

HOW FAR AWAY MUST FORCED LETTERS BE SO THAT SQUARES ARE STILL AVOIDABLE?

MATTHIEU ROSENFELD

ABSTRACT. We describe a new nonconstructive technique to show that squares are avoidable by an infinite word even if we force some letters from the alphabet to appear at certain occurrences. We show that as long as forced positions are at a distance at least 19 (resp., 3, resp., 2) from each other, then we can avoid squares over 3 letters (resp., 4 letters, resp., 6 or more letters). We can also deduce exponential lower bounds on the number of solutions. For our main theorem to be applicable, we need to check the existence of some languages and we explain how to verify that they exist with a computer. We hope that this technique could be applied to other avoidability questions where the good approach seems to be nonconstructive (e.g., the Thue-list coloring number of the infinite path).

1. INTRODUCTION

A square is a word of the form uu where u is a nonempty word. We say that a word is square-free (or avoids squares) if none of its factors is a square. For instance, *hotshots* is a square while *minimize* is square-free. In 1906, Thue showed that there are arbitrarily long ternary words avoiding squares [8]. This result is often regarded as the starting point of combinatorics on words, and the generalizations of this particular question received a lot of attention. The authors of [2] studied three such questions asked by Harju [5]. They also introduced a stronger version of the third problem.

Problem 1 ([2, Problem 4]). Let $p \geq 2$ be an integer, and let $v = v_1v_2v_3 \dots$ be any infinite ternary word. Does there exist an infinite ternary square-free word $w = w_1w_2w_3 \dots$ such that for all i , $w_{p \cdot i} = v_i$?

They give a partial solution to this question and show that the answer is yes for any v if $p \geq 30$. In fact, they showed something slightly stronger. Let $d(\Sigma)$ be the smallest integer such that for all $v \in \Sigma^\omega$ and for all sequence of indices $(p_i)_{1 \leq i}$ such that $\forall i$, $p_{i+1} - p_i \geq d(\Sigma)$, there is an infinite square-free word $u \in \Sigma^\omega$ such that $v_i = u_{p_i}$. They showed that $6 \leq d(\{0, 1, 2\}) \leq 30$. Moreover, the fact that squares are avoidable over 3 letters can be used to show that $d(\{0, 1, 2, 3\}) \leq 7$. We show that $7 \leq d(\{0, 1, 2\}) \leq 19$, $d(\{0, 1, 2, 3\}) = 3$, $d(\{0, 1, 2, \dots, k\}) = 2$ for $k \geq 6$.

The main theorem of this paper gives sufficient conditions for the existence of square-free languages that fulfill some constraints. Kolpakov showed that there are

Received by the editor May 31, 2019, and, in revised form, October 22, 2019, January 7, 2020, and January 30, 2020.

2010 *Mathematics Subject Classification.* Primary 68R15.

Computational resources were provided by the Consortium des Équipements de Calcul Intensif (CÉCI), funded by the Fonds de la Recherche Scientifique de Belgique (F.R.S.-FNRS) under Grant No. 2.5020.11 and by the Walloon Region.

more than 1.30125^n square-free words of length n over a ternary alphabet using a new nonconstructive technique [6]. One of the ideas behind Kolpakov's result is roughly to approximate (using a computer) the language of square-free words by the language of words avoiding squares of period less than l for large l , and to show that we do not lose too many words if we remove the larger squares from this language. We use a similar idea in this paper. We also use ideas from the power series method (see for instance [1, 7]) even if we do not explicitly manipulate any power series. It seems to be a good approach to show that the Thue-list number of paths is 3 (see [3, 4] for definitions and conjectures on this topic) or to tackle other problems that might require a nonconstructive approach.

This paper is organized as follows. We start by fixing some notation in Section 2. In Section 3, we give a weaker version of Theorem 4 to present the ideas of the theorem without some of the technicalities. Then in Section 4, we give the proof of Theorem 4, our main theorem. In Section 5, we explain how to verify with a computer the existence of some languages that are required to apply Theorem 4. Finally, in Section 6, we use Theorem 4 to bound the values of d for different alphabet sizes.

2. DEFINITIONS AND NOTATIONS

We denote the set of nonnegative integers (resp., positive integers) by \mathbb{N}_0 (resp., $\mathbb{N}_{>0}$). For any word $w \in \Sigma^*$, we denote the i th letter of w by w_i and the length of w by $|w|$. Then for any $w \in \Sigma^*$, $w = w_1 w_2 \dots w_{|w|}$. For any set of nonempty words W , we let W^* (resp., W^ω) be the set of words obtained by catenation of finitely many (resp., infinitely many) elements of W . A *language* over an alphabet is a set of finite words over this alphabet. We use the convention that $\prod_{x \in \emptyset} x = 1$ and $\max_{x \in \emptyset} x = 0$ (we could use $-\infty$ for the second one, but it is slightly less convenient for the implementation).

A *partial word over* Σ is a (possibly infinite) word over the alphabet $\Sigma \cup \{\diamond\}$. For any partial word $\mu \in (\Sigma \cup \{\diamond\})^* \cup (\Sigma \cup \{\diamond\})^\omega$ and word $v \in \Sigma^* \cup \Sigma^\omega$, we say that v is *compatible with* μ if $|v| \leq |\mu|$ and $\mu_i \neq \diamond \implies \mu_i = v_i$ for all i such that v_i and μ_i are defined. We denote by $S(\mu)$ the set of square-free words that are compatible with the partial word μ .

3. IDEA OF THEOREM 4

The main theorem of this paper is Theorem 4. The main idea of this theorem is that if a language avoids short squares and is large enough, then it contains square-free words of any length. The statement and proof of this theorem are rather difficult to follow, so in this section we give a version of the theorem for the case where the set W is a singleton $\{w\}$. We hope that this helps to convey the ideas of the proof of Theorem 4. This is, in fact, really similar to the ideas of [7], but instead of building the word letter by letter, we construct it factor by factor. For that we fix one size of a factor and look at the number of words whose length corresponds to multiples of this size.

Theorem 2. *Let Σ be an alphabet, let $w \in (\Sigma \cup \{\diamond\})^*$ be a finite partial word, and let $p \geq 2|w|$ such that $|w|$ divides p . Suppose that there are $C \in \mathbb{N}_{>0}$ and L a language such that:*

$$(I) \quad \varepsilon \in L.$$

- (II) For all $u \in L$, u avoids squares of period less than p .
- (III) For any $u \in L$ there are at least C different words $v \in \Sigma^{|w|}$ compatible with w such that $uv \in L$.
- (IV) There exists $x \in]0, 1[$ such that

$$C \left(1 - \frac{x^{\frac{p}{|w|}-1} |w|^2}{1-x} \right) \geq x^{-1}.$$

Then $S(w^\omega)$ is infinite.

Proof. Let $\mu = w^\omega$. Let $L(\mu)$ be a set of words from L that are compatible with μ such that, for any $u \in L(\mu)$ of length divisible by $|w|$, there are exactly C different words $v \in \Sigma^*$ compatible with w with $uv \in L(\mu)$. Conditions (I), (II), and (III) imply that such a set can be obtained by removing words from L . For all nonnegative i , let $s_i = |S(\mu) \cap L(\mu) \cap \{u \in \Sigma^* : |u| = i|w|\}|$ be the number of square-free words of $L(\mu)$ of length $i|w|$.

