \mapsto 120 \\ 2 \mapsto 201 \end{matrix}.$$

For differences of the form $d = 3^n - 1$, Proposition 30 and Corollary 26 imply that, for all $n \geqslant 1$, we have $3^n \leqslant A(3^n - 1) \leqslant 3^{n+4}$. For differences of the form $d = 3^{2n} + 3^n + 1$, it follows directly from Proposition 8 that, for all $n \geqslant 1$, we have

$$A\big(3^{2n} + 3^n + 1\big) \geqslant 3^n;$$

see Figure 3 for a plot of $A(d)$ for differences up to 2200, for the ternary Thue–Morse sequence.

We can give an alternative visual approach by identifying a long monochromatic arithmetic progression across a diagonal of a block substitution, as in the proof of Proposition 30, but now in three dimensions. As in Figure 1, we can consider the word $\varphi^{3n}(0)$ and arrange it inside a block. The only difference is now we arrange it in a three-dimensional cube of side-length 3^n. There is no fixed choice of fitting the word inside a cube, and one must only be consistent when going up and through a layer.

In our choice depicted in Figure 2 below, we start from the lower left corner of the cube, traverse along the x-direction, then go up the next row. Once all rows in the bottom-most layer are filled, one moves one layer up and starts directly above the point where the origin is. The red shaded squares precisely correspond to the monochromatic arithmetic progression that starts at the origin, with difference $d = 3^{2n} + 3^n + 1$ and has length 3^n. $\diamond$

FIGURE 2. A monochromatic arithmetic progression of difference $d = 3^{2n} + 3^n + 1$ in the word $\varphi^{3n}(0)$, with $n = 2$.

FIGURE 3. Plot of the exact values of $A(d)$ for $1 \leqslant d \leqslant 2200$ for the ternary Thue–Morse sequence.

3.4.2. *Other bijective substitutions.* Below, we give more examples of bijective substitutions and the corresponding explicit bounds from the results in the previous sections.

Example 32 ($G = A_4$)**.** Consider the following substitution of length $L = 3$,

$$\varrho: \begin{aligned} 0 &\mapsto 011 \\ 1 &\mapsto 120 \\ 2 &\mapsto 203 \\ 3 &\mapsto 332 \end{aligned} \quad,$$

with $\varrho_0 = \mathrm{id}$, $\varrho_1 = (012)$ and $\varrho_2 = (01)(23)$. The group G generated by ϱ_1 and ϱ_2 is the alternating group A_4, which consists of $|G| = 12$ elements. Proposition 8 then implies that $A(d) \geqslant 3^k \geqslant d^{1/11}$, for all $k \in \mathbb{N}$ and all $d = (3^{12k} - 1)/(3^k - 1)$. However, Remark 9 shows that, instead of the group order $|G|$, we can use the least common multiple of $|\varrho_1| = 3$ and $|\varrho_2| = 2$, which is 6. Consequently, $A(d) \geqslant 3^k \geqslant d^{1/5}$, for all $k \in \mathbb{N}$ and all $d = (3^{6k} - 1)/(3^k - 1)$. $\quad\diamond$

Example 33 (Inverse-palindromic Abelian)**.** Consider the three-letter substitution

$$\varrho: \begin{aligned} 0 &\mapsto 02010 \\ 1 &\mapsto 10121 \\ 2 &\mapsto 21202 \end{aligned} \quad,$$

with $\varrho_0 = \varrho_2 = \varrho_4 = \mathrm{id}$, $\varrho_1 = (021)$ and $\varrho_3 = (012) = \varrho_1^{-1}$, and hence ϱ is inverse palindromic. The group generated by the columns is $G = C_3$, which is Abelian. It follows from Proposition 12 and Remark 13 that $A(5^k - 1) \geqslant 5^k + 2$, for $k \in \mathbb{N}$. Corollary 26 with $L = 5$, $|G| = 3$ and $N = 2$ yields, for every positive integer k,

$$5^k \leqslant A(25^k + 5^k + 1) \leqslant 25 \cdot 5^{3k}.$$

$\diamond$

15

Example 34 (Inverse-palindromic non-Abelian)**.** Here we demonstrate why the Abelian assumption is necessary in Proposition 12. Consider

$$\varrho: \begin{array}{l} 0 \mapsto 01120 \\ 1 \mapsto 12001 \\ 2 \mapsto 20212 \end{array} \quad .$$

This substitution is inverse palindromic with $\varrho_0 = \mathrm{id} = \varrho_4, \varrho_2 = (01)$ and $\varrho_1 = (012) = \varrho_3^{-1}$ and $G = S_3$, which is non-Abelian. Consider ϱ^2 and let us compute $(\varrho^2)_7$ and $(\varrho^2)_{17}$. Since $7 = [2, 1]$, we have $(\varrho^2)_7 = \varrho_2 \varrho_1 = (12)$. Similarly $17 = [2, 3]$ so $(\varrho^2)_{17} = \varrho_2 \varrho_3 = (02)$. Since $(\varrho^2)_7 \neq (\varrho^2)_{17}^{-1}$, ϱ^2 is not inverse palindromic. $\diamond$

3.5. **Non-bijective substitutions with super-substitution structure.**

Definition 35. Let $\mathcal{A}$ be a finite alphabet and consider a constant-length substitution ϱ of length L over $\mathcal{A}$. A partition

$$\mathcal{A} = \mathcal{A}_1 \cup \mathcal{A}_2 \cup \cdots \cup \mathcal{A}_n$$

of $\mathcal{A}$ is said to induce a *supersubstitution* for ϱ if, for each $\ell \in \{0, 1, \ldots L - 1\}, i \in \{1, 2, \ldots, n\}$ and $a, b \in \mathcal{A}_i$, there exists $j \in \{1, 2, \ldots, n\}$ such that $\varrho_\ell(a)$ and $\varrho_\ell(b)$ are both in $\mathcal{A}_j$. In such a case, we can define a map $\theta: \mathcal{A} \to \{1, 2, \ldots, n\}$ by defining $\theta(a) = i$, where $a \in \mathcal{A}_i$.

This allows one to define a substitution ξ of length L on the alphabet $\{1, 2, \ldots, n\}$ via $\xi_\ell(i) = j$ for each $\ell \in \{0, 1, \ldots, L - 1\}$ and $i \in \{1, 2, \ldots, n\}$, where j is such that $\varrho_\ell(a) \in \mathcal{A}_j$ for each $a \in \mathcal{A}_i$. By definition, we have $\theta \circ \varrho = \xi \circ \theta$.

In what follows, we let ϱ be a constant-length substitution of length L over $\mathcal{A}$ which admits a partition $\mathcal{A} = \mathcal{A}_1 \cup \mathcal{A}_2 \cup \cdots \cup \mathcal{A}_n$ that induces a supersubstitution ξ. Without loss of generality, let v be the fixed point for ϱ starting with some $a_1 \in \mathcal{A}_1$ and w be the fixed point of ξ starting with 1. We have $\theta(v) = w$, because we have $\theta \circ \varrho = \xi \circ \theta$ and so

$$\theta \circ \varrho^n(a_1) = \xi^n \circ \theta(a_1) = \xi^n(1)$$

for each $n \geqslant 1$.

