Unavoidable Sets of Partial Words

By: F. Blanchet-Sadri, N.C. Brownstein, Andy Kalcic, Justin Palumbo, T. Weyand

F. Blanchet-Sadri, N.C. Brownstein, A. Kalcic, J. Palumbo and T. Weyand, "Unavoidable Sets of Partial Words." Theory of Computing Systems, Vol. 45, No. 2, 2009, pp 381-406.

Made available courtesy of Springer Verlag: http://dx.doi.org/10.1007/s00224-008-9106-1

***** The original publication is available at www.springerlink.com**

Abstract:
The notion of an unavoidable set of words appears frequently in the fields of mathematics and theoretical computer science, in particular with its connection to the study of combinatorics on words. The theory of unavoidable sets has seen extensive study over the past twenty years. In this paper we extend the definition of unavoidable sets of words to unavoidable sets of partial words. Partial words, or finite sequences that may contain a number of "do not know" symbols or "holes," appear naturally in several areas of current interest such as molecular biology, data communication, and DNA computing. We demonstrate the utility of the notion of unavoidability of sets of partial words by making use of it to identify several new classes of unavoidable sets of full words. Along the way we begin work on classifying the unavoidable sets of partial words of small cardinality. We pose a conjecture, and show that affirmative proof of this conjecture gives a sufficient condition for classifying all the unavoidable sets of partial words of size two. We give a result which makes the conjecture easy to verify for a significant number of cases. We characterize many forms of unavoidable sets of partial words of size three over a binary alphabet, and completely characterize such sets over a ternary alphabet. Finally, we extend our results to unavoidable sets of partial words of size k over a k-letter alphabet.

Keywords: Combinatorics on words, Partial words, Unavoidable sets

Article:

1 Introduction

An *unavoidable* set of words X over an alphabet A is a set for which any sufficiently long word over A will have a factor in X. It is clear from the definition that from each unavoidable set we can extract a finite unavoidable subset, so the study can be reduced to finite unavoidable sets. This concept was explicitly introduced in 1983 in connection with an attempt to characterize the rational languages among the context-free ones [1 1]. Since then it has been consistently studied by researchers in both mathematics and theoretical computer science. There is a vast literature on unavoidable sets of words and we refer the reader to [9, 10, 13, 15] for more information.

Another concept relevant to this paper is that of a *partial word*, or a finite sequence of symbols over a finite alphabet that may contain a number of "do not know" symbols or "holes". Partial words appear in natural ways in several areas of current interest such as molecular biology, data communication, and DNA computing [2, 3, 12]. While a word can be described as a total function whose range is the alphabet, a partial word is described as a partial function. More precisely a partial word of length n over a finite alphabet A is a partial function from $\{0,...,n-1\}$ into A. Elements of $\{0, \ldots, n-1\}$ with no image are called holes (a word is then just a partial word without holes). To emphasize this distinction we will often refer to words without holes as full words. In this paper, we introduce unavoidable sets of partial words. In terms of unavoidability, sets of partial words serve as efficient representations of sets of full words. This is strongly analogous to the study of unavoidable patterns, in which sets of patterns are used to represent infinite sets of full words [13]. The main goal here is to demonstrate that the study of unavoidable sets of partial words leads to new insights both on the theory of unavoidable sets and on the combinatorial structure of the set of words A^* as a whole. In accomplishing this we mainly focus on the problem of classifying unavoidable sets of partial words of small cardinality and in particular those with two or three elements.

Efficient algorithms to determine if a finite set of full words X is unavoidable are well known [13, 14]. For example, we can check whether there is a loop in the finite automaton of Aho and Corasick [1] recognizing $A^* \setminus A^*XA^*$. These same algorithms can be used to decide if a finite set of partial words X is unavoidable by determining the unavoidability of $\hat{X}$, the set of all full words *compatible* with an element of X. However this incurs a dramatic loss in efficiency, as each partial word u in X can contribute as many as $\|A\|^{\|H(u)\|}$ elements to $\hat{X}$ (here $H(u)$ denotes the set of holes of u). In [5], Blanchet-Sadri, Jungers and Palumbo showed that the problem of testing the unavoidability of a finite set of partial words over any alphabet of size larger or equal to two is NP-hard by using techniques similar to those used in a recent paper of Blondel, Jungers and Protasov on the complexity of computing the capacity of codes that avoid forbidden difference patterns [6].

The contents of our paper are summarized as follows. In Sect. 2, we review some basic definitions related to words and partial words. In Sect. 3, we recall the definition of unavoidable sets of words and some useful elementary properties. We present our definition of unavoidable sets of partial words and discuss testing unavoidability of such sets. There, we introduce the problem of classifying unavoidable sets of small cardinality. In Sect. 4, we restrict ourselves to sets with two elements. In particular, we propose in Sect. 4.1 a conjecture characterizing two-word unavoidable sets and prove that verifying this conjecture is sufficient for solving the size two problem. There, we also prove one direction of our conjecture. In Sect. 4.2, we give partial results towards the other direction of our conjecture and in particular prove that it is easy to verify in a large number of cases. In Sect. 5, we restrict ourselves to sets with three elements. In Sect. 5.1 we discuss the case of a binary alphabet, while in Sect. 5.2 we characterize the unavoidable sets of size three over a ternary alphabet. Finally in Sect. 6, we pose several questions related to unavoidable sets of partial words.

2 Preliminaries

Throughout this paper A is a fixed nonempty finite set called an *alphabet* whose elements we call *letters*. A *word* of length n over A is a finite sequence of elements of A and will be denoted by $a_0a_1 \cdots a_{n-1}$ where $a_i \in A$ for every $0 \leq i < n$. We use A^* (respectively, A^n) to denote the set of finite words over A (respectively, the set of words of length n over A). For $u \in A^*$, we write $|u|$ for the length of u. Under the concatenation operation of words, A^* forms a free monoid whose identity is the empty word which we denote by ε. We will abbreviate $A^*\setminus\{\varepsilon\}$ by A^+. If there exist $x, y \in A^*$ such that $u = xvy$, then we say that v is a *factor* of u.

A *two-sided infinite word* w is a function $w\colon \mathbb{Z} \to A$. A finite word u is a factor of w if u is a finite subsequence of w, that is, if there exists some $i \in \mathbb{Z}$ such that $u = w(i) \cdots w(i + |u| - 1)$. For a positive integer p, we say that w has *period* p, or that w is *p-periodic*, if $w(i) = w(i + p)$ for all $i \in \mathbb{Z}$. If w has period p for some p, then we call w *periodic*. If v is a nonempty finite word, then we denote by $v^{\mathbb{Z}}$ the unique two-sided infinite word w such that w has period $|v|$ and $w(0) \cdots w(|v| - 1) = v$.

A word of finite length n over an alphabet A can be defined as a total function $w : \{0,..., n - 1\} \to A$. Analogously a *partial word* of length n over A is a partial function $u : \{0,..., n - 1\} \to A$. For $0 \leq i < n$, if $u(i)$ is defined, then we say that i belongs to the *domain* of u (denoted by $i \in D(u)$). Otherwise we say that i belongs to the *set of holes* of u (denoted by $i \in H(u)$). In cases where $H(u)$ is empty, we say that u is a *full word*.

If u is a partial word of length n over A, then the *companion* of u is the total function $u_\diamond\colon \{0,..., n - 1\} \to A_\diamond$ defined by

$$u_\diamond(i) = \begin{cases} u(i) \text{ if } i \in D(u) \\ \diamond \text{ otherwise} \end{cases}$$

where $A_\Diamond = A \cup \{\Diamond\}$. Throughout this paper we identify a partial word with its companion. We reserve the term *letter* for elements of A. We will refer to an occurrence of the symbol $\Diamond$ in a partial word as a *hole*. We will denote by $A_\Diamond^*$ the set of all partial words over A with an arbitrary number of holes. A partial word v is a *factor* of a partial word u if there exists partial words x, y (possibly equal to ε) such that $u = xvy$. We denote the set of all factors of u by $F(u)$. The partial word v is a *prefix* (respectively, *suffix*) of u if $x = \varepsilon$ (respectively, $y = \varepsilon$). The reverse of u, denoted rev(u), is the partial word u written backwards. If u is defined over the binary alphabet $\{a, b\}$, then the complement of u, denoted $\overline{u}$, is the partial word formed by replacing all the a's with b's and vice versa.

Two partial words u and v of equal length are said to be *compatible*, denoted by $u \uparrow v$, if $u(i) = v(i)$ for every $i \in D(u) \cap D(v)$. If X is a set of partial words, then we use $\hat{X}$ to denote the set of all full words compatible with an element of X. For example, if our alphabet is $\{a, b\}$ and $X = \{\Diamond a, b\Diamond\}$ then $\hat{X} = \{aa, ba, bb\}$. The partial word u is said to be *contained* in v, denoted by $u \subset v$, if $|u| = |v|$ and $u(i) = v(i)$ for all $i \in D(u)$. If a partial word u can be written as $u = u_1\Diamond u_2\Diamond \cdots u_{n-1}\Diamond u_n$, then the set $\{u_1a_1u_2a_2 \cdots u_{n-1}a_{n-1}u_n \mid a_1, a_2,\dots, a_{n-1} \in A\}$ is called a *partial expansion* on u (note that $u_1, u_2,\dots, u_n$ are partial words that may contain holes, and also note that $u \subset v$ for every member v of a partial expansion on u).

3 Unavoidable Sets

We first recall the definition of an unavoidable set of full words and some relevant properties. Let $X \subset A^*$. A two-sided infinite word w over A *avoids* X if no factor of w is a member of X. We say that X is *unavoidable* if no two-sided infinite word over A avoids X. In other words X is unavoidable if every two-sided infinite word over A has a factor in X. For example, over the binary alphabet $\{a, b\}$ the set $\{a, bbb\}$ is unavoidable. Indeed, if w did not have a as a factor we necessarily have $w = b^{\mathbb{Z}}$ and w thus has bbb as a factor.

Following are two useful facts giving alternative characterizations of unavoidable sets: (1) The set $X \subset A^*$ is unavoidable if and only if there are only finitely many words in A^* with no member of X as a factor; and (2) If the set $X \subset A^*$ is finite, then X is unavoidable if and only if no periodic two-sided infinite word avoids it. Proofs can be found in [13].

We now give our extension of the definition of unavoidable sets of words to unavoidable sets of partial words.

Definition 1 Let $X \subset A_\Diamond^*$. A two-sided infinite word w over A *avoids* X if no factor of w is a member of $\hat{X}$. We say that X is *unavoidable* if no two-sided infinite word over A avoids X. In other words X is unavoidable if every two-sided infinite word over A has a factor compatible with a member of X.

For example, if our alphabet is $\{a, b\}$ then the set $X = \{aa, b\Diamond b\}$ is unavoidable. Clearly $b^{\mathbb{Z}}$ does not avoid X. Thus if there were a two-sided infinite word w avoiding X it would have an a as a factor. Without loss of generality $w(0) = a$. Then since w avoids aa, $w(-1) = w(1) = b$. Then $bab \in \hat{X}$ is a factor of w.

Clearly if all the members of X are full words, then the new definition of unavoidable set is equivalent to the old one. There is a simple connection between sets of partial words and sets of full words that is worth noting. By the definition of $\hat{X}$, w has a factor in $\hat{X}$ if and only if that same factor is compatible with a member of X. Thus the two-sided infinite words which avoid X are exactly those which avoid $\hat{X}$, and X is unavoidable if and only if $\hat{X}$ is unavoidable.