We will show by induction on i that for all positive i , $s_{i+1} \geq x^{-1}s_i$. Let n be a positive integer such that

$$(IH1) \quad \forall 0 \leq i < n, s_{i+1} \geq x^{-1}s_i.$$

By the definition of $L(\mu)$, for any word w of $S(\mu) \cap L(\mu)$ there are exactly C different factors v of length $|w|$ such that wv is in $L(\mu)$. Let F be the set of words in $L(\mu) \setminus S(\mu)$ of length $(n+1)|w|$ whose prefix of length $n|w|$ is in $S(\mu) \cap L(\mu)$. Then by definition

$$(1) \quad s_{n+1} \geq Cs_n - |F|.$$

In order to bound $|F|$, let us introduce for all $i < n+1$, $F_i = \{uvvy \in F : |w|(i-1) < |uv| \leq i|w|, |y| < |w|\}$. That is, F_i is the set of words of F that contain a square whose midpoint (the middle of the square) is located between the positions $(i-1)|w|$ and $i|w|$ in the word. Clearly $|F| \leq \sum_{i=1}^n |F_i|$, so our next task is to compute bounds on $|F_i|$ for all i .

Lemma 3. *We have the following inequalities:*

- for all $i > n+1 - \frac{p}{|w|}$, $|F_i| = 0$,
- for all $i \leq n+1 - \frac{p}{|w|}$, $|F_i| \leq s_i C |w|^2$.

Proof. If $i > n+1 - \frac{p}{|w|}$, then $(i-1)|w| + p \geq (n+1)|w|$. Since, L does not contain squares of period less than p , $F_i = \emptyset$.

Now, let $i \leq n+1 - \frac{p}{|w|}$. For any i and $z \in S(\mu) \cap L(\mu) \cap \{u \in \Sigma^* : |u| = i|w|\}$, let $F_i(z)$ be the set of words of F_i that admit z as a prefix.

By definition of F_i , any word $F_i(z)$ contains a square whose second half starts in position $a+1$ and ends in position b where $(i-1)|w| < a \leq i|w|$ and $n|w| < b \leq (n+1)|w|$. Given z , a , and b , we know the first half of the square, and thus the word is known at least up to position $n|w|$. By definition of $L(\mu)$ there are at most C possible values for the remaining $|w|$ letters. By summing over all the values of a and b one gets $|F_i(z)| \leq C|w|^2$. By summing over all the values of z , we finally get $|F_i| \leq s_i C |w|^2$. \square

Now, by (IH1), for all i , $|F_i| \leq C|w|^2 x^{n-i} s_n$, and thus

$$|F| \leq \sum_{i=1}^n |F_i| \leq \sum_{i=1}^{n+1-\frac{p}{|w|}} C|w|^2 x^{n-i} s_n \leq C|w|^2 s_n \sum_{i=\frac{p}{|w|}-1}^{\infty} x^i,$$

$$|F| \leq C|w|^2 s_n \frac{x^{\frac{p}{|w|}-1}}{1-x}.$$

We can use this bound in inequality (1) and we get

$$s_{n+1} \geq C s_n - C|w|^2 s_n \frac{x^{\frac{p}{|w|}-1}}{1-x},$$

$$s_{n+1} \geq s_n C \left(1 - \frac{x^{\frac{p}{|w|}-1} |w|^2}{1-x} \right),$$

$$s_{n+1} \geq x^{-1} s_n \quad (\text{by theorem hypothesis IV}).$$

This concludes the proof that for all positive i , $s_{i+1} \geq x^{-1} s_i$. Since $s_0 = 1$, we deduce that s_i is unbounded and thus $S(w^\omega)$ is infinite. \square

4. THE MAIN THEOREM

This section is devoted to the proof of the main theorem. As already mentioned, the ideas of the proof are the same as for the proof of Theorem 2. However, this is more technical because W is no longer a singleton. Moreover, we want the equivalent of condition (IV) to be as general as possible, and for that we need to bound the size of $|F|$ as tightly as possible. Thus the equivalent of Lemma 3 (Lemma 5) is much more technical and we delay its proof to a later subsection.

Theorem 4. *Let Σ be an alphabet, let $W \subseteq (\Sigma \cup \{\diamond\})^*$ be a finite set of finite partial words, and let $p \geq 2 \max\{|w| : w \in W\}$ be an integer. Suppose that there is a language L and a function $f : \mathbb{N}_{>0} \rightarrow \mathbb{N}_{>0}$ such that:*

- (I) $\varepsilon \in L$.
- (II) *For any $u \in L$ and $w \in W$ there are at least $f(|w|)$ different words $v \in \Sigma^{|w|}$ compatible with w and such that $uv \in L$.*
- (III) *For all $u \in L$, u avoids squares of period less than p .*
- (IV) *For all $u, v \in W$ and integer $1 \leq i \leq |v|$, let*

$$\alpha(|u|, |v|) = \sum_{m=1}^{|u|} \sum_{j=0}^{\lfloor \frac{|v|-1}{m} \rfloor} \min \left\{ f(|v|), (|\Sigma| - 1)^{|v|-1-jm} \right\}$$

and

$$\alpha'(i, |v|) = \sum_{m=0}^{i-1} \min \{ f(|v|), (|\Sigma| - 1)^m \}.$$

There exist $x_1, x_2, \dots, x_{\max\{|w|:w \in W\}} \in]0, 1[$ and $\beta : \{0, \dots, p\} \rightarrow [0, 1]$ solutions of the following system:

$$\left\{ \begin{array}{l} \forall w \in W, \\ f(|w|) - \max_{\substack{u, v \in W \\ 1 \leq r \leq |v|}} \left\{ \beta(r + p - |w| - |v|) \left(\alpha'(r, |w|) + \frac{x_{|v|} \alpha(|u|, |w|)}{1 - x_{|u|}} \right) \right\} \geq x_{|w|}^{-1} \\ \\ \forall j \leq p, \beta(j) = \max \left\{ \prod_{i \in \{|u|:u \in W\}} x_i^{n_i} \mid \begin{array}{l} \forall i \in \{|u|:u \in W\}, n_i \in \mathbb{N}_0, \\ \text{and} \\ \sum_{i \in \{|u|:u \in W\}} i \cdot n_i = j \end{array} \right\} \end{array} \right. .$$

Then for any infinite partial word $\mu \in W^\omega$, $S(\mu)$ is infinite.

Proof. Let $\mu \in W^\omega$ and $(\mu_i)_{i \in \mathbb{N}_{>0}} \in W^{\mathbb{N}_{>0}}$ be a sequence of elements of W such that $\mu = \mu_1 \mu_2 \mu_3 \dots$. For any integer i , let $l(i) = |\mu_1 \dots \mu_i|$. Let $L(\mu)$ be a set of words from L that are compatible with μ such that, for any $u \in L(\mu)$ of length $|\mu_1 \dots \mu_j|$, there are exactly $f(|\mu_{j+1}|)$ different words $v \in \Sigma^*$ compatible with μ_{j+1} with $uv \in L(\mu)$. That is, we remove words from L in order to replace the “at least $f(|\mu_{j+1}|)$ ” by “exactly $f(|\mu_{j+1}|)$ ”. For all nonnegative i , let $s_i = |S(\mu) \cap L(\mu) \cap \{v \in \Sigma^* : |v| = i\}|$ be the number of square-free words of $L(\mu)$ of length i .

