INFORMATIQUE THÉORIQUE ET APPLICATIONS

PHAN TRUNG HUY
NGUYEN HUONG LAM

Unavoidable set : extension and reduction

Informatique théorique et applications, tome 33, nº 3 (1999), p. 213-225

<http://www.numdam.org/item?id=ITA_1999__33_3_213_0>

Theoretical Informatics and Applications
Theoret. Informatics Appl. **33** (1999) 213–225

UNAVOIDABLE SET: EXTENSION AND REDUCTION

PHAN TRUNG HUY[1] AND NGUYEN HUONG LAM[2]

Abstract. We give an explicit criterion for unavoidability of word sets. We characterize extendible, finitely and infinitely as well, elements in them. We furnish a reasonable upper bound and an exponential lower bound on the maximum leghth of words in a reduced unavoidable set of a given cardinality.

AMS Subject Classification. 68R15, 68S05.

1. INTRODUCTION

A subset of words, or a *language*, of the monoid A^* on a finite alphabet A is *unavoidable* if all but finitely many words have a factor (or subword) in it. Unavoidable languages, having an interesting history of recent origin dated from the decade 1960s in the work of Schützenberger [1], explicitly emerged in a joint paper of Ehrenfeucht, Haussler and Rozenberg in the mid 1980s [3]. They are futher the sole subject of intensive treatment by Choffrut anf Culik [4]. The reader should find a concise and pleasant overview on the subject in Rosaz [5].

It is eligible and convenient to deal with unavoidable sets from the point of irredundancy: that is when we cannot discard any element of the set without affecting the unavoidability status, namely, an unavoidable set is said X to be *minimal* provided for every word x of it $X - \{x\}$ is no longer unavoidable. Minimal unavoidable sets have been studied in the framework of extension and reduction in the following sense. If in each word of the set we select a subword then the collection of these subwords obviuosly forms an unavoidble set; that is to say, passing from set of words to set of subwords, *reduction* [5], presreves unavoidability. A reduction does not increase cardinality of the original set and happens not to

Keywords and phrases: Unavoidable set, reduced unavoidable set, extension, reduction.

[1] Department of Applied Mathematics, Hanoi Institute of Technology, 1 Dai Co Viet, Hanoi, Vietnam.
[2] Hanoi Institute of Mathematics, P.O. Box 631, Bo Ho, 10 000 Hanoi, Vietnam;
e-mail: nhlam@thevinh.ac.vn

retain minimality; but we are interested in those reductions that preserve both cardinality and minimality. An unavoidable set that does not admit such reduction is in some sense minimal in the hyerarchy "word – subword", will be called *reduced.* Although the number of minimal unavoidable set of a given cardinality is infinite, Choffrut has conjectured that the number of reduced ones is finite. In Rosaz [5] the answer is claimed to be affirmative and it is asked about a bound on the length of longest words in a reduced unavoidable set.

The opposite to reduction is extension. Let X be a minimal unavoidable set and let $wa \in X$ for a word w and a letter a and consider the particular reduction taking the subword w of wa and leaving the other words unchanged which leads to the unavoidable set $X' = X - \{wa\} + \{w\}$. Are all unavoidbable sets obtained by this way? That is, given an unavoidable set X', does there exist a word $w \in X'$ and a letter $a \in A$ such that $X = X' - \{w\} + \{wa\}$ is unavoidable? If yes, the word w is *extendible* by the letter a. The affirmative claim for *every* X' is called the Word Extension Conjecture [5]. This conjecture is actually equivalent to the conjecture about infinite extension: every unavoidable set has an infinitely extendible element, *i.e.* elemement w for which there exists an infinite sequence of letters $a_1, a_2, ..., a_n, ...$ such that $X - \{w\} + \{wa_1a_2...a_n\}$ is unavoidable for every $n = 1, 2, ...$ But Rosaz has shown in [5] that the word extension conjecture generally fails by exposing a counterexample. Consequently, there exists an unavoidable set having no infinitely extendible elements at all.

In this paper, first we give characterizations of unavoidable languages, extendible elements in general and infinitely extendible ones in particular (in Sect. 3). In Section 4, we apply these results in an analysis yielding an upper bound which is about $O(n^n)$, on the word lengths of reduced unavoidable sets of cardinality n. We conclude, in Section 5, with saying some comments on how much plausible the given estimate is by providing an exponential lower bound (to the base 3). Now the subsequent section is devoted to the basic notations and preliminary results used in the text.

2. Preliminaries

We specify a minimum amount of customary notations used in this article. Hereafter for two sets S, T we use $S - T$ and $S + T$ to denote their difference and union respectively. Throughout A is a fixed finite alphabet of at least two symbols (letters). We denote by A^* the free monoid of words on A, by ϵ the unit (empty word) of A^* and by A^+ the set of non-empty words: $A^+ = A^* - \{\epsilon\}$. For any word w we denote $|w|$ its length and for any set S we denote $|S|$ its cardinality.