Proposition 36. *Suppose $\mathcal{A}_1$ is singleton $\{a_1\}$ and that ξ satisfies property $(*)$ in Section 3 with column group G. Then we have*

$$A_v \left(\frac{L^{k|G|} - 1}{L^k - 1} \right) \geqslant L^k,$$

for all $k \geqslant 1$.

Proof. By Proposition 8, there is an arithmetic progression of $\mathcal{A}_1$ of difference $\frac{L^{k|G|} - 1}{L^k - 1}$ of length L^k in w. In the preimage v, there is an arithmetic progression of a_1 of the same difference and length, since $\theta^{-1}(1) = \{a_1\}$. $\square$

Example 37. Consider the alphabet $\mathcal{A} = \{a, b, c, d, e\}$ with the partition $\mathcal{A}_1 = \{a\}, \mathcal{A}_2 = \{b, c\}, \mathcal{A}_3 = \{d, e\}$. Let $G = S_3$. We start with a bijective substitution ξ with letters $\mathcal{A}_1, \mathcal{A}_2, \mathcal{A}_3$ such that

$$\xi: \begin{array}{l} \mathcal{A}_1 \mapsto \mathcal{A}_1 \mathcal{A}_2 \mathcal{A}_3 \mathcal{A}_1 \mathcal{A}_3 \mathcal{A}_2 \\ \mathcal{A}_2 \mapsto \mathcal{A}_2 \mathcal{A}_1 \mathcal{A}_2 \mathcal{A}_3 \mathcal{A}_1 \mathcal{A}_3 \\ \mathcal{A}_3 \mapsto \mathcal{A}_3 \mathcal{A}_3 \mathcal{A}_1 \mathcal{A}_2 \mathcal{A}_2 \mathcal{A}_1 \end{array} \quad .$$

16

We then construct a substitution ϱ on $\mathcal{A}$ which is compatible with the supersubstitution ξ. For example, consider

$$\varrho: \begin{aligned} a &\mapsto acdaec \\ b &\mapsto babead \\ c &\mapsto bacead. \\ d &\mapsto ddabca \\ e &\mapsto edabca \end{aligned}$$

By Proposition 36, for the fixed point v of ϱ starting with a, one has $A\left(\frac{L^{6k}-1}{L^k-1}\right) \geqslant L^k$. $\qquad \Diamond$

One can relax the singleton criterion in Proposition 36 and replace it with some restrictions on the columns of the original substitution.

Proposition 38. *Let ϱ be a length-L substitution with a supersubstitution structure ξ. Assume there are $0 \leqslant c_1 < c_2 < \cdots < c_k < L$ such that $\varrho_{c_j}(a) = a_1$ for each j, for all $a \in \mathcal{A}_1$. Consider $n \geqslant 1$ with L-adic expansion $n = [n_m, n_{m-1}, \ldots, n_1, n_0]$, with $n_0 \in \{c_1, c_2, \ldots, c_k\}$. Set n' to be $n' = [n_m, n_{m-1}, \ldots, n_1]$. If $w_{n'} = 1$, then $v_n = a_1$.*

Proof. Since $\theta(v_{n'}) = w_{n'} = 1$, $v_{n'} \in \mathcal{A}_1$ and we have $v_n = \varrho_{n_0}(v_{n'}) = a_1$. $\qquad \square$

The previous result allows one to construct differences which correspond to long arithmetic progressions.

Example 39. Consider $\mathcal{A} = \{a, b, c, d, e, f\}$ with $\mathcal{A}_1 = \{a, b\}, \mathcal{A}_2 = \{c, d\}, \mathcal{A}_3 = \{e, f\}$. Fix the length of the substitution to be $L = 6$ and consider the substitution ϱ given by

$$\begin{aligned} a &\mapsto abbabd & c &\mapsto cddcce & e &\mapsto effeea \\ b &\mapsto aabaac & d &\mapsto dccddf & f &\mapsto fefefb. \end{aligned}$$

We see that, here $k = 2$ and $c_1 = 0, c_2 = 3$. The supersubstitution is

$$\xi: \begin{aligned} \mathcal{A}_1 &\mapsto \mathcal{A}_1\mathcal{A}_1\mathcal{A}_1\mathcal{A}_1\mathcal{A}_1\mathcal{A}_2 \\ \mathcal{A}_2 &\mapsto \mathcal{A}_2\mathcal{A}_2\mathcal{A}_2\mathcal{A}_2\mathcal{A}_2\mathcal{A}_3. \\ \mathcal{A}_3 &\mapsto \mathcal{A}_3\mathcal{A}_3\mathcal{A}_3\mathcal{A}_3\mathcal{A}_3\mathcal{A}_1 \end{aligned}$$

For each $n \geqslant 1$, set

$$d_n = 3 + 3 \cdot 6^n + 3 \cdot 6^{2n} = [3, 0, 0, \ldots, 0, 0, 3, 0, 0, \ldots, 0, 0, 3],$$

where $[\cdot]$ is the base-6 expansion. If $0 \leqslant k \leqslant 4 \cdot 6^{n-1} - 1$, then $3k < 2 \cdot 6^n$ and the base-6 expansion for $3k$ is $[i_n, i_{n-1}, \cdots, i_0]$ with $i_n < 2$ and $i_0 \in \{0, 3\}$. For all such k, the base-6 expansion of $kd_n = 3k + 3k \cdot 6^n + 3k \cdot 6^{2n}$ is

$$[i_n, i_{n-1}, \ldots, i_1, i_0 + i_n, i_{n-1}, \ldots i_1, i_0 + i_n, \ldots i_{n-1}, \ldots i_1, i_0].$$

We now consider $k'_n := (kd_n)'$ to be the number whose base-6 expansion is the same as kd_n with the last digit i_0 omitted. Let w be the fixed point of ξ starting at 1. We show that $w_{k'_n} = 1$ for all $0 \leqslant k \leqslant 4 \cdot 6^{n-1} - 1$. From the supersubstitution, we get that

$$w_{k'_n} = \left(\xi_{i_n} \xi_{i_0+i_n}^2 \prod_{j=1}^{n-1} \xi_{i_1}^3\right)(1) = 1.$$

Here, $\xi_{i_n} = \mathrm{id}$ because $i_n < 2$, $\xi_{i_0+i_n} = \mathrm{id}$ because $i_0 + i_n < 5$, and the last factor in the product is also the identity because all ξ_i are cube roots of unity. We can now apply the previous result to the original sequence kd_n. Since $i_0 \in \{0, 3\}$, by Proposition 38, the (kd_n)th letter in the fixed

17

point v starting with a is $\varrho_{i_0}(\mathcal{A}_1) = a$, which yields an arithmetic progression of as with length $4 \cdot 6^{n-1}$. $\diamond$

Remark 40. Example 39 shows how Proposition 38 can be used to establish lower bounds for $A(d)$ for some differences d appropriately chosen. For differences of the form $d = \frac{L^{k|G|}-1}{L^k - 1}$, one can use Propositions 36 and 38 to show that $A_v(Ld) \geqslant L^k$, for all $k \geqslant 1$. Furthermore, if we assume that $c_1 = 0$ in Proposition 38, then $A_v(L^m d) \geqslant L^k$, for every $m \geqslant 1$ and all $k \geqslant 1$ (see [1] for details). $\diamond$

4. Spin substitutions

We consider monochromatic arithmetic progressions in infinite words arising from spin substitutions, which are generalisations of the Rudin–Shapiro substitution. A *spin substitution* θ is a special type of constant-length substitution. A finite set $\mathcal{D}$ of digits is considered, each of which, carrying a spin, can be in a finite number of distinct states. The spin states are represented using a finite Abelian group G, called the *spin group*. This results in the alphabet $\mathcal{A} = \mathcal{D} \times G$. The substitution θ is then completely determined using a $|\mathcal{D}| \times |\mathcal{D}|$ matrix V with entries in G, which is called the *spin matrix*. The matrix V encodes, for each digit $d \in \mathcal{D}$, the spin state of the letters of the image of d under the substitution. For background on spin substitutions and generalisations, we refer the reader to [3, 10, 20, 38].