With regards to unavoidability X is then essentially a representation of a set of full words. This representation makes possible new approaches to unavoidable sets of full words. It is easier to consider the two-sided infinite words avoiding $X = \{aa, b\Diamond^3b\}$ as those without an occurrence of aa and no two occurrences of b separated by three letters rather as the words avoiding

$$\hat{X} = \{aa, baaab, baabb, babab, babbb, bbaab, bbabb, bbbab, bbbbb\}$$

It is also often useful to view a set of partial words as a collection of rules about the full words that avoid it. For example say that A is the binary alphabet $\{a, b\}$ and $X = \{a\Diamond^n b, b\Diamond^n a\}$ for some $n \in \mathbb{N}$. Let w be a word avoiding X. Whenever $w(i + n + 1) = b$, because w avoids $a\Diamond^n b$ we necessarily have that $w(i) = b$. Similarly whenever $w(i + n + 1) = a$, $w(i) = a$. Thus we can characterize the words avoiding X as exactly those with period $n + 1$.

The following remark concerning symmetry will be used to allow us to focus on one set and apply the results to all symmetric sets.

Remark 1 Say that w avoids X. The reverse word, rev(w), avoids $Y = \bigcup_{x\in X} rev\,(x)$ and $\bar{w}$ avoids $Z = {\sim} x {\in} X \,\bar{\ }\, x$.

An algorithm for determining if a set of full words is unavoidable or not, along with a proof of the correctness of the algorithm, can be found in [13]. To summarize that algorithm, if a set of full words X can be reduced to $\{\varepsilon\}$ by elementary derivations, X is unavoidable. The elementary derivations from a set X to a set Y are defined as follows:

1. Factoring—if there exist words $x, y \in X$ such that y is a factor of x, then $Y = X \setminus \{x\}$.
2. Prefix-Suffix—if there exists a word $x = ya \in X$ with $a \in A$ such that for every $b \in A$ there exists a suffix z of y where $zb \in X$, then $Y = (X \setminus \{x\}) \cup \{y\}$.
3. Shortcut—if there exists a word y such that $ya \in X$ for all $a \in A$, then $Y = (X \setminus \bigcup_{a\in A} ya) \cup \{y\}$.

Note that if a Shortcut operation is valid, then a Prefix-Suffix operation is as well. We adapt these elementary derivations to accommodate partial words as follows:

1. Factoring—if there exist partial words $x, y \in X$ and $y' \in F(x)$ such that $y \subset y'$, then $Y = X \setminus \{x\}$.
2. Prefix-Suffix—if there exists a partial word $x = ya \in X$ with $a \in A$ such that for every $b \in A$ there exists a suffix z of y and a partial word $v \in X$ with $v \subset zb$, then $Y = (X \setminus \{x\}) \cup \{y\}$.
3. Hole Truncation—if $x\Diamond^n \in X$ for some positive integer n, then $Y = (X \setminus \{x\Diamond^n\}) \cup \{x\}$.
4. Expansion—$Y = (X \setminus \{x\}) \cup \{x_1, x_2, \ldots, x_n\}$ where $\{x_1, x_2, \ldots, x_n\}$ is a partial expansion of $x \in X$.

Theorem 1 *The operations* 1–4 *as defined above preserve avoidability*, *that is*, *X is avoidable if and only if Y is avoidable*.

Proof Details follow for the Prefix-Suffix operation. If a two-sided infinite word w does not avoid x, then w does not avoid y. Now suppose, for contradiction, that w avoids X but not y. Then there exists a factor y' of w such that $y \subset y'$. Since w avoids x, let $b \neq a$ be the letter following the y' factor in w. Then by assumption, there exists a suffix z of y and a partial word $v \in X$ with $v \subset zb$. Thus, there exists a suffix z' of y' satisfying $z \subset z'$, and so $v \subset zb \subset z'b$. Therefore, $z'b$ is a factor of w compatible with $v \in X$ implying that w does not avoid X, a contradiction. ■

It was shown in [13] that the Prefix-Suffix operation for full words can also be performed on the front of the word, that is, if there exists $x = ay \in X$ with $a \in A$ such that for every $b \in A$, there exists a prefix z of y where $bz \in X$, then $Y = (X \setminus \{x\}) \cup \{y\}$ has the same avoidability as X. This is true for partial words as well and will be called the Suffix-Prefix operation.

Algorithm 1 Let $X \subset A_\Diamond^*$. Then X is unavoidable if and only if it can be transformed to $\{\varepsilon\}$ by the elementary derivations defined above. To improve the efficiency of the algorithm, the Expansion operation is only considered valid if there exists $x = ya \in X$ with $a \in A$ such that for every $b \in A$ there exist a suffix z of y and a partial word $v \in X$ where $zb \uparrow v$, and the positions chosen to expand are such that a Prefix-Suffix operation might become valid.

Lemma 1 *If Algorithm* 1 *reduces the original set down to* Y *and no more operations are valid, then there exist no valid Prefix-Suffix operations on the set* $\hat{Y}$.

Proof Suppose that there exists a word $x' = y'a \in \hat{Y}$ with $a \in A$ such that for every $b \in A$ there exists a suffix z' of y' where $v' = z'b \in \hat{Y}$. We know that x' and v' came from partial words $x, v \in Y$ and that $x = ya$ for some partial word y satisfying $y \subset y'$ (otherwise, x would end with a hole and the algorithm would not have terminated at Y since a Hole Truncation operation would be valid). There exists a suffix z of y such that $z \subset z'$, and $zb \uparrow v$. This implies that a valid Expansion operation of x exists in the set Y, so Algorithm 1 would not have terminated at Y, a contradiction. ■

Theorem 2 *Whenever Algorithm 1 is unable to reduce a set* X *to* $\{\varepsilon\}$, X *is avoidable.*

Proof First note that because every operation performed by Algorithm 1 either reduces the size of X, shortens a word in X, or reduces the number of holes in a word in X, it must eventually terminate. We know from Lemma 1 that when Algorithm 1 terminates after reducing a set X to Y, the only operations that the full word algorithm could perform on $\hat{Y}$ are Factoring or Shortcut operations. However, we know that if a Shortcut operation is valid, then a Prefix-Suffix operation must be as well, a contradiction, leaving only Factoring operations as a possibility. A Factoring operation can never make a Prefix-Suffix operation valid, as it results in the removal of a word from the set, rather than a modification of a word in the set. Also, Factoring operations alone can never reduce a set to $\{\varepsilon\}$ unless ε is already contained in the set. Since ε cannot be in Y (otherwise, Algorithm 1 would have been able to perform Factoring operations), ε cannot be in $\hat{Y}$ and thus, $\hat{Y}$ cannot be reduced to $\{\varepsilon\}$ by the full-word algorithm, and is therefore avoidable. This implies that Y, as well as the original set X, is also avoidable. Thus, whenever Algorithm 1 cannot reduce a set X to $\{\varepsilon\}$, X is avoidable. ■

The following example demonstrates how the partial word algorithm is used. The number above each arrow corresponds to the operation used to transform the set.

Example 1 Starting with the set $X = \{a\Diamond\Diamond\Diamond a, b\Diamond\Diamond\Diamond b, a\Diamond a\Diamond b\}$, Algorithm 1 may perform the following chain of derivations: $X \xrightarrow{2} \{a\Diamond\Diamond\Diamond a, b\Diamond\Diamond\Diamond b, a\Diamond a\Diamond\} \xrightarrow{3} \{a\Diamond\Diamond\Diamond a, b\Diamond\Diamond\Diamond b, a\Diamond a\} \xrightarrow{4} \{a\Diamond a\Diamond a, a\Diamond b\Diamond a, b\Diamond\Diamond\Diamond b, a\Diamond a\} \xrightarrow{1} \{a\Diamond b\Diamond a, b\Diamond\Diamond\Diamond b, a\Diamond a\} \xrightarrow{4} \{a\Diamond b\Diamond a, b\Diamond b\Diamond b, b\Diamond a\Diamond b, a\Diamond a\} \xrightarrow{2} \{a\Diamond b\Diamond a, b\Diamond b\Diamond b, b\Diamond a\Diamond, a\Diamond a\} \xrightarrow{3} \{a\Diamond b\Diamond a, b\Diamond b\Diamond b, b\Diamond a, a\Diamond a\} \xrightarrow{1} \{b\Diamond b\Diamond b, b\Diamond a, a\Diamond a\} \xrightarrow{2\&3} \{b\Diamond b, b\Diamond a, a\Diamond a\} \xrightarrow{2\&3} \{b, b\Diamond a, a\Diamond a\} \xrightarrow{1} \{b, a\Diamond a \xrightarrow{2\&3} \{b, a \xrightarrow{2} \{\varepsilon, a\} \xrightarrow{1} \{\varepsilon\}$. The operations reduced the set X to $\{\varepsilon\}$, so X is unavoidable.

Is there an efficient algorithm to determine if a finite set of partial words is unavoidable? In [5], it was shown that testing unavoidability on sets of partial words is much harder to handle than the similar problem for full words. Indeed, using a reduction from the 3SAT problem that is known to be NP-complete, the problem of determining if a finite set of partial words over a k-letter alphabet where $k \geq 2$ is unavoidable was shown to be NP-hard.

It is most natural to look first for the unavoidable sets of partial words that have small cardinality. Insight into the structure of A^* can be gained by identifying an unavoidable set, especially if that set contains few elements. For example, in Sect. 4.1 we will show that for the binary alphabet $\{a, b\}$ the set $\{a\Diamond^7 a, b\Diamond b\Diamond^3 b\}$ is unavoidable. Thus in any sufficiently long binary word either two a's occur separated by seven letters, or three b's occur for which the first two are separated by a single letter and the second two are separated by three letters.

Any set of partial words containing the empty word or $\Diamond^n$ for some $n \in \mathbb{N}$ will be called a *trivial* unavoidable set. If a set of partial words is unavoidable, then it must have an element compatible with a factor of each two-sided infinite unary word. In particular no nontrivial unavoidable set can have fewer elements than the alphabet. Every word avoids the empty set, so there are no unavoidable sets of size 0. It is also clear that unless the alphabet is unary the only unavoidable sets of size 1 are the trivial ones. If the alphabet is unary, then every nonempty set is unavoidable and in that case there is only one two-sided infinite word. We will not consider the unary alphabet further in this paper.

References [7, 8, 16] contain studies on classical pattern avoidability related to our considerations.

4 Two-Word Unavoidable Sets

To find nontrivial unavoidable sets of size 2 we may assume that $A = \{a, b\}$. Classifying the unavoidable sets of size 2 is a daunting task and is the focus of this section.

Say $X = \{x_1, x_2\}$ is unavoidable. As mentioned before if X is nontrivial it must be that one member of X is compatible with a power of a and the other is compatible with a power of b, as that is the only way to guarantee that both $a^{\mathbb{Z}}$ and $b^{\mathbb{Z}}$ will not avoid X. So in order to classify the unavoidable sets of size 2, it is sufficient to determine for which nonnegative integers $m_1, m_2,..., m_k$ and $n_1, n_2,..., n_l$ the set

$$Xm_1,...,m_k|n_1,...,n_l = \{a\Diamond^{m_1}a \cdots a\Diamond^{m_k}a,\ b\Diamond^{n_1}b \cdots b\Diamond^{n_l}b\}$$

is unavoidable. We can in fact simplify the situation a little further. The following lemma tells us that it is enough to solve the problem for cases where $m_1 +1, m_2 + 1,..., m_k + 1$ and $n_1 + 1 , n_2 + 1 ,..., n_l + 1$ are relatively prime.