Let $w = a_1a_2...a_n$ be a word written with letters $a_1, a_2, ...a_n \in A$. We say that a word u *occurs* in, or is a *factor*, or a *subword*, of w whenever there are integers i, k such that $1 \leq i \leq i + k - 1 \leq n$ and $u = a_ia_{i+1}...a_{i+k-1}$; the pair (i, k) is an *occurence* of u. As a matter of fact $k = |u|$. Most of times, we identify an occurence with the interval $a_ia_{i+1}...a_{i+k-1}$ bearing the mark i in mind. An occurence is *prefix* if $i = 1$; *suffix* if $i+k-1 = n$; *internal* if $1 < i$ and $i+k-1 < n$.

A factor is accordingly prefix, suffix or internal if it has a prefix, a suffix or an internal occurence. A factor is *proper* if it is not w itself.

Now we present specific notions to the subject. A subset of words is called *normal* if no word in it is a proper factor of another. We say that a word w avoids the set X if w has no subwords in X. We use the paper [4] as a basic reference source. We summarize some fundamental facts from there.

Proposition 2.1. *Every unavoidable set contains a finite subset which is unavoidable. In particular, minimal unavoidable sets are always finite.*

Let now A^ω be the set of left infinite words on A. Withtout being too formal, we can say that it it the set of infinite sequences of letters

$$A^\omega = \{a_0a_1a_2... : a_i \in A, i = 0,1,2,\}\cdot$$

Symmetrically, we can say of the set of right infinite words

$${}^\omega A = \{...a_{-2}a_{-1}a_0 : a_i \in A, i = 0,-1,-2,...\}\cdot$$

Extending to both directions, we have the set bi-infinite words

$${}^\omega A^\omega = \{...a_{-2}a_{-1}a_0a_1a_2... : a_i \in A, i = ...,-2,-1,0,1,2,...\}\cdot$$

Given $u \in A^+$, let ${}^\omega u, u^\omega$ and ${}^\omega u^\omega$ be successively the *periodic* infinite words $...uuu$, $uuu...$ and bi-infinite word $...uuu.....$ We state the following important result.

Proposition 2.2. *A (finite) set of words is unavoidable if and only if every (resp. periodic) left infinite word has a factor in it, or equivalently, every (resp. periodic) right infinite word has a factor in it, or equivalently, every (resp. periodic) bi-infinite word has a factor in it.*

3. Unavoidable set and extendibility

Let X be a subset of A^+ and x, y be two words of X. Define $S(x, y)$ as the set of words not including x, y, having x as a prefix, y as a suffix and no internal factors in X, precisely

$$S(x,y) = (xA^+ \cap A^+y) - A^+XA^+.$$

If there is a need to refer to X we write $S(x, y, X)$ instead.

In order to prove a criterion for unavoidability, we need an ad hoc technical notion. Let $a_ia_{i+1}...a_{i+k-1}$ and $a_ja_{j+1}...a_{j+l-1}$ be occurences in $w = a_1a_2...a_n$. We say that the two occurences *overlap*, or more precisely, $a_ia_{i+1}...a_{i+k-1}$ *overlaps* $a_ja_{j+1}...a_{j+l-1}$ if $i < j \leq i + k - 1$. Two factors overlap if they have, one by each factor, two occurences one overlapping the other.

Theorem 3.1. *A set X is unavoidable if and only if the set $S(x,y)$ is finite for all pairs x,y in X and an unavoidable set X is minimal if and only if X is normal and the set $S(x,x)$ is not empty for all x in X.*

Proof. Let X be unavoidable and $x,y \in X$. Then $S(x,y)$ must be finite otherwise it contains an infinite sequence of words $xu_1y, xu_2y, \ldots$ with the infinitely many words $u_1, u_2, \ldots$ all avoid X. Conversely, suppose that $S(x,y)$ is finite for all $x,y \in X$. Put $H = \max\{|x| : x \in X\}$, $K = \max\{|s| : s \in S(x,y), x,y \in X\}$ and $L = \max(H,K)$. We show that every sufficiently long word w of A^* has a factor in X. Take arbitrarily $x,y \in X$ and $w \in A^*$ with $|w| \geq L + 2H$ and consider the word xwy. Let $x_1w_1y_1$ be a subword of xwy such that $x_1, y_1 \in X$ and w_1 as short as possible. In fact $|w_1| \leq |w|$. We rule out the impossible outcome for w.

If, as occurences in xwy, x_1 and y_1 overlap respectively x and y then $|w_1| > |w| - |x_1| - |y_1| \geq L + 2H - 2H = L$. Let further $x_2w_2y_2$ be a subword of $x_1w_1y_1$ the shortest possible length for some $x_2, y_2 \in X$. Indeed we have $|w_2| \geq |w_1|$ because of the minimality of $|w_1|$. Since $|x_2w_2y_2| > |w_2| \geq |w_1| > L$ we conclude. Now that $x_2w_2y_2 \in XA^+ \cap A^+X$ we get $x_2w_2y_2 \notin S(x_2,y_2)$. Therefore, $x_2w_2y_2 \in A^+XA^+$ showing that $x_2w_2y_2$ has an internal factor in X which is unable to overlap x_2 and y_2 at the same time because of $L \geq H$. But it is quite a contradiction with the minimality of $|x_2w_2y_2|$.

The case when x_1 occurs as a factor of x and y_1 overlaps y (or *vice versa*) is also not possible by the minimality of w_1. So, as $|w_1| \leq |w|$, we get the remaining possibility: x_1 or y_1 is a factor of w.