In Sections 4.1 and 4.2, we study $A(d)$ for the Rudin–Shapiro sequence. We give lower bounds for $A(d)$ for two sequences of differences along which $A(d)$ grows at least linearly in d, in analogy to the classical Thue–Morse case studied in [2]. In Section 4.3 we extend these results to Vandermonde sequences.

4.1. **The Rudin–Shapiro sequence.** Consider a spin substitution θ with digit set $\mathcal{D} = \{0, 1\}$, spin group $G = C_2$ and spin matrix $V = \left(\begin{smallmatrix} 1 & 1 \\ 1 & -1 \end{smallmatrix} \right)$. The resulting alphabet is $\mathcal{A} = \mathcal{D} \times G = \{0, 1, \tilde{0}, \tilde{1}\}$, where 'tilded' letters have non-trivial spin. The spin matrix determines the positions of the tildes in $\theta(0)$ and $\theta(1)$; the positions of the tildes in $\theta(\tilde{0})$ and $\theta(\tilde{1})$ are determined via the invariance relation $\theta(\tilde{a}) = \widetilde{\theta(a)}$ with number of tildes modulo 2, for $a \in \mathcal{D}$. The resulting substitution is

$$\theta: \quad \begin{array}{ll} 0 \mapsto 01 & \tilde{0} \mapsto \tilde{0}\tilde{1} \\ 1 \mapsto 0\tilde{1} & \tilde{1} \mapsto \tilde{0}1 \end{array}.$$

The *Rudin–Shapiro sequence* u over the alphabet $\{1, -1\}$ is obtained from the fixed point of θ starting with 0 under the projection

$$\pi_G: \quad 0, 1 \mapsto 1, \quad \tilde{0}, \tilde{1} \mapsto -1;$$

see [7, Section 7.7.1]. The first few terms of u (with commas inserted for the sake of clarity) are

$$u = 1, 1, 1, -1, 1, 1, -1, 1, 1, 1, 1, -1, -1, -1, 1, -1, \cdots.$$

The nth element of u can be derived from V as

$$u_n = \prod_{i=0}^{k-1} V(n_{i+1}, n_i) = V(n_k, n_{k-1}) \ \cdots \ V(n_2, n_1) \, V(n_1, n_0), \tag{3}$$

where $[n_k, \ldots, n_1, n_0]$ is the binary representation of n with n_0 the least significant digit and n_k the most significant digit, and $V(i, j)$ is the (i, j)th entry of V [3, 20]. Alternatively, u_n can be obtained as $u_n = (-1)^{t(n)}$, where $t(n)$ counts the number of (possibly overlapping) occurrences

18

of the word 11 in the binary representation of the integer n; see [4]. The following recurrence relations can easily be obtained from Eq. (3)

$$u_{2n} = u_n, \qquad u_{2n+1} = (-1)^n\, u_n. \tag{4}$$

The following simple argument invoking Proposition 6 shows that $A(d) < \infty$, for all $d \in \mathbb{N}$, for the sequence u.

Proposition 41. *There is no infinite monochromatic arithmetic progression in the sequence u.*

Proof. Let v be the fixed point of θ starting with 0. It follows from the definition of θ that $v_{2n+a} \in \{a, \tilde{a}\}$, for all $a \in \mathcal{D}$ and $n \in \mathbb{N}$. Then, since $u = \pi_G(v)$, we know that $u_{2n+a} = 1$ (resp. -1) implies $v_{2n+a} = a$ (resp. $\tilde{a}$), for all $a \in \mathcal{D}$ and $n \in \mathbb{N}$. The proof is by contradiction. Assume there exist $s \in \mathbb{N}$ and $d \geqslant 1$ such that $u_{s+nd} = 1$ (resp. -1), for all $n \in \mathbb{N}$. This implies that $v_{s+nd} = a$ (resp. $\tilde{a}$) for all $n \in 2\mathbb{N}$, where $a \equiv s \bmod 2$. But this is a contradiction because θ is an aperiodic, primitive, constant-length substitution of height 1 and so, by Proposition 6, v does not contain infinite monochromatic arithmetic progressions. $\qquad \square$

It is not difficult to show that $A(2^n d) = A(d)$, for all $d, n \geqslant 1$ (similar to the Thue–Morse case [2]), and from here that $A(2^n) = 4$, for all $n \in \mathbb{N}$. The next two propositions, where we find sequences of long monochromatic arithmetic progressions for differences of the form $2^n \pm 1$, are an analog of Proposition 8 for bijective substitutions.

Proposition 42. *The sequence u satisfies $A(2^n + 1) \geqslant 2^{n-1} + 2$, for all $n \geqslant 1$.*

Proof. It is easy to see, by direct inspection of u, that the result holds for $n = 1$. For $n > 1$, we will show that $u_k = 1$ with $k = 2^{2n+1} + m(2^n + 1)$, for all $-1 \leqslant m \leqslant 2^{n-1}$. Fix n. For $m = -1$, $k = 2^{2n+1} - 2^n - 1$ with binary representation given by $[1, \ldots, 1, 0, 1, \ldots, 1]$, consisting of two sequences of n consecutive 1's separated by a single 0. Then, by Eq. (3), $u_k = 1$. For $0 \leqslant m \leqslant 2^{n-1}$, let the binary representation of m be $[m_r, \ldots, m_1, m_0]$, where $0 \leqslant r \leqslant n - 1$. Then the binary representation of $k = 2^{2n+1} + m(2^n + 1)$ is

$$[\, 1, \overset{n-r}{\overbrace{0, \ldots, 0}}, m_r, \ldots, m_1, m_0, \overset{n-r-1}{\overbrace{0, \ldots, 0}}, m_r, \ldots, m_1, m_0 \,].$$

For $0 \leqslant r < n - 1$, we have $n - r - 1 \geqslant 1$ and then, by Eq. (3), $u_k = 1$. For $r = n - 1$, we have $n - r - 1 = 0$ and then, by Eq. (3), $u_k = V(m_0, m_r)$. But, for $r = n - 1$, we also have $m = 2^{n-1} = [1, 0, 0, \ldots, 0]$ and so, $u_k = V(m_0, m_r) = V(0, 1) = 1$. $\qquad \square$

