RAIRO-Theor. Inf. Appl. 50 (2016) 117–133
DOI: 10.1051/ita/2016014

Available online at:
www.rairo-ita.org

COMPUTING DEPTHS OF PATTERNS

F. Blanchet-Sadri[1] and Andrew Lohr[2]

Abstract. Pattern avoidance is an important research topic in combinatorics on words which dates back to Thue's construction of an infinite word over three letters that avoids squares, *i.e.*, a sequence with no two adjacent identical factors. This result finds applications in various algebraic contexts where more general patterns than squares are considered. A more general form of pattern avoidance has recently emerged to allow for undefined positions in sequences. New concepts on patterns such as depth have been introduced and a number of questions have been raised, some of them we answer. In the process, we prove a *strict* bound on the number of square occurrences in an unavoidable pattern, and consequently, any pattern with more square occurrences than distinct variables is avoidable over three letters. We also provide an algorithm that determines whether a given pattern is of bounded depth, and if so, computes its depth.

Mathematics Subject Classification. 68R15.

1. Introduction

A *pattern* is a word (or sequence) over an alphabet Δ of *variables*, denoted by $A, B, C, \ldots$ A pattern p is *avoidable* over some finite alphabet Σ if there exists an infinite word over Σ with no occurrence of p. The terminology of avoidable pattern, although studied by Thue at the beginning of the twentieth century, was introduced much later by Bean *et al.* [1] and by Zimin [18] who described a simple procedure to decide avoidability. The problem of deciding whether a pattern is k-avoidable, *i.e.*, avoidable over k letters, has however remained open. Thus the problem of classifying the avoidability indices of all patterns over a fixed number of variables has become subject of investigation [9, 16] (the smallest k such that a pattern is k-avoidable is its *avoidability index*). Chapter 3 of [14] provides background on avoidable patterns. Currie [11] formulates a number of open problems in pattern avoidance. Among recent results, let us mention that an eighteen-year old conjecture from Cassaigne [9] regarding strict bounds for pattern avoidance got finally settled independently by Blanchet-Sadri and Woodhouse [8] and by Ochem and Pinlou [17].

A more general form of pattern avoidance has recently emerged to allow for undefined positions. In this context, *partial words* are sequences that may have such positions, called don't care symbols or holes, that match any letter of the alphabet (partial words without holes are *total words*). The occurrences of the same variable in a pattern are replaced with pairwise "compatible" partial words. For example, an occurrence of the pattern AAA

Keywords and phrases. Formal languages, combinatorics on words, pattern avoidance, partial words, depth of pattern.

[1] Department of Computer Science, University of North Carolina, P.O. Box 26170, Greensboro, NC 27402–6170, USA.
blanchet@uncg.edu

[2] Department of Mathematics, Rutgers University, 110 Frelinghuysen Rd., Piscataway, NJ 08854–8019, USA.
ajl213@scarletmail.rutgers.edu

Article published by EDP Sciences

has the form uvw where u is compatible with both v and w, and v is compatible with w. Constructing an infinite partial word with infinitely many holes that avoids a given pattern amounts to constructing an infinite set of infinite total words that avoid the pattern. New research topics are being developed such as pattern avoidance with respect to hole sparsity [2], abelian pattern avoidance [3], pattern avoidance using entropy compression [12], pattern avoidance in partial permutations [10], to name a few.

Clearly AA is unavoidable due to occurrences of *trivial* squares of the form $a\diamond$ or $\diamond a$, where a is a letter and $\diamond$ is the hole symbol. In [15], it was shown that there exists a partial word with infinitely many holes over two letters that avoids the pattern A^n, $n \geq 3$, and so its avoidability index in partial words is 2. Reference [4, 5, 7] provide, using "division" of patterns, the avoidability indices of *all* binary patterns, those over A and B, and *almost all* ternary patterns, those over A, B and C, except for four patterns whose avoidability index was shown to be between 2 and 5. To calculate the avoidability index of a pattern p, the lower bound is usually computed using backtracking. For the upper bound, a HD0L system is built that consists of an inner morphism ϕ and of an outer morphism ψ. Then $\psi(\phi^\omega(a))$ is shown to avoid p, for some letter a.

In the process of classifying the ternary patterns with respect to partial word avoidability, new concepts such as depth and shallowness, were introduced. More precisely, a k-unavoidable pattern p is (h, k)-*deep* if there exists some $m \in \mathbb{N}$ such that every partial word w over a k-letter alphabet meets p whenever w has at least h holes separated pairwise from each other and from the first and final position of w by factors of length m or greater. A function $\delta : \mathbb{N} \setminus \{0, 1\} \to \mathbb{N}$ is the depth function of an unavoidable pattern p if for all k the pattern p is $(\delta(k), k)$-deep and p is not (h, k)-deep for any $h < \delta(k)$. When the depth function of p is bounded, its supremum is the *depth* of p. A pattern p is k-*shallow* if p is $(0, k)$-deep or $(1, k)$-deep. If p is k-shallow for all k, then p is *shallow*.

A number of questions were raised [5]. Among them are the following:

1. If p_1Ap_2 is k-shallow and p_1 and p_2 are (h_1, k)-deep and (h_2, k)-deep respectively, is p_1Ap_2 $(h_1 + h_2, k)$-deep? In general, what relation does the depth of p_1Ap_2 have with the depth of p_1 and p_2?
2. Can every partial word unavoidable pattern, that is not total word unavoidable, be written in the form of ([5], Cor. 2)? More precisely, let p be a pattern of only distinct variables over Δ and let $0 \leq i < |p|$. Define $\mathrm{dig}_i(p)$ as a partial pattern that matches p except at position i where it is a hole. The corollary states that if $p_0, p_1, \ldots, p_n \in \Delta^*$ are compatible with factors of some $\mathrm{dig}_i(p)$ and $A_1, \ldots, A_n$ are distinct variables not in Δ, then $p_0A_1p_1 \ldots A_np_n$ is partial word unavoidable.

In relation to Question 1, it was mentioned that the classification of the depths of patterns may give insight; this classification was completed in [5] though the problem remained open. In this paper, among other things, we answer these questions.

The contents of our paper are as follows. In Section 2, we review a few basic concepts and notations. In Section 3, we prove, in particular, a *strict* bound on the number of square occurrences in a pattern that is partial word unavoidable, and consequently, any pattern with more square occurrences than distinct variables is 3-avoidable in partial words. In Section 4, we exhibit an unavoidable pattern that cannot be written in the form of ([5], Cor. 2), negatively answering Question 2 above. In Section 5, we answer Question 1 above. We also provide an algorithm that determines if a given pattern has bounded depth, and if so, outputs its depth. Finally in Section 6, we conclude with some open problems.

2. Basic concepts and notations

Let Σ be a finite alphabet of letters. Define $\Sigma_\diamond = \Sigma \cup \{\diamond\}$, where $\diamond \notin \Sigma$ represents an undefined position or a hole. A *partial word* over Σ is a sequence constructed from the concatenation of symbols from $\Sigma_\diamond$ while a *total word* over Σ is a partial word over Σ with no $\diamond$'s. The symbol at position i of partial word w is denoted by $w[i]$, where the labelling of positions starts at 0. The *length* of a partial word w over Σ is the number of symbols from $\Sigma_\diamond$ that it contains; it is denoted by $|w|$. The *empty word* is denoted by ε; it is the sequence of length zero.

The set of all total words (respectively, non-empty total words) over Σ is denoted by Σ^* (respectively, Σ^+), while the set of all partial words (respectively, non-empty partial words) over Σ by $\Sigma_\diamond^*$ (respectively, $\Sigma_\diamond^+$). The sets Σ^* and $\Sigma_\diamond^*$ equipped with the associative operation of concatenation form monoids, where ε acts as identity. Similarly, the sets Σ^+ and $\Sigma_\diamond^+$ equipped with the associative operation of concatenation form semigroups.

If u, v are partial words over Σ of equal length, then u, v are *compatible*, denoted $u \uparrow v$, if $u[i] = v[i]$ for all i such that $u[i], v[i] \in \Sigma$. If u, v are non-empty and compatible, then uv is a *square*. For example, $ab\diamond b$ is a square.

A partial word u is a *factor* of a partial word v if there exist x, y such that $v = xuy$. We denote by $v[i..j]$ (respectively, $v[i..j)$) the factor $v[i] \dots v[j]$ (respectively, $v[i] \dots v[j-1]$). A total word is a *subword* of v if it is compatible with a factor of v. For example if we consider the partial word $v = abcacb\diamond bcbac$ over the alphabet $\{a, b, c\}$, then $u = cb\diamond b$ is a factor of v and $cbab, cbbb, cbcb$ are the three subwords of v compatible with u. A *completion* of a partial word is a total word compatible with it. Returning to our example, $cbab$ is one of the three completions of $cb\diamond b$.