In order to prove the second claim, note that by Propositions 2.1 and 2.2, an unavoidable set X is minimal iff for each $x \in X$ there exists a bi-infinite word μ avoiding $X - \{x\}$ and, hence, having (infinitely many) occurences of x as the only factors in X. Since X is normal the existence of such words is equivalent to $S(x,x) \neq \emptyset$. Infact, every subword of μ having in turn two consecutive occurences of x as prefix and suffix is clearly in $S(x,x)$. Conversely, let $s = xu = vx \in S(x,x)$. As $u, v \neq \epsilon$, consider the periodic bi-infinite word ${}^\omega vxu^\omega = {}^\omega u^\omega = {}^\omega v^\omega$, in which the subwords having an occurence of x as prefix and the next one as suffix are all the same s. If μ has a factor in $X - \{x\}$, then it neither contains an occurence of x nor is a subword of x (normality!), hence it must be an internal factor of s that contradicts $s \in S(x,x)$. Thus μ has infitely many occurences of x and has no factor in $X - \{x\}$. The proof is complete.

Now we characterize extendible elements in terms of the set $S(x,x)$. Recall that $x \in X$ is *extendible* by the word t if $X - \{x\} + \{xt\}$ is unavoidable. The following assertion is a more precise reformulation of Lemma 3.3 of Choffrut and Culik [4] and its proof requires more intensified argumentation.

Proposition 3.2. *Let X be a minimal unavoidable set. Then $x \in X$ is extendible by t if and only if t is a common prefix of all u^ω such that $xu \in S(x,x)$. More precisely, if and only if t is a common prefix of all u such that $xu \in S(x,x)$ when $|S(x,x)| \geq 2$ and t is a prefix of u^n for some positive integer n when $S(x,x)$ is a singleton $\{xu\}$.*

Proof. Suppose x is extendible by t. Take an arbitrary word $s = xu = vx \in X$ with $u, v \in A^+$ and consider the periodic bi-infinite word ${}^{\omega}vxu^{\omega}$ already handled in the proof of the previous theorem. As said therein ${}^{\omega}vxu^{\omega}$ has no factors in $X - \{x\}$ thus it must have xt as a factor since $X - \{x\} + \{xt\}$ is unavoidable. Since every occurence of x in this word is invariably followed by the infinite word u^{ω} we get that t is a prefix of every u^{ω}.

Next, a prefix of u^{ω} is indeed prefix of u^n for some possitive integer n. Finally, note that in $S(x,x)$ no word is a prefix of another and, consequently, that in the set $\{u : xu \in S(x,x)\}$ no word is a prefix of another either. Thus when $|S(x,x)| \geq 2$ a common prefix of all u^{ω} is then a common prefix of all u.

Conversely, let t be a common prefix of all u when $|S(x,x)| \geq 2$ and be a prefix of u^n when $S(x,x) = \{xu\}$. Observe that in the first case every word in $xA^+ \cap A^+x$ contains always a factor xt and in the second instance, it equals xu^n when it contains a total of n consecutive internal occurences of x.

Now put $X' = X - \{x\} + \{xt\}$ and put $K = \max\{|s| : s \in S(y,z,X), y, z \in X\}$. For every $x', y' \in X'$ and $s \in S(x', y', X')$ we see that s contains no internal occurences from $X - \{x\} \subseteq X'$ and, if in addition to this s also avoids $\{x\}$ then $s \in S(y,z,X)$ except for $y' = xt$, in which case $s \in S(y,z,X)t$, hence $|s| < K + |t|$.

Further, by the observation, we see that s contains at most $n \geq 1$ internal occurences of x, otherwise it contains an occurence of xu^n and then it contains xt as internal occurence. Thus in this case $|s| < nK + |t|$. Summing up, we see that the word length of (x', y', X') is bounded by a constant: $S(x', y', X')$ is finite for all $x', y' \in X'$. By the previous theorem X' is unavoidable.

The preceding theorem tells us that when $|S(x,x)| > 1$ the element x cannot extend infinitely, thus the possiblity of infinite extention then falls upon the remaining case $|S(x,x)| = 1$. We show in the following proposition that this indeed characterizes the infinite extendibility. Recall that element $x \in X$ is infinitely extendible if there are infinitely many words t such that $X - \{x\} + \{xt\}$ is unavoidable, or equivalently, there exists an infinite sequence of letters $a_1, a_2, ..., a_n, ...$ satisfying $X - \{x\} + \{xa_1a_2...a_n\}$ is unavoidable for all $n \geq 1$.

Proposition 3.3. *For a minimal unavoidable set X and an element $x \in X$ the following assertions are equivalent:*

(a) x is infinitely extendible;
(b) there is only one bi-infinite word avoiding $X - \{x\}$;
(c) x is extendible to a word w having x as proper suffix, i.e. $X - \{x\} + \{w\}$ is unavoidable for some word $w \in xA^+ \cap A^+x$;
(d) $S(x,x)$ is a singletone set.