Proposition 43. *The sequence u satisfies, for all $n \geqslant 1$,*

$$A(2^n - 1) \geqslant \begin{cases} 2^{n-1} + 1, & \text{if n is even,} \\ 2^{n-1} + 3, & \text{otherwise.} \end{cases}$$

Proof. The result holds if $n = 1$, so we assume that $n \geqslant 2$. We will first show that, for every $n \geqslant 2$ and all $0 \leqslant m \leqslant 2^{n-1}$, there exists $a \in \{1, -1\}$ such that $u_k = a$, where $k = 2^{2n} + (m+1)(2^n - 1)$. Fixing n and writing the binary representation of m as $[m_{n-1}, \ldots, m_1, m_0]$, where $m_i \in \{0, 1\}$ for all $0 \leqslant i \leqslant n - 1$, the binary representation of k takes the form $[1, m_{n-1}, \ldots, m_1, m_0, \overline{m_{n-1}}, \ldots, \overline{m_1}, \overline{m_0}]$, where $\overline{m_i} = 1 - m_i$. By Eq. (3) and given that, for each $0 \leqslant i \leqslant n - 2$, $V(m_{i+1}, m_i)V(\overline{m_{i+1}}, \overline{m_i})$ is equal to -1 if $m_{i+1} = m_i$, and to 1 otherwise, we see that

$$u_k = V(1, m_{n-1})\, V(m_0, \overline{m_{n-1}})\, (-1)^{n-1+m_{n-1}-m_0}$$

19

If $m_{n-1} \neq m_0$, $u_k = (-1)^{n-1}$. If $m_{n-1} = m_0$, $m_{n-1} = m_0 = 0$ because $m \leqslant 2^{n-1}$, and again $u_k = (-1)^{n-1}$. If n is even, this implies that $u_k = -1$, which completes the proof for the even cases. If n is odd, it implies that $u_k = 1$. In this case, one can easily further check that $u_k = 1$ for $m = -1$ and $m = -2$, which completes the proof for the odd cases. $\qquad\square$

Corollary 44. *For all $d = 2^n \pm 1$ with $n \geqslant 1$, the sequence u satisfies $A(d) \gtrsim d/2$.*

Proof. The claim follows directly from Propositions 42 and 43. $\qquad\square$

By computer experiments we have verified the preceding results for $1 \leqslant d \leqslant 4200$. In fact, we have seen that the inequalities in Propositions 42 and 43 are equalities, if $n \geqslant 4$ and if $n \geqslant 5$, respectively. Moreover, the differences of the form $2^n \pm 1$ are those for which the Rudin–Shapiro sequence has the longest monochromatic arithmetic progressions, in the sense that $A(d)$ has local maxima at these differences. A plot of $A(d)$ similar to that in [2] for the Thue–Morse sequence can be obtained in this case for the Rudin–Shapiro sequence.

Remark 45. Sobolewski's paper [39] concerns the computation of upper bounds of $A_w(d)$, for sequences $w \in \mathcal{A}^{\mathbb{N}}$ defined using a block-counting function. More precisely, given a binary block $v \in \mathcal{A}^+$, the digit w_n is given by the sum mod 2 of (possibly overlapping) occurrences of v in the binary representation of n, for all $n \in \mathbb{N}$. The author focuses most of his attention on the $v = 11$ case, for which w is the Rudin–Shapiro sequence. In this case, an upper bound of the maximum length of monochromatic arithmetic progressions starting at position 0 is given, and exact values of $A(d)$ are determined for differences of the form $2^n \pm 1$. $\qquad\Diamond$

Remark 46. The arguments for the Rudin–Shapiro sequence can be extended to the case when the spin matrix is the Hadamard matrix [19]

$$V = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 \end{pmatrix}$$

and hence, θ is a substitution of the eight-letter alphabet $\mathcal{A} = \mathcal{D} \times C_2$, where $\mathcal{D} = \{0, 1, 2, 3\}$. The studied sequence u arises as the image of the fixed point of θ starting with 0 under the coding π_G mapping untilded letters to 1 and tilded letters to -1. In this case $A(4^n d) = A(d)$; in particular, $A(4^n) = 6$ for all $n \in \mathbb{N}$. Results similar to Propositions 42 and 43 can also be derived. On the one hand, $A(4^n + 1) \geqslant 4^{n-1} + 2$, for all $n \geqslant 1$, the proof of which is similar to the proof of Proposition 42. More precisely, it can be shown that, for all $-1 \leqslant m \leqslant 4^{n-1}$, $u_k = 1$ with $k = 4^{2n+1} + m(4^n + 1)$. On the other hand, $A(4^n - 1) \geqslant 4^{n-1} + 3$, for all $n \geqslant 1$, the proof of which is analogous to the proof of Proposition 43. More precisely, it can be shown that, for all $-2 \leqslant m \leqslant 4^{n-1}$, $u_k = 1$ with $k = 3 \cdot 4^{2n+2} + 4^n - 1 + m(4^{n+1} - 1)$. $\qquad\Diamond$

4.2. **An alternative approach for the Rudin–Shapiro sequence.** The Rudin–Shapiro sequence can alternatively be obtained from a staggered substitution or by a substitution acting on an alphabet consisting of pairs of letters in $\{1, -1\}$, say $\{v, \tilde{v}, w, \tilde{w}\}$ with $v = 11$, $w = 1{-}1$, where $\tilde{1} = -1$ and $\widetilde{-1} = 1$ swaps the two letters. The substitution reads

$$\varrho: \quad \begin{aligned} v &\mapsto vw \\ w &\mapsto v\tilde{w} \end{aligned} \qquad \begin{aligned} \tilde{w} &\mapsto \tilde{v}w \\ \tilde{v} &\mapsto \tilde{v}\tilde{w} \end{aligned}.$$

FIGURE 4. The square array of letters in $v^{(2n-1)}$, for $n \in \{2, 3, 4\}$. The highlighted letters on the diagonals form arithmetic progressions of length 2^{n-1} for distances $d = 2^n + 1$ (upper left quadrant) or $d = 2^n - 1$ (upper right quadrant).

Note that this substitution is exactly the same as the original four-letter substitution, except that we now interpret v and w as two-letter words in $\{1, -1\}$.

Note that the substitution is invariant under the letter exchange in the sense that $\varrho(\widetilde{a}) = \widetilde{\varrho(a)}$ for all $a \in \{v, w, \widetilde{w}, \widetilde{v}\}$. Moreover, the first part of $\varrho(a)$ for any $a \in \{v, w, \widetilde{w}, \widetilde{v}\}$ is either v or $\widetilde{v}$, and the last is either w or $\widetilde{w}$. By induction, this structure is preserved for larger superwords as follows.