Lemma 2 *Let p be a positive integer. The set $X = X_{m_1,...,m_k|n_1,...,n_l}$ is unavoidable if and only if the set*

$$Y = \{a\Diamond^{p(m1+1)-1}a \cdots a\Diamond^{p(mk+1)-1}a,\ b\Diamond^{p(n1+1)-1}b \cdots b\Diamond^{p(nl+1)-1}b\}$$

is unavoidable.

Proof In terms of notation it will be helpful to define

$$M_j = \sum_{i=1}^{j}(m_i + 1)$$

Now suppose the two-sided infinite word w avoids X, and let

$$v = \cdots (w(-1))^p\ (w(0))^p\ (w(1))^p \cdots$$

We claim that v avoids Y. Suppose otherwise. Then v has a factor compatible with some $x \in Y$. Without loss of generality say that

$$x = a\Diamond^{p(m1+1)-1}a \cdots a\Diamond^{p(mk+1)-1}a$$

Then to say that v has a factor compatible with x is equivalent to saying that there exists $i \in \mathbb{Z}$ for which

$$v(i) = v(i + pM_1) = \cdots = v(i + pM_k) = a$$

But if we set $h = \lfloor \frac{i}{p} \rfloor$ then this implies that

$$w(h) = w(h + M_1) = \cdots = w(h + M_k) = a$$

contradicting the fact that w avoids X.

We prove the other direction analogously. Suppose now that the two-sided infinite word w avoids Y, and set $v = \cdots w(-p)w(0)w(p) \cdots$. We claim that v avoids X.

Otherwise v has a factor compatible with some $x \in X$ which we may suppose without loss of generality is $a\Diamond^{m_1}a \cdots a\Diamond^{m_k}a$. Then there exists $i \in \mathbb{Z}$ for which

$$v(i) = v(i + M_1) = \cdots = v(i + M_k) = a$$

but this implies that

$$w(pi) = w(pi + pM_1) = \cdots = w(pi + pM_k) = a$$

which contradicts the fact that w avoids Y. ■

By Remark 1, two simple facts of symmetry are also worth noting. Say that w avoids

$$X_{m_1,\dots,m_k|n_1,\dots,n_l}$$

The reverse word $\cdots w(1)w(0)w(-1) \cdots$ avoids $X_{m_k,\dots,m_1|n_l,\dots,n_1}$, and the word obtained from w by swapping the a's and b's avoids $X_{n_1,\dots,n_l|m_1,\dots,m_k}$. Hence one of these sets is unavoidable precisely when all three of them are.

In order to solve the problem of identifying when $X_{m_1,\dots,m_k|n_1,\dots,n_l}$ is unavoidable we start with small values of k and l. The set $\{a, b\Diamond^{n_1}b \cdots b\Diamond^{n_l}b\}$ is unavoidable for if w is a two-sided infinite word which lacks a factor compatible with a it must be $b^{\mathbb{Z}}$. This handles the case where $k = 0$ (and symmetrically $l = 0$).

4.1 Results for $k > 0$ and $l > 0$

We first consider the case where $k = 1$ and $l = 1$, that is, we consider the set $X_{m|n} = \{a\Diamond^m a, b\Diamond^n b\}$. In this case, we can give an elegant characterization of which integers m, n make this set unavoidable.

Theorem 3 *Write $m + 1 = 2^s r_0$, $n + 1 = 2^t r_1$ where r_0, r_1 are odd. Then $X_{m|n}$ is unavoidable if and only if $s \neq t$.*

Proof Let w be a two-sided infinite word avoiding $X_{m|n}$. Then w also avoids $b\Diamond^m b$. Otherwise for some $i \in \mathbb{Z}$, $w(i) = b$ and $w(i + m + 1) = b$. Since w avoids $b\Diamond^n b$ we must have that $w(i + n + 1) = a$ and $w(i + m + 1 + n + 1) = a$, which contradicts the fact that w avoids $a\Diamond^m a$. A symmetrical argument shows that w avoids $a\Diamond^n a$.

For ease of notation, write $\bar{a} = b$ and $\bar{b} = a$. Let $p \in \mathbb{N}$. We will say that a two- sided infinite word is p-alternating if for all $i \in \mathbb{Z}$, $w(i) = \overline{w(i + p)}$. By our previous observation w avoids $X_{m|n}$ if and only if w is m+1-alternating and n+1-alternating. Thus to prove the theorem it is sufficient to show that a two-sided infinite word exists which is p-alternating and q-alternating if and only if $s = t$ where $p = 2^s r_0$ and $q = 2^t r_1$ with r_0 and r_1 odd. Notice that if w is p-alternating then it has period $2p$.

Suppose $s \neq t$. Without loss of generality say $s < t$. Then $s + 1 \leq t$. Let l be the least common multiple of p and q. The prime factorization of l must have no greater power of 2 than the prime factorization of q. Thus there exists an odd number k such that $kq \equiv 0 \bmod 2p$ (here $k = r_0$). If there were a two-sided infinite word w which was p-alternating and q-alternating we would have $w(0) = w(2p) = w(kq)$ since w has period $2p$. But since k is odd and w is q-alternating we also have $w(0) = \overline{w(kq)}$. This is a contradiction. We have half of the necessary implication.

Suppose $s = t$. Then $p = 2^s r_0$, $q = 2^s r_1$. We only need to prove that there exists some w which is p-alternating and q-alternating and we do this by induction on s. If $s = 0$, then p and q are odd. Then the word $\cdots ababab \cdots$ is p-alternating and q-alternating. This handles our base case. Now say w is $2^s r_0$ and $2^s r_1$-alternating. Then

$$v = \cdots w(-1)w(-1)w(0)w(0)w(1)w(1) \cdots$$

is $2^{s+1}r_0$ and $2^{s+1}r_1$-alternating. This finishes the induction and our proof. ■

We next consider the case where $k = 1$ and $l = 2$, that is, sets of the form $X_{m|n_1,n_2} = \{a\Diamond^m a, b\Diamond^{n_1}b\Diamond^{n_2}b\}$.

Proposition 1 *Suppose either* $m = 2n_1 + n_2 + 2$ *or* $m = n_2 - n_1 - 1$, *and* $n_1 + 1$ *divides* $n_2 + 1$. *Then* $X_{m|n_1,n_2}$ *is unavoidable if and only if* $\{a\Diamond^m a, b\Diamond^{n_1}b\}$ *is unavoidable.*

Proof If a two-sided infinite word w avoids $\{a\Diamond^m a, b\Diamond^{n_1}b\}$ then it also avoids $X_{m|n_1,n_2}$. Now we suppose instead that $\{a\Diamond^m a, b\Diamond^{n_1}b\}$ is unavoidable.

We will just consider the case $m = 2n_1 + n_2 + 2$, the situation where $m = n_2 - n_1 - 1$ is similar. Suppose for contradiction that the two-sided infinite word w avoids X. Since $\{a\Diamond^m a, b\Diamond^{n_1}b\}$ is unavoidable and w avoids $a\Diamond^m a$, w must have a factor compatible with $b\Diamond^{n_1}b$. Suppose without loss of generality that $w(0) = w(n_1 + 1) = b$. We must have that $w(n_1 + n_2 + 2) = a$ which immediately gives us

$$w(n_1 + n_2 + 2 - m - 1) = w(n_1 + n_2 + 1 - 2n_1 - n_2 - 2) = w(-n_1 - 1) = b$$

Since $w(-n_1 - 1) = w(0) = b$, we must have $w(n_2 + 1) = a$. By an easy induction we can verify that this process continues, and we ultimately find that

$$a = w(n_2 + 1) = w(n_2 + 1 - (n_1 + 1)) = w(n_2 + 1 - 2(n_1 + 1)) = \cdots$$

Since $n_1 + 1$ divides $n_2 + 1$ we find that $w(0) = a$, a contradiction. ■

One notable consequence of Proposition 1 is that if m is odd, then both $\{a\Diamond^m a, bb\Diamond^{m+1}b\}$ and $\{a\Diamond^m a, bb\Diamond^{m-2}b\}$ are unavoidable.

The next theorem takes advantage of the perpetuating pattern phenomenon in a more complicated context. Proposition 1 held because each a forced a b into the next position of an occurrence of $w(i) = w(i + n_1 + 1) = b$, which in turn forced a new a in w. This created a single traveling sequence of a's and b's, causing an a to overlap with the b at $w(0)$, yielding a contradiction. In the next argument, we take notice of the fact that each a occurring in w may contribute to two occurrences of $w(i) = w(i + n_1 + 1) = b$ simultaneously so that a contradiction will occur after many traveling sequences of letters appear and overlap.

Theorem 4 *Say that* $m = n_2 - n_1 - 1$ *or* $m = 2n_1 + n_2 + 2$, *and that the highest power of* 2 *dividing* $n_1 + 1$ *is less than the highest power of* 2 *dividing* $m + 1$. *Then* $X_{m|n_1,n_2}$ *is unavoidable.*

Proof Since the highest power of 2 dividing $n_1 + 1$ is different than the highest power of 2 dividing $m + 1$, we have that the set $Y = \{a\Diamond^m a, b\Diamond^{n_1}b\}$ is unavoidable. Consider first the case where $m = n_2 - n_1 - 1$ and suppose for contradiction that there exists a two-sided infinite word w that avoids X. Then w has no factor compatible with $a\Diamond^m a$, and so since Y is unavoidable it must have a factor compatible with *bon*1 *b*. Assume without loss of generality that $w(0) = b$ and $w(n_1 + 1) = b$.

We now generate an infinite table of facts about w. Two horizontally adjacent entries in the table will represent positions in w which are $n_1 + 1$ letters apart. Two vertically adjacent entries in the table will represent positions in w which are $m + 1 = n_2 - n_1$ letters apart. The two upper left entries of our table are $w(0) = b$ and $w(n_1 + 1) = b$, two facts we have already assumed. Since w avoids $X_{m|n_1,n_2}$ we have more information relevant to the table: two horizontally adjacent b entries force an a entry diagonally down and to the right from them, and an a entry forces a b entry in the vertically adjacent positions. From these rules we can build the following table, labeling the columns C_0, C_1,...:

C_0	C_1	C_2	C_3	$\cdots$
$w(0)=b$	$w(n_1+1)=b$	$w(2n_1+2)=b$	$w(3n_1+3)=b$	
		$w(n_1+n_2+2)=a$	$w(2n_1+n_2+3)=a$	
		$w(2n_2+2)=b$	$w(n_1+2n_2+3)=b$	

For $i \in \mathbb{N}$, we shall define v_i to be the factor of w represented by C_i. If i is odd then C_i has i entries, and if i is even then C_i has $i + 1$ entries. Thus we define

$$v_i \begin{cases} \mathrm{w(in1\ +\ i)w(in1\ +\ i\ +\ 1)\cdots\ w(in2\ +\ i)} & \text{if i even} \\ \mathrm{w((in1\ +\ i)w(in1\ +\ i\ +\ 1)\ \cdots\ w(n1 + (i-1)n2 + i)} & \text{if } i \text{ odd} \end{cases}$$

Two adjacent entries in C_i represent a distance of $m + 1$ positions between letters in v_i. Thus for i even we have that $|v_i| = im + 1$ and for i odd we have that $|v_i| = (i - 1)m + 1$. We can also use the table to get some partial information about the positions of a's and b's in v_i. For $j \in \mathbb{N}$, $v_i(j) = b$ if $j \equiv 0 \bmod (2m + 2)$, and $v_i(j) = a$ if $j \equiv (m + 1) \bmod (2m+2)$.