Let Δ be an alphabet of variables, $\Sigma \cap \Delta = \emptyset$, and let $p = A_0 \dots A_{n-1}$, where $A_i \in \Delta$, be a pattern. The set of distinct variables that occur in p is denoted by $\alpha(p)$. If a variable occurs only once in p, it is a *singleton* variable. Define an *occurrence* of p in a partial word w over an alphabet Σ as a factor $u_0 \dots u_{n-1}$ of w, where for all i, $u_i \neq \varepsilon$, and for all i, j, if $A_i = A_j$, then $u_i \uparrow u_j$. In other words, $u_0 \dots u_{n-1} \uparrow \varphi(p)$, where φ is any non-erasing morphism from Δ^* to Σ^*. We call such φ a *meeting morphism*. The partial word w *meets* the pattern p, or p *occurs* in w, if for some factorization $w = xuy$, we have that u is an occurrence of p in w; otherwise, w *avoids* p or w is *p-free*. For instance, $ab\underline{\diamond b}\underline{a\diamond}\underline{bb}a$ meets $ABBA$ (take the morphism $\varphi(A) = bb$ and $\varphi(B) = a$), while $\diamond babb$ avoids $ABBA$. These definitions also apply to (one-sided) infinite partial words over Σ, which are functions from $\mathbb{N}$ to $\Sigma_\diamond$.

A pattern $p \in \Delta^*$ is *k-avoidable in partial words* if for every integer $h > 0$ there is a partial word with h holes over a k-letter alphabet that avoids p. If there is an infinite partial word over a k-letter alphabet with infinitely many holes that avoids p, then p is obviously k-avoidable. On the other hand, if, for some integer $h \geq 0$, every long enough partial word in $\Sigma_\diamond^*$ with h holes meets p, then p is *k-unavoidable* (it is unavoidable over Σ). Finally, a pattern which is k-avoidable for some k is *avoidable*, and a pattern which is k-unavoidable for every k is *unavoidable*. The *avoidability index* of a pattern p is the smallest integer k such that p is k-avoidable, or is ∞ if p is unavoidable. Note that k-avoidability implies $(k+1)$-avoidability.

If a pattern p occurs in a pattern q, then p *divides* q and denote this by $p \mid q$; for instance, $AA \nmid ABA$ but $AA \mid ABAB$. Note that if $p \mid q$ and an infinite partial word avoids p then it also avoids q, and so the avoidability index of q is less than or equal to the avoidability index of p.

Throughout the paper, avoidable means avoidable in partial words unless otherwise stated.

3. Avoiding patterns

Any infinite partial word with at least one hole must meet A^2, so A^2 is clearly unavoidable in partial words. The theorem below addresses the avoidability of all other patterns where each variable occurs at least twice.

Theorem 3.1. *Let p be a pattern with $|p| > 2$ such that each variable in p occurs at least twice. Then p can be avoided by an infinite total word over k letters, for some k, and there exists a partial word with infinitely many holes over an alphabet of size $k+5$ that avoids p. Moreover, if there are no squares of length two in p, there exists a partial word with infinitely many holes over an alphabet of size $k+3$ that avoids p.*

Proof. By ([14], Cor. 3.2.10), p can be avoided by an infinite total word if each of its variables occurs at least twice. Therefore, let w be an infinite total word over an alphabet Σ of cardinality k, such that w avoids p. Take $k' = 1$ if there are no squares of length two in p and $k' = 2$ otherwise. There exist some $a_0, a_1, \dots, a_{2k'} \in \Sigma$ (not necessarily distinct) such that $a_0 a_1 \dots a_{2k'}$ occurs infinitely often as a factor of w. We create a sequence of integers $\{k_j\}$ as follows. Let k_0 be the smallest positive integer where $a_0 a_1 \dots a_{2k'} = w[k_0 - k']w[k_0 - (k'-1)] \dots w[k_0 + k']$. Define k_j recursively so that k_{j+1} is the smallest integer with $k_{j+1} > 4k_j$ and $a_0 a_1 \dots a_{2k'} = w[k_{j+1} - k']w[k_{j+1} - (k'-1)] \dots w[k_{j+1} + k']$.

Define the alphabet $\Sigma' = \Sigma \cup \{b_0, b_1, \ldots, b_{2k'}\}$, where $b_i \notin \Sigma$ for all i. We define the partial word w' as follows. If $j \equiv 0 \mod 6|p|$, for $0 \leq i \leq k'-1$ let $w'[k_j+i+1] = b_{k'+i+1}$ and $w'[k_j-i-1] = b_{k'-i-1}$; also define $w'[k_j] = \diamond$. If $j \not\equiv 0 \mod 6|p|$, let $w'[i] = b_{k'}$ if $i = k_j$, and let $w'[i] = w[i]$ otherwise. Note that w' is basically w, except the factor $b_0 \ldots b_{k'-1} \diamond b_{k'+1} \ldots b_{2k'}$ is *inserted* infinitely often, and between each two occurrences of this factor, there are $6|p|-1$ instances of $b_{k'}$, where the distance between any two such instances is greater than or equal to the distance from the first $b_{k'}$ to the beginning of the partial word w'. This construction also guarantees that for any i with $w'[i] = b_0$, we must have $w'[i+k'] = \diamond$. Likewise, for any i with $w'[i] = b_{2k'}$, $w'[i-k'] = \diamond$. Thus b_0 and $b_{2k'}$ can be viewed as "sentinel" letters on the left and right of the holes in w'.

The partial word w' is well-defined, and its letters come from an alphabet of size $k+2k'+1$. We show that w' avoids p by assuming that w' meets p and reaching a contradiction. Set $p = A_0 \ldots A_{|p|-1}$, where each A_i is a variable in Δ. Define j_0 and j_1 so that $u = u_0 \ldots u_{|p|-1} = w'[j_0..j_1]$ is a factor of w' such that if $A_i = A_j$ then $u_i \uparrow u_j$, *i.e.*, u is an occurrence of p in w'.

Two occurrences of the same variable A in p, say A_i and $A_{i'}$, where $i < i'$, correspond to partial words $u_i, u_{i'}$ such that $u_i \uparrow u_{i'}$. Moreover, there exist s, t, and n, $s \leq s+n < t \leq t+n$, so that $u_i = w'[s] \ldots w'[s+n]$ and $u_{i'} = w'[t] \ldots w'[t+n]$. Let $J_1 = \{j \mid s \leq k_j \leq s+n\}$ and $J_2 = \{j \mid t \leq k_j \leq t+n\}$. We show that $|J_2| \leq 1$, thus $|J_1| \leq 2$. Assume for the sake of contradiction that $|J_2| > 1$, so there exists $j \in J_2$ such that $j+1 \in J_2$. However,

$$n = t+n-t \geq k_{j+1} - k_j > k_j > s+n \geq n,$$

a contradiction. For the second inequality, if we assume $|J_1| > 2$, there are at least two occurrences of the letter $b_{k'}$ in u_i, and for each such occurrence there must also be an occurrence of $b_{k'}$ or $\diamond$ in $u_{i'}$. This would imply the contradiction $|J_2| > 1$, therefore $|J_1| \leq 2$.

Since each variable in p occurs at least twice, $|J_1| \leq 2$ and $|J_2| \leq 1$ imply there are at most $2|p|$ non-negative integers j with $j_0 \leq k_j \leq j_1$. By construction of w', there are $6|p|-1$ integers j such that $w'[k_j] = b_{k'}$ between any two holes in w'. Thus u contains at most one hole. It remains to show that u actually contains no holes.

First, suppose that $|u_i| > 1$ and u_i contains a hole. Then either $b_{k'-1}$ or $b_{k'+1}$ must occur in u_i. Assume $b_{k'-1}$ occurs in u_i (the other case is similar). Then $u_{i'}$, with $u_{i'} \uparrow u_i$, must contain $b_{k'-1}$ or $\diamond$ in the corresponding position. If $u_{i'}$ contains $b_{k'-1}$, by construction of w', the next letter in $u_{i'}$ must be a hole. So $u_{i'}$ contains $\diamond$. But then u contains two holes, contradicting the above statement that u contains at most one hole.

Now, suppose $u_i = \diamond$ and $k' = 1$. Note that the theorem assumes $|p| > 2$. Either $i > 0$ or $i < |p|-1$. Assume $i > 0$ (the other case is similar). Then, since $k' = 1$ indicates that there are no squares of length two in p, $A_{i-1} \neq A_i$, so u_{i-1} must end with $b_{k'-1} = b_0$. Consider one other instance of variable A_{i-1}, say A_j, with $j \neq i$. Then u_j must end with b_0 or a hole. If u_j ends with a hole, there are two holes in u, a contradiction. If u_j ends with b_0 (a sentinel letter for a hole by construction of w'), the letter in w' following u_j must be a hole and cannot be the hole in A_i. Thus $w'[j_0..j_1+1]$ contains at least two holes. We showed above that there are at most $2|p|$ non-negative integers j with $j_0 \leq k_j \leq j_1$, hence there are at most $2|p|+1$ non-negative integers j with $j_0 \leq k_j \leq j_1+1$. But by construction of w', there are $6|p|-1$ integers j such that $w'[k_j] = b_{k'}$ between any two holes in w', a contradiction.

Finally, suppose $u_i = \diamond$ and $k' = 2$. Assume $i > 0$ as before (the other case is similar). If $A_{i-1} \neq A_i$, the same argument as in the $k' = 1$ case leads to a contradiction. Thus, assume $A_{i-1} = A_i$. Then $i > 1$ or $i < |p|-1$, so assume $i > 1$ (the other case is similar). If $A_{i-2} = A_{i-1}$, an argument similar to the one in the $k' = 1$ case leads to a contradiction. Thus, assume $A_{i-2} \neq A_{i-1} = A_i$. Since $u_i = \diamond$, $|u_i| = |u_{i-1}| = 1$, hence u_{i-2} must end with $b_{k'-2} = b_0$. Consider one other instance of variable A_{i-2}, say A_j, with $j \notin \{i-1, i\}$. Then u_j must end with b_0 or a hole. If u_j ends with a hole, there are two holes in u, a contradiction. If u_j ends with b_0, the letter in w' two positions from the end of u_j must be a hole and cannot be the hole in A_i. Thus $w'[j_0..j_1+2]$ contains at least two holes. We showed above that there are at most $2|p|$ non-negative integers j with $j_0 \leq k_j \leq j_1$, hence there are at most $2|p|+2$ non-negative integers j with $j_0 \leq k_j \leq j_1+2$. But by construction of w', there are $6|p|-1$ integers j such that $w'[k_j] = b_{k'}$ between any two holes in w', a contradiction.