We will show by induction on i that for all positive i , $s_{l(i+1)} \geq x_{|\mu_{i+1}|}^{-1} s_{l(i)}$.

Let n be a positive integer such that

$$(IH1) \quad \forall 0 \leq i < n, s_{l(i+1)} \geq x_{|\mu_{i+1}|}^{-1} s_{l(i)}.$$

By the definition of $L(\mu)$, for any word w of $S(\mu) \cap L(\mu)$ of length $l(n)$ there are exactly $f(|\mu_{n+1}|)$ different factors v of length $|\mu_{n+1}|$ such that wv is in $L(\mu)$. Let F be the set of words in $L(\mu) \setminus S(\mu)$ of length $l(n+1)$ whose prefix of length $l(n)$ is in $S(\mu) \cap L(\mu)$. Then by definition

$$(2) \quad s_{l(n+1)} \geq f(|\mu_{n+1}|) s_{l(n)} - |F|.$$

In order to bound $|F|$, let us introduce for all $i < n+1$, $F_i = \{uvvw \in F : l(i-1) < |uv| \leq l(i), |w| < |\mu_{n+1}|\}$. That is, F_i is the set of words of F that contain a square whose midpoint (the middle of the square) is located between the positions $|\mu_1 \dots \mu_i|$ and $|\mu_1 \dots \mu_{i-1}|$ in the word. Clearly $|F| \leq \sum_{i=1}^n |F_i|$, so our next task is to compute the values of $|F_i|$ for all i . Let d be the smallest integer such that $|\mu_{d+1} \dots \mu_{n+1}| \leq p$, and let $r = |\mu_d \mu_{d+1} \dots \mu_{n+1}| - p$. Recall that $r > 0$.

Lemma 5. *We have the following inequalities:*

- for all $i > d$, $|F_i| = 0$,
- $|F_d| \leq s_{l(d)} \alpha'(r, |\mu_{n+1}|)$,
- for all $i \leq d$, $|F_i| \leq s_{l(i)} \alpha(|\mu_i|, |\mu_{n+1}|)$.

The proof of this lemma is not really informative and is mostly a rather technical counting argument, so we moved it to Section 4.1.

We can use the bounds on the sizes of the F_i 's to bound $|F|$ as in the following lemma.

Lemma 6. *We have*

$$|F| \leq s_{l(d)} \max_{u \in W} \left\{ \alpha'(r, |\mu_{n+1}|) + \frac{x_{|\mu_d|} \alpha(|u|, |\mu_{n+1}|)}{1 - x_{|u|}} \right\}.$$

Proof. First, let us show by induction on i that for all $0 \leq i < d$

$$(IH2) \quad \sum_{j=0}^i s_{l(j)} \alpha(|\mu_j|, |\mu_{n+1}|) \leq s_{l(i)} \max_{u \in W} \left\{ \frac{\alpha(|u|, |\mu_{n+1}|)}{1 - x_{|u|}} \right\}.$$

Let us first show that this is true with $i = 0$, using the fact that $\mu_0 \in]0, 1[$:

$$s_{l(0)} \alpha(|\mu_0|, |\mu_{n+1}|) \leq s_{l(0)} \frac{\alpha(|\mu_0|, |\mu_{n+1}|)}{1 - x_{|\mu_0|}} \leq s_{l(0)} \max_{u \in W} \left\{ \frac{\alpha(|u|, |\mu_{n+1}|)}{1 - x_{|u|}} \right\}.$$

Now, let $i + 1$ be an integer such that (IH2) is true for i :

$$\begin{aligned} \sum_{j=0}^{i+1} s_{l(j)} \alpha(|\mu_j|, |\mu_{n+1}|) &\leq s_{l(i+1)} \alpha(|\mu_{i+1}|, |\mu_{n+1}|) + s_{l(i)} \max_{u \in W} \left\{ \frac{\alpha(|u|, |\mu_{n+1}|)}{1 - x_{|u|}} \right\} \\ &\leq s_{l(i+1)} \left(\alpha(|\mu_{i+1}|, |\mu_{n+1}|) + x_{|\mu_{i+1}|} \max_{u \in W} \left\{ \frac{\alpha(|u|, |\mu_{n+1}|)}{1 - x_{|u|}} \right\} \right) \quad (\text{by (IH1)}) \\ &\leq s_{l(i+1)} \max_{u \in W} \left\{ \frac{\alpha(|u|, |\mu_{n+1}|)}{1 - x_{|u|}} \right\}. \end{aligned}$$

Thus equation (IH2) is true for all $i \leq d$ and, in particular, for $i = d - 1$, and we get

$$\begin{aligned} |F| &\leq |F_d| + s_{l(d-1)} \max_{u \in W} \left\{ \frac{\alpha(|u|, |\mu_{n+1}|)}{1 - x_{|u|}} \right\}, \\ |F| &\leq s_{l(d)} \alpha'(r, |\mu_{n+1}|) + s_{l(d)} x_{|\mu_d|} \max_{u \in W} \left\{ \frac{\alpha(|u|, |\mu_{n+1}|)}{1 - x_{|u|}} \right\}, \\ |F| &\leq s_{l(d)} \max_{u \in W} \left\{ \alpha'(r, |\mu_{n+1}|) + \frac{x_{|\mu_d|} \alpha(|u|, |\mu_{n+1}|)}{1 - x_{|u|}} \right\}. \end{aligned}$$

This concludes the proof of this lemma. \square

By induction hypothesis (IH1) $s_{l(d)} \leq s_{l(n)} \prod_{i=d+1}^n x_{|\mu_j|}$. Let us bound the product on the right hand side:

$$\begin{aligned} \prod_{i=d+1}^n x_{|\mu_j|} &\leq \max \left\{ \prod_{i \in \{ |u| : u \in W \}} x_i^{n_i} \mid \begin{array}{l} \forall i \in \{ |u| : u \in W \}, n_i \in \mathbb{N}_0 \\ \text{and} \\ \sum_{i \in \{ |u| : u \in W \}} i \cdot n_i = l(n+1) - l(d) - \mu_{n+1} \end{array} \right\}, \\ \prod_{i=d+1}^n x_{|\mu_j|} &\leq \beta (l(n+1) - l(d) - \mu_{n+1}), \\ \prod_{i=d+1}^n x_{|\mu_j|} &\leq \beta (r + p - \mu_{n+1} - \mu_d). \end{aligned}$$

Now, using this equation with Lemma 6 gives

$$|F| \leq s_{l(n)} \beta (r + p - \mu_{n+1} - \mu_d) \max_{u \in W} \left\{ \alpha'(r, |\mu_{n+1}|) + \frac{x_{|\mu_d|} \alpha(|u|, |\mu_{n+1}|)}{1 - x_{|u|}} \right\}.$$