Because the highest power of 2 dividing $n_1 + 1$ is no greater than the highest power of 2 dividing $m_1 + 1$, there exists some k for which $k(n_1 + 1) \equiv (m + 1) \bmod (2m + 2)$. Take i sufficiently large so that $|v_i| > kn_1 + k$. Because of how k was chosen, we have that $v_i(kn_1 + k) = a$. However examining the table we see that

$$w((i + k)n_1 + i + k) = vi(kn_1 + k) = v_{i+k}(0) = b$$

a contradiction. This handles the situation where $m = n_2 - n_1 - 1$. The proof for the case where $m = 2n_1 + n_2 + 2$ is similar, the only difference is that the table will represent increasingly negative positions of w, rather than increasingly positive ones. ■

As an application of Theorem 4, take $m = 1$. Let us see for which $n_1 \in \mathbb{N}$ the hypotheses of the theorem hold to make $X_{m|n_1,n_2}$ unavoidable. The highest power of 2 dividing $n_1 + 1$ should be less than the highest power of 2 dividing $m + 1 = 2$. Thus $n_1 + 1$ must be odd, n_1 is even. Since $m = 1$ we cannot have $m = 2n_1 + n_2 + 2$. Say we have $m = n_2 - n_1 - 1$. Then $n_2 = n_1 + 2$. So we have that for any even n_1, the set $\{a\Diamond a, b\Diamond^n b\Diamond^{n+2} b\}$ is unavoidable. We will prove in Sect. 4.2 that this is a complete characterization of unavoidability of $X_{m|n_1,n_2}$ for $m = 1$.

The next proposition identifies another large class of unavoidable sets using a modification of the strategies discussed so far.

Proposition 2 *Suppose* $n\ 1 < n_2$, $2m = n_1 + n_2$ *and* $|m - n_1|$ *divides* $m + 1$. *Then* $X_{m|n_1,n_2}$ *is unavoidable*.

Proof Say for contradiction that the two-sided infinite word w does avoid $X_{m|n_1,n_2}$. An a occurs in w, say without loss of generality that $w(0) = a$. Then $w(-m - 1) = w(m + 1) = b$. We have $n_1 + 1 + n_2 + 1 = 2m + 2$, and so since w avoids $b\Diamond^{n_1}b\Diamond^{n_2}b$ we necessarily have $w(-m - 1 + n_1 + 1) = w(n_1 - m) = a$. Repeating this argument,

$$w(2(n_1 - m)) = w(3(n_1 - m)) = \cdots = a$$

But since $n_1 - m$ divides $m + 1$, $w(-m - 1) = a$, a contradiction. ■

By taking $n_1 = m - 1$ and $n_2 = m + 1$, Proposition 2 yields a nice fact: the set $\{a\Diamond^m a, b\Diamond^{m-1}b\,\Diamond^{m+1} b\}$ is unavoidable for all $m > 0$.

We believe that together Lemma 2, Propositions 1, 2, and Theorem 4 nearly give a complete characterization of when $X_{m|n_1,n_2}$ is unavoidable. Following is what we believe to be the only exception.

Proposition 3 *The set* $X_{6|1,3} = \{a\Diamond^6 a, b\Diamond b\Diamond^3 b\}$ *is unavoidable*.

Proof We first claim that any two-sided infinite word w which avoids $X_{6|1,3}$ must also avoid $b\Diamond b\Diamond b$. Suppose otherwise. Then w avoids $X_{6|1,3}$ but has a factor compatible with $b\Diamond b\Diamond b$. Without loss of generality say that $w(0) = w(2) = w(4) = b$. Then we have $w(6) = w(8) = a$, which in turn implies that $w(-1) = w(1) = b$. This implies that $w(5) = a$, which tells us that $w(-2) = b$. Since $w(-2) = w(0) = b$, $w(4) = a$, a contradiction.

Now suppose for contradiction that the two-sided infinite word w avoids $X_{6|1,3}$. It must avoid $a\Diamond^6 a$, and since $\{a\Diamond^6 a, b\Diamond b\}$ is unavoidable it has a factor compatible with $b\Diamond b$. Say without loss of generality that $w(0) = w(2) = b$. The reader may verify that this ultimately leads to a contradiction, using the fact that w avoids $X_{6|1,3}$ and $\{b\Diamond b\Diamond b\}$. ■

We now state our conjecture.

Conjecture 1 *The set $X_{m|n_1,n_2}$ is unavoidable precisely when the hypotheses of at least one of Lemma* 2, *Propositions* 1, 2, 3 *or Theorem* 4 *hold. Restated,*

$X_{m|n_1,n_2}$ is unavoidable for relatively prime $m + 1$, $n_1 +1$ and $n_2 +1$ with $n_1 \le n_2$ if and only if one of thefollowing conditions (or their symmetric equivalents) hold:

- *Proposition* 1: *The case where the set $\{a\Diamond^m a, b\Diamond^{n_1} b\}$ is unavoidable, $m = 2n_1 + n_2 + 2$ or $m = n_2 - n_1 - 1$, and $n_1 + 1$ divides $n_2 + 1$.*
- *Theorem* 4: *The case where $m = n_2 - n_1 - 1$ or $m = 2n_1 + n_2 + 2$, and the highest power of* 2 *dividing $n_1 + 1$ is less than the highest power of* 2 *dividing $m + 1$.*
- *Proposition* 2: *The case where $n_1 < n_2$, $2m = n_1 + n_2$ and $|m - n_1|$ divides $m + 1$.*
- *Proposition* 3: *The case where $m = 6$, $n_1 = 1$ and $n_2 = 3$.*

The reader may verify that for any fixed m the only one of the above conditions that contributes infinitely many unavoidable sets to $X_{m|n_1,n_2}$ is Theorem 4, and that this theorem never applies to even m. Thus the conjecture states that there are only finitely many values of m, n_1, n_2 with m fixed and even and $X_{m|n_1,n_2}$ unavoidable. We will prove in Sect. 4.2 that this is indeed the case.

Using Lemma 2 we may assume without loss of generality that $m + 1$, $n_1 + 1$, $n_2 + 1$ are relatively prime. An important consequence of the conjecture is that in order for $X_{m|n_1,n_2}$ to be unavoidable it is necessary that either $m = 6$ and $\{n_1, n_2\} = \{1, 3\}$, or that one of the following equations hold:

$$m = 2n_1 + n_2 + 2 \tag{1}$$

$$m = 2n_2 + n_1 + 2 \tag{2}$$

$$m = n_1 - n_2 - 1 \tag{3}$$

$$m = n_2 - n_1 - 1 \tag{4}$$

$$2m = n_1 + n_2 \tag{5}$$

Using this fact we can show that an affirmative proof of the conjecture has a powerful consequence.

We end this section with the following proposition which implies that if Conjecture 1 is true then we have completely classified the unavoidable sets of size 2.

Proposition 4 *If Conjecture* 1 *holds, then $X_{m_1,\ldots,m_k|n_1,\ldots,n_l}$ is avoidable for all $k \ge 2$ and $l \ge 2$, and for all $k \ge 1$ and $l \ge 3$.*

Proof Assuming Conjecture 1 holds, it is enough to prove that both

$$X_{m_1 \mathrm{m}_2|n_1,n_2} \text{ and } X_{m|n_1,n_2 n_3}$$

are avoidable for all m_1,m_2,n_1,n_2. We handle the case of $X_{m_1 \mathrm{m}_2|n_1,n_2}$. Assume without loss of generality that m_1, m_2, n_1, n_2 are relatively prime. In order for this set to be unavoidable it is necessary that the sets

$$\{a\Diamond^{m_1}a \;,\; b\Diamond^{n_1}\; b\Diamond^{n_2}b\},\; \{a\Diamond^{m_2}a,\; b\Diamond^{n_2}b\Diamond^{n_2}b\},\; \{a\Diamond^{m_1}a\Diamond^{m_2}a,\; b\Diamond^{n_1}b\}$$

and the set $\{a\Diamond^{m_1}a\Diamond^{m_2}a,\; b\Diamond^{n_2}b\}$ are unavoidable as well. For each of these sets, Conjecture 1 gives a necessary condition: either $m = 6$ and $n_1 = 1$, $n_2 = 3$ (or symmetrically $n_1 = 3$, $n_2 = 1$) or one of (1), (2), (3), (4) or (5) must hold. Consider the following tables:

$m_1 = 2n_1 + n_2 + 2$	$m_2 = 2n_1 + n_2 + 2$
$m_1 = 2n_2 + n_1 + 2$	$m_2 = 2n_2 + n_1 + 2$
$m_1 = n_1 - n_2 - 1$	$m_2 = n_1 - n_2 - 1$
$m_1 = n_2 - n_1 - 1$	$m_2 = n_2 - n_1 - 1$
$m_1 = 6, n_1 = 1, n_2 = 3$	$m_2 = 6, n_1 = 1, n_2 = 3$
$m_1 = 6, n_2 = 1, n_1 = 3$	$m_2 = 6, n_2 = 1, n_1 = 3$
$2m_1 = n_1 + n_2$	$2m_2 = n_1 + n_2$
$n_1 = 2m_1 + m_2 + 2$	$n_2 = 2m_1 + m_2 + 2$
$n_1 = 2m_2 + m_1 + 2$	$n_2 = 2m_2 + m_1 + 2$
$n_1 = m_1 - m_2 - 1$	$n_2 = m_1 - m_2 - 1$
$n_1 = m_2 - m_1 - 1$	$n_2 = m_2 - m_1 - 1$
$n_1 = 6, m_1 = 1, m_2 = 3$	$n_2 = 6, m_1 = 1, m_2 = 3$
$n_1 = 6, m_2 = 1, m_1 = 3$	$n_2 = 6, m_2 = 1, m_1 = 3$
$2n_1 = m_1 + n_2$	$2n_2 = m_1 + m_2$

In order for $X_{m_1 m_2|n_1,n_2}$ to be unavoidable it is necessary that at least one equation from each column be satisfied. It is easy to verify using a computer algebra system that this is impossible except in the case where the last equation in each column is satisfied. However in this case $m_1 = m_2 = n_1 = n_2$ and so by Theorem 3 the set is avoidable.