Thus u contains no holes. Define $\varphi : (\Sigma')^* \to \Sigma^*$ with $\varphi(a) = a$ if $a \in \Sigma$ and $\varphi(b_i) = a_i$ for $0 \leq i \leq 2k'$. By construction $\varphi(u)$ is a factor of w that represents an occurrence of p, contradicting the fact that w avoids p. $\square$

The next theorem provides a bound on the number of square occurrences in a pattern that is partial word unavoidable. This bound cannot be improved. For a variable alphabet of size n, the pattern

$$A_0A_0A_1A_0A_0A_2A_0A_0 \dots A_0A_0A_{n-1}A_0A_0$$

has n square occurrences, and is unavoidable in partial words (there has to be a factor "$a\diamond$" that occurs infinitely often for some letter a).

Theorem 3.2. *The number of square occurrences in a pattern that is partial word unavoidable is less than or equal to the number of distinct variables used. Moreover, any pattern with more square occurrences than distinct variables is 3-avoidable.*

Proof. If any square occurrence in the pattern is of length greater than two, it is divisible by $ABAB$, which is 3-avoidable [7], so, we restrict to only when we have a single variable squared. Also, no square occurrences are adjacent, otherwise the pattern would be divisible by $AABB$, which is 3-avoidable [7]. Lastly, no square occurrences overlap, because an overlap of two length two square occurrences is an occurrence of AAA, which is 2-avoidable [7].

Suppose that p is a partial word unavoidable pattern over an alphabet of n variables Δ. Proceeding by contradiction, write

$$p = A_1A_1p_1A_2A_2 \dots A_nA_np_nA_{n+1}A_{n+1}$$

where, $A_i \in \Delta$ and $p_i \in \Delta^+$ for all i (we can ignore the ends of p before A_1A_1 and after $A_{n+1}A_{n+1}$). Let $\Delta_1 = \{A_1, A_2, \dots, A_{n+1}\}$.

First, we claim that $p_i \notin {\Delta_1}^+$ for all i. Suppose towards a contradiction that there is an i such that $p_i \in {\Delta_1}^+$. Let $\Sigma = \{a, b, c\}$ and let $\theta : \Sigma^* \to \Sigma^*$ be the morphism defined by $\theta(a) = abc$, $\theta(b) = ac$, and $\theta(c) = b$. Define the morphism $\phi : \Sigma^* \to \Sigma_\diamond^*$ as θ^3 with the factor bab of $\theta^3(a)$ changed to $b\diamond b$, *i.e.*,

$$\phi(\ell) = \begin{cases} abcacb\diamond bcbac, & \text{if } \ell = a; \\ abcacbac, & \text{if } \ell = b; \\ abcb, & \text{if } \ell = c. \end{cases}$$

Let $w = \phi \circ \theta^\omega(a)$ and let $\langle i_m \rangle$ be the sequence of indices of holes of w, *i.e.*, $w[j] = \diamond$ if and only if $j \in \langle i_m \rangle$.

Let $\langle j_m \rangle$ be any subsequence of $\langle i_m \rangle$ such that $j_{m+1} > 2j_m + 7 + 2|p_i|$. We construct a partial word w' from w by replacing, for all i_m, $w[i_m - 1..i_m + 1]$ with $b\diamond b$ if $i_m \in \langle j_m \rangle$ or with bab if not. Note that $\text{fill}_a(w) = \text{fill}_a(w') = \theta^\omega(a)$ which is known to be square-free [13] (here fill_a fills the holes with letter a). It follows that any square-compatible factor of w', *i.e.*, a factor of w' compatible with a total word that is a square, must contain a $\diamond$.

We show that the set of square subwords of w' is exactly $\{bb, cbcb, bcbc\}$. Note that any length five or greater factors of w' containing $\diamond$ are always equal whenever they are compatible, as the length five factors of w' containing $\diamond$ are

$$cacb\diamond, acb\diamond b, cb\diamond bc, b\diamond bcb, \diamond bcba,$$

which are all pairwise incompatible. It follows that if there exists any length ten or more square-compatible factor $y = y_1y_2$ where $y_1 \uparrow y_2$, it satisfies $y_1 = y_2$ which implies $\text{fill}_a(y_1) = \text{fill}_a(y_2)$, so $\text{fill}_a(y)$ is a square factor of $\theta^\omega(a)$, a contradiction. Therefore, every square-compatible factor has length less than ten and must be a factor of $\phi(\ell a)$ or $\phi(a\ell)$ where $\ell \in \{b, c\}$. It is easy to see that the only such square subwords have length two or four and belong to $\{bb, cbcb, bcbc\}$.

Since we assumed that the pattern p is partial word unavoidable, the partial word w' meets p, or p occurs in w'. Let φ be a non-erasing morphism from Δ^* to Σ^* such that $\varphi(p)$ is a subword of w'. In particular, $\varphi(A_jA_j)$ is a square subword of w' for all j. This implies that $|\varphi(A_j)| \leq 2$ for all j. So, we have $|\varphi(p_i)| \leq 2|p_i|$ since we assumed that $p_i \in \Delta_1^+$. And, since for all j, $\varphi(A_jA_j)$ must have its ends within three positions of a $\diamond$, there is some m such that $|\varphi(p_i)| \geq j_{m+1} - j_m - 6$, which implies that $2|p_i| \geq j_{m+1} - j_m - 6 > j_m + 1 + 2|p_i|$, which in turn implies, $0 > j_m + 1$, a contradiction. So $p_i \notin {\Delta_1}^+$ for all i, proving our claim.

Now, because $|\Delta \setminus \Delta_1| < n$, there must be some i such that there is some mapping f from $\alpha(p_i) \setminus \Delta_1$ to $\{1, \ldots, i-1\}$ such that for every $A \in \alpha(p_i) \setminus \Delta_1$, the membership $A \in \alpha(p_{f(A)})$ holds. In other words, every variable that occurs in p_i has to either appear in a square occurrence or at some point further left in p. We know such an i exists because with each j, p_j can either use a variable from $\Delta \setminus \Delta_1$ that has not been used before, or it can only use variables that occurred before. Because there are more p_j's than variables in $\Delta \setminus \Delta_1$ to introduce for the first time, at least one of the p_j's has to not introduce any new variables from $\Delta \setminus \Delta_1$, this is the p_i we want. Let $\Delta_2 = \alpha(p_i) \setminus \Delta_1$ and let $g : \Delta \to \mathbb{N}$ map A to the number of times A appears in p_i.

Next, we construct a partial word w'' with infinitely many holes over three letters whose construction is similar to the one of the partial word w'. Instead of requiring that $\langle j_m \rangle$ be any subsequence of $\langle i_m \rangle$ such that $j_{m+1} > 2j_m + 7 + 2|p_i|$, we require that $\langle j_m \rangle$ be any subsequence of $\langle i_m \rangle$ such that

$$j_{m+1} > \left(1 + \sum_{A \in \Delta_2} g(A)\right) j_m + 6 + 2 \sum_{A \in \Delta_1} g(A). \tag{1}$$

We construct w'' from w by replacing, for all i_m, $w[i_m - 1..i_m + 1]$ with $b\diamond b$ if $i_m \in \langle j_m \rangle$ or with bab if not.

Since p is partial word unavoidable, the partial word w'' meets p, or p occurs in w''. Let φ be a non-erasing morphism from Δ^* to Σ^* such that $\varphi(p)$ is a subword of w''. Note that, for every $A \in \Delta_2$, $p_{f(A)}$ contains A means $|\varphi(p_{f(A)})| \geq |\varphi(A)|$. Then because each square occurrence in p has its ends within three positions of a $\diamond$, there is a function gap mapping p_j, $j \leq i$, to $j_{m'} - j_m$ where the two terms j_m and $j_{m'}$, which belong to the selected subsequence of $\langle i_m \rangle$, are the positions of the two holes that the ends of p_j are near, *i.e.*, j_m is the position of the last hole in A_jA_j and $j_{m'}$ is the position of the first hole in $A_{j+1}A_{j+1}$. This means $\text{gap}(p_j) \geq |\varphi(p_j)| \geq \text{gap}(p_j) - 6$ for all $j \leq i$.

Note that for every $j' < j \leq i$, $\text{gap}(p_{j'}) + 6 < \text{gap}(p_j)$. So for every $A \in \Delta_2$, if $f(A) \neq i-1$ then

$$|\varphi(A)| \leq |\varphi(p_{f(A)})| \leq \text{gap}(p_{f(A)}) < \text{gap}(p_{i-1}) - 6 \leq |\varphi(p_{i-1})|.$$

Let m and m' be such that $\text{gap}(p_i) = j_{m'} - j_m$. Then using Equation (1),

$$\begin{aligned} |\varphi(p_i)| \geq j_{m+1} - j_m - 6 &> \left(\textstyle\sum_{A \in \Delta_2} g(A)\right) |\varphi(p_{i-1})| + 2 \textstyle\sum_{A \in \Delta_1} g(A) \\ &\geq \left(\textstyle\sum_{A \in \Delta_2} g(A)|\varphi(A)|\right) + 2 \textstyle\sum_{A \in \Delta_1} g(A), \end{aligned}$$

which contradicts the fact that

$$|\varphi(p_i)| = \sum_{A \in \Delta} g(A)|\varphi(A)| \leq \left(\sum_{A \in \Delta_2} g(A)|\varphi(A)|\right) + 2 \sum_{A \in \Delta_1} g(A).$$

Thus we have proved that the number of square occurrences in a pattern that is partial word unavoidable is less than or equal to the number of distinct variables used.