Proof. (a) implies (c). Being infinitely extendible x is in particular extendible by an arbitrarily long word t. Since X is minimal, the set $X' = X - \{x\} + \{xt\}$ is minimal unavoidable, hence normal. When long enough, t must admit a factor in X that will be x by the normality of X'. We write $t = rxs$ and $xt = xrxs$. Consequently $X - \{x\} + \{xrx\}$ is unavoidable because xrx is a subword of xt. Thus $w = xrx$ is a wanted word.

(c) implies (d). Let $w = xu = vx$ with $u, v \neq \epsilon$ and put $X' = X - \{x\} + \{w\}$. We can assume that $w \in S(x, x)$, or which amounts to the same, w has no internal occurence in X, otherwise we take the shortest prefix, not x, of w with this property instead. Since x is extendible by u, by Proposition 3.2, $xu = w$ is a prefix of every word s of $S(x, x)$ if $|S(x, x)| \geq 2$ but this is impossible, for x is not an internal factor of s. So $|S(x, x)| = 1$.

(d) implies (b). Let $S(x, x) = \{s\}$ with $s = xu = vx$ and let μ be a bi-infinite word avoiding $X - \{x\}$. As said before, each factor of μ having an occurence of x as prefix and the next one as suffix is in $S(x, x)$, hence they must be all the same s. But x occurs infinitely often in μ, thus $\mu = {}^{\omega}vxu^{\omega}$ is a unique bi-infinite word avoiding $X - \{x\}$.

(b) implies (a). Let μ be the unique by-infinite word avoiding $X - \{x\}$. Clearly μ has x as a factor by unavoidability of X. Then for any factor xt of μ, the set $X_t = X - \{x\} + \{xt\}$ is evidently unavoidable because the only word that avoids $X - \{x\}$ has a factor xt. The number of such factors is indeed infinite and we get (a). The proposition is proved.

Example 3.4. Let $A = \{a, b\}$ and $X = \{aa, bb\}$. Then $S(aa, bb) = aa(ba)^*bb$ is infinite, hence X is not unavoidable. Direct verification: $(ab)^n$ avoids X for all n.

Let $X = \{aa, ab, bb\}$. Note that every word of length 5 has an internal factor in X, hence every word of the sets S is of length 3 or 4. Direct calculation: $S(aa, aa) = \{aaa\}$, $S(aa, ab) = \{aab\}$, $S(aa, bb) = \emptyset$; $S(ab, aa) = \{abaa\}$, $S(ab, ab) = \{abab\}$, $S(ab, bb) = \{abb\}$; $S(bb, aa) = \emptyset$, $S(bb, ab) = \{bbab\}$, $S(bb, bb) = \{bbb\}$. We see that X is minimal unavoidable. Direct verification: $\{ab, bb\}$, $\{aa, bb\}$ and $\{aa, ab\}$ are all not unavoidale. Every word in X is infinitely extendible. Direct verification: $\{a^n, ab, bb\}$, $\{aa, (ab)^n, bb\}$ and $\{aa, ab, b^n\}$ are all unavoidale for $n \geq 1$.

Let $X = \{a, b^n\}$, $n \geq 1$. We have $S(a, a) = \{aa, aba, ..., ab^{n-1}a\}$, $S(a, b^n) = \{ab^n\}$, $S(b^n, b^n) = \emptyset$. If $n \geq 2$, $a, ba, ...$ and $b^{n-1}a$ have no prefix in common: the word a in X is not extendible.

Propositions 3.2 and 3.3 provide a means to decide if a given element x is extendible, finitely or infinitely, by inspecting $S(x, x)$. Although this could be efficiently done by constructing an automaton recognizing $S(x, x)$ but here we are interested in the quantitative aspect which will be used in the next section. We recast Aho-Corasick's argument into a reasoning in combinatorics on words to bound above the length of words avoiding a given unavoidable set. We should mention a result of Shützenberger and Crochemore, le Rest and Wender: in case an unavoidable set consists of words of the same length m, the longest word avoiding it has length $|A|^{m-1} + m - 2$ at the most and there exists an unavoidable set for which this bound is attained (*cf.* [4], Th. 2.1).

Lemma 3.5. *Let X be an unavoidable set consisting of N words of length at most H. Then every word avoiding X has length at most $N(H - 1)$.*

Proof. Suppose on the contrary that the word $w = a_1 a_2 ... a_n$ of length $n > N(H - 1)$ avoids X. We show that there are then infinitely many words avoiding X. Note first that X, being unavoidable, contains powers of every letter.

For every $i = 1, 2, ..., n$ we define $p(i)$ as the smallest possitive integer so that $a_{p(i)} a_{p(i)+1} .... a_i$ is a proper prefix of words of X. Such $p(i)$ always exists since w avoids X, we have $a_i \notin X$ and a_i is a proper prefix of a power of a_i that is in X. Because the number of distinct proper prefices of X does not exceed $N(H - 1)$ and $n > N(H - 1)$ there are then two indices $s < t$ with $a_{p(s)} ... a_s = a_{p(t)} ... a_t$. Now the pumping technique will bring on an infinity of desired words.