Now we consider $X_{m|n_1 n_2,n_3}$. In order for this set to be unavoidable, it is necessary that $\{a\Diamond^m a, b\Diamond^{n_1}b\Diamond^{n_2}b\}$, $\{a\Diamond^m a, b\Diamond^{n_2}b\Diamond^{n_3}b\}$, $\{a\Diamond^m a, b\Diamond^{n_1+n_2+1}b\Diamond^{n_3}b\}$ and $\{a\Diamond^m a,\; b\Diamond^{n_1}b\Diamond^{n_2+n_3+1}b\}$ be unavoidable as well. Again, for each of these sets Conjecture 1 gives a necessary condition: either $m = 6$ and $\{n_1, n_2\} = \{1, 3\}$ or one of (1), (2), (3), (4) or (5) must hold. Consider now the following tables:

$m = 2n_1 + n_2 + 2$	$m = 2n_2 + n_3 + 2$
$m = 2n_2 + n_1 + 2$	$m = 2n_3 + n_2 + 2$
$m = n_1 - n_2 - 1$	$m = n_2 - n_3 - 1$
$m = n_2 - n_1 - 1$	$m = n_3 - n_2 - 1$
$2m = n_1 + n_2$	$2m = n_2 + n_3$
$m = 6, n_1 = 1, n_2 = 3$	$m = 6, n_2 = 1, n_3 = 3$
$m = 6, n_2 = 1, n_1 = 3$	$m = 6, n_3 = 1, n_2 = 3$
$m = 2(n_1 + n_2 + 1) + n_3 + 2$	$m = 2n_1 + n_2 + n_3 + 3$
$m = 2n_3 + (n_1 + n_2 + 1) + 2$	$m = 2(n_2 + n_3 + 1) + n_1 + 2$
$m = (n_1 + n_2 + 1) - n_3 - 1$	$m = n_1 - (n_2 + n_3 + 1) - 1$
$m = n_3 - (n_1 + n + 2 + 1) - 1$	$m = (n_2 + n_3 + 1) - n_1 - 1$
$2m = (n_1 + n_2 + 1) + n_3$	$2m = n_1 + (n_2 + n_3 + 1)$
$m = 6, n_1 + n_2 + 1 = 1, n_3 = 3$	$m = 6, n_1 = 1, n_2 + n_3 + 1 = 3$
$m = 6, n_3 = 1, n_1 + n_2 = 3$	$m = 6, n_2 + n_3 = 1, n_1 = 3$

Again unavoidability of $X_{m|n_1n_2,n_3}$ requires that one equation from each column be satisfied. It is easy to verify that no such system of equations has a nonnegative solution. ■

4.2 Other Results for $k = 1$ and $l = 2$

In order to prove the conjecture, only one direction remains. We must show that if none of the hypotheses of Lemma 2, Propositions 1, 2, 3 or Theorem 4 hold then $X_{m|n_1n_2}$ is avoidable. In this section we give partial results towards this goal.

We have found that in general identifying sets of the form $X_{m|n_1n_2}$ as avoidable tends to be a more difficult task than identifying them as unavoidable. In the case of unavoidability we needed only consider a single word then derive a contradiction from its necessary structural properties. To find a class of avoidable sets we must invent some general procedure for producing a two-sided infinite word which avoids each such set. This is precisely what we move towards in the following propositions in which we verify that the conjecture holds for certain values of m and n_1.

It is easy to see that none of (1), (2), (3), (4) or (5) are satisfied when $\max(n_1, n_2) < m \leq n_1 + n_2 + 2$. Thus the conjecture for such values is that $X_{m|n_1n_2}$ is avoidable. The following fact verifies that this is indeed the case.

Proposition 5 *If* $\max(n_1, n_2) < m < n_1 + n_2 + 2$ *then* $X_{m|n_1n_2}$ *is avoidable.*

Proof Let $v = a^m b^{m+1}$ and $w = v^{\mathbb{Z}}$. We claim that w avoids $X_{m|n_1n_2}$. Clearly it avoids $a\Diamond^m a$. Let $i \in \mathbb{Z}$. If $w(i) = w(i + n_1 + 1) = b$, then the gap in $b\Diamond^{n_1}b$ cannot straddle a block of a's, since $n_1 < m$ and these blocks come in sequences m letters long. Thus we must have $w(i) \cdots w(i + n_1 + 1) = b^{n_1+2}$. Similarly if $w(i) = w(i + n_2 + 1) = b$ since $n_2 < m$ we have $w(i) \cdots w(i + n_2 + 1) = b^{n_2+2}$. Hence if therewere $i \in \mathbb{Z}$ with $w(i)=w(i + n_1 +1)= w(i + n_1 + n_2 + 2) = b$, we would have $w(i) \cdots w(i + n_1 + n_2 + 2) = b^{n_1+n_2+3}$ which is impossible since $m + 1 < n_1 + n_2 + 3$. ■

The next proposition makes the conjecture easy to verify for a large number of even values of m.

Proposition 6 *Assume m is even and that* $2m \leq \min(n_1, n_2)$. *Then* $X_{m|n_1n_2}$ *is avoidable.*

Proof If either n_1 or n_2 is even then either $\{a\Diamond^m a, b\Diamond^{n_1}b\}$ or $\{a\Diamond^m a, b\Diamond^{n_2}b\}$ is avoidable by Theorem 3. Both situations imply that $X_{m|n_1n_2}$ is avoidable. Thus we only need to prove that $X_{m|n_1n_2}$ is avoidable for n_1, n_2 odd. Say without loss of generality that $n_1 \leq n_2$.

Let $v = b^m$, and let $u = baba \cdots ab$ with $|u| = n_2 + 2$. We claim that $w = (uv)^{\mathbb{Z}}$ avoids $X_{m|n_1n_2}$. Clearly it avoids $a\Diamond^m a$. Because of periodicity, it is enough to prove that for any $i \in \{0,..., n_2 + 2 + m - 1\}$ if $w(i - n_1 - 1) = b$ and $w(i) = b$ then $w(i + n_2 + 1) = a$. We claim that such an i must be greater than m. Suppose for contradiction $i \leq m$. Then $-|uv| = -n_2 - 2 - m < i - n_1 - 1 < -m = -|v|$. Thus $w(i - n_1 - 1)$ occurs in the repetition of u at $w(-n_2 - 1 - m) \cdots w(-m - 1)$, and so since $i - n_1 - 1$ is an even number, $w(i - n_1 - 1) = a$ which is a contradiction. Thus we indeed have that $i > m$.

Since $i > m$ we have that

$$|uv| = n_2 + 2 + m \leq i + n_2 + 1 \leq n_2 + 2 + m - 1 + n_2 + 2$$
$$= 2n_2 + m + 3 = |uv| + |u| - 1$$

Thus $w(i + n_2 + 1)$ occurs in the second repetition of u at $w(n_2 + 2 + m) \cdots w(2n_2 + m + 3)$, and so since $i + n_2 + 1$ is an even number, $w(i + n_2 + 1) = a$. ■

For any fixed even m there are then only finitely many values of n_1, n_2 which might be unavoidable. The reader may verify that this is consistent with the conjecture. The reader may also verify that the conjecture for $m = 0$ is that $X_{0|n_1n_2}$ is always avoidable, and indeed this is given by Proposition 6. Similarly the conjecture for $m = 2$ is that $X_{2|n_1n_2}$ is avoidable except for $n_1 = 1$, $n_2 = 3$ or $n_2 = 3$, $n_1 = 1$. It is easy to find avoiding two-sided infinite words for other values of n_1 and n_2 less than 5 when $m = 2$. By Proposition 6 this is all that is necessary to confirm the conjecture for $m = 2$. In this way we have been able to verify the conjecture for all even m up to very large values. The odd values of m seem to be much more difficult and will most likely require more sophisticated techniques. The following proposition gives our confirmation of the conjecture for $m = 1$.

Proposition 7 *The conjecture holds for $m = 1$, that is, $X_{1|n_1n_2}$ is unavoidable if and only if n_1 and n_2 are even numbers with $|n_1 - n_2| = 2$.*

Proof That $X_{1|n_1n_2}$ is unavoidable for n_1 and n_2 even with $|n_1 - n_2| = 2$ is a direct consequence of Theorem 4 and was explained in Sect. 4.1. Thus we only need to prove that the set is avoidable for other values of n 1 and $n2$. We divide these values of n_1 and n_2 into cases and prove that $X_{1|n_1n_2}$ is avoidable in each case. By symmetry we may assume that $n_1 \leq n_2$ throughout.

Claim 1. For n_1 or n_2 odd $X_{1|n_1n_2}$ is avoidable. If both n_1 and n_2 are odd then $m + 1$, $n_1 + 1$, $n_2 + 1$ are all divisible by 2. Thus by applying Lemma 2 with Proposition 6 we find that $X_{1|n_1n_2}$ is avoidable. If either n_1 or n_2 is equivalent to 1 mod 4 then $X_{1|n_1n_2}$ is avoidable by Theorem 3. The last case to consider is when n_1 or n_2 is equivalent to 3 mod 4. By symmetry we may suppose that $n_1 \equiv 3 \bmod 4$ and n_2 is even. Thus we divide into two further cases.

First say $n_2 \equiv 0 \bmod 4$ and write $n_2 = 4k$. Let $v = (aabb)^k\, abb$. We prove that $w = v^{\mathbb{Z}}$ avoids $X_{1|n_1n_2}$. Certainly it avoids $a\Diamond a$. By periodicity we may assume without loss of generality that $i \in \{0,..., |v| - 1\}$ and $w(i + n_1 + 1) = w(i + n_1 + n_2 + 2) = b$. We then only need to prove that $w(i) = a$. Since w is $|v| = (4k + 3)$-periodic, we have that $w(i + n_1 + n_2 + 2) = w(i + n1 + n2+2-n2-3) = w(i +n1 - 1)$. Examining v we see that $w(i +n_1 - 1) = w(i + n_1 + 1) = b$ can only occur if $i + n_1 + 1 = 4k + 1$. It is easy to see that since $n_1 + 1 \equiv 0 \bmod 4$ and n $1 \leq n_2$ that $w\,(i + n_1 + 1 - n_1 - 1) = a$, and so $w\,(i) = a$.

For the second case say $n_2 \equiv 2 \bmod 4$ and write $n_2 = 4k + 2$. The reader may verify using a similar argument that $((aabb)^k\, aabbb)^{\mathbb{Z}}$ avoids $X_{1|n_1n_2}$ in this case and the claim is proved.

Claim 2. If $n_1 < n_2 - 2$ and either $n_1 \equiv 0 \bmod 4$ and $n_2 \equiv 2 \bmod 4$, or $n_1 \equiv 2 \bmod 4$ and $n_2 \equiv 0 \bmod 4$, then $X_{1|n_1n_2}$ is avoidable. Take the first case, $n_1 \equiv 0 \bmod 4$ and $n_2 \equiv 2 \bmod 4$. Write $n_2 = 4k + 2$, with $k \geq 1$ which is valid since we have assumed $n_1 < n_2 - 2$. Let $v = (aabb)^{k-1}\, aabbb$. Our argument is similar to those used for the last claim. In particular we show that $w = v^{\mathbb{Z}}$ avoids $X_{1|n_1n_2}$. Certainly it avoids $a\Diamond a$. By periodicity we may assume without loss of generality that $i \in \{0,..., |v| - 1\}$ and $w(i + n_1 + 1) = w(i + n_1 + n_2 + 2) = b$. We then only need to prove that $w\,(i) = a$. Since w is $|v| = (4k + 1)$-periodic, we have that $w\,(i + n_1 + n_2 + 2) = w(i + n_1 + n_2 + 2 - n_2 +1) = w(i + n_1 + 3)$. Examining v we see that $w(i +n_1 + 1) = w(i +n_1 +3) = b$ can only occur for $w(i +n_1 + 1) = 4k - 2$. It is easy to see that since $n_1 + 1 \equiv 1 \bmod 4$ and $n_1 < n_2 - 2$ that $w(i + n_1 + 1 - n_1 - 1) = a$, and so $w\,(i) = a$.

For the second case, where $n_1 \equiv 2 \bmod 4$ and $n_2 \equiv 0 \bmod 4$, write $n_2 = 4k$. Then $((aabb)^{k-1}\, abb)$ avoids $X_{1|n_1n_2}$ and the claim is proved.

The only possible values of n_1 and n_2 left to consider are where $n_1, n_2 \equiv 0 \bmod 4$ or $n_1, n_2 \equiv 2 \bmod 4$.

Claim3. If $n_1 < n_2 - 2$ and either $n_1, n_2 \equiv 0 \bmod 4$ or $n_1, n_2 \equiv 2 \bmod 4$ then $X_{1|n_1n_2}$ is avoidable. First say $n_1, n_2 \equiv 0 \bmod 4$ and write $n_2 = 4k$. In this case $(aabb)^k\, abb$ avoids $X_{1|n_1n_2}$. Secondly suppose $n_1, n_2 \equiv 2 \bmod 4$ and write $n_2 = 4k + 2$. In this case $(aabb)^k\, aabbb$ avoids $X_{1|n_1n_2}$. ■

The following and final proposition says that if m and n_1 are close enough in value then $X_{1|n_1 n_2}$ is avoidable for large enough n_2.