We have also proved that the partial word w'' with infinitely many holes over three letters avoids p, so any pattern with more square occurrences than distinct variables is 3-avoidable. □

The proof of our next theorem refers to avoidability of *simple formulas*. We extend this concept, defined for total words in ([13], Problem 3.1.2), to partial words.

Definition 3.3. Let Σ be an alphabet of letters and Δ be an alphabet of variables such that $\Sigma \cap \Delta = \emptyset$.

- A *simple formula* f is a finite set of patterns $p_1, \ldots, p_n$ over Δ, denoted by $f = p_1 \cdot \ldots \cdot p_n$, where the order of the patterns $p_1, \ldots, p_n$ is not important.

- A partial word w over Σ *meets* a simple formula $f = p_1 \cdot \ldots \cdot p_n$ if there exists a non-erasing morphism φ from Δ^* to Σ^* such that all the total words $\varphi(p_1), \ldots, \varphi(p_n)$ are compatible with factors of w.
- Avoidability and k-avoidability in partial words of simple formulas is defined as for patterns.

We next state a couple of lemmas regarding avoidability of simple formulas.

Lemma 3.4. *The simple formula $f = p_1 \cdot \ldots \cdot p_n$ is avoidable in total (respectively, partial) words over Σ if and only if the pattern $p_1 A_1 \ldots p_{n-1} A_{n-1} p_n$ is avoidable in total (respectively, partial) words over Σ, where $A_1, \ldots, A_{n-1}$ are distinct variables that do not occur in f.*

Note that the pattern $p_1 A_1 \ldots p_{n-1} A_{n-1} p_n$ in Lemma 3.4 is said to be obtained from the simple formula $p_1 \cdot \ldots \cdot p_n$ (and *vice versa*).

Lemma 3.5. *Let f be a simple formula consisting of patterns of length at most two, none of which are squares, and every variable in f occurs at most twice. Then f is unavoidable.*

Proof. The proof is by induction on the number of distinct variables in f. For the basis, the result holds for a number of two distinct variables A and B in f because $AB \cdot BA$ is unavoidable (we can remove the patterns of length one). For the inductive step, construct the adjacency graph G of the pattern p obtained from the simple formula f. This undirected graph G is the bipartite graph with two copies of $\alpha(p)$ as vertices, denoted by $\alpha(p)_L = \{A_L \mid A \in \alpha(p)\}$, called the set of left vertices, and $\alpha(p)_R = \{A_R \mid A \in \alpha(p)\}$, called the set of right vertices, and with an edge between A_L and B_R if and only if AB is a factor of p.

We show that there is a free set for p. Recall that a non-empty subset of $\alpha(p)$ is a *free set* if there exists no path in G connecting a left vertex A_L to a right vertex B_R with A and B in the free set. Finding the connected components of G helps us find the free sets, so edges involving vertices corresponding to singleton variables can be ignored.

Say we are left with $2n$ vertices corresponding to n distinct variables. Each pattern in f contributes at most one edge to G, and there are at most n patterns in f (this comes from the fact that every variable occurs at most twice, and we also remove the patterns of length one). There are at least n connected components in G because at least $2n - 1$ edges are needed to have it connected, and we have $n - 1$ fewer edges than that. The only way we can partition the $2n$ vertices into n connected components so that there is no free set is if for every variable A, the left vertex A_L is connected to the right vertex A_R by an edge. This would mean however that each of the length two patterns in f is a square, which we ruled out any of them from being. So we have a free set F containing one variable. Delete all occurrences of this variable from the pattern p to obtain a pattern q, then use the inductive hypothesis on q. Thus, q is unavoidable. For sake of completeness, we recall the arguments from ([13], Lem. 3.2.2) to show that p is also unavoidable (and so is f).

To show this, we introduce some notation. Given a set X of vertices of G, we denote by $C(X)$ the set of vertices of G that belong to the same connected component as an element of X, and by $C_L(X)$ (respectively, $C_R(X)$) the set of variables $A \in \alpha(p)$ such that $A_L \in C(X)$ (respectively, $A_R \in C(X)$). The set F being a free set translates as $F \subseteq C_L(F_L) \setminus C_R(F_L)$, where F_L is the singleton set consisting of the left vertex of the variable in F.

We now show that p is unavoidable over any alphabet Σ by induction on $|\Sigma|$. The basis $|\Sigma| = 1$ is obvious. For the inductive step, suppose that p is unavoidable over an alphabet Σ', and let $\Sigma = \Sigma' \cup \{a\}$ where a is a new letter not in Σ'. Let Γ be a new alphabet, whose letters are words in Σ^*, defined as follows:

$$\Gamma = \{a^i w a^j \mid w \in (\Sigma')^+, w \text{ avoids } p, 0 < i < |p|, 0 \leq j < |p|\}.$$

Note that Γ is finite since p is unavoidable over Σ'. Each word over Σ, that avoids p and that starts with a, is either a non-empty power of a or is a non-empty concatenation of letters in Γ. Let $\iota : \Gamma^* \to \Sigma^*$ be the identity morphism.

Let C be a variable that is not in $\alpha(p)$. Then the pattern qC is unavoidable. So for every $w \in \Gamma^*$ sufficiently long, there exists a non-erasing morphism $\theta : (\alpha(p) \cup \{C\})^* \to \Gamma^*$ such that $\theta(qC)$ is a factor of w. For any variable A in $\alpha(p) \cup \{C\}$, since $\theta(A) \in \Gamma^+$ we have $\iota(\theta(A)) \in a\Sigma^+$.

Let $\psi : \alpha(p)^* \to \Sigma^*$ be the non-erasing morphism defined as follows:

- If $A \in \alpha(p) \setminus (C_L(F_L) \cup C_R(F_L))$, then $\psi(A) = \iota(\theta(A))$.
- If $A \in C_R(F_L) \setminus C_L(F_L)$, then $a\psi(A) = \iota(\theta(A))$.
- If $A \in C_L(F_L) \setminus (C_R(F_L) \cup F)$, then $\psi(A) = \iota(\theta(A))a$.
- If $A \in C_L(F_L) \cap C_R(F_L)$, then $a\psi(A) = \iota(\theta(A))a$.
- If $A \in F$, then $\psi(A) = a$.

It is easy to check that ψ is well-defined. We claim that $\psi(p)$ is a factor of $\iota(\theta(qC))$.

For $1 \le k \le |p|$, let q_k be the prefix of q obtained by deleting from $p[0..k)$ all occurrences of variables from F. By induction on k, we prove that $b\psi(p[0..k)) = \iota(\theta(q_k))c_k$, where $b = a$ if $p[0] \in C_R(F_L)$ or $b = \varepsilon$ otherwise and where $c_k = a$ if $p[k-1] \in C_L(F_L)$ or $c_k = \varepsilon$ otherwise. The basis $k = 1$ follows from the definition of ψ. For the inductive step, let $p[0..k+1) = p[0..k)B$. Setting $p[k-1] = A$, there is an edge from A_L to B_R in G. We have

$$b\psi(p[0..k+1)) = b\psi(p[0..k))\psi(B) = \iota(\theta(q_k))c_k\psi(B).$$

Observe that $c_k = a$ if and only if $A \in C_L(F_L)$ if and only if $B \in C_R(F_L)$. This implies that $c_k\psi(B) = \iota(\theta(B))c_{k+1}$ if $B \notin F$ or $c_k\psi(B) = c_{k+1}$ otherwise. So $b\psi(p[0..k+1)) = \iota(\theta(q_{k+1}))c_{k+1}$, where $c_{k+1} = a$ if $B \in C_L(F_L)$ or $c_{k+1} = \varepsilon$ otherwise. This shows our claim.

Therefore, the word $\iota(w)$ over Σ meets p, which means that the set of words over Σ that avoid p and start with a is finite. So p is unavoidable over Σ. □

The next theorem will be useful for computing the depth of a given pattern in Section 5. It is based on the following concept of holeboundedness.

Definition 3.6. Let $u_0 \dots u_{|p|-1}$ be an occurrence of a pattern p in a partial word. Let the function f map each u_i to the variable of p that corresponds to u_i. A non-singleton variable A is *holebound to position* j or j*-holebound* if the only u_i with $f(u_i) = A$ corresponding to any factor other than the factor $\diamond$ is u_j. If a variable is j-holebound for some j, it is *holebound*.