Lemma 47. *Let $n \geqslant 1$ and set $v^{(0)} = v$, $w^{(0)} = w$, $v^{(n)} := \varrho^n(v)$, and $w^{(n)} := \varrho^n(w)$. Then, $v^{(n)} = v^{(n-1)} w^{(n-1)}$ and $w^{(n)} = v^{(n-1)} \widetilde{w^{(n-1)}}$.*

Proof. Clearly, this is true for $n = 1$. Assuming the structure holds for n, we find that

$$v^{(n+1)} = \varrho(v^{(n)}) = \varrho(v^{(n-1)} w^{(n-1)}) = \varrho(v^{(n-1)}) \varrho(w^{(n-1)}) = v^{(n)} w^{(n)}$$

and

$$w^{(n+1)} = \varrho(w^{(n)}) = \varrho(v^{(n-1)} \widetilde{w^{(n-1)}}) = \varrho(v^{(n-1)}) \varrho(\widetilde{w^{(n-1)}}) = v^{(n)} \widetilde{w^{(n)}},$$

which completes the proof. $\square$

For the rest of the section, we consider the superword $v^{(2n-1)} = \varrho^{2n-1}(v)$ for $n \geqslant 2$, and write the resulting word in the alphabet $\{1, -1\}$, which has 2^{2n} letters, as a square array of letters with 2^n rows of length 2^n. For $n \in \{2, 3, 4\}$, they are shown in Figure 4. Let us denote the (i, j)th entry of this matrix by $a_{i-1,j-1}$. With this notation, one has $v^{(2n-1)} = a_{0,0}, a_{0,1} \ldots a_{2^n-1, 2^n-1}$. In particular, the sequence $a_{0,0}, a_{0,1} \ldots a_{0, 2^n-1}$ corresponds to the topmost row of the block (or the first row of the matrix). In what follows, $v_j^{(n)}$ (resp. $w_j^{(n)}$) denotes the jth letter of $v^{(n)}$ (resp. $w^{(n)}$) seen as a word over $\{1, -1\}$.

Lemma 48. *Let $v^{(n)}$ and $w^{(n)}$ be as in Lemma 47. We then have the following.*

(1) $v_0^{(n)} = v_{2^n}^{(n)} = 1,$ $\quad v_{2^n-1}^{(n)} = \begin{cases} -1, & n \in 2\mathbb{Z} \\ 1, & n \in 2\mathbb{Z}+1 \end{cases},$ $\quad v_{2^{n+1}-1}^{(n)} = \begin{cases} 1, & n \in 2\mathbb{Z} \\ -1, & n \in 2\mathbb{Z}+1 \end{cases}.$

(2) $w_0^{(n)} = 1, w_{2^n}^{(n)} = -1,$ $\quad w_{2^n-1}^{(n)} = \begin{cases} -1, & n \in 2\mathbb{Z} \\ 1, & n \in 2\mathbb{Z}+1 \end{cases},$ $\quad w_{2^{n+1}-1}^{(n)} = \begin{cases} -1, & n \in 2\mathbb{Z} \\ 1, & n \in 2\mathbb{Z}+1 \end{cases}.$

Proof. Note that the word $v^{(n)}$ always starts with 1, and ends with 1 if n is even or -1 if n is odd. Similarly, the word $w^{(n)}$, starts with 1 and ends with -1 if n is even or 1 is n is odd. The properties above then follow from Lemma 47 by induction. $\qquad\square$

From the substitution structure, one has

$$\varrho^n(a_{0,2i}a_{0,2i+1}) = a_{2i,0}a_{2i,1}\cdots a_{2i,2^n-1}a_{2i+1,0}a_{2i+1,1}\cdots a_{2i,2^n-1}.$$

Together with Lemma 48, we get the following.

Lemma 49. *Let $v^{(2n-1)} = a_{0,0}, a_{0,1}\ldots a_{2^n-1,2^n-1}$ as above.*

(1) *(From top to left) If $a_{0,2i}a_{0,2i+1} = v$, then $a_{2i,0}a_{2i+1,0} = v$. If $a_{0,2i}a_{0,2i+1} = w$, then $a_{2i,0}a_{2i+1,0} = w$.*

(2) *(From top to right) If $a_{0,2i}a_{0,2i+1} = v$, then $a_{2i,2^n-1}a_{2i+1,2^n-1}$ is w if n is odd and $\tilde{w}$ if n is even. If $a_{0,2i}a_{0,2i+1} = w$, then $a_{2i,2^n-1}a_{2i+1,2^n-1}$ is v if n is odd and $\tilde{v}$ if n is even.*

with obvious extensions to the case when $a_{0,2i}a_{0,2i+1}$ is $\tilde{v}$ or $\tilde{w}$.

The previous lemma relates words in the topmost row of the matrix to words found along the leftmost and the rightmost columns. We can define the following maps which convert words in the topmost row in the matrix to the words along the rightmost column. Note that v read backwards is still v and that w read backwards is $\tilde{w}$.

Definition 50. Let

$$I_{\text{odd}}: \begin{matrix} v \mapsto \tilde{w} & w \mapsto v \\ \tilde{v} \mapsto w & \tilde{w} \mapsto \tilde{v} \end{matrix}, \qquad\qquad I_{\text{even}}: \begin{matrix} v \mapsto w & w \mapsto \tilde{v} \\ \tilde{v} \mapsto \tilde{w} & \tilde{w} \mapsto v \end{matrix},$$

and set

$$I_*(b_1 b_2 \cdots b_n) = I_*(b_n)I_*(b_{n-1})\cdots I_*(b_1)$$

for $b_i \in \{v, w, \tilde{v}, \tilde{w}\}$ and $* \in \{\text{even}, \text{odd}\}$.

From (2) in Lemma 49, the rightmost column, read from bottom to top, is $I_{\text{even}}(v^{(n-1)})$ if n is even and $I_{\text{odd}}(v^{(n-1)})$ if n is odd. We now express $I_*(v^{(n-1)})$ as a level-n superword.

Lemma 51. *We have*

$$I_*(v^{(n-1)}) = \varrho^{n-1}(\tilde{w}) \qquad and \qquad I_*(w^{(n-1)}) = \varrho^{n-1}(v)$$

for $* \in \{\text{even}, \text{odd}\}$.

Proof. Since the proof for the case $* = $ odd is similar, we omit it and only present the one for the even case. We proceed by induction. The statement is clear for $n = 2$. If the statement

holds for some even n, then for $n + 2$ we have:

$$
\begin{aligned}
I_{\text{even}}(\varrho^{n+1}(v)) &= I_{\text{even}}(\varrho^{n-1}\varrho^2(v)) \\
&= I_{\text{even}}(\varrho^{n-1}(vwv\tilde{w})) \\
&= I_{\text{even}}(\varrho^{n-1}(\tilde{w}))I_{\text{even}}(\varrho^{n-1}(v))I_{\text{even}}(\varrho^{n-1}(w))I_{\text{even}}(\varrho^{n-1}(v)) \\
&= \varrho^{n-1}(\tilde{v})\varrho^{n-1}(\tilde{w})\varrho^{n-1}(v)\varrho^{n-1}(\tilde{w}) \\
&= \varrho^{n-1}(\tilde{v}\tilde{w}v\tilde{w}) \\
&= \varrho^{n+1}(\tilde{w}).
\end{aligned}
$$

By a similar computation, we have $I_{\text{even}}\varrho^{n-1}(w) = \varrho^{n-1}(v)$. $\qquad\square$

We now have the following result.