Now recall that $r = |\mu_d \mu_{d+1} \dots \mu_{n+1}| - p$ and thus by definition of d , $1 \leq r \leq \mu_d$. We deduce that

$$|F| \leq s_{l(n)} \max_{\substack{u, v \in W \\ 1 \leq r \leq |v|}} \left\{ \beta(r + p - \mu_{n+1} - |v|) \left(\alpha'(r, |\mu_{n+1}|) + \frac{x_{|v|} \alpha(|u|, |\mu_{n+1}|)}{1 - x_{|u|}} \right) \right\}.$$

We can finally replace $|F|$ by this bound in inequality (2), and we get

$$\begin{aligned} s_{l(n+1)} &\geq s_{l(n)} \left(f(|\mu_{n+1}|) \right. \\ &\quad \left. - \max_{\substack{u, v \in W \\ r \in \{1, \dots, |v|\}}} \left\{ \beta(r + p - \mu_{n+1} - |v|) \left(\alpha'(r, |\mu_{n+1}|) + \frac{x_{|v|} \alpha(|u|, |\mu_{n+1}|)}{1 - x_{|u|}} \right) \right\} \right), \\ s_{l(n+1)} &\geq s_{l(n)} x_{|\mu_{n+1}|}^{-1} \quad (\text{by theorem hypothesis (IV)}). \end{aligned}$$

Moreover, $s_0 = 1$, and thus for all i , $s_{|\mu_1 \dots \mu_i|} \geq \prod_{j=1}^i x_{|\mu_j|}^{-1}$. For all j , $x_{|\mu_j|}^{-1} > 1$, so we conclude that $S(\mu)$ is infinite. \square

Remark that Theorem 4 is far from sharp. One could improve the bounds given by Lemma 5. This could be done by lowering α and α' or by introducing a third coefficient α'' for the second nonempty F_i . However, we were not able to obtain significant improvements that were worth the additional technicalities.

In Section 5 we explain how to verify with a computer that there exists a language L that satisfies conditions (I), (II), and (III). We also need a way to verify condition (IV). In order to compute β , we can use that $\beta(0) = 1$ and, for all $j \in \{1, \dots, p\}$, $\beta(j) = \max \{x_{|u|} \beta(j - |u|) : u \in W, |u| \leq j\}$. Thus given the values of the x_i , one can compute β using a dynamic algorithm and all the rest is straightforward to compute. Thus it is easy to verify with a computer whether or not a given set of values of x_i is a solution. We provide a C++ program that takes as input $|\Sigma|$, k , p , f , and x_1, \dots, x_k and verifies whether this is a solution of the equations of condition (IV).

4.1. Proof of Lemma 5. This subsection is dedicated to the proof of Lemma 5. Recall that the statement and proof are not self-contained since some of the notation are defined in the proof of Theorem 4.

Lemma 5. *We have the following inequalities:*

- for all $i > d$, $|F_i| = 0$,
- $|F_d| \leq s_{l(d)} \alpha'(r, |\mu_{n+1}|)$,
- for all $i \leq d$, $|F_i| \leq s_{l(i)} \alpha(|\mu_i|, |\mu_{n+1}|)$.

Proof. If $i > d$, then by definition $|\mu_i \dots \mu_{n+1}| \leq p$. Moreover, L does not contain squares of period less than p , and thus $F_i = \emptyset$.

Now, let $i \leq d$. By definition, any word from F_i can be written $uvvy$ with $l(i-1) < |uv| \leq l(i)$, $|y| < |\mu_{n+1}|$. For any i and $z \in S(\mu) \cap L(\mu) \cap \{u \in \Sigma^* : |u| = l(i)\}$, let $F_i(z)$ be the set of words of F_i that admit z as a prefix. Clearly $F_i = \sum_{z \in S(\mu) \cap L(\mu) \cap \{u \in \Sigma^* : |u| = l(i)\}} F_i(z)$.

Let a, b (resp., a', b') be integers such that there is an element $z' \in F_i(z)$ that contains a square starting at a (resp., a') and of period b (resp., b') with $l(i-1)+1 < a' + b' = a + b \leq |z| + 1$ and $a > a'$. Because of the square in z' , we know that for

all $0 \leq j \leq |z| - a - b$, $z_{a+j} = z_{a+b+j} = z_{a'+b'+j} = z_{a'+j}$. If $a \leq a' + |z| - a - b + 1$, then z contains a square which is not possible. Hence

$$\begin{aligned} a &> a' + |z| - a - b + 1, \\ a + b + b' &> a' + b + b' + |z| - a - b + 1, \\ (3) \quad a' + 2b' &> a + 2b + |z| - a - b + 1 \quad \text{since } a + b = a' + b'. \end{aligned}$$

Let $u \in F_i(z)$ be a word that contains a square starting at a and of period b ; then we know its suffix of size $a + 2b - 1 > l(n)$. Thus there are at most $f(|\mu_{n+1}|)$ possibilities. Moreover, since the size of the unknown suffix is $l(n+1) + 1 - a - 2b < p$, it is square-free and there are at most $(|\Sigma| - 1)^{l(n+1) + 1 - a - 2b}$ possibilities. Thus for a fixed z and value of $a + b$, the number of ways to add a suffix to z to obtain an element of F_i that contains a square of period b starting at a is

$$\begin{aligned} &\max_{\substack{j \in \mathbb{N}_0, l(n)+2 \leq p_1, \dots, p_j \leq l(n+1)+1 \\ \forall i < j, p_{i+1} > p_i + |z| - a - b + 1}} \left\{ \sum_{i=1}^j \min \left\{ f(|\mu_{n+1}|), (|\Sigma| - 1)^{l(n+1) + 1 - p_i} \right\} \right\} \\ &= \max_{\substack{j \in \mathbb{N}_0, 0 \leq p_1, \dots, p_j \leq |\mu_{n+1}| - 1 \\ \forall i < j, p_{i+1} < p_i + |z| - a - b + 1}} \left\{ \sum_{i=1}^j \min \left\{ f(|\mu_{n+1}|), (|\Sigma| - 1)^{|\mu_{n+1}| - 1 - p_i} \right\} \right\}. \end{aligned}$$

But since $\min \left\{ f(|\mu_{n+1}|), (|\Sigma| - 1)^{|\mu_{n+1}| - 1 - p_i} \right\}$ is a nonincreasing function in p_i , the maximum is reached when we pack the p_i as much as we can on the lowest values of p_i . Thus this quantity is in fact equal to

$$\sum_{j=0}^{\left\lfloor \frac{|\mu_{n+1}| - 1}{|z| - a - b + 2} \right\rfloor} \min \left\{ f(|\mu_{n+1}|), (|\Sigma| - 1)^{|\mu_{n+1}| - 1 - j(|z| - a - b + 2)} \right\}.$$

Now we can sum over all the values of $a + b$ and we get

$$\begin{aligned} |F_i(z)| &\leq \sum_{a+b=|z|+2-|\mu_i|}^{|z|+1} \sum_{j=0}^{\left\lfloor \frac{|\mu_{n+1}| - 1}{|z| - a - b + 2} \right\rfloor} \min \left\{ f(|\mu_{n+1}|), (|\Sigma| - 1)^{|\mu_{n+1}| - 1 - j(|z| - a - b + 2)} \right\} \\ &\leq \sum_{a+b=0}^{|\mu_i|-1} \left(\sum_{j=0}^{\left\lfloor \frac{|\mu_{n+1}| - 1}{|\mu_i| - a - b} \right\rfloor} \min \left\{ f(|\mu_{n+1}|), (|\Sigma| - 1)^{|\mu_{n+1}| - 1 - j(|\mu_i| - a - b)} \right\} \right) \\ &\leq \sum_{m=1}^{|\mu_i|} \left(\sum_{j=0}^{\left\lfloor \frac{|\mu_{n+1}| - 1}{m} \right\rfloor} \min \left\{ f(|\mu_{n+1}|), (|\Sigma| - 1)^{|\mu_{n+1}| - 1 - jm} \right\} \right), \end{aligned}$$

$$|F_i(z)| \leq \alpha(|\mu_i|, |\mu_{n+1}|).$$

Summing over all the possible z yields $|F_i| \leq s_{l(i)} \alpha(|\mu_i|, |\mu_{n+1}|)$.