Proposition 8 *Let* $s \in \mathbb{N}$ *with* $s < m - 2$. *Then for* $n > 2(m + 1)^2 + m - 1$, $X_{m|m+s,n} = \{a\Diamond^m a, b\Diamond^{m+s} b\Diamond^n b\}$ *is avoidable*.

Proof For any $p \in \mathbb{N}$ all integers greater than p^2 can be written as $pq + (p + 1) r$ for some $q, r \in \mathbb{N}$. This is because

$$pp, (p - 1)p + p + 1, (p - 2)p + 2(p + 1), \ldots, p + (p - 1)(p + 1)$$

is a sequence of consecutive integers with p members.

Now let $C = \{b^{m+1}a^{m+1}, b^{m+2}a^{m+1}\}$. There exists $u \in C^*$ with $|u| = n - m - 2$. We claim that $w = u^{\mathbb{Z}}$ avoids $X_{m\,|m+s,n}$. It certainly avoids $a\Diamond^m a$. We need to verify that whenever $w(i - m - s - 1) = b$ and $w(i) = b$ that $w(i + n + 1) = a$. Examining C we see that the only i's for which this is possible are those for which $w(i)$ is part of an initial segment of s b's in a sequence of b's. Say without loss of generality that

$$w(0)w(1) \cdots w(s) \cdots w(m) = b^m \text{ and } w(m+1) = a$$

Since w is $(n - m - 2)$-periodic, $w(s + n + 1) = w(s + m + 3)$, $s + m + 3 < m - 2 + m + 3 = 2m + 1$ so $w(s + n + 1) = a$. Similarly $w(0 + n + 1) = w(m + 3) = a$. Thus w has no factor compatible with $b\Diamond^{m+s} b\Diamond^n b$ and so w avoids $X_{m|m+s,n}$. ■

5 Three-Word Unavoidable Sets

In this section, we restrict ourselves to three-element sets. As discussed earlier, if the cardinality of an alphabet is larger than the number of words in the set, the set is always avoidable. Since the set here has size 3, the alphabet is unary, binary or ternary. As previously stated, the unary alphabet is trivial. We discuss three-word unavoidable sets over a binary alphabet in Sect. 5.1 and over a ternary alphabet in Sect. 5.2.

5.1 The Case of a Binary Alphabet

Here, we assume that the alphabet A is binary with distinct letters a and b. We will be considering nontrivial sets of size 3 over A that have the form

$$X = \{a\Diamond^m a, b\Diamond^n b, u\} \qquad (6)$$

where u is a partial word over A, and no two-word subset of X is unavoidable. The question is: What forms can the partial word u take so that the set X is unavoidable?

We first investigate for which integers m, n and l the set

$$X = \{a\Diamond^m a, b\Diamond^n b, a\Diamond^l x\}$$

is unavoidable where $x \in A$.

Theorem 5 *If* $X = \{a\Diamond^m a, b\Diamond^n b, a\Diamond^l b\}$ *for some integers m, n and l, then X is unavoidable if and only if* $\gcd(l + 1, m + n + 2)$ *is odd or there exists some integer j such that* $(m + 1) \equiv j(l + 1) \bmod (m + n + 2)$.

Proof Let $d = \frac{m+n+2}{\gcd(l+1,m+n+2)}$. First note that any two-sided infinite word w which avoids X must be $(m + n + 2)$-periodic in order to avoid the words $a\Diamond^m a$ and $b\Diamond^n b$, and we let z denote one period of w. For the remainder of the proof we consider any position i of z to be i mod $(m + n + 2)$ as any index which is bigger than $|z|$ correlates to a position in another occurrence of z in w. We can assume without loss of generality that position 0 of z is an a, as we know w must contain at least one a. We know that in order to avoid $a\Diamond^l b$, position $i\,(l + 1)$ of z must also be an a for all $i \in \mathbb{N}$. This accounts for exactly d of the $m + n + 2$ positions of z. For each a we have placed, we must place a b in position $i\,(m + l + 2)$ of z in order to avoid $a\Diamond^m a$. If there exists some integer j such that $(m + 1) \equiv j(l + 1) \bmod (m + n + 2)$, then these b's will need to be at the same positions as the a's already placed, resulting in a contradiction, indicating that X is unavoidable. Otherwise, all of these b's will be placed in unfilled positions of z, bringing the total number of determined positions in z to $2d$.

Notice that at least one of the unfilled positions of z must be filled with an a. To show this, first note that every a already placed has a b $m + 1$ spaces after it, which, due to the length of z being $m + n + 2$, results in the same b being $n + 1$ spaces before it. Since every a has the position $n + 1$ spaces behind it filled, any unfilled position in z either has a b or an unfilled position $n + 1$ spaces after it. In either case, there is an unfilled position in z which must be filled with an a to avoid $b\Diamond^n b$. Furthermore, placing a single a in any of the unfilled positions of z amounts to filling $2d$ of the unfilled positions of z, as the same restrictions which generated the first group from the original a can be applied to this new a.

If $\gcd(l + 1, m + n + 2)$ is odd, then we can never entirely fill z in groups of $2d$ positions, meaning that there is no word w which avoids X and thus X is unavoidable.

If $\gcd(l + 1, m + n + 2)$ is even, then using the same construction as before, we can completely fill z in groups of $2d$ positions, forming a word $w = z^{\mathbb{Z}}$ which avoids X, so X is avoidable. ■

Theorem 6 *If $X = \{a\Diamond^m a, b\Diamond^n b, a\Diamond^l a\}$ for some integers m, n and l, then X is unavoidable if and only if some proper subset of X is unavoidable*.

Proof Let $d = \frac{2(m+n+2)}{\gcd(l+n+2,m+n+2)}$. Clearly if some proper subset of X is unavoidable, X is also unavoidable, and so we only need to show that if every proper subset of X is avoidable, then X is avoidable. Assume that X contains no unavoidable proper subsets. Note that, similar to Theorem 5, any two-sided infinite word w which avoids X must be $(m + n + 2)$ -periodic. Let z denote one such period of w, and let the zeroth position of z be an a, and again consider any position i of z to be i mod $(m + n + 2)$. To avoid the words $a\Diamond^l a$ and $b\Diamond^n b$, we can fill in every position p of z with an a, where $p \equiv i\,(l + n + 2) \bmod (m + n + 2)$ for some $i \in \mathbb{N}$, and every position q of z with a b, where $q \equiv (i\,(l + n + 2) + l + 1) \bmod (m + n + 2)$. This will fill in exactly d positions of z. Furthermore, every a placed in z has a b $m + 1$ places after it, ensuring that z does not contain a factor compatible with $a\Diamond^m a$. To see this, note that the positions of z that are $m + 1$ places away from an a, which have the form $(i\,(l + n + 2) + m + 1) \bmod (m + n + 2)$, are exactly the spaces already filled in with b's. Comparing the two forms, we see that if there exists an integer k such that

$$(m + 1) \equiv (k(l + n + 2) + l + 1) \bmod (m + n + 2)$$

then the indices are identical. If $k = m + n + 1$, this condition is satisfied, and so we can conclude that z does not contain a factor compatible with $a\Diamond^m a$.

Now we only need to show that we can fill in the remaining positions of z with this same pattern. We know that if $\gcd(l + n + 2, m + n + 2)$ is even, we will be able to completely fill in z with these groups of d positions. Furthermore, because every proper subset of X is avoidable, we know from Theorem 3 that m, n and l all have the same parity. Thus, both $l + n + 2$ and $m + n + 2$ are even, so $\gcd(l + n + 2, m + n + 2)$ must be even, allowing us to make a full word z such that $w = z^{\mathbb{Z}}$ avoids X when no proper subset of X is unavoidable. ■

Theorems 5 and 6 completely characterize all three-word unavoidable sets where each word has only two defined positions and provide us with some conditions that more complex sets must satisfy in order to be unavoidable.

The following proposition provides some necessary conditions for the unavoidability of a three-word set X with one word having three defined positions.

Proposition 9 *If $X = \{a\Diamond^m a, b\Diamond^n b, a\Diamond^l x\Diamond^p y\}$ for some integers m, n, l, p and some letters $x, y \in A$, then X is unavoidable only if the conditions from Theorems* 5 *and* 6 *hold for all the substring sets of X.*

Proof For X to be avoidable, it is sufficient that one of the sets

$$\{a\Diamond^m a, b\Diamond^n b, a\Diamond^l x\}, \{a\Diamond^m a, b\Diamond^n b, x\Diamond^p y\} \text{ or } \{a\Diamond^m a, b\Diamond^n b, a\Diamond^{l+p+1} y\}$$

be avoidable. We call these sets the *substring sets* of X. So all the substring sets of X must be unavoidable in order for X to be unavoidable, and the substring sets have the same forms as in Theorems 5 and 6. ■

We obtain the following two corollaries.

Corollary 1 *The set $X = \{a\Diamond^n a, b\Diamond^n b, a\Diamond^m b\}$ is unavoidable for some integers m and n if and only if $L = \frac{lcm(n+1,m+1)}{n+1}$ is odd.*

Proof To show the result, we will first show that the conditions of Theorem 5 imply that L is odd.

Suppose that $d = \gcd(m + 1, 2(n + 1))$ is odd. Set $m + 1 = k_1 d$ and $n + 1 = k_2 d$ for some k_1, k_2. For d to be odd, $m + 1$ must be odd so k_1 is odd. Since k_1 is odd, $L = \frac{k1k2d}{k2d} = k1$ is odd.

Suppose that there exists some integer j such that $(n + 1) \equiv j(m + 1) \bmod (2n + 2)$. Let $d = \gcd(m + 1, n + 1)$ and set $m + 1 = k_1 d$ and $n + 1 = k_2 d$, where k_1, k_2 must be relatively prime. We can then write $j(k_1 d) \equiv k_2 d \bmod (2k_2 d)$, which is equivalent to $jk_1 = (2i + 1)k_2$ for some i. If k_2 is even, then k_1 is odd since this is the only possible way that both k_1 and k_2 be relatively prime. If k_2 is odd, then $\frac{(2i+1)k_2}{n+1} = k_1$ is also odd. Similar to the previous case, k_1 is odd means that L is odd.