Theorem 3.7. *If p is a pattern with no squares that is k-avoidable in total words, then, for every positive integers m and h, there is an infinite partial word over $k + 4h$ letters, with h holes each at least m positions away from each other and the beginning of the partial word, that avoids p.*

Proof. Let m and h be positive integers. We do induction on the number of distinct variables in a pattern p with no squares that is k-avoidable in total words. For the basis, the smallest size of an alphabet of variables Δ over which it is possible to have a square-free pattern that is avoidable by infinite total words is $|\Delta| = 3$. The only such ternary pattern with a singleton variable is $ABACBAB$ which, by [5], can be avoided by an infinite partial word having infinitely many holes over only three letters, using the HD0L system given by $\phi(\theta^\omega(a))$ where

$$\theta(\ell) = \begin{cases} ad, & \text{if } \ell = a; \\ ab, & \text{if } \ell = b; \\ db, & \text{if } \ell = c; \\ c, & \text{if } \ell = d; \end{cases} \quad \text{and} \quad \phi(\ell) = \begin{cases} bb, & \text{if } \ell = a; \\ caabc, & \text{if } \ell = b; \\ aab\diamond acbaabc, & \text{if } \ell = c; \\ ac, & \text{if } \ell = d. \end{cases}$$

Then we just fill in all but h of the holes, each of these h holes at least m positions away from each other and the beginning of the partial word. In the case of a pattern p with no singleton variables so each variable in p occurs at least twice, by Theorem 3.1, p can be avoided by an infinite partial word having infinitely many holes over $k + 3$ letters, and we just fill in all but h of the holes that are far enough apart as before. We thus have our basis with $|\alpha(p)| = 3$.

For the inductive hypothesis, assume that if p' is a pattern with no squares, with $|\alpha(p')| \geq 3$, that is k-avoidable in total words, then, for every positive integers m and h, there is an infinite partial word over $k+4h$ letters, with h holes each at least m positions away from each other and the beginning of the partial word, that avoids p'.

For the inductive step, let w be an infinite total word over k letters that avoids a pattern p with no squares and $|\alpha(p)| \geq 4$. Let y be a length five factor of w that occurs infinitely often. Let $x_0, x_1, \ldots, x_{h-1}$ be h disjoint occurrences of y in w that appear in order at least m positions apart and m positions away from the beginning of the word. Let $\{a_i, b_i, c_i, d_i\}_{0 \leq i < h}$ be a collection of $4h$ distinct letters that do not appear in w. Then, for each i, replace x_i with $a_i b_i \diamond c_i d_i$; call the resulting partial word w'. Our claim is that w' avoids p because under the mapping $\theta(\ell) = y[0]$ if $\ell = a_i$ for some i; $y[1]$ if $\ell = b_i$ for some i; $y[2]$ if $\ell = \diamond$; $y[3]$ if $\ell = c_i$ for some i; $y[4]$ if $\ell = d_i$ for some i; and ℓ otherwise, $\theta(w') = w$ avoids p. Assume to the contrary that w' meets p with pattern occurrence $u_0 \ldots u_{|p|-1}$. Then there must be some hole in $u_0 \ldots u_{|p|-1}$.

First, suppose towards a contradiction that for every i such that u_i contains a hole, *i.e.*, $\diamond$ is a factor of u_i, the variable in p corresponding to u_i is a singleton variable. Thus, if u_i contains a hole, u_i is not required to be compatible with any u_j, $j \neq i$, so, the pattern occurrence $u_0 \ldots u_{|p|-1}$ is preserved under applying θ to w', contradicting the fact that w is p-free. So let u_j be such that it contains a hole and it corresponds to a non-singleton variable of p.

Second, suppose towards a contradiction that $\diamond$ is at some position other than either the first or last position of u_j. Then there is some j' such that $b_{j'} \diamond c_{j'}$ is a factor of u_j, but any possible completion of $b_{j'} \diamond c_{j'}$ is a subword that only appears in u_j, contradicting the fact that u_j corresponds to a non-singleton variable. So, u_j either starts or ends with a $\diamond$.

Third, suppose towards a contradiction that $|u_j| > 2$. Then there is some j' such that u_j has either $a_{j'} b_{j'} \diamond$ or $\diamond c_{j'} d_{j'}$ as a suffix or prefix, respectively. Any completion of either however is not a subword that appears anywhere else. So, u_j is $\diamond$ (Case 1), $\diamond c_{j'}$ for some j' (Case 2), or $b_{j'} \diamond$ for some j' (Case 3). Note that Case 3 is symmetric to Case 2.

The rest of the proof is based on the concept of holeboundedness according to Definition 3.6.

Case 1: $u_j = \diamond$

Let $X_0 = f(u_j)$, and say that position j of $u_0 \ldots u_{|p|-1}$ is visited. Construct a sequence of variables $X_0, X_1, \ldots$ from the pattern p where, for each $i \geq 0$, X_{i+1} is a $(i'+1)$-holebound variable for some i' such that $f(u_{i'}) = X_i$, $f(u_{i'+1}) = X_{i+1}$, $u_{i'} = \diamond$, and $u_{i'+1} \neq \diamond$, or X_{i+1} is a $(i'-1)$-holebound variable for some i' such that $f(u_{i'}) = X_i$, $f(u_{i'-1}) = X_{i+1}$, $u_{i'} = \diamond$, and $u_{i'-1} \neq \diamond$; in either case, say that position i' of $u_0 \ldots u_{|p|-1}$ is visited. In other words, X_{i+1} is a holebound variable whose only occurrence that corresponds to a non-hole is next to an occurrence of X_i that corresponds to a hole. If ever there are two such holebound variables, we claim that only one of them generates a path that visits an occurrence of X_0 that corresponds to a non-hole (which means that the resulting sequence of variables would repeat the variable X_0). To show our claim, suppose towards a contradiction that the two of them generate paths that visit an occurrence of X_0 that corresponds to a non-hole. Then X_0 is holebound to two different positions, a contradiction with the definition of holeboundedness. So, if there are two occurrences of holebound variables that correspond to a non-hole next to occurrences of X_i that correspond to a hole, take as X_{i+1} a holebound variable that does not generate a path that visits an occurrence of X_0 that corresponds to a non-hole.

Each variable in $X_0, X_1, \ldots$, with the exception of X_0, is holebound to a position next to a position that corresponds to an occurrence corresponding to a hole of a variable before it. For each i, X_{i+1} cannot be holebound to an already visited position, because this occurrence of X_{i+1}, appearing next to an occurrence of X_i that corresponds to a hole, cannot correspond to a hole. This implies that either the sequence $X_0, X_1, \ldots$ repeats X_0, and we are in the case below where there exists an integer i, $i > 0$, such that $X_0 = X_i$, or the sequence $X_0, X_1, \ldots$ does not repeat X_0, and we are in the case below where there exists a non-singleton variable surrounded by singleton variables, because there are only finitely many variables and X_0 is the only variable in X_0, $X_1, \ldots$ that can repeat.

We illustrate these two cases with the following diagrams:

$$\begin{array}{ccccccccccccccc} \dots & b_{i_1} & \diamond & c_{i_1} & \dots & b_{i_2} & \diamond & c_{i_2} & \dots & b_{i_3} & \diamond & c_{i_3} & \dots \\ & & X_2 & X_3 & & & X_0 & X_1 & & & X_2 & X_1 & \\ & & & & & & X_3 & & & & & & \end{array}$$

$$\begin{array}{ccccccccccccccc} \dots & b_{i_1} & \diamond & c_{i_1} & \dots & b_{i_2} & \diamond & c_{i_2} & \dots & b_{i_3} & \diamond & c_{i_3} & \dots \\ & X_2 & X_1 & & & & X_0 & X_1 & & C & X_2 & D & \end{array}$$

In the first diagram, the sequence $X_0, X_1, \dots$ repeats X_0, *i.e.*, $X_0 = X_3$. In the second diagram, there is no occurrence of a holebound variable next to the occurrence of X_2 corresponding to a hole, so singleton variables C and D must correspond to words ending in b_{i_3} and beginning in c_{i_3}, respectively (the variables C and D are singletons since the letters b_{i_3} and c_{i_3} are unique).

First, suppose there exists some integer i, $i > 0$, such that $X_0 = X_i$. Because there was never a choice to select a path that did not visit an occurrence of X_0 that corresponds to a non-hole, every time we found the next variable in the sequence, there was only one singleton variable next to the occurrences that correspond to a hole of the current variable, and only one occurrence that corresponds to a hole of each variable in the sequence $X_0, X_1, \dots$. So, there is a set of i variables $\Delta' = \{A_0, \dots, A_{i-1}\}$ each occurring exactly twice in the pattern p, and $2i$ singleton variables $\{B_0, C_0, \dots, B_{i-1}, C_{i-1}\}$ such that there are i length two patterns over Δ', call them $p_0, \dots, p_{i-1}$, none of which are squares, and

$$B_0 p_0 C_0, \dots, B_{i-1} p_{i-1} C_{i-1}$$

occur in p. We define a pattern p' such that $|\alpha(p')| = |\alpha(p)| - 1$. To do this, we use $i-1$ variables $D_0, \dots, D_{i-2}$ that appear nowhere in p, replace $p_0, \dots, p_{i-2}$ with them, and delete p_{i-1}. This way, the i variables $A_0, \dots, A_{i-1}$ present in $p_0, \dots, p_{i-1}$ get removed and the $i-1$ variables $D_0, \dots, D_{i-2}$ get added. In other words, we define the pattern p' by replacing the $i-1$ factors $B_0 p_0 C_0, \dots, B_{i-2} p_{i-2} C_{i-2}$ with $B_0 D_0 C_0, \dots, B_{i-2} D_{i-2} C_{i-2}$, respectively, and by replacing the factor $B_{i-1} p_{i-1} C_{i-1}$ with $B_{i-1} C_{i-1}$. Note that by divisibility, any total word that avoids p' avoids p. Because the simple formula obtained from $\{p_0, \dots, p_{i-1}\}$ is unavoidable by Lemma 3.5 and over a variable alphabet disjoint from the rest of the pattern p, its removal does not affect the total word avoidability index. This means that we can apply the inductive hypothesis. Since p' is a pattern with no squares, with $|\alpha(p')| \geq 3$, that is k-avoidable in total words, there is an infinite partial word over $k + 4h$ letters, with h holes each at least m positions away from each other and the beginning of the partial word, that avoids p'. This partial word with h holes avoids p.