Consider the word

$$w' = a_1 ... a_s a_{s+1} ... a_t a_{s+1} ... a_t a_{t+1} ... a_n \tag{1}$$

$$= a_1 ... a_{p(t)-1} a_{p(t)} ... a_t a_{s+1} ... a_t a_{t+1} ... a_n \tag{2}$$

or, replacing $a_{p(t)} ... a_t$ with $a_{p(s)} ... a_s$

$$w' = a_1 ... a_{p(t)-1} a_{p(s)} ... a_s a_{s+1} ... a_t a_{t+1} ... a_n. \tag{3}$$

We claim that w' avoids X. If, otherwise, w' contains a factor $x \in X$ then x must not be a factor of $a_1 ... a_s a_{s+1} ... a_t$ and $a_{p(s)} ... a_s a_{s+1} ... a_t a_{t+1} ... a_n$ (both are factors of w). Therefore, x must overlap $a_{s+1} ... a_t a_{t+1} ... a_n$, in view of (1), and $a_1 ... a_{p(t)-1}$ in view of (3). This already means, in view of (2), that in w the word $a_1 ... a_t$ has a suffix that begins from ahead of $a_{p(t)}$ that is a *proper* prefix of x. But this fact is in a contradition with the minimality $p(t)$. So w' has no factor in X. Now that w' is longer than w, apply the argument over and over to obtain infinitely many words that avoid X. By contradiction this accomplishes the proof.

In the follwoing immediate corollary we bound the maximal length of words in $S(x, x)$.

Proposition 3.6. *Let X be an unavoidable set of N words of maximum length H and x an element of X. Then the words in $S(x, x)$ are of length $N(H - 1) + 2$ at the most.*

Proof. Let $s \in S(x, x)$ then, by definition, the internal factor of length $|s| - 2$ of s indeed avoids X. Therefore $|s| - 2 \leq N(H - 1)$ and $|s| \leq N(H - 1) + 2$.

4. Reduction and reduced unavoidable set

We conceive reduction as an operation picking a subword from words. We should formulate a formal definition of reduction — the notion due to Choffrut (1985) — as follows. Let X be a finite set of words, a *reduction* of X is a mapping $\theta : X \to A^*$ such that $\theta(x)$ is a subword of x for every $x \in X$. We say also that the set $\theta(X) = \{\theta(x) : x \in X\}$ is a reduction of X. Clearly $|\theta(X)| \leq |X|$ for any reduction of X, and if X is unavoidable $\theta(X)$ is also unavoidable. We focus chiefly on those reductions that preserve cardinality and minimality of unavoidable sets. We say that a reduction is *trivial* if $\theta(x) = x$ for all $x \in X$, or else, *non-trivial*; *proper* if $|\theta(X)| = |X|$.

Let now X be unavoidable set, X is said to be *reduced* if it is minimal and it admits only the trivial reduction as proper reduction that is minimal unavoidable.

Example 4.1. In a binary alphabet $A = \{a, b\}$, the two-element unavoidable set is $\{a^n, b\}$ or $\{a, b^n\}$ for $n \geq 1$. Then the only reduced unavoidable set of 2 elements is $\{a, b\}$ since $\{a^n, b\}$ is a non-trivial proper reduction of $\{a^m, b\}$ whenever $m > n$.

We have the following, simple and evident, characterization of reduced unavoidable sets [5].

Proposition 4.2. *Let X be an unavoidable set then X is reduced if and only if for each word w and letter a such that $wa \in X$ or $aw \in X$ the set $X - \{wa\} + \{w\}$, or respectively $X - \{wa\} + \{w\}$, is not minimal.*

Example 4.3. Let $A = \{a, b\}$. The set $X = \{aa, bbb, bbabb, bbaba, ababa\}$ is minimal unavoidable (direct verification or by Theorem 2.1); it is reduced since no set of the form $X - \{wa\} + \{w\}$ for $wa \in X$, $a \in A$ or $X - \{aw\} + \{w\}$ for $aw \in X$, $a \in A$ is minimal: $\{a, bbb, bbabb, bbaba, ababa\}$ (not normal); $\{aa, bb, bbabb, bbaba, ababa\}$ (not normal); $\{aa, bbb, babb, bbaba, ababa\}$ ($bbaba$ extra), $\{aa, bbb, bbab, bbaba, ababa\}$ (not normal); $\{aa, bbb, bbabb, baba, ababa\}$ (not normal), $\{aa, bbb, bbabb, bbab, ababa\}$ (not normal); $\{aa, bbb, bbabb, bbaba, baba\}$ (not normal) and $\{aa, bbb, bbabb, bbaba, abab\}$ ($bbaba$ extra).

If we order the minimal unavoidable sets of a given cardinality by the relation "being a proper reduction of" then reduced sets are minimal in this ordering. This meaning is reflected in the following

Proposition 4.4. *Every minimal unavoidable set has a proper reduction that is a reduced unavoidable set.*

Proof. We order two proper reductions θ_1, θ_2 as $\theta_1 \leq \theta_2$ iff $\theta_1(x)$ is a subword of $\theta_2(x)$ for all $x \in X$. The collection of minimal proper reductions of X is not empty (it contains the trivial one) and is indeed finite, hence it has a reduction θ_0 minimal by the above ordering. Then $\theta_0(X)$ is a reduced unavoidable set which is a proper reduction X. In fact, for any proper reduction θ' of $\theta_0(X)$, the mapping $\theta'\theta_0$ is a proper reduction of X hence $\theta'\theta_0 = \theta_0$ because of minimality (in the current order) of θ_0. Consequently, θ' is the trivial reduction of $\theta_0(X)$ showing that $\theta_0(X)$ is a reduced set which completes the proof.