The remaining case is $i = d$. Once again, let a, b (resp., a', b') be integers such that there is an element of $F_d(z)$ that contains a square starting at a (resp., a') and of period b (resp., b') with $l(d-1) + 1 < a' + b' = a + b \leq l(n+1) + 1 - p$ and $a > a'$. By the definition of d , $l(n+1) - l(d) \leq p$. We can use equation (3) again and we get

$$a' + 2b' > b + |z| + 1 \geq p + |z| + 1 \geq l(n+1) + 1.$$

This is a contradiction with the fact that $a' + 2b' \leq l(n+1) + 1$. Thus given the value of $a + b$ there is at most one possible value for a and b . The number of ways for a fixed z and value of $a + b$ to complete z with a suffix into an element of F_d is at most

$$\begin{aligned} & \max \left\{ \min \left\{ f(|\mu_{n+1}|), (|\Sigma| - 1)^{l(n+1)+1-s} \right\} : \begin{array}{l} s \in \mathbb{N}_{>0}, s \leq l(n+1) + 1, \\ a + b + p \leq s, \end{array} \right\} \\ & \leq \min \left\{ f(|\mu_{n+1}|), (|\Sigma| - 1)^{l(n+1)+1-a-b-p} \right\}. \end{aligned}$$

Then by summing over all the possible values of $a + b$, we get

$$|F_d(z)| \leq \sum_{a+b=l(d-1)+2}^{l(n+1)+1-p} \min \left\{ f(|\mu_{n+1}|), (|\Sigma| - 1)^{l(n+1)+1-a-b-p} \right\}.$$

We can use the variable substitution $m = l(n+1) + 1 - a - b - p$ and remark that $l(n+1) + 1 - p - (l(d-1) + 2) = |\mu_d \mu_{d+1} \dots \mu_{n+1}| - p - 1 = r - 1$, and we get

$$|F_d(z)| \leq \sum_{m=0}^{r-1} \min \{ f(|\mu_{n+1}|), (|\Sigma| - 1)^m \} \leq \alpha'(r, |\mu_{n+1}|).$$

We conclude that $|F_d| \leq s_{l(d)} \alpha'(r, |\mu_{n+1}|)$ by summing over all z . \square

5. FINDING A SET L THAT SATISFIES THEOREM 4

In this section, we explain how to verify the existence of a language that fulfills conditions (I), (II), and (III) of Theorem 4.

We consider some particular directed labeled graphs: $G(V, A)$ is a set V of vertices together with a set $A \subseteq (V \times V \times \Sigma)$ of labeled arcs. For any $u, v \in V$ and $a \in \Sigma$, $(u, v, a) \in A$ is an arc from u to v with label a . These graphs could also be seen as finite state machines where all the states are initial and final.

The Rauzy graph of length n of a factorial language L over Σ is the graph $G(V, A)$ where $V = L \cap \Sigma^n$ and $E = \{(au, ub, b) : aub \in L, a, b \in \Sigma\}$. For any graph $G(V, A)$ and any set $X \in V$, we denote by $G[X]$ the subgraph induced by X .

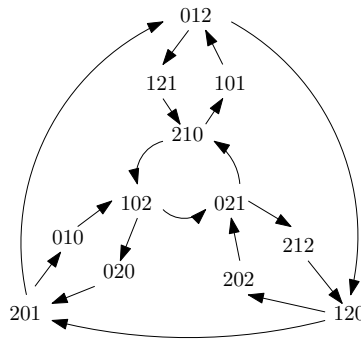
Let $R_p(\Sigma)$ be the Rauzy graph of length $2p - 3$ of the square-free words over Σ . Recall, that the factors of length $2p - 2$ of any walk on $R_p(\Sigma)$ correspond to edges of $R_p(\Sigma)$ and that by definition they are square free. Thus, the sequence of labels of any walk on $R_p(\Sigma)$ avoids squares of period less than p , but they can contain longer squares. We let $S_p(\Sigma)$ be the set of words that contains no square of period less than p (from the previous remark, $S_p(\Sigma)$ can also be seen as the set of walks on $R_p(\Sigma)$).

As an illustration, we give $R_3(\{0, 1, 2\})$ in Figure 1 without the arc labels.

For this section, we abuse the notation and allow ourself to identify words and sequences.

For any graph $G(V, A)$ and partial word $w \in (\Sigma \cup \{\diamond\})^*$, we inductively define for any integer $i \in \{0, \dots, |w|\}$ and vertex $v \in V$:

$$p_{i,w,G}(v) = \begin{cases} 1 & \text{if } i = 0, \\ \sum_{\substack{(v,u,a) \in A \\ a \in \Sigma}} p_{i-1,w,G}(u) & \text{if } w_{|w|+1-i} = \diamond, \\ \sum_{(v,u,w_{|w|+1-i}) \in A} p_{i-1,w,G}(u) & \text{otherwise.} \end{cases}$$

FIGURE 1. The Rauzy graph $R_3(\{0, 1, 2\})$

Intuitively, $p_{i,w,G}(v)$ gives the number of walks of length i starting from v that are compatible with the i last letters of w . Indeed, there is one walk of length 0, and we always take the transition that is labeled by the current letter of w and any transition if this letter is \diamond . Remark that in the third case, there are, in fact, either 0 or 1 summands in the sum.

Lemma 7. *Let $W \subseteq (\Sigma \cup \{\diamond\})^*$, let $G(V, A) = R_p(\Sigma)$, let $f : \mathbb{N}_{>0} \rightarrow \mathbb{N}_{>0}$, and let there be a non-empty set $X \subseteq V$. If for all $v \in X$ and $w \in W$, $p_{|w|,w,G[X]}(v) \geq f(|w|)$, then there exists a language L such that:*

- $\varepsilon \in L$,
- for all $u \in L$, u avoids squares of period less than p ,
- for any $u \in L$ and $w \in W$ there are at least $f(|w|)$ different words $v \in \Sigma^{|w|}$ compatible with w such that $uv \in L$.

Proof. Let L be the set of sequences of labels that correspond to a walk in $G[X]$. By definition, the two first conditions on L are fulfilled.