Proposition 52. *The word which appears on the leftmost column of the top left quadrant is always $v^{(n-2)}$. The word on the rightmost column of the top-right quadrant, read from bottom to top, is $w^{(n-2)}$.*

Proof. The first claim follows immediately from (1) of Lemma 49. The second claim follows from (2) of Lemma 49 , Lemma 51 and the fact that the word on the upper half of the rightmost quadrant (read from bottom to top) is the second half of $\varrho^{n-1}(\tilde{w}) = \varrho^{n-2}(\tilde{v}w) = \varrho^{n-2}(\tilde{v})\varrho^{n-2}(w)$, which is $w^{(n-2)}$; see Figure 5 for an illustration. $\qquad\square$

FIGURE 5. Schematic representation of block arrangement of the superword $v^{(2n-1)}$. The dashed lines indicate the reflection symmetries of the upper left and upper right quadrants.

Now we can prove the existence of monochromatic diagonal and anti-diagonals.

Proposition 53. *We have $a_{j,j} = 1$ for $0 \leqslant j \leqslant 2^{n-1} - 1$.*

Proof. Each left half-row $a_{i,0}a_{i,1}\cdots a_{i,2^{n-1}-1}$ is either $v^{(n-2)}$ or $\widetilde{v^{(n-2)}}$ and the topmost half-row (with $i = 0$) is always $v^{(n-2)}$. If $a_{0,j} = 1$, then by Proposition 52, $a_{j,0} = 1$, which implies that the jth row is $v^{(n-2)}$. This implies $a_{j,j} = a_{0,j} = 1$. If $a_{0,j} = -1$, then by Proposition 52, $a_{j,0} = -1$, which implies that the jth row is $\widetilde{v^{(n-2)}}$. We see $a_{j,j} = \widetilde{a_{0,j}} = 1$. $\qquad\square$

Proposition 54. *If n is odd, then we have $a_{j,2^n-j-1} = 1$ for $0 \leqslant j \leqslant 2^{n-1} - 1$. If n is even, then we have $a_{j,2^n-j-1} = -1$ for $0 \leqslant j \leqslant 2^{n-1} - 1$.*

Proof. The ith right half-row $a_{i,2^{n-1}}a_{i,2^{n-1}+1}\cdots a_{i,2^n-1}$ is either $w^{(n-2)}$ or $\widetilde{w^{(n-2)}}$ and the topmost right half-row is always $w^{(n-2)}$. Assume that n is odd. We have $a_{0,2^n-1}=1$. If $a_{0,2^n-j-1}=1$, then by Proposition 52, $a_{j,2^n-1}=1$, which implies that the jth right half-row is $w^{(n-2)}$ and is the same as the topmost right half-row. It follows that $a_{j,2^n-j-1}=a_{0,2^n-j-1}=1$. If $a_{0,2^n-j-1}=-1$, then by Proposition 52, $a_{j,2^n-1}=-1$, which implies that the jth right half-row is $\widetilde{w^{(n-2)}}$. We then have $a_{j,2^n-j-1}=\widetilde{a_{0,2^n-j-1}}=1$. The case when n is even admits a completely analogous proof, which we leave to the reader. $\qquad\square$

Note that Proposition 53 proves the existence of a monochromatic arithmetic progression of difference $d=2^n+1$ while Proposition 54 yields one with difference $d=2^n-1$. We conclude with the following comparable version of Propositions 42 and 43 in the previous section.

Corollary 55. *For the binary Rudin–Shapiro sequence, one has*

$$A(2^n-1)\geqslant 2^{n-1}\quad and\quad A(2^n+1)\geqslant 2^{n-1}.$$

4.3. **Vandermonde sequences.** In this section, we consider general Vandermonde substitutions, the simplest of which is the Rudin–Shapiro substitution. Let be a spin substitution with digit set $\mathcal{D}=\{0,1,\dots,L-1\}$ and spin group $G=\{1,\omega,\omega^2,\dots,\omega^{L-1}\}$, where $\omega=\mathrm{e}^{-2\pi\mathrm{i}/L}$, resulting in the alphabet $\mathcal{A}=\mathcal{D}\times G$. We will consider the *digit projection* $\pi_{\mathcal{D}}\colon\mathcal{A}\to\mathcal{D}$, defined by $\pi_{\mathcal{D}}(a)$ to be the digit of a, and the *spin projection* $\pi_G\colon\mathcal{A}\to G$, defined by $\pi_G(a)$ to be spin of a. For each letter $a\in\mathcal{A}$, let $s_a\in\{0,1,\dots,L-1\}\subseteq\mathbb{N}$ be the *spin number* of a, given by the exponent of ω in $\pi_G(a)$. Let the spin matrix V of the spin substitution be a *Vandermonde matrix*, given by $V(i,j)=\omega^{ij\bmod L}$, for $0\leqslant i,j\leqslant L-1$. In matrix form

$$V=\begin{pmatrix}1 & 1 & 1 & \cdots & 1 & 1\\ 1 & \omega & \omega^2 & \cdots & \omega^{L-2} & \omega^{L-1}\\ 1 & \omega^2 & \omega^4 & \cdots & \omega^{2(L-2)} & \omega^{2(L-1)}\\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots\\ 1 & \omega^{L-1} & \omega^{(L-1)2} & \cdots & \omega^{(L-1)(L-2)} & \omega^{(L-1)(L-1)}\end{pmatrix}.$$

Now, let the spin substitution $\theta\colon\mathcal{A}\to\mathcal{A}^L$ be defined, for each $a\in\mathcal{A}$, by $\theta(a)=a_0a_1\cdots a_{L-1}$, where, for each $0\leqslant i\leqslant L-1$, $a_i\in\mathcal{A}$ is such that $\pi_{\mathcal{D}}(a_i)=i$ and $\pi_G(a_i)=V(i,\pi_{\mathcal{D}}(a))\,\pi_G(a)=\omega^{i\cdot\pi_{\mathcal{D}}(a)+s_a\bmod L}$. We call θ a *Vandermonde substitution*, and the infinite word u obtained from the fixed point of θ starting with 0 under the projection π_G a *Vandermonde sequence*. The first few terms of u (with commas inserted for the sake of clarity) are

$$u=\overset{L}{\overbrace{1,1,\dots,1}},1,\omega,\omega^2,\dots,\omega^{L-2},\omega^{L-1},1,\omega^2,\omega^3,\dots,\omega^{L-1},\omega,1,\omega^3,\omega^4,\cdots.$$

The nth entry of u can be obtained from V using again Eq. (3), namely, $u_n=\prod_{i=0}^{k-1}V(n_{i+1},n_i)$, where $[n_k,\dots,n_1,n_0]$ is now the base-L representation of n. Using this, one can easily prove the following lemma (the proof of which we omit), which gives analogous recurrence relations to those in Eq. (4).

Lemma 56. *The Vandermonde sequence u satisfies, for all $n\in\mathbb{N}$ and each $0\leqslant a\leqslant L-1$, the recurrence relation $u_{Ln+a}=V(a,b)\,u_n$, where $0\leqslant b\leqslant L-1$ is such that $n\equiv b\bmod L$.*

A simple argument, similar to that used in Proposition 41, can be used to show that the Vandermonde sequence u satisfies $A(d)<\infty$, for all $d\geqslant 1$.