Second, suppose there exists a non-singleton variable surrounded by two singleton variables. Call the non-singleton variable A and singleton variables C and D, such that CAD appears as a factor of p. Construct the pattern p' from p by replacing the factor CAD with C. This implies that if $p = p_1 CAD p_2$, then $p' = p_1 C p_2$. The patterns p and p' correspond to the simple formulas $f_p = p_1 \cdot A \cdot p_2$ and $f_{p'} = p_1 \cdot p_2$ respectively; note that f_p is k-avoidable in total words if and only if $f_{p'}$ is. Thus, because of Lemma 3.4, p' has the same total word avoidability index as p. Since $|\alpha(p')| = |\alpha(p)| - 1$, apply the inductive hypothesis. Since p' is a pattern with no squares, with $|\alpha(p')| \geq 3$, that is k-avoidable in total words, there is an infinite partial word over $k+4h$ letters, with h holes each at least m positions away from each other and the beginning of the partial word, that avoids p'. This partial word with h holes that avoids p' definitely avoids p since $p' \mid p$.

Case 2: $\boldsymbol{u_j = \diamond c_{j'}}$ for some j'

Let $X_0 = f(u_j)$. This case is similar to Case 1, except, the sequence $X_0, X_1, \dots$ constructed can never repeat X_0 because u_j having length two means X_0 cannot be holebound. So, there is a non-singleton variable surrounded by two singleton variables. ∎

4. Answering a conjecture

We settle a conjecture from [5] that every partial word unavoidable pattern, that is not total word unavoidable, can be written in the form of ([5], Cor. 2).

First, let p be a pattern of only distinct variables over Δ and let $0 \leq i < |p|$. Define $\mathrm{dig}_i(p)$ as a partial pattern with exactly one hole over Δ, such that $\mathrm{dig}_i(p)[i] = \diamond$ and $\mathrm{dig}_i(p)[j] = p[j]$ if $0 \leq j < |p|, j \neq i$. For example, if $p = ABCDEF$, then $\mathrm{dig}_2(p) = AB\diamond DEF$.

Now, recall the corollary.

Corollary 4.1 (Cor. 2, [5])**.** *Let p be a pattern of only distinct variables over Δ and let $p_0, p_1, \ldots, p_n \in \Delta^*$ be compatible with factors of some $\mathrm{dig}_i(p)$. If $A_1, \ldots, A_n$ are distinct variables not in Δ, then $p_0A_1p_1 \ldots A_np_n$ is partial word unavoidable.*

Next, let B_m, for $m \in \mathbb{N}$ be different variables. Let $Z_0 = \varepsilon$, and for all $m \in \mathbb{N}$, let $Z_{m+1} = Z_mB_{m+1}Z_m$; the Z_m's are the Zimin words well-known to be unavoidable in total words [14]. For example, $Z_4 = B_1B_2B_1B_3B_1B_2B_1B_4B_1B_2B_1B_3B_1B_2B_1$. Note that Z_m is over m distinct variables and $|Z_m| = 2^m - 1$.

Next, recall the conjecture.

Conjecture 4.2 ([5])**.** *Every partial word unavoidable pattern, that is not total word unavoidable, can be written in the form of Corollary 4.1.*

Finally, let us prove negatively the conjecture.

Theorem 4.3. *Conjecture 4.2 is false.*

Proof. It suffices to provide a pattern that is partial word unavoidable and that is neither total word unavoidable nor of the form of Corollary 4.1. The pattern

$$q = Z_4EFF = ABACABADABACABAEFF$$

satisfies such property. It is clearly not total word unavoidable because it is divisible by AA and AA is 3-avoidable in total words. It is also not of the form of Corollary 2. To see this, the only possibilities for $A_1, \ldots, A_n$ are $A_1 = D$ and $A_2 = E$ as they are the only variables occurring only once. This means that $p_0 = p_1 = ABACABA$ and $p_2 = FF$. However, since p_0 has A occurring more than twice, it cannot be compatible with a factor of some $\mathrm{dig}_i(p)$, which is a partial pattern with exactly one hole constructed from a pattern p of only distinct variables, as is the restriction on p in Corollary 4.1.

To see that q is partial word unavoidable, first note that Z_4 is unavoidable in partial words (as mentioned above, Z_4 is unavoidable even in total words). Since Z_4 is unavoidable in partial words, there must be some hole occurring at least two positions to the right of an occurrence of Z_4 in any infinite partial word w with infinitely many holes. Let F map to a letter compatible with the symbol occurring immediately to the right of the hole, and E map to a total word compatible with the factor of w between the occurrence of Z_4 and the hole. $\square$

5. Computing depths of patterns

Recall the definitions of *depth* and *shallowness*.

Definition 5.1 ([5])**.**

- A k-unavoidable pattern p is (h,k)-*deep* if there exists some $m \in \mathbb{N}$ such that every partial word w over a k-letter alphabet meets p whenever w has at least h holes separated pairwise from each other and from the first and final position of w by factors of length m or greater.
- A function $\delta : \mathbb{N} \setminus \{0,1\} \to \mathbb{N}$ is the depth function of an unavoidable pattern p if for all k the pattern p is $(\delta(k), k)$-deep and p is not (h,k)-deep for any $h < \delta(k)$.
- When the depth function of p is bounded, its supremum d is the *depth* of p and p is d-deep.

Definition 5.2 ([5]).

- A pattern p is *k-shallow* if p is $(0,k)$-deep or $(1,k)$-deep.
- If p is k-shallow for all k, then p is *shallow*.
- The pattern p is *k-non-shallow* if it is k-unavoidable but not k-shallow.

Shallow patterns have some properties in common with total word unavoidable patterns that higher-depth patterns do not have.

A use of shallowness from [5] states that if p_1, p_2 are k-unavoidable patterns over an alphabet of variables Δ and A is a variable which does not appear in p_1 or p_2, *i.e.*, $A \in \Delta \setminus (\alpha(p_1) \cup \alpha(p_2))$, then the pattern $p_1 A p_2$ is k-unavoidable if there exists some k-shallow pattern p such that p_1 and p_2 are factors of p. Note that it is also much easier to check that a given pattern is shallow for a given k than to check that it has higher depth. This is done just by starting with a hole then trying to add a letter on each end, backtracking if no letter works. If there are only finitely many such partial words, then the pattern is unavoidable with depth 1. This does not work as easily for higher depths because if the backtracking came up finite, then it could be that the two holes starting the backtracking were not far enough apart.

The classification of the depths of the 2-unavoidable binary patterns has been completed.

Theorem 5.3 ([5]). *The 2-unavoidable binary patterns in partial words fall into five categories with respect to depth (up to reversal and complement):*

(1) *The patterns ε, A, AB, and ABA are shallow with depth 0;*
(2) *The patterns AA and AAB are $(0,2)$-deep and $(1,k)$-deep for all $k \geq 3$;*
(3) *The pattern $AABA$ is $(0,2)$-deep, $(1,3)$-deep, and $(2,k)$-deep for all $k \geq 4$;*
(4) *The pattern $AABAA$ has depth function δ satisfying $\delta(2) = 0$ and, for all $k \geq 3$, $\delta(k) = k+1$;*
(5) *The patterns $AABAB$, $AABB$, $ABAB$, $ABBA$ are $(0,2)$-deep.*

For example, consider the depth function δ of the pattern $AABAA$. To have an occurrence, the same square must occur twice, separated by at least one symbol. First, $AABAA$ is 2-unavoidable in total words, so it is $(0,2)$-deep and $\delta(2) = 0$. Now, let $k \geq 3$. If a partial word w over a k-letter alphabet $\{a_1, \ldots, a_k\}$ has $k+1$ holes far enough apart, one of the k letters occurs next to two distinct holes. So the same trivial square occurs twice in w, which means w meets the pattern. So $AABAA$ is $(k+1,k)$-deep and $\delta(k) \leq k+1$. An avoiding partial word with $k-1$ holes can be constructed by surrounding them like $a_k a_1 \diamond a_1 a_k$, $a_k a_2 \diamond a_2 a_k$, $\ldots$, $a_k a_{k-1} \diamond a_{k-1} a_k$ which avoids the pattern, to show that $AABAA$ is not $(k-1,k)$-deep. Moreover, an avoiding partial word with k holes can also be constructed by starting with the fixed point at a of the morphism mapping a to abc, b to ac, and c to b to show that $AABAA$ is not (k,k)-deep and $\delta(k) \geq k+1$. Details appear in [5].

The next theorem describes the form of all 1-deep patterns, knowing that the variable that appears squared cannot appear anywhere else, and the variables appearing around the square occurrence must be singleton variables. The rest of the pattern must be 0-deep, once the square surrounded by singleton variables is replaced with a single singleton variable. The proof relies on the following two lemmas.