Let $u \in L$, and let $w \in W$. If $|u| \geq 2p - 3$, we let u' be the suffix of length $2p - 3$ of u . Otherwise, we let $u' \in L$ such that $|u'| = 2p - 3$ and w is a suffix of u' (there is such an element in L). Each walk of length $|w|$ starting in u' gives a unique sequence of labels u'' such that uu'' contains no square of period less than p .

We easily deduce, by induction on i , that for all v the number of walks of length i starting at v that are compatible with $w_{|w|-i+1}w_{|w|-i+2} \dots w_{|w|}$ is at least $p_{i,w,G[X]}(v)$. So, in particular, the number of walks of length $|w|$ starting at u' and compatible with w is at least $p_{|w|,w,G[X]}(u') \geq f(|w|)$. This concludes the proof. \square

In fact, we need something stronger because for the values of p that we use, the graphs R_p are too big to fit in a computer. We can exploit symmetries of $R_p(\Sigma)$ to work on a smaller equivalent graph.

For any square-free word $w \in \Sigma^*$, we let $\Psi(w)$ be the shortest suffix of w such that for all $i \in \{\lceil \frac{|\Psi(w)|}{2} \rceil + 1, \dots, p - 1\}$ there exists $k \in \{0, 1, \dots, |\Psi(w)| - i - 1\}$ with $\Psi(w)_{|\Psi(w)|-k} \neq \Psi(w)_{|\Psi(w)|-k-i}$. If $|w| = 2p - 3$, then w is such a suffix of itself (since $\{\lceil \frac{|w|}{2} \rceil + 1, \dots, p - 1\}$ is empty), and thus there is always a shortest suffix.

For instance, with $p = 5$ we have $\Psi(0210120) = 10120$. For any letter α , the word 0210120α is square free if and only if 10120α is square free. Indeed, a new square is necessarily a suffix of 0210120α . It is enough to look at the two letters in bold in 10120α to deduce that there is no square of length 4 and all the other possible suffixes of even length of 0210120α are also suffixes of 10120α . In fact, for any w of size $2p - 3$, the word wa avoids squares if and only if $\Psi(w)a$ avoids squares (this is proven in the next lemma), and this is the main motivation behind the definition of Ψ (in particular, words with the same image by Ψ can be extended in the same way).

For any graph $G(V, A)$, let $\Psi(G)(\Psi(V), A')$ be the graph such that $A' = \Psi(A) = \{(\Psi(a), \Psi(b), c) : (a, b, c) \in A\}$. The next lemma tells us that we only need to consider the walks on $\Psi(G)$ instead of the walks on G .

Lemma 8. *Let p be a positive integer, and let $w \in (\Sigma \cup \{\diamond\})^*$, let $G(V, A) = R_p(\Sigma)$, and let $X \subseteq \Psi(V)$. Also, let $\Psi^{-1}(X) = \{x \in V : \Psi(x) \in X\}$. Then for all $v \in \Psi^{-1}(X)$, $p_{|w|, w, G[\Psi^{-1}(X)]}(v) = p_{|w|, w, \Psi(G)[X]}(\Psi(v))$.*

Proof. Let us first show that for any $a \in \Sigma$ and $v \in S_p(\Sigma)$ with $|v| = 2p - 3$ if $\Psi(v)a \in S_p(\Sigma)$, then $va \in S_p(\Sigma)$. Let us show that under these assumptions, for any i , va avoids squares of period i . Since v is square free, we only need to show that no suffix of va is a square. We have to distinguish between two cases:

- $2i \leq |\Psi(v)| + 1$. Suppose for the sake of contradiction that there is a square of period i in va . We deduce that the suffix of length $|\Psi(v)| + 1$ of va contains a square of period i . That is, $\Psi(v)a$ contains a square of period i , which is a contradiction.
- $2i \geq |\Psi(v)| + 2$. Since i is an integer we get $i \geq \lceil \frac{|\Psi(v)|}{2} \rceil + 1$. Moreover, $|va| = 2p - 2$ and thus $i \leq p - 1$. Thus by the definition of $\Psi(v)$, there exists $k \in \{0, 1, \dots, |\Psi(v)| - i - 1\}$ such that $\Psi(v)_{|\Psi(v)|-k} \neq \Psi(v)_{|\Psi(v)|-k-i}$. Thus there is $k \in \{1, \dots, |\Psi(v)a| - i - 1\}$ such that $(\Psi(v)a)_{|\Psi(v)a|-k} \neq (\Psi(v)a)_{|\Psi(v)a|-k-i}$. Recall that $|\Psi(v)a| - i - 1 = |\Psi(v)| - i \leq 2i - 2 - i = i - 2$. We conclude that there is $k \in \{0, \dots, i - 2\}$ such that $(va)_{|va|-k} \neq (va)_{|va|-k-i}$. This implies that the suffix of length $2i$ of va is not a square of period i .

We deduce that for any $a \in \Sigma$ and $v \in S_p(\Sigma)$ with $|v| = 2p - 3$ if $\Psi(v)a \in S_p(\Sigma)$, then $va \in S_p(\Sigma)$.

Let $u \in V$, $v \in \Psi(V)$ and $a \in \Sigma$ such that $(\Psi(u), v, a) \in \Psi(A)$. By the definition of $\Psi(A)$ this implies that there is $(u', v', a) \in A$ with $\Psi(u') = \Psi(u)$ and $\Psi(v') = v$. Thus $u'a \in S_p(\Sigma)$ and $\Psi(u)a = \Psi(u')a \in S_p(\Sigma)$. From the previous paragraph, it implies that ua is square free. Let us show that $\Psi(u_2u_3 \dots u_{|u|}a) = v$. By definition, for all $i \in \{\lceil \frac{|\Psi(u')|}{2} \rceil + 1, \dots, p - 1\}$ there exists $k \in \{0, 1, \dots, |\Psi(u')| - i - 1\}$ with $\Psi(u')_{|\Psi(u')|-k} \neq \Psi(u')_{|\Psi(u')|-k-i}$. We easily deduce that for all $i \in \{\lceil \frac{|\Psi(u')a|}{2} \rceil + 1, \dots, p - 1\}$ there exists $k \in \{0, 1, \dots, |\Psi(u'a)| - i - 1\}$ with $\Psi(u'a)_{|\Psi(u'a)|-k} \neq \Psi(u'a)_{|\Psi(u'a)|-k-i}$. This implies that $v = \Psi(v')$ is a suffix of $\Psi(u')a$. Since $\Psi(u')a = \Psi(u)a$, we deduce that v is also a suffix of $u_2u_3 \dots u_{|u|}a$. Since $v = \Psi(v')$ and v is a suffix of $u_2u_3 \dots u_{|u|}a$, we get that $\Psi(u_2u_3 \dots u_{|u|}a) = v$. We showed that if there are $u \in V$, $v \in \Psi(V)$, and $a \in \Sigma$ such that $(\Psi(u), v, a) \in \Psi(A)$, then there exists v'' such that $(u, v'', a) \in A$ and $\Psi(v'') = v$.