Proposition 57. *There is no infinite monochromatic arithmetic progression in the sequence u.*

Proof. Assume there exist $r \in \mathbb{N}$ and $d \geqslant 1$ such that $u_{r+nd} = g \in G$, for all $n \in \mathbb{N}$. Then $u_{r+nLd} = g$, for all $n \in \mathbb{N}$, and writing r as $mL + k$, where $m \in \mathbb{N}$ and $0 \leqslant k \leqslant L - 1$, we have $u_{sL+k} = g$, where $s = m + nd$, for all $n \in \mathbb{N}$. Let $v \in \mathcal{A}^{\mathbb{N}}$ be the fixed point of θ starting with 0, hence $u = \pi_G(v)$. Then $\pi_G(v_{sL+k}) = g$ and, by the definition of θ, $\pi_{\mathcal{D}}(v_{sL+k}) = k$, for all s. Therefore, v contains an infinite monochromatic arithmetic progression. But, since θ is an aperiodic, primitive, constant-length substitution of height 1, this is a contradiction, by Proposition 6. $\qquad\square$

It can be shown that $A(L^n d) = A(d)$ for all $n \in \mathbb{N}$ and, in particular, $A(L^n) = A(1) = L + 2$. As an analogy with the Rudin–Shapiro sequence, in the following proposition we look at the differences of the form $\frac{L^{nL}-1}{L^n-1} = L^{(L-1)n} + L^{(L-2)n} + \ldots + L^n + 1$.

Proposition 58. *The sequence u satisfies* $A\left(\frac{L^{nL}-1}{L^n-1}\right) \geqslant L^{n-1} + 1$, *for all* $n \geqslant 1$.

Proof. We omit the details for the case $n = 1$. Let be $n > 1$. To prove the claim, we will show that $u_k = 1$ with $k = L^{Ln+1} + m(L^{(L-1)n} + \ldots + L^n + 1)$, for all $0 \leqslant m \leqslant L^{n-1}$. Let the base-$L$ representation of m be $[m_r, \ldots, m_1, m_0]$, where $0 \leqslant r \leqslant n - 1$. The base-$L$ representation of k is then given by

$$[\,1, 0, \overbrace{0, \ldots, 0}^{n-r-1}, m_r, \ldots, m_1, m_0, \overbrace{\ldots \ldots, \underbrace{0, \ldots, 0}_{n-r-1}}^{L\ \text{times}}, m_r, \ldots, m_1, m_0\,].$$

For $0 \leqslant m \leqslant L^{n-1} - 1$, we have $n - r - 1 \geqslant 1$ and, by Eq. (3), $u_k = \prod_{i=0}^{r-1}(V(m_{i+1}, m_i))^L = 1$. For $m = L^{n-1}$, we have $n - r - 1 = 0$, but also $m_i = 0$, for all $0 \leqslant i \leqslant r - 1$. Consequently, the base-L representation of k consists of $L + 1$ isolated digits equal to 1 separated by sequences of digits equal to 0, which, by Eq. (3), implies that $u_k = 1$. $\qquad\square$

It is easy to check that the arithmetic progression of 1's found in the proof of Proposition 58 cannot be extended to the right, and that it can neither be extended to the left, except if $L = 2$ (thus yielding as a corollary Proposition 42 for the Rudin–Shapiro sequence). The following is immediate from Proposition 58 and it implies Theorem 2.

Corollary 59. *For all* $d = \frac{L^{kL}-1}{L^k-1}$ *with* $k \geqslant 1$, $A(d) \gtrsim \frac{d^\alpha}{L}$, *where* $\alpha = (L - 1)^{-1}$.

Note that if $L = 2$, we recover Corollary 44 for the Rudin–Shapiro sequence, for which $A(d)$ grows linearly in d.

5. Outlook

It is not obvious how to extend Proposition 8 to the general constant-length case. Unlike in the bijective setting, the columns generate a semigroup rather than a group, and we no longer necessarily have an identity column. Even the task of finding a suitable subsequence with growing arithmetic progressions becomes nontrivial, as the following example illustrates.

Let $\mathcal{A}$ be the 6-letter alphabet and ϱ be the length 2 substitution

$$
\begin{array}{llll}
a \mapsto ad & \quad c \mapsto ea & \quad e \mapsto bf & \qquad (5) \\
b \mapsto bc & \quad d \mapsto ab & \quad f \mapsto ba.
\end{array}
$$

The graph of sets in Figure 6 traces which letters occur in the columns. This follows a modified version of the graph in [12] and incorporates the subsets of $\mathcal{A}$ which appear as columns. The

sets which appear at the lowermost level are called *minimal sets*; see also [31]. These are the subsets of the alphabet that appear as columns in a large enough power of the substitution. Thus, the size of the minimal sets corresponds to the column number $c(\varrho)$. In this example, $c(\varrho) = 2$.

FIGURE 6. Graph of sets for ϱ

This graph incorporates many interesting paths. Any path starting from a minimal set leads only to other minimal sets. The relation between this graph and arithmetic progressions found within the fixed points of the substitution can be seen through the following observations together with Fact 3, which does not require the substitution to be bijective. The graph helps narrow down the scope of differences d for which suitable long arithmetic progressions may be found. For example, there is no arithmetic progression that includes both positions 4 and 37. Converting to binary and following the path from the topmost level, we obtain the disjoint minimal sets $\{a, b\}$ and $\{f, d\}$; see the columns highlighted in blue in Figure 7.

For the next example, let us restrict ourselves to the minimal set $\{a, b\}$ to illustrate how the paths in the graph represent the columns. Note that for a difference $d = 4 = [1, 0, 0]$, the path indexed by $d, 2d, 3d, \ldots$ always returns to the minimal set $\{a, b\}$. However, this condition does not guarantee a large arithmetic progression at these positions, as the corresponding letter might be b, instead, as the following picture illustrates via the columns highlighted in red.

FIGURE 7. The fixed points of ϱ starting from the seeds a and b, with positions $0, 4, 8, 12$ highlighted in red and places 4 and 37 highlighted in blue.

Since the substitution is primitive, the subgraph of minimal sets is strongly connected; in particular we can find a cycle starting from the set $\{a, b\}$ that visits every other minimal set. This path is indexed by the edges 1, 0, 1, 0 and might be a natural candidate for difference of a long arithmetic progression.

It would be interesting to find out if one can use these graphs to bound $A(d)$ or gather more information on its behaviour. In particular, it is currently not certain whether there exist a sequence of differences for which $A(d)$ grows polynomially. Numerical data suggests that $A(d)$ grows polynomially along the subsequence $d_n = 2^n + 2$; see Figure 8.

FIGURE 8. $A(d)$ for $1 \leqslant d \leqslant 2300$, for the substitution in Eq (5).

REFERENCES

[1] I. Aedo, Forward limit sets of semigroups of substitutions and arithmetic progressions in automatic sequences, PhD thesis, The Open University, in preparation.

[2] I. Aedo, U. Grimm, Y. Nagai, P. Staynova, Monochromatic arithmetic progressions in binary Thue–Morse-like words, Theor. Comp. Sci. 934 (2022), 65–80.

[3] J.-P. Allouche, P. Liardet, Generalized Rudin–Shapiro sequences, Acta Arith. 60 (1991) 1–27.

[4] J.-P. Allouche, J. Shallit, *Automatic Sequences: Theory, Applications, Generalizations*, Cambridge University Press, Cambridge, 2003.