In 1985 Choffrut raised a question about the quantity of reduced unavoidable sets of a given cardinality as minimal elements in the above ordering. The answer is hopefully to be finite, and it is really so [5]. Subsequently we undertake another approach by placing a bound on the longest length of words in a reduced set.

Let N be a positive integer greater ≥ 2; we denote by H_N the maximal length of words in all reduced unavoidable set of cardinaliy $\leq N$. For a few values of N, the number H_N is easily computed; Example 4.1 shows $H_2 = 1$; [5] gives $H_3 = 2$. We have the following recursive inequality.

Proposition 4.5. $H_{N+1} < 2NH_N$.

Proof. Let $X = \{w_1, w_2, ..., w_N, w_{N+1}\}$ be a reduced unavoidable set of $N+1$ elements and let w_1 be one of the longest words. We write $w_1 = w_1'a$ for a letter $a \in A$. By Propositon 4.2 the set $\{w_1', w_2, ..., w_N, w_{N+1}\}$ is not minimal, therefore it contains a proper subset which is minimal. Let for instance

$$X' = \{w_1', w_2, ..., w_k\}, \qquad k < N+1$$

be such a minimal unavoidable set. By Proposition 4.4, X' has a reduced proper reduction. Explicitly, we get a reduced unavoidable set

$$X'' = \{w_1'', w_2', ..., w_k'\}$$

of such k distinct words $\{w_1'', w_2', ..., w_k'\}$ that they correspondingly are subwords of $w_1', w_2, ..., w_k$. In particular w_1' has an occurence of w_1''. We show that w_1' may not contain more than one occurence of w_1''.

Suppose that w_1' has two occurences of w_1'' and let w'' be the subword of w_1' with an occurence having these two occurences of w_1'' as prefix and suffix respectively. Note that if an unavoidable set has a minimal proper reduction then it is minimal itself. Hence the set $\{w'', w_2, ..., w_k\}$ and $\{w_1'', w_2, ..., w_k\}$, which are unavoidable as reductions of X', are both minimal since they have X'' as a minimal proper reduction. Then Proposition 3.3 (c) and (b), by the form of w'', states that the word w_1'' in $\{w_1'', w_2, ..., w_k\}$ is infinitely extendible and there is exactly one bi-infinite word, say μ, avoiding $\{w_2, ..., w_k\}$. The set of bi-infinite words that avoid $\{w_2, ..., w_k, w_{k+1}, ..., w_{N+1}\}$ is non-empty (X is minimal) and is clearly included in the set of words avoiding $\{w_2, ..., w_k\}$ which is the singleton $\{\mu\}$, therefore it must be also $\{\mu\}$. This imlies that μ has a factor w_1, as X is unavoidable. But then $\{w_1, w_2, ..., w_k\}$ is unavoidable since μ being the unique word avoids $\{w_2, ..., w_k\}$ has a factor w_1. This is impossible by minimality of X.

So w_1' has only one occurence of the factor w_1''. Also w_1' has no factors in $\{w_2', ..., w_k'\}$ as $\{w_1', w_2', ..., w_k'\}$ is normal (minimal) by admitting X'' as reduction. Hence w_1' has exactly one factor in the reduced unavoidable set X'' of $k \leq N$ words. In view of Proposition 3.5, $|w_1'| \leq 2N(H_N - 1)$ and $|w_1| \leq 2N(H_N - 1) + 1 < 2NH_N$. As X varies among the set of reduced sets of $N+1$ elements, we get $H_{N+1} < 2NH_N$. The proof is completed.

We have immediately the estimate $H_{N+1} < 2^N N!$ and by Stirling's approximation one can figure out that the order of magnitude of H_{N+1} is $O(N^N)$. In fact, by induction we can show that $N! < \left(\frac{N}{2}\right)^N$ when $N > 5$ which is left to the reader.

Proposition 4.6. $H_{N+1} < 2^N N!$. *In particular* $H_{N+1} < N^N$ *for* $N > 5$.

This eatimate is apparently rude but is fit surely for one thing: the reduced sets of a given cardinality are finite in number.

In the last section we present some arguments on how close this estimate could be to the best possible.

5. H_N: Some examples

We feature some aspects in the behaviour of the H_N relative to N by investigating certain special families of reduced unavoidable sets and in this way we can place some lower bound on H_N. Let $A = \{a, b\}$ a binary alphabet.

For instance we see that the value $H_N - N$ may tend to infinity. Consider the following sets

$$A_n = \{aa, bbb, (ab)^n a, (ab)^n b, bbabb, ..., bba(ba)^{n-2}bb\}$$

for $n = 2, 3, ...$ Detail in checking that each A_n is unavoidable, minimal unavoidable and then reduced unavoidable is left to the reader. For them the difference $H_n - n = 2n + 1 - n - 3 = n - 2$ grows boundlessly as a linear function.

Moreover, we indicate a family in which N behaves like a linear function, while H_N grows exponentially. But first we need some auxiliary facts.