We deduce that for all $u \in V$, $\{(\Psi(u), \Psi(v), a) \in \Psi(A)\} \subseteq \{(\Psi(u), \Psi(v), a) : (u, v, a) \in A\}$. The other inclusion is clear from the definition of $\Psi(A)$ and we get

for all $u \in V$ that

$$(4) \quad \{(\Psi(u), \Psi(v), a) \in \Psi(A)\} = \{(\Psi(u), \Psi(v), a) : (u, v, a) \in A\}.$$

By the definition of $R_p(\Sigma)$, for any u there is at most one outgoing arc for every label in the set on the right. Since the two sets are equals, we deduce that every vertex of the graph $\Psi(G)$ has at most one outgoing arc for any label. Intuitively, (4) implies (by induction on the length of the walk) that for any $u, v \in V$ the set of labeled walks from u to v in G is equal to the set of labeled walks from $\Psi(u)$ to $\Psi(v)$ in $\Psi(G)$. We are now ready to show by induction on i that for all $i \in \{0, \dots, |w|\}$ and $v \in \Psi^{-1}(X)$, $p_{i,w,G[\Psi^{-1}(X)]}(v) = p_{i,w,\Psi(G)[X]}(\Psi(v))$. By definition $p_{0,w,G[\Psi^{-1}(X)]}(v) = 1 = p_{0,w,\Psi(G)[X]}(\Psi(v))$.

Let n be a positive integer such that for all $v \in \Psi^{-1}(X)$,

$$(IH) \quad \forall i < n, p_{i,w,G[\Psi^{-1}(X)]}(v) = p_{i,w,\Psi(G)[X]}(\Psi(v)).$$

Then, for all $v \in \Psi^{-1}(X)$, if $w|_{w|+1-i} \neq \diamond$ we get

$$\begin{aligned} p_{i,w,G[\Psi^{-1}(X)]}(v) &= \sum_{\substack{(v,u,w|_{w|+1-i}) \in A \\ u \in \Psi^{-1}(X)}} p_{i-1,w,G[\Psi^{-1}(X)]}(u), \\ p_{i,w,G[\Psi^{-1}(X)]}(v) &= \sum_{\substack{(v,u,w|_{w|+1-i}) \in A \\ \Psi(u) \in X}} p_{i-1,w,\Psi(G)[X]}(\Psi(u)) \quad (\text{from (IH)}), \\ p_{i,w,G[\Psi^{-1}(X)]}(v) &= \sum_{\substack{(\Psi(v), \Psi(u), w|_{w|+1-i}) \in \Psi(A) \\ \Psi(u) \in X}} p_{i-1,w,\Psi(G)[X]}(\Psi(u)) \quad (\text{from (4)}), \\ p_{i,w,G[\Psi^{-1}(X)]}(v) &= \sum_{\substack{(\Psi(v), u, w|_{w|+1-i}) \in \Psi(A) \\ u \in X}} p_{i-1,w,\Psi(G)[X]}(u), \\ p_{i,w,G[\Psi^{-1}(X)]}(v) &= p_{i,w,\Psi(G)[X]}(\Psi(v)). \end{aligned}$$

The case where $w|_{w|+1-i} = \diamond$ is similar. □

Using Lemma 7 together with Lemma 8, we get the following lemma.

Lemma 9. *Let $W \subseteq (\Sigma \cup \{\diamond\})^*$, let $G(V, A) = R_p(\Sigma)$, let $f : \mathbb{N}_{>0} \rightarrow \mathbb{N}_{>0}$, and let $X \subseteq \Psi(V)$ be a nonempty set. If for all $v \in X$ and $w \in W$, $p_{|w|,w,\Psi(G)[X]}(v) \geq f(|w|)$, then there exists a language L such that:*

- $\varepsilon \in L$,
- for all $u \in L$, u avoids squares of period less than p ,
- for any $u \in L$ and $w \in W$ there are at least $f(|w|)$ different words $v \in \Sigma^{|w|}$ compatible with w and such that $uv \in L$.

The graph $\Psi(R_p(\Sigma))$ is much smaller than $R_p(\Sigma)$, and we can use a computer to check the conditions of this lemma for the values of p that we used. One should first find the graph $\Psi(R_p(\Sigma))$. The following fact allows us to easily compute the set of vertices of $\Psi(R_p(\Sigma))$ without computing $R_p(\Sigma)$.

Lemma 10. *Let $w \in S_p(\Sigma)$. Then $w \in \Psi(S_p(\Sigma) \cap \Sigma^{2p-3})$ if and only if w is the smallest nonempty suffix of w such that $\Psi(w) = w$.*

Moreover, given a graph G , the definition of $p_{|w|,w,G}$ gives a trivial dynamic algorithm that computes $p_{|w|,w,G}$ in time $O(|\Sigma| \cdot |w| \cdot |G|)$. Starting with $X = \Psi(R_p(\Sigma))$ and inductively removing from X all the vertices for which $p_{|w|,w,\Psi(R_p(\Sigma))[X]} < f(|w|)$ gives the largest subgraph that meets the conditions of Lemma 9. As long as this subgraph is not empty one can then apply Lemma 9. Algorithm 1 computes the largest subgraph of $\Psi(G)$ with the required property.

Algorithm 1: How to compute the subgraph of $\Psi(G)$.

Input : The graph $\Psi(G)$, the set W
Output: The largest set $X \subseteq \Psi(V)$ such that for all $v \in X$ and $w \in W$,
 $p_{|w|,w,\Psi(G)[X]}(v) \geq f(|w|)$
 $X = \Psi(V)$;
 $todo := true$;
while $todo$ **do**
 $todo := false$;
 foreach $w \in W$ **do**
 compute $p_{|w|,w,\Psi(G)[X]}$;
 $X' := \{v \in X : p_{|w|,w,\Psi(G)[X]}(v) \geq f(|w|)\}$;
 if $X \neq X'$ **then**
 $X := X'$;
 $todo := true$;
return X ;

6. APPLICATION OF THEOREM 4

In this section we apply Theorem 4. We provide a C++ implementation of Algorithm 1 that verifies the existence of the language that fulfill conditions (I), (II), and (III) from Theorem 4. Condition (IV) can be easily verified (as long as solutions are given), and we also provide a C++ code to do that.¹

Theorem 11. *For any alphabet Σ , let $d(\Sigma)$ be the smallest integer such that for all $v \in \Sigma^\omega$ and for all sequences $(p_i)_{1 \leq i}$ such that $\forall i, p_{i+1} - p_i \geq d(\Sigma)$, there is an infinite square-free word $u \in \Sigma^\omega$ such that $v_i = u_{p_i}$. Then:*

- $7 \leq d(\{0, 1, 2\}) \leq 19$,
- $d(\{0, 1, 2, 3\}) = 3$,
- $2 \leq d(\{0, 1, 2, 3, 4\}) \leq 3$,
- *if* $|\Sigma| \geq 6$, $d(\Sigma) = 2$.

Proof. For any alphabet Σ , we have $d(\Sigma) \geq 2$. Moreover, d is a decreasing function of the size of the alphabet. Thus the third statement can easily be deduced from the second one. We show the remaining statements independently of each other. We will show the upper bounds using Algorithm 1, Lemma 9, and Theorem 4. The lower bounds are verified by an exhaustive search.