[5] S.V. Avgustinovich, J. Cassaigne, A.E. Frid, Sequences of low arithmetical complexity, RAIRO-Theor. Inf. Appl. 40 (2006) 569–582.

[6] S.V. Avgustinovich, D.G. Fon-Der-Flaass, A.E. Frid, Arithmetical complexity of infinite words, in: M. Ito, T. Imaoka (eds.), *Words, Languages & Combinatorics III*, World Scientific, Singapore, 2003, pp. 51–62.

[7] M. Baake, U. Grimm, *Aperiodic Order. Vol. 1: A Mathematical Invitation*, Cambridge University Press, Cambridge, 2013.

[8] M. Baake, J. A. G. Roberts, R. Yassawi, Reversing and extended symmetries of shift spaces, Discrete Contin. Dynam. Syst. 38 (2018) 835–866.

[9] A. Bustos, D. Luz, N. Mañibo. Admissible reversing and extended symmetries for bijective substitutions, Discrete Comput. Geom. *in press* (2022).

[10] L. Chan, U. Grimm, I. Short, Substitution-based structures with absolutely continuous spectrum, Indag. Math. 29 (2018) 1072–1086.

[11] J. Coquet, T. Kamae, and M. Mendès France, Sur la mesure spectrale de certaines suites arithmétiques, Bull. Soc. Math. France 105 (1977) 369–384.

[12] E. Coven, A. Quas, R. Yassawi, Computing automorphism groups of shifts using atypical equivalence classes, Discrete Analysis. (2016) 611 (24 pp.)

[13] F.M. Dekking, The spectrum of dynamical systems arising from substitutions of constant length, Z. Warscheinlichkeitsth. verw. Geb. 41 (1978) 221-239.

[14] F. Durand, A characterization of substitutive sequences using return words, Discrete Math. 179 (1998) 89–101.

[15] F. Durand, B. Host, C. Skau, Substitutive dynamical systems, Bratteli diagrams and dimension groups, Ergod. Th. & Dynam. Sys. 19 (1999) 953–993.

[16] F. Durand, J. Leroy, Decidability of isomorphism and factorization between minimal substitution subshifts, Discrete Anal. (2022) (65 pp.)

[17] G. Fici, A. Restivo, M. Silva, L. Zamboni, Anti-powers in infinite words, J. Combin. Theor. A 157 (2018) 109–119.

[18] N.P. Frank, Multidimensional constant-length substitution sequences, Topol. Appl. 152 (2005) 44–69.

[19] N.P. Frank, Substitution sequences in $\mathbb{Z}^d$ with a nonsimple Lebesgue component in the spectrum, Ergodic Th. & Dynam. Syst. 23 (2003) 519–532.

[20] N.P. Frank, N. Mañibo, Spectral theory of spin substitutions, Discrete Contin. Dyn. Syst. 42 (2022) 5399–5435.

[21] A.E. Frid, Sequences of linear arithmetical complexity, Theor. Comput. Sci. 339 (2005) 68–87.

[22] A.E. Frid, Arithmetical complexity of symmetric D0L words, Theor. Comput. Sci. 306 (2003) 535–542.

[23] I. Goldstein, Asymptotic subword complexity of fixed points of group substitutions, Theor. Comput. Sci 410 (2009) 2084–2098.

[24] T. Gowers, A new proof of Szemerédi's Theorem, Geom. Funct. Anal. 11 (2001) 465–588.

[25] R.L. Graham, B.L. Rothschild, Ramsey's Theorem for n-Parameter Sets, Trans. Amer. Math. Soc. 159 (1971) 257—292.

[26] J. Kellendonk, R. Yassawi, The Ellis semigroup of bijective substitutions, Groups Geom. Dyn. 16 (2022) 29–73.

[27] A. Klick, N. Strungaru, On higher dimensional arithmetic progressions in Meyer sets, J. Austral. Math. Soc. *in press* (2021).

[28] A. Klick, N. Strungaru, A. Tcaciuc, On arithmetic progressions in model sets, Discrete Comput. Geom. 67 (2022) 930—946.

[29] B. M. Landman, A. Robertson, *Ramsey Theory on the Integers*, American Mathematical Society, USA, 2004.

[30] J.-Y. Lee, R.V. Moody, B. Solomyak, Consequences of pure point diffraction spectra for multiset substitution systems, Discrete Comput. Geom. 29 (2003) 525–560.

[31] M. Lemańczyk, C. Müllner, Automatic sequences are orthogonal to aperiodic multiplicative functions, Discrete Contin. Dynam. Syst. 40 (2020) 6877–6918.

[32] A. de Luca, E.V. Pribavkina, L.Q. Zamboni, A coloring problem for infinite words, J. Combin. Theor. A 125 (2014) 306–332.

[33] J.F. Morgenbesser, J. Shallit, T. Stoll, Thue–Morse at multiples of an integer, J. Number Theory 131 (2011) 1498–1512.

[34] Y. Nagai, S. Akiyama, J.-Y. Lee. On arithmetic progressions in non-periodic self-affine tilings. Ergod. Th. & Dynam. Sys. 42 (2022) 2957 - 2989.

[35] O.G. Parshina, Homogeneous arithmetic progressions in the Thue–Morse word, International Scientific Student Conference, Novosibirsk, Russia, April 2015.

[36] O.G. Parshina, On arithmetic progressions in the generalized Thue–Morse word, In: *Combinatorics on Words: WORDS 2015 Proceedings*, F. Manea, D. Nowotka (eds.), Springer, Cham, 2015, pp. 191–196.

[37] O.G. Parshina, On arithmetic index in the generalized Thue–Morse word, In: *Combinatorics on Words: WORDS 2017 Proceedings*, S. Brlek, F. Dolce, C. Reutenauer, É. Vandomme (eds.) Springer, Cham, 2017, pp. 121–131.

[38] M. Queffélec, *Substitution Dynamical Systems — Spectral Analysis, 2nd ed.*, Springer, Berlin, 2010.

[39] B. Sobolewski, On monochromatic arithmetic progressions in binary words associated with block-counting functions, *preprint*, `arXiv:2204.05287` (2022).

[40] B.L. van der Waerden, Beweis einer Baudetschen Vermutung, Nieuw. Arch. Wisk. 15 (1927) 212–216.

[41] H. Wielandt, Unzerlegbare nichtnegative Matrizen, Math. Z. 52 (1950) 642–645.

[42] C. Wojcik, L.Q. Zamboni, Colouring problems for infinite words, In: *Sequences, Groups, and Number Theory*, V. Berthé, M. Rigo (eds.), Birkhäuser, Basel, 2018, pp. 213–231.

School of Mathematics and Statistics, The Open University, Walton Hall, Milton Keynes MK7 6AA, UK
Email address: ibai.aedo@open.ac.uk

Fakultät für Mathematik, Universität Bielefeld, Postfach 100131, 33501 Bielefeld, Germany
Email address: cmanibo@math.uni-bielefeld.de

School of General Education, Shinshu University, 3-1-1 Asahi, Matsumoto, Nagano, 390-8621, Japan
Email address: ynagai@shinshu-u.ac.jp

School of Computing and Engineering, University of Derby, Kedleston Road, Derby DE22 1GB, UK.
Email address: petra.staynova@gmail.com