We will now show that L is odd implies the conditions of Theorem 5. Let $d = \gcd(m + 1, n + 1)$ and set $m + 1 = k_1 d$ and $n + 1 = k_2 d$ for some k_1, k_2. Note that because L is odd, k_1 must also be odd. Also, notice that $(n + 1) \equiv j(m + 1) \bmod (2n + 2)$ can be written as $jd(k_1 - \frac{k_2}{j}) \equiv 0 \bmod (2k_2 d)$, which is true when $j = k_2$ and k_1 is odd. ■

Corollary 2 *The set $X = \{a\Diamond^n a, b\Diamond^n b, a\Diamond^m a\}$ where m, n are integers satisfying $m < n$ is unavoidable if and only if $\frac{lcm(n+1,n-m)}{n+1}$ is odd.*

Proof If we look at the set $Y = \{a\Diamond^n a, b\Diamond^n b, a\Diamond^m a\Diamond^{n-m-1} b\}$, we see that a valid Prefix-Suffix operation followed by Hole Truncations can reduce Y to X, indicating that the two sets have the same avoidability. However, we could also perform a Suffix-Prefix operation and Hole Truncations on Y to reduce it to $Z = \{a\Diamond^n a, b\Diamond^n b, a\Diamond^{n-m-1} b\}$, and so X and Z have the same avoidability. We know from $\frac{lcm(n+1,n-m)}{n+1}$ is odd, and since the Corollary 1 that Z is unavoidable if and only if avoidability of Z is the same as the avoidability of X, this completes our proof. ■

We are now interested in three-word unavoidable sets where all words have length n and every proper subset is avoidable. It is easy to verify that the set $\{a\Diamond^{n-2} a, b\Diamond^{n-2}b\}$ is avoidable, as the word $(a^{n-1}b^{n-1})^{\mathbb{Z}}$ avoids it. Thus, we describe a set of words, W, each of length n such that when one is added to the set $\{a\Diamond^{n-2} a, b\Diamond^{n-2}b\}$, the set becomes unavoidable and both $\{a\Diamond^{n-2}a, u\}$ and $\{b\Diamond^{n-2}b, u\}$ are avoidable for u E W. Notice that because u has length n, if it were to begin and end with the same letter, then either $a\Diamond^{n-2} a$ or $b\Diamond^{n-2}b$ would be compatible with it, and the set would be avoidable. Thus u must have the form $a\Diamond^{n-2}b$ with some number of holes filled in. We consider the word $a\Diamond^{n-2} b$ with no holes filled in trivial. Although words of the form $b\Diamond^{n-2}a$ with some (or none) of the holes filled in may also form unavoidable sets, we also consider these trivial due to symmetry. During the analysis of these sets, some unique patterns were found.

Conjecture 2 *Every word in W has exactly two, three, or four defined positions.*

Theorem 7 *Every word $u \in W$ has both a's and b's. Additionally, if we consider only those words that begin with an a, every a in u appears before any b's.*

Proof Suppose, for contradiction, that there exists a word $u \in W$ that contains only a's and holes. Then, u is compatible with $a\Diamond^{n-2} a$ so a Factoring operation can be performed, leaving the set $\{a\Diamond^{n-2}a, b\Diamond^{n-2}b\}$, which is avoidable as stated above.

For the second part, suppose, for contradiction, that there exists a word $u \in W$ such that a b occurs before an a. If that is the case, the word $(a^{n-1}b^{n-1})$Z avoids the set $\{u, a\Diamond^{n-2}a, b\Diamond^{n-2}b\}$,which contradicts the definition of W. ■

The next result is an extension of Corollary 1, and characterizes all words in W that have three defined positions.

Corollary 3 *Let x* E l*a, b*I. *The word $a\Diamond^{m}x\Diamond^{n-m-3}b$ is in W if and only if $\frac{lcm(n-1,m+1)}{n-1}$ is odd if x = b or $\frac{lcm(n-1,n-m-2)}{n-1}$ is odd if x = a.*

Proof Notice that there exists a valid Prefix-Suffix operation on the set $\{a\Diamond n\text{-}2\ a, b\Diamond n\text{-}2b, a\Diamond mx\Diamond n\ \text{-}m\text{-}3\ b\}$ that, along with Hole Truncation operations, can reduce it to $\{a\Diamond^{n-2}a, b\Diamond^{n-2}b, a\Diamond^{m}x\}$. In the case where $x = a$, the set has the form covered by Corollary 2, and when $x = b$, it has the form covered by Corollary 1. In either case, the conditions for unavoidability are identical to those stated above. ■

In the remainder of this section, we partially characterize words in W that have four defined positions.

Proposition 10 *The set $X = \{a\Diamond^{n-2}a, b\Diamond^{n-2}b, a\Diamond^{l}x\Diamond^{p}y\Diamond^{n-l-p-4}b\}$ where $x, y \in A$ and n, l, p are integers is unavoidable only if Proposition* 9 *holds for the sets*

$$Y1 = \{a\Diamond^{n-2}a, b\Diamond^{n-2}b, a\Diamond^{l}x\Diamond^{p}y\}$$
$$Y2 = \{a\Diamond^{n-2}a, b\Diamond^{n-2}b, x\Diamond^{p}y\Diamond^{n-l-p-4}b\}$$

Proof Notice that a Prefix-Suffix operation can be performed to reduce X to $Y1$ and a Suffix-Prefix operation can be performed to reduce X to $Y2$. Hence, the sets have the same avoidability. By Proposition 9, X is unavoidable only if certain conditions are met by $Y1$ and $Y2$. ■

Proposition 11 *The words of the form $aab\Diamond^{n-4}b$, $a\Diamond^{n-4}abb$, $a\Diamond^{n-4}aab$, and $abb\Diamond^{n-4}b$ are in the set W if and only if n > 4 is odd.*

Proof First, notice that by Remark 1 we only need to consider $aab\Diamond^{n-4}b$. We transform the set $\{a\Diamond^{n-2}a, b\Diamond^{n-2}b, aab\Diamond^{n-4}b\}$ using the algorithm defined in Sect. 3. By a Prefix-Suffix operation followed by Hole Truncation operations, it can be seen that this set has the same avoidability as $\{a\Diamond^{n-2}a, b\Diamond^{n-2}b, aab\}$. By Expanding the word $a$on-2 a at the positions of the last two holes and then performing a Prefix-Suffix operation, we get the set $\{a\Diamond^{n-4}aa, a\Diamond^{n-4}aba, a\Diamond^{n-4}baa, a\Diamond^{n-4}bba, b\Diamond^{n-2}b, aab\}$. We can now perform an Expansion on the word $a\Diamond^{n-4}aa$ one hole at a time, starting with the last, and then perform a Prefix-Suffix operation each time. We are then left with the set

$$\bigcup_{1\le i\le n-4} \{aa, aab, a\diamond^{n-4}aba, a\diamond^{n-4}baa, a\diamond^{n-4}bba, b\diamond^{n-2}b, a\diamond^{n-4-i}baa\}$$

Then, Factoring operations result in the set $\{aa, a\Diamond^{n-4}aba, a\Diamond^{n-4}bba, b\Diamond^{n-2}b\}$.Expansions on the word $b\Diamond^{n-2}b$ and Prefix-Suffix operations transform the set to $\{aa, a\Diamond^{n-4}aba, a\Diamond^{n-4}bba, b\Diamond^{n-3}a, bb\Diamond^{n-4}b, ba\Diamond^{n-4}bb\}$. A Prefix-Suffix operation, followed by a Hole Truncation, results in the set $\{aa, a\Diamond^{n-4}aba, a\Diamond^{n-4}bba, b\Diamond^{n-3}a, bb, ba\Diamond^{n-4}bb\}$. Furthermore, Factoring operations can result in the set $\{aa, a\Diamond^{n-4}aba, b\Diamond^{n-3}a, bb\}$.

Notice that if $a\Diamond^{n-4}aba$ or $b\Diamond^{n-3}a$ is expanded such that it has an aa or bb as a factor, it will be removed by a Factoring operation. If all Expansions have aa or bb, we will be left with the set $\{aa, bb\}$, which means the original set is avoidable since it could not be transformed to $\{\varepsilon\}$.

Suppose n is even. Then $a\Diamond^{n-4}aba$ has even length. Since we want the a's and b's to alternate, the full expansion word should have the form $(ab)^{\frac{n}{2}}$ or $(ba)^{\frac{n}{2}}$. However, notice that neither of these start and end with an a, so every full expansion of $a\Diamond^{n-4}aba$ will have a factor of aa or bb and thus, it can be eliminated from the set. Likewise, $b\Diamond^{n-3}a$ has odd length and the full expansion word should be of the form $a(ba)^{\frac{n-2}{2}}$ or $b(ab)^{\frac{n-2}{2}}$ to ensure that a's and b's alternate. However, none of these start with a b and end with an a so every full expansion of $b\Diamond^{n-3}a$ has aa or bb as a factor. This word can also be deleted by a Factoring operation, leaving the set $\{aa, bb\}$. Therefore, the set is avoidable for n even.

Suppose n is odd. Then $a\Diamond^{n-4}aba$ has odd length. Notice that the full expansion $a(ba)^{\frac{n-1}{2}}$ is compatible with this word and alternates a's and b's. Additionally, $b\Diamond^{n-3}a$ also has one full expansion that alternates letters, specifically, $(ba)^{\frac{n-1}{2}}$.Thus, Expansion and Factoring operations get us to the set $\{aa, bb, (ba)^{\frac{n-1}{2}}, a(ba)^{\frac{n-1}{2}}\}$. Prefix-Suffix operations can then be done on the two longest words several times to transform the set to $\{aa, bb, a, b\}$, which can be reduced by Prefix-Suffix operations to $\{\varepsilon\}$. Therefore, the set is unavoidable for n odd.

Thus, the set is unavoidable if and only if $n > 4$ is odd. ■

Proposition 12 *The words of the form $aa\Diamond^{n-4}ab$ and $ab\Diamond^{n-4}bb$ are in the set W if and only if $n > 4$ is odd.*

Proof We consider the set $X = \{aa\Diamond n{-}4\ ab, a\Diamond n\ {-}2\ a, b\Diamond n{-}2\ b\}$. The unavoidability of $\{ab\Diamond^{n-4}bb, a\Diamond^{n-2}a, b\Diamond^{n-2}b\}$ can then be easily seen by symmetry.

If n is odd, then we show that no two-sided infinite word w avoids X. Otherwise, one position of w must be an a, so we will arbitrarily label $w(0) = a$. To avoid $a\Diamond^{n-2}a$, $w(n-1) = b$. Notice that to avoid $aa\Diamond^{n-4}ab$, $w(1)$ or $w(n-2)$ must be a b. Suppose $w(1) = b$ (the other case is similar). Then $w(n) = a$ to avoid $b\Diamond^{n-2}b$.

Suppose $w(2) = b$. Then, $w(n+1) = a$ to avoid $b\Diamond^{n-2}b$. To avoid $a\Diamond^{n-2}a$ and $b\Diamond^{n-2}b$, $w(2n-2) = a$ and $w(2n-1) = w(2n) = b$. Thus, w has the substring $w(n)w(n+1) \cdots w(2n-2)w(2n-1)$ compatible with $aa\Diamond^{n-4}ab$, so w does not avoid $aa\Diamond^{n-4}ab$.

Suppose $w(2) = a$. Then, $w(n+1) = b$ to avoid $a\Diamond^{n-2}a$. Notice that $w(3) \neq a$ because then w would contain the substring $w(2)w(3) \cdots w(n)w(n+1)$ compatible with $aa\Diamond^{n-4}ab$. Hence, $w(3) = b$. If $w(4) = b$, then we will have the same problem noted above starting at position $n+2$. To avoid $aa\Diamond^{n-4}ab$, we must continue to alternate a's and b's (the way we started numbering the word w, every even position should be an a and every odd position a b). However, we also know that $w(n-1) = b$ so two a's or two b's will appear next to each other, which will cause the string $aa\Diamond^{n-4}ab$ to be compatible with a factor in w while trying to avoid $a\Diamond^{n-2}a$ and $b\Diamond^{n-2}b$.