If $|\Sigma| \geq 6$, $d(\Sigma) = 2$: Let $\Sigma = \{0, 1, 2, 3, 4, 5\}$, and let $W = \{\diamond\} \cup \{\diamond a \diamond : a \in \Sigma\} \cup \{\diamond a \diamond b : a, b \in \Sigma\}$. We can use Algorithm 1 to check that we can apply Lemma 9

¹The codes can be found in the ancillary files of <https://arxiv.org/abs/1903.04214>.

with $f(1) = 3$, $f(3) = f(4) = 6$, $p = 12$. Thus conditions (I), (II), and (III) of Theorem 4 are fulfilled. We can check with a computer that condition (IV) of Theorem 4 is also fulfilled with $x_1 = \frac{2}{5}$ and $x_3 = x_4 = \frac{1}{4}$. This implies that for any $\mu \in W^\omega$ there are infinite square-free words over Σ compatible with μ . Since $\{\diamond^i a : i \geq 1, a \in \Sigma\}^\omega \subseteq W^\omega$, we deduce that for any $\mu \in \{\diamond^i a : i \geq 1, a \in \Sigma\}^\omega$ there are infinite square-free words over Σ compatible with μ . We get $d(\{0, 1, 2, 4, 5, 6\}) \leq 2$.

$\mathbf{d}(\{0, 1, 2, 3\}) = 3$: Let $w = (0\diamond 1\diamond 2\diamond 3)^\omega$. An exhaustive search confirms that there are only 636 square-free words over $\{0, 1, 2, 3\}$ compatible with w . Thus $d(\{0, 1, 2, 3\}) \geq 3$.

Let $\Sigma = \{0, 1, 2, 3\}$, and let $W = \{\diamond\} \cup \{\diamond a \diamond : a \in \Sigma\} \cup \{\diamond a \diamond b : a, b \in \Sigma\}$. We can use Algorithm 1 to check that we can apply Lemma 9 with $f(1) = 2$, $f(4) = 5$ and $f(6) = 8$, $p = 18$. We can then apply Theorem 4 with $x_1 = \frac{11}{20}$, $x_4 = \frac{1}{4}$, and $x_6 = \frac{1}{5}$, and we deduce that for any $\mu \in W^\omega$ there are infinite square-free words over Σ compatible with μ . Moreover, $\{\diamond^i a : i \geq 2, a \in \Sigma\}^\omega \subseteq W^\omega$. We deduce that for any $\mu \in \{\diamond^i a : i \geq 2, a \in \Sigma\}^\omega$ there are infinite square-free words over Σ compatible with μ . Thus $d(\{0, 1, 2, 3\}) \leq 3$.

$7 \leq \mathbf{d}(\{0, 1, 2\}) \leq 19$: Let $w = (0\diamond^5 1\diamond^5 2\diamond^5)^\omega$. An exhaustive search confirms that there are only 4281 square-free words over $\{0, 1, 2\}$ compatible with w . Thus $d(\{0, 1, 2\}) \geq 7$.

Let $\Sigma = \{0, 1, 2\}$, and let $W = \{\diamond^9\} \cup \{\diamond^i a : i \in \{18, \dots, 26\}, a \in \Sigma\}$. We can use Algorithm 1 to check that we can apply Lemma 9 with $p = 61$ and the values of f given in Table 1. Thus conditions (I), (II), and (III) of Theorem 4 are fulfilled. We can also check that the values of $x_{|w|}$ given in Table 1 fulfill condition (IV) of Theorem 4. We deduce that for any $\mu \in W^\omega$ there are infinite square-free words over Σ compatible with μ . Moreover, $\{\diamond^i a : i \geq 18, a \in \Sigma\}^\omega \subseteq W^\omega$. We deduce that for any $\mu \in \{\diamond^i a : i \geq 18, a \in \Sigma\}^\omega$ there are infinite square-free words over Σ compatible with μ . Thus $d(\{0, 1, 2\}) \leq 19$. \square

TABLE 1. The values of $f(|w|)$ and $x_{|w|}$ for the computation of $d(\{0, 1, 2\})$

$ w $	9	19	20	21	22	23	24	25	26	27
$f(w)$	4	19	22	28	36	50	63	88	118	148
$x_{ w }$	$\frac{27}{100}$	$\frac{7}{100}$	$\frac{13}{200}$	$\frac{11}{200}$	$\frac{9}{200}$	$\frac{1}{25}$	$\frac{3}{100}$	$\frac{1}{40}$	$\frac{1}{40}$	$\frac{1}{50}$

The three applications of Algorithm 1 require between 30 and 100GB of RAM (and around 5 hours of computation). We had to optimize the way strings are stored in memory in order to be able to compute the graphs for large enough values of p . The rest of the computations (finding the solution to the system and the exhaustive search) easily run on a laptop in a few milliseconds. Recall that we showed something slightly stronger since the results would still hold if an adversary was to tell us at every choice of letter only the next 5 forced letters with their positions (that is, we know the next element of W).

Experimental computations suggest that $d(\{0, 1, 2\})$ is closer to 7 than to 19 and that $d(\{0, 1, 2, 3, 4\}) = 2$.

REFERENCES

- [1] J. P. Bell and T. L. Goh, *Exponential lower bounds for the number of words of uniform length avoiding a pattern*, Inform. and Comput. **205** (2007), no. 9, 1295–1306, DOI 10.1016/j.ic.2007.02.004. MR2334234
- [2] J. Currie, T. Harju, P. Ochem, and N. Rampersad, *Some further results on squarefree arithmetic progressions in infinite words*, Theoret. Comput. Sci. **799** (2019), 140–148, DOI 10.1016/j.tcs.2019.10.006. MR4029029
- [3] S. Czerwiński and J. Grytczuk, *Nonrepetitive Colorings of Graphs*, 6th Czech-Slovak International Symposium on Combinatorics, Graph Theory, Algorithms and Applications, Electron. Notes Discrete Math., vol. 28, Elsevier Sci. B. V., Amsterdam, 2007, pp. 453–459, DOI 10.1016/j.endm.2007.01.063. MR2324051
- [4] J. Grytczuk, J. Kozik, and P. Micek, *New approach to nonrepetitive sequences*, Random Structures Algorithms **42** (2013), no. 2, 214–225, DOI 10.1002/rsa.20411. MR3019398
- [5] T. Harju, *On square-free arithmetic progressions in infinite words*, Theoret. Comput. Sci. **770** (2019), 95–100, DOI 10.1016/j.tcs.2018.09.032. MR3944738
- [6] R. M. Kolpakov, *On a bound for the number of repetition-free words* (Russian, with Russian summary), Diskretn. Anal. Issled. Oper. Ser. 1 **13** (2006), no. 2, 21–37, DOI 10.1134/S1990478907040084; English transl., J. Appl. Ind. Math. **1** (2007), no. 4, 453–462. MR2289336
- [7] P. Ochem, *Doubled patterns are 3-avoidable*, Electron. J. Combin. **23** (2016), no. 1, Paper 1.19, 8. MR3484724
- [8] A. Thue, *Über unendliche Zeichenreihen*, Norske Vid. Selsk. Skr. I. Mat. Nat. Kl. Christiania **7** (1906), 1–22.

DEPARTMENT MATHÉMATIQUE, UNIVERSITÉ DE LIÈGE, AILÉE DE LA DÉCOUVERTE 12, 4000 LIÈGE, BELGIUM

Email address: `matthieu.rosenfeld@gmail.com`