Lemma 5.1. *Let X be a minimal unavoidable set containing aa and, besides, wab for some nonempty word w. Then the set $X' = X - \{wab\} + \{wa\}$ is also minimal unavoidable.*

Proof. Clearly, X' is unavoidable. We must show that for each $x \in X'$ the set $X' - \{x\}$ is no longer unavoidable. This is valid if $x = aa$ since $w \neq \epsilon$ and $X' - \{aa\}$ contains no subwords of a^+. This is also valid if $x = wa$: the set $X' - \{wa\} = X - \{wab\}$ is avoidable by minimality of X.

Let $x \neq aa, wa$ we have then $x \in X$ and $wab, aa \in X - \{x\}$. Let μ be a bi-infinite word avoiding $X - \{x\}$. Then μ has not wa as factor otherwise it would have either factor wab or factor aa that is impossible. So μ avoids wa and, altogether, μ avoids $X' - \{x\} \subseteq X - \{x\} + \{wa\}$ that is to be proved.

Lemma 5.2. *Let X be a minimal unavoidable set containing bbb and, besides, $wbba$ for some nonempty word w. Then the set $X' = X - \{wbba\} + \{wbb\}$ is also minimal unavoidable.*

Proof. Similar to the preceding proof with bbb playing the role of aa.

Lemma 5.3. *Let X be a minimal unavoidable set containing $aa, ababa$ and, besides, $wababb$ for some word w. Then the set $X' = X - \{wababb\} + \{waba\}$ is also minimal unavoidable.*

Proof. As before we must show that $X' - \{x\}$ is not unavoidable for every $x \in X'$. Note that $aa, ababa \in X'$. First, $X' - \{waba\} = X - \{wababb\}$ is not unavoidable. Next, since X is normal aa is the unique word of a^+ in X, thus, in X', therefore $X' - \{aa\}$ is not unavoidable. Next, $ababa$ is the unique factor of ${}^{\omega}(ab)^{\omega}$ in X. Otherwise suppose that X posseses another factor ub of ${}^{\omega}(ab)^{\omega}$ then $u \neq ababa$ by normality of X. Lemma 5.1 shows that $X_1 = X - \{ub\} + \{u\}$ is also a minimal unavoidable set with $u, ababa \in X_1$. Repeated application of Lemma 5.1 gives us a minimal unavoidable set containing $u, ababa$ with u a factor of ${}^{\omega}(ab)^{\omega}$, $u \neq ababa$, beginning and ending with the letter a. But this contradicts the normality (minimality). So $ababa$ is the unique factor of ${}^{\omega}(ab)^{\omega}$ in X, hence in X'. Then $X' - \{ababa\}$ is not unavoidable, as ${}^{\omega}(ab)^{\omega}$ avoids it.

Now let $x \in X'$ and $x \neq aa, ababa, waba$. Then $x \in X$ and $aa, ababa$, $wababb \in X - \{x\}$. Every bi-infinite word μ that avoids $X - \{x\}$ has no factor $waba$. In fact, otherwise, as in the proof of Lemma 5.1, μ has a factor $wabab$ and further, a factor $wababb$: a contradiction. This implies that every word avoiding $X - \{x\}$ avoids $X - \{x\} + \{waba\}$, and *a fortiori*, avoids $X' - \{x\}$ which completes the proof.

Define the substitution $\eta : A \to A^*$ by

$$\eta(a) = a, \quad \eta(b) = aba$$

and the mapping $\lambda : A^* \to A^*$ for $x_1, ..., x_t \in A, t \geq 0$:

$$\begin{aligned}
\lambda(a) &= bb\eta(a)bb \\
\lambda(b) &= bb\eta(b) \\
\lambda(ax_1...x_ta) &= bb\eta(a)bb\eta(x_1)...\eta(x_t)bb\eta(a)bb \\
\lambda(bx_1...x_tb) &= \eta(b)bb\eta(x_1)...\eta(x_t)bb\eta(b) \\
\lambda(ax_1...x_tb) &= bb\eta(a)bb\eta(x_1)...\eta(x_t)bb\eta(b) \\
\lambda(bx_1...x_ta) &= \eta(b)bb\eta(x_1)...\eta(x_t)bb\eta(a)bb
\end{aligned}$$

and for any bi-infinite word $\mu = ...x_{-2}x_{-1}x_0x_1x_2...$

$$\lambda(\mu) = ...bb\eta(x_{-2})bb\eta(x_{-1})bb\eta(x_0)bb\eta(x_1)bb\eta(x_2)bb...$$

Observe that $\lambda({}^{\omega}A^{\omega})$ is the set of bi-infinte words with infinitely many occurences of bb and any consecutive two of them are separated by a or aba. It follows that $\lambda({}^{\omega}A^{\omega})$ is exactly the set of bi-infinite words avoiding aa, bbb and $ababa$. Note also that λ is injective. Now for a subset $X \subseteq A^*$ consider the set $\{aa, bbb, ababa\} \cup \lambda(X)$.