If n is even, then the two-sided infinite word $w = (ab)^{\mathbb{Z}}$ avoids X. ■

5.2 The Case of a Ternary Alphabet

Here, we assume that the alphabet A is ternary with distinct letters a, b, and c. Thus, in order for a set X to be unavoidable, one element must be compatible with a factor of $a^{\mathbb{Z}}$, another with $b^{\mathbb{Z}}$, and the remaining element with $c^{\mathbb{Z}}$ since this is the only way to guarantee that $a^{\mathbb{Z}}$, $b^{\mathbb{Z}}$, and $c^{\mathbb{Z}}$ will not avoid X. Hence, an unavoidable set of size 3 has the form

$$\{a\Diamond^{m_1}a \cdots a\Diamond^{m_i}a,\ b\Diamond^{n_1}b \cdots b\Diamond^{n_j}b,\ c\Diamond^{l_1}c \cdots c\Diamond^{l_k}c\} \quad (7)$$

for some nonnegative integers $m_1,\dots,m_i$, $n_1,\dots,n_j$, and $l_1,\dots,l_k$, which will be denoted by

$$X_{m_1,\dots,m_i|n_1,\dots,n_j|l_1,\dots,l_k}$$

The goal is to find which values of $m_1,\dots,m_i$, $n_1,\dots,n_j$, and $l_1,\dots,l_k$ make the set $X_{m_1,\dots,m_i|n_1,\dots,n_j|l_1,\dots,l_k}$ unavoidable.

Notice that if i, j or k is zero, then X contains a word that is a single letter and any word w which avoids X cannot contain that letter. Thus, the problem of finding the avoidability of X reduces to determining whether or not the set containing the remaining two words is unavoidable over a binary alphabet. We consider such cases to be trivial, as they reduce to two-word sets over a binary alphabet. For the rest of this section we will only consider sets where i, j and k are all nonzero.

Remark 2 For $Xm1,\dots,mi\ |n1,\dots,nj\ |l1,\dots,lk$ to be avoidable, it is sufficient that one of the substring sets

$$\{a\Diamond^{m_1}a \cdots a\Diamond^{m_{i'}}a\ ,\ b\Diamond^{n_1}b \cdots b\Diamond^{n_{j'}}b,\ c\Diamond^{l_1}c \cdots c\Diamond^{l_{k'}}c\}$$

for some $i' \le i$, $j' \le j$ and $k' \le k$ be avoidable.

Note that because every two-sided infinite word must have a factor compatible with an element of an unavoidable set, we can gather restrictions on unavoidable sets by examining specific two-sided infinite words. For example, in order for a set of the form $X = \{a\Diamond^m a, b\Diamond^n b, c\Diamond^l c\}$ to be unavoidable, the two-sided infinite word $w = (aabc)^{\mathbb{Z}}$ must contain a factor compatible with one of the elements of X. Along with all the possible reletterings of w, we get the condition that either any of l, m and n is congruent to 3 mod 4 or all of l, m and n are congruent to either 0 mod 4 or 2 mod 4. In general, for a two-sided infinite word with period p, we get the condition that either any of l, m and n is congruent to $(p-1)$ mod p or another condition C involving the congruence modulo p of some of l, m and n. If C includes a condition requiring any of l, m and n to be congruent to i mod p, we say *C includes i*. For example, if there are two a's 3 positions apart in a period of length p, then the corresponding C will involve m being congruent to 3 mod p, and so C includes 3. Any unavoidable set must either satisfy C or have at least one of l, m and n be congruent to $(p-1)$ mod p, otherwise the two-sided infinite word that generated C will avoid the set.

Lemma 3 *There exists a two-sided infinite word w over the ternary alphabet $\{a, b, c\}$ with period p that avoids the set $Y = \{a\Diamond^i a, b\Diamond^i b, c\Diamond^i c\}$ for every positive integer i such that $i \le$ L $\lfloor\frac{p-2}{2}\rfloor$.*

Proof Consider the word $v = a^{i+1} (b^{i+1}c^{i+1})^{\mathbb{Z}}$. Let vp be the prefix of length p of v, and let w^1 be $v^{\mathbb{Z}}$. Note that w_1 avoids Y when $i < \lfloor\frac{p-2}{2}\rfloor$ because no two occurrences of the same letter are i positions apart. Let $w_2 = (a^{i+1}b^i\, c^q)^{\mathbb{Z}}$, where $q = 1$ if p is even, and 2 otherwise. If $i = \lfloor\frac{p-2}{2}\rfloor$, then w_2 avoids Y, completing the proof. ■

Theorem 8 *For any unavoidable set of the form* $X = \{a\Diamond^m a, b\Diamond^n b, c\Diamond^l c\}$ *where m, n and l are integers, for every* $p \geq 8$ *one of l, m or n must be congruent to* $(p - 1) \bmod p$.

Proof Let w_i, $i \leq \lfloor\frac{p-2}{2}\rfloor$, denote a two-sided infinite word with period p that avoids the set $Y = \{a\Diamond^i a, b\Diamond^i b, c\Diamond^i c\}$, which we know exists from Lemma 3. From w_i we get the restriction on X that either any of l, m and n must be congruent to $(p - 1) \bmod p$ or some other condition C_i on the congruence modulo p of l, m and n. Since w_i avoids Y, no two occurrences of the same letter in w_i are i or $p - i - 2$ positions apart, so C_i does not include i or $p - i - 2$. Notice that all of the restrictions generated by the w_i's can be satisfied if any of l, m and n is congruent to $(p - 1) \bmod p$.

For contradiction, assume that the restrictions generated by all w_i's can be satisfied without any of l, m and n being congruent to $(p - 1) \bmod p$. This requires that all C_i's be simultaneously satisfiable. Notice that for any $i \neq j$, the only way for both C_i and C_j to be satisfied is if both include some k distinct from both i and j, or at least one of l, m and n be congruent to $i \bmod p$ to satisfy C_j and at least one be congruent to $j \bmod p$ to satisfy C_i. When $p \geq 8$, there are at least four distinct C_i's and because for all $j \leq p - 2$, some C_i does not include j, the only way all of the C_i's can be simultaneously satisfied is that for every i at least one of l, m and n be congruent to $i \bmod p$. However, because there are at least four distinct values of i and only three variables, there must be at least one C_i that cannot be satisfied at the same time as the others, resulting in a contradiction. Thus, the only way for X to be unavoidable is that one of l, m or n be congruent to $(p - 1) \bmod p$. ■

Corollary 4 *There exist no nontrivial three-word sets over a ternary alphabet that are unavoidable.*

Proof Let $X = \{a\Diamond^m a, b\Diamond^n b, c\Diamond^l c\}$ where m, n and l are integers be unavoidable. From Theorem 8, we know that for every $p \geq 8$ one of l, m or n must be congruent to $(p - 1) \bmod p$. Let $q = \max(l, m, n) + 8$. Clearly none of l, m, n are congruent to $(q - 1) \bmod q$, which is a contradiction. ■

We end this section by extending Corollary 4 to unavoidable sets of k words over an alphabet of cardinality k.

Corollary 5 *There exist no nontrivial k-word sets over a k-letter alphabet where* $k \geq\geq 3$ *that are unavoidable*.

Proof The proof is done by induction on k. The base case, $k = 3$, is true by Corollary 4. For the inductive step, the $(k + 1)$-word set must have the form

$$\{a_1\Diamond^{m_{1,1}}a_1\cdots a_1\Diamond^{m_{1,i_1}}a_1,\ldots,a_{k+1}\Diamond^{m_{k+1,1}}a_{k+1}\cdots a_{k+1}\,\Diamond^{m_{k+1,i_{k+1}}}a_{k+1}\} \qquad (8)$$

where $a_1,\ldots,a_{k+1}$ are the distinct elements of the alphabet. According to the inductive hypothesis, the subset

$$\{a_1\Diamond^{m_{1,1}}a_1\cdots a_1\Diamond^{m_{1,i_1}}a_1,\ldots,a_k\Diamond^{m_{1,k}}a_k\cdots a_k\Diamond^{m_{k,i_k}}a_k\} \qquad (9)$$

is avoidable over the alphabet $\{a_1,\ldots,a_k\}$. The word that avoids (9) will also avoid (8), as the word does not contain the letter a_{k+1}. ■

6 Conclusion

Conjecture 1, although tested in numerous cases via computer, and verified for $m = 1$ and a large number of even values of m, still remains to be proven. As it has been shown in Sect. 4.1, an affirmative answer to this question would imply that $X_{m_1,\ldots,m_k|n_1,\ldots,n_l}$ is avoidable for all $k \geq 2$ and $l \geq 2$ and for all $k \geq 1$ and $l \geq 3$, and would settle the classification of *all* unavoidable two-word sets.

In Theorems 5 and 6, we have completely characterized all three-word unavoidable sets over a binary alphabet where each word has at most two defined positions. We have discussed some special cases concerning three-word unavoidable sets over a binary alphabet where one word has more than two defined positions, but general criteria for these sets have not been found.

We have also completely characterized all k-word unavoidable sets over a k-letter alphabet for $k \geq 3$. Indeed, in Corollary 5 we have proved that there are no nontrivial such sets, but k-word sets where k is larger than the alphabet size have not been considered for $k > 3$.

References

1. Aho, A.V., Corasick, M.J.: Efficient string machines, an aid to bibliographic research. Commun. ACM **18**,333–340(1975)
2. Berstel, J., Boasson, L.: Partial words and a theorem of Fine and Wilf. Theor. Comput. Sci. **218**, 135–141(1999)
3. Blanchet-Sadri, F.: Algorithmic Combinatorics on Partial Words. Chapman & Hall/CRC Press, Boca Raton (2007)
4. Blanchet-Sadri, F., Brownstein, N.C., Palumbo, J.: Two element unavoidable sets of partial words. In: Harju, T., Karhumäki, J., Lepistö, A. (eds.) DLT 2007, 1 1th International Conference on Developments in Language Theory, July 3–6, 2007, Turku, Finland. Lecture Notes in Computer Science, vol. 4588, pp. 96–107. Springer, Berlin (2007). www.uncg.edu/mat/research/unavoidablesets
5. Blanchet-Sadri, F., Jungers, R., Palumbo, J.: Testing avoidability of sets of partial words is hard. Preprint
6. Blondel, V.D., Jungers, R., Protasov, V.: On the complexity of computing the capacity of codes that avoid forbidden difference patterns. IEEE Trans. Inf. Theory **52**, 5122–5127 (2006)
7. Cassaigne, J.: Unavoidable binary patterns. Acta Inform. **30**, 385–395 (1993)
8. Cassaigne, J.: An algorithm to test if a given circular HDOL-language avoids a pattern. In: IFIP World Computer Congress 1994. Technology and Foundations, vol. 1, pp. 459–464. Elsevier, Amsterdam (1994)
9. Choffrut, C., Culik II, K.: On extendibility of unavoidable sets. Discrete Appl. Math. **9**, 125–137 (1984)
10. Choffrut, C., Karhumäki, J.: Combinatorics of words. In: Rozenberg, G., Salomaa, A. (eds.) Handbook of Formal Languages, vol. 1, pp. 329–438. Springer, Berlin (1997). Chap. 6
11. Ehrenfeucht, A., Haussler, D., Rozenberg, G.: On regularity of context-free languages. Theor. Comput. Sci. **27**, 311–322 (1983)
12. Leupold, P.: Partial Words for DNA Coding. Lecture Notes in Computer Science, vol. 3384, pp. 224–234. Springer, Berlin (2005)
13. Lothaire, M.: Algebraic Combinatorics on Words. Cambridge University Press, Cambridge (2002)
14. Rosaz, L.: Unavoidable languages, cuts and innocent sets of words. RAIRO Theor. Inform. Appl. **29**, 339–382(1995)
15. Rosaz, L.: Inventories of unavoidable languages and the word-extension conjecture. Theor. Comput. Sci. **201**, 151–170 (1998)
16. Roth, P.: Every binary pattern of length six is avoidable on the two-letter alphabet. Acta Inform. **29**, 95–106(1992)