Lemma 5.4. *If the patterns p_1 and p_2 are (h_1,k)-deep and (h_2,k)-deep respectively, then $p = p_1 A p_2$ is not (h,k)-deep for any $h < h_1 + h_2$.*

Proof. Let $h < h_1 + h_2$. Suppose that the pattern p needs a minimum hole spacing of m to achieve being (h,k)-deep, *i.e.*, every partial word over k letters, with h holes separated pairwise from each other and from the ends of the partial word by factors of length m or greater, must meet p. Since p_1 is (h_1,k)-deep, there is a partial word w' avoiding p_1 over k letters, with $h_1 - 1$ holes at least m positions away from each other and the ends of the partial word. Similarly since p_2 is (h_2,k)-deep, there is a partial word w'' avoiding p_2 over k letters, with $h_2 - 1$ holes at least m positions away from each other and the ends of the partial word.

We claim that the partial word $w = w'\diamond w''$ avoids p despite it having $h_1 + h_2 - 1$ holes, each at least m positions away from each other and from the ends of the partial word. To show our claim, suppose towards a contradiction that there is a non-erasing morphism φ such that $\varphi(p)$ is compatible with a factor of w. Because w' avoids p_1, $\varphi(p_1)$ cannot be compatible with a factor of w'. Similarly, $\varphi(p_2)$ cannot be compatible with a factor of w''. This leads to a contradiction since we only added a single $\diamond$ between w' and w''.

Now, let the partial word v be the prefix of w that ends m positions after the hth hole. Because w avoids p, so does v. This contradicts p being (h, k)-deep with achieved minimum distance between holes m. □

This implies that p is either k-avoidable or (h, k)-deep with $h \geq h_1+h_2$. We do not necessarily have $h = h_1+h_2$, as demonstrated by $ABACAA$ which is 3-deep even though ABA has depth 0 and AA has depth 1.

Lemma 5.5. *If the patterns p_1 and p_2 are both 0-deep, then, taking A to be a variable not appearing in either p_1 or p_2, we have $p = p_1Ap_2$ is not h-deep for any $h > 0$.*

Proof. Because p_1 and p_2 are total word unavoidable, they have no squares, so p has no squares. If p is not 0-deep, it has a total word avoidability index of k, say. Therefore, by Theorem 3.7, for any $h > 0$, there is a partial word over $k + 4h$ letters with h holes spaced arbitrarily far apart that avoids p meaning that p is not $(h, k + 4h)$-deep, so, p is not h-deep. □

Theorem 5.6. *The patterns that are 1-deep have exactly one square occurrence.*

Proof. We prove our result by induction on the number of distinct variables in the patterns. For the basis, *i.e.*, patterns over a single variable, the only 1-deep pattern is AA by Theorem 5.3. For the inductive step, let p be a 1-deep pattern. Note that p cannot have more than one square occurrence, otherwise it would be $(h, 4)$-deep with $h > 1$ (this follows from Thm. 5.3 and arguments similar to those following it).

So, suppose towards a contradiction that p has no square occurrences. Because p is unavoidable, it must have a singleton variable, say A (recall that by ([14], Cor. 3.2.10), p can be avoided by an infinite total word if each of its variables occurs at least twice). By Lemma 5.4, either one end is 1-deep and the other is 0-deep, or both ends are 0-deep. But not both ends are 0-deep by Lemma 5.5. Let p_1 be the end that has non-zero depth, *i.e.*, we either label $p = p_1Ap_2$ or $p = p_2Ap_1$. Note that p_1 has one less variable than p, it is 1-deep, and it has no square occurrences because p has no square occurrences. This implies a contradiction with the inductive hypothesis. □

We now have the necessary machinery to describe Algorithm 1, which finds the depth of an arbitrary pattern p. Recall from Definition 3.6 that a variable A of p is *holebound* if all but a single occurrence of A must map to a $\diamond$ in any meeting morphism. This concept is used in the proof of Theorem 5.8 in which we insert h factors

$$a_{0,0}\cdots a_{0,|p|-1}\diamond a_{0,|p|}\cdots a_{0,2|p|-1},\ldots,a_{h-1,0}\cdots a_{h-1,|p|-1}\diamond a_{h-1,|p|}\cdots a_{h-1,2|p|-1}$$

arbitrarily far apart, where each of the $a_{i,j}$'s that are used are unique to the h holes.

To help understand our algorithm, consider $p = AABCECDFBGD$ of length 11. Let us discuss an occurrence that would contain some holes with the following diagram:

$$\begin{array}{ccccccccccc} \cdots & a_{i_1,10} & \diamond & a_{i_1,11} & a_{i_1,12} & \cdots & \diamond & a_{i_2,11} & \cdots & \diamond & \cdots & \diamond & \cdots \\ & A & A & B & C & & C & D & & B & & D & \end{array}$$

The pattern p has a square occurrence AA that corresponds to the factor $a_{i_1,10}\diamond$. Since the letter $a_{i_1,10}$ is unique, the variable A is holebound. Since the letter $a_{i_1,11}$ is also unique and the variable B appears again, B is holebound and there is an occurrence that must correspond to a hole. The same is true with the variables C and D. On the other hand, the variables E, F, and G are singleton variables that can correspond to factors of length larger than one.

The following example illustrates Algorithm 1.

Algorithm 1. Determine if a pattern has bounded depth, if so, find its depth.

Require: p is a pattern
Ensure: the depth of p if p has bounded depth, FALSE otherwise
1: $V \leftarrow \emptyset$
2: $S \leftarrow \emptyset$
3: $S_f \leftarrow \emptyset$
4: **for** variables A that appear in a square occurrence in p **do**
5:    **if** A has two or more square occurrences in p **then**
6:       return FALSE
7:    $S \leftarrow \{$all maximal occurrences of powers of $A\} \cup S$
8:    $V \leftarrow \{A\} \cup V$
9: **while** $S \neq \emptyset$ **do**
10:    remove an occurrence O from S
11:    $S_f \leftarrow \{O\} \cup S_f$
12:    **for** occurrences O_B of variables B between O and either the end of the pattern or a singleton variable **do**
13:       **if** $B \in V$ **then**
14:          return FALSE
15:       $V \leftarrow \{B\} \cup V$
16:       $S \leftarrow \{$all occurrences of B other than $O_B\} \cup S$
17: $f \leftarrow$ simple formula obtained by removing all occurrences in S_f from p
18: **if** f is total word avoidable (using Zimin's procedure) **then**
19:    return FALSE
20: return $|S_f|$

Example 5.7. Let us determine if the pattern $p = AABCECDFBGD$ has bounded depth.

The sets V, S, S_f are initialized with $\emptyset$ in lines 1–3. In line 4, the only variable that appears in a square occurrence in p is A. The variable A has only one square occurrence in p, so S is updated to $\{p[0]p[1] = AA\}$ in line 7 and V to $\{A\}$ in line 8.

For the first pass through the while loop in lines 9–16, the algorithm removes the occurrence $O = p[0]p[1] = AA$ from S in line 10 and S_f is updated to $\{p[0]p[1]\}$ in line 11. There are two occurrences of variables between O and the singleton variable E, *i.e.*, $O_B = p[2]$ and $O_C = p[3]$, so the **for** loop in lines 12–16 has two passes. After the first pass, V is updated to $\{A, B\}$ and S to $\{p[8] = B\}$, and after the second pass, V is updated to $\{A, B, C\}$ and S to $\{p[8] = B, p[5] = C\}$.

For the second pass through the while loop, the algorithm removes the occurrence $O = p[8] = B$ from S in line 10 and S_f is updated to $\{p[0]p[1], p[8]\}$ in line 11. There is no occurrence of variables between O and the singleton variable G.

For the third pass through the while loop, the algorithm removes the occurrence $O = p[5] = C$ from S in line 10 and S_f is updated in line 11 to $\{p[0]p[1], p[8], p[5]\}$. There is one occurrence of variables between O and the singleton variable F, *i.e.*, $O_D = p[6]$, so the **for** loop in lines 12–16 has one pass. After this pass, V is updated to $\{A, B, C, D\}$ and S to $\{p[10] = D\}$.

For the fourth pass through the while loop, the algorithm removes the occurrence $O = p[10] = D$ from S in line 10 and S_f is updated in line 11 to $\{p[0]p[1], p[8], p[5], p[10]\}$. There is no occurrence of variables between O and either the end of the pattern or a singleton variable.

In line 17, f is the simple formula obtained by removing all occurrences in $S_f = \{p[0]p[1], p[8], p[5], p[10]\}$ from p, *i.e.*, $f = BCE \cdot DF \cdot G$. Since f is not total word avoidable, the algorithm returns $|S_f| = 4$ in line 20. Thus, the given pattern p is unavoidable with depth four.

We now prove that Algorithm 1 behaves as desired.

Theorem 5.8. *Given as input a pattern p, Algorithm 1 determines if p has bounded depth, and if so, it outputs its depth; otherwise, it returns FALSE.*

Proof. If the pattern p is total word unavoidable, then it has bounded depth 0. So, suppose that p is total word avoidable over k letters, and consider an infinite total word w over k letters that avoids p.

First, consider the case where p does not have at least one square occurrence. By Theorem 3.7, for every positive integers m and h, there is an infinite partial word over $k+4h$ letters, with h holes each at least m positions away from each other and the beginning of the partial word, that avoids p. Thus, p cannot have bounded depth, because it cannot be $(h, k+4h)$-deep for any h. In this case, by lines 17–19, f is the simple formula obtained from p and is total word avoidable, and the algorithm returns FALSE.

Now, consider the case where p has at least one square occurrence. If the same variable A has at least two square occurrences (not only one occurrence of a power of A), then it appears squared at least twice and p is divisible by $AABAA$. This implies that the depth function of p is unbounded by Theorem 5.3, in which case the algorithm returns FALSE (see lines 5–6). So, assume that each variable has at most one square occurrence in p.

Let y be a factor of w of length $2|p|+1$ that occurs infinitely often. Let $x_0, x_1, \dots, x_{h-1}$ be h disjoint occurrences of y in w that appear in order arbitrarily far apart and away from the beginning of the word. Let $\{a_{0,0}, \dots, a_{0,2|p|-1}, \dots, a_{h-1,0}, \dots, a_{h-1,2|p|-1}\}$ be a collection of $2h|p|$ distinct letters that do not appear in w. Then, replace x_0 with $a_{0,0} \dots a_{0,|p|-1} \diamond a_{0,|p|} \dots a_{0,2|p|-1}$, x_1 with $a_{1,0} \dots a_{1,|p|-1} \diamond a_{1,|p|} \dots a_{1,2|p|-1}$, and so on; call the resulting partial word w'. Since the total word w avoids p, any occurrence of p in the partial word w' must contain a hole.

Any square occurrence AA in p must correspond to a factor of w' of the form $a\diamond$ or $\diamond a$, where a is a letter in the alphabet, meaning the variable A is holebound because any letter that appears adjacent to a hole never appears again in w'. Here V serves to keep track of exactly those variables which have been holebound within two positions of occurrences of variables already in either S or S_f, *i.e.*, the variables have already been considered and these occurrences correspond to factors containing holes. This means that if a variable that is in V ever appears again when considering some different hole-containing occurrence, then that variable is holebound to two different positions, a contradiction with the definition of holeboundedness. So, w' would avoid p.

Each time we remove an occurrence O from S in line 10, it either corresponds exactly to a hole, or if it comes from a square occurrence in the first **for** loop, it corresponds to a factor of the form $a\diamond$ or $\diamond a$ with a a letter in the alphabet. Because each of the $2h|p|$ letters surrounding the holes are distinct, the neighbors of the O occurrence are either holebound variables or singleton variables. To see this, note that no subword of length greater than two ever appears again. Holebound variables correspond to factors of length one meaning that we must eventually reach a singleton variable or the end of the pattern. For the non-singleton variables considered before reaching a singleton variable, their other occurrences must correspond to holes, so they are added to S in line 16.

Note then, that splitting p on its singleton variables to create a simple formula, say $q_1 \cdot \dots \cdot q_n$, each q_i must consist entirely of variables in V, or have no variables in V. After removing all such chunks of the holebound variables to create the simple formula f, the corresponding pattern, say p_f, is square-free, and if f is total word avoidable (and so is p_f by Lem. 3.4), there is an avoiding partial word over $4h$ additional letters by Theorem 3.7, meaning that p is not of bounded depth and the algorithm returns FALSE in line 19. On the other hand, if f is total word unavoidable, then there is a way of spacing the holes far enough apart so that each hole corresponds to some occurrence that is in S_f and occurrences of the patterns whose variables are not in V must appear between the holes. Because the only holes that are used are for occurrences in S_f, and only one hole for each such occurrence, the depth of p is $|S_f|$ which the algorithm returns in line 20. $\square$

Note that in the above proof, if the pattern p_f, after deleting variables occurring in V, is entirely composed of singleton variables, then we are in the interesting case where p is unavoidable even over an infinite alphabet so long as there are $|S_f|$ holes spaced far enough apart.

6. Conclusion and open problems

In Section 3, we proved, in particular, a strict bound on the number of square occurrences in a pattern that is partial word unavoidable, and consequently, any pattern with more square occurrences than distinct variables is 3-avoidable in partial words. In Section 4, we exhibited an unavoidable pattern that cannot be written in the form of Corollary 4.1, settling a conjecture from [5]. In Section 5, we answered a number of questions regarding the concept of depth of patterns that were raised in [5]. In particular, we examined the relation between the depth of a pattern of the form p_1Ap_2, where A is a variable, and the depth of p_1 and p_2. We also provided an algorithm that determines if a given pattern has bounded depth, and if so, outputs its depth.

Dealing with unavoidable patterns with unbounded depth functions is much more complicated than dealing with patterns with bounded ones because letters around holes must be reused at some point. Because our algorithm uses a construction that introduces $2|p|$ new letters per hole, every depth function is either bounded or is in $\Omega(k)$ where k is the alphabet size.

Open problem 6.1. *Study unavoidable patterns that have an unbounded depth function.*

For any n, the pattern

$$p = A_0A_0A_1A_2 \dots A_{n-2}A_{n-1}A_0A_0A_1A_2 \dots A_{n-2}$$

over n variables has depth function in $\Theta(k^{n-1})$. In fact, it has depth function at least $(k-3)^{n-1}+1$. To see this, we construct a partial word w' over k letters with $h = (k-3)^{n-1}$ holes that avoids p. Since p has avoidability index 3 in total words, we start with a square-free total word w over 3 letters. There are $(k-3)^{n-1}$ distinct total words of length $n-1$ over the remaining $k-3$ letters, say $x_0, \dots, x_{h-1}$. To build w', we insert the h factors $\diamond x_0, \dots, \diamond x_{h-1}$ exponentially far apart in w, where each of the x_i's are unique to the h holes. Each occurrence of A_0A_0 in p would map to a square in any meeting morphism, so each occurrence of A_0A_0 would correspond to a factor of w' that contains a hole, and for every $0 \le i < n-1$, A_i would have an image of length one.

Open problem 6.2. *Determine whether there are patterns that have a depth function that grows more quickly than the one of p.*

Acknowledgements. This material is based upon work supported by the National Science Foundation under Grant Nos. DMS–0754154 and DMS–1060775. The Department of Defense is gratefully acknowledged. A research assignment from the University of North Carolina at Greensboro for the first author is also gratefully acknowledged. Part of this paper was presented at LATA 2014 [6]. We wish to thank Sean Simmons and Brent Woodhouse for their contributions. We also wish to thank the referees of preliminary versions of this paper for their very valuable comments and suggestions that have helped improved the presentation.

References

[1] D.R. Bean, A. Ehrenfeucht and G. McNulty, Avoidable patterns in strings of symbols. *Pacific J. Math.* **85** (1979) 261–294.
[2] F. Blanchet-Sadri, K. Black and A. Zemke, Unary pattern avoidance in partial words dense with holes. In *LATA 2011, 5th International Conference on Language and Automata Theory and Applications*. Vol. 6638 of *Lect. Notes Comput. Sci.*, edited by A.-H. Dediu, S. Inenaga and C. Martín-Vide. Springer-Verlag. Berlin, Heidelberg. (2011) 155–166.
[3] F. Blanchet-Sadri, B. De Winkle and S. Simmons, Abelian pattern avoidance in partial words. *RAIRO-Theoretical Informatics and Applications* **48** (2014) 315–339.
[4] F. Blanchet-Sadri, A. Lohr and S. Scott, Computing the partial word avoidability indices of binary patterns. *J. Discrete Algorithms* **23** (2013) 113–118.
[5] F. Blanchet-Sadri, A. Lohr and S. Scott, Computing the partial word avoidability indices of ternary patterns. *J. Discrete Algorithms* **23** (2013) 119–142.
[6] F. Blanchet-Sadri, A. Lohr, S. Simmons and B. Woodhouse, Computing depths of patterns. In *LATA 2014, 8th International Conference on Language and Automata Theory and Applications*. Vol. 8370 of *Lect. Notes Comput. Sci.*, edited by A.-H. Dediu, C. Martín-Vide, J.-L. Sierra-Rodriguez and B. Truthe. Springer-Verlag. Berlin, Heidelberg (2014) 173–185.

[7] F. Blanchet-Sadri, R. Mercaş, S. Simmons and E. Weissenstein, Avoidable binary patterns in partial words. *Acta Inf.* **48** (2011) 25–41.
[8] F. Blanchet-Sadri and B. Woodhouse, Strict bounds for pattern avoidance. *Theor. Comput. Sci.* **506** (2013) 17–28.
[9] J. Cassaigne, *Motifs évitables et régularités dans les mots*, Ph.D. thesis, Paris VI (1994).
[10] A. Claesson, V. Jelínek, E. Jelínková and S. Kitaev, Pattern avoidance in partial permutations. *Electron. J. Combin.* **18** (2011) P25.
[11] J.D. Currie, Pattern avoidance: themes and variations. *Theoret. Comput. Sci.* **339** (2005) 7–18.
[12] A. Gagol, Pattern avoidance in partial words over a ternary alphabet. *Ann. Univ. Mariae Curie-Sklodowska* Lublin-Polonia **LXIX** (2015) 73–82.
[13] M. Lothaire, *Combinatorics on Words*. Cambridge University Press, Cambridge (1997).
[14] M. Lothaire, *Algebraic Combinatorics on Words*. Cambridge University Press, Cambridge (2002).
[15] F. Manea and R. Mercaş, Freeness of partial words. *Theoret. Comput. Sci.* **389** (2007) 265–277.
[16] P. Ochem, A generator of morphisms for infinite words. *RAIRO: ITA* **40** (2006) 427–441.
[17] P. Ochem and A. Pinlou, Application of entropy compression in pattern avoidance. *Electron. J. Combin.* **21** (2014) P2.7.
[18] A.I. Zimin, Blocking sets of terms. *Mathematics of the USSR-Sbornik* **47** (1984) 353–364.

Communicated by J. Kari.
Received January 22, 2016. Accepted June 13, 2016.