Proposition 5.4. *Let X be a subset of A^*. Then X is unavoidable, minimal unavoidable or reduced unavoidable if and only if $\{aa, bbb, ababa\} \cup \lambda(X)$ is unavoidable, minimal unavoidable or reduced unavoidable correspondingly.*

Proof. The set $\{aa, bbb, ababa\} \cup \lambda(X)$ is unavoidable iff every bi-infinite word avoiding $\{aa, bbb, ababa\}$, that is every word of $\lambda({}^{\omega}A^{\omega})$, has a factor in $\lambda(X)$. For

$x_1...x_t$ and $y_1...y_s \in A^*$, the word $\eta(x_1)bb...bb\eta(x_t)$ is a factor of $bb\eta(y_1)bb...bb\eta(y_s)bb$ if and only if $x_1...x_t$ is a factor of $y_1...y_s$. It follows that for a bi-infinite word μ of $\lambda({}^\omega A^\omega)$ and a finite word w, $\lambda(w)$ is a factor of $\lambda(\mu)$ if and only if w is a factor of μ. Consequently, $\lambda(X)$ is unavoidable iff X is unavoidable.

For the next claim, if X is not minimal unavoidable then $\{aa, bbb, ababa\} \cup \lambda(X)$ is also not minimal unavoidable in view of the preceding claim and injectivity of λ. Conversely, let $\{aa, bbb, ababa\} \cup \lambda(X)$ not be minimal unavoidable then it contains a proper subset that is still unavoidable. This subset must not be deprived of aa, bbb or $ababa$, otherwise ${}^\omega a^\omega$, ${}^\omega b^\omega$ or ${}^\omega(ab)^\omega$ would respectively avoid it as they all avoid $\lambda(X)$. Hence it has the form $\{aa, bbb, ababa\} \cup \lambda(X')$ for some proper subset X' of X. Again by the first claim X' is unavoidable and X is not minimal.

Lastly, suppose that $Y = \{aa, bbb, ababa\} \cup \lambda(X)$ is minimal but not reduced. There exists then a word wx in Y with $w \in A^*$, $x \in A$ such that $Y - \{wx\} + \{w\}$ is minimal unavoidable. We see that $wx \neq aa$ as a occurs in $ababa$ (normality!); also $wx \neq bbb$, since X is not empty and bb is always a subword of $\lambda(X)$. If $wx = ababa$ the we have $\{aa, bbb, abab\} \cup \lambda(X)$ is minimal and as a consequence X is minimal which imply $X = \{a^n, b\}$ for $n \geq 1$, since if a word u has an occurence of b which is not the last letter then $\lambda(u)$ has $abab$ as a proper factor. If $m = 1$, $\{aa, bbb, abab\} \cup \lambda(X) = \{aa, bbb, abab\} \cup \{bbabb, bbaba\}$ is not minimal (the word $bbaba$ is extra). Hence we get $n > 1$ and $X = \{a^n, b\}$ is not reduced (Ex. 4.2). Now if $wx = \lambda(x_1...x_t) \in \lambda(X)$. two cases are possible regarding the last letter x_t.

(a) $x_t = a$ then $wx = ubbabb$ for $u \in A^*$. Hence $w = ubbab$ and $Y - \{ubbabb\} + \{ubbab\}$ is minimal. Applying Lemmas 5.1 and 5.2 we see that $Y - \{ubbabb\} + \{ubb\}$ is also minimal. If $x_{t-1} = a$ we have $ubb = \lambda(x_1...x_{t-1})$, therefore $\{aa, bbb, ababa\} \cup (\lambda(X) - \{\lambda(x_1...x_t)\} + \{\lambda(x_1...x_{t-1})\}) = \{aa, bbb, ababa\} \cup \lambda(X - \{x_1...x_t\} + \{x_1...x_{t-1}\})$ is minimal. Consequently $X - \{x_1...x_t\} + \{x_1...x_{t-1}\}$ is minimal and X is not reduced. If now $x_{t-1} = b$ then $ubb = u'bbababb$. By Lemma 5.3, $Y - \{ubb\} + \{u\}$ is minimal. As $u = u'bbaba = \lambda(x_1...x_{t-1})$ we have $X - \{x_1...x_t\} + \{x_1...x_{t-1}\}$ is minimal and X is not reduced.

(b) $x_t = b$ then $wx = ubbaba$ for $u \in A^*$. Hence $w = ubbab$ and and $Y - \{ubbaba\} + \{ubbab\}$ is minimal. Applying Lemmas 5.1 and 5.2 we see that $Y - \{ubbaba\} + \{ubb\}$ is minimal. Futher we proceed just as in the above case. If $x_{t-1} = a$ we have $ubb = \lambda(x_1...x_{t-1})$ and if $x_{t-1} = b$ then $ubb = u'bbababb$ and $u = u'bbaba = \lambda(x_1...x_{t-1})$. Again, in both instances we equally get that $X - \{x_1...x_t\} + \{x_1...x_{t-1}\}$ is minimal which indeed proves X not to be reduced.

The reverse implication is evident. Suppose that X is not reduced: $X' = X - \{wa\} + \{w\}$ is minimal for some $wa \in X$ and $a \in A$. Then $Y' = \{aa, bbb, ababa\} \cup \lambda(X') = \{aa, bbb, ababa\} \cup (\lambda(X) - \{\lambda(wa)\} + \{\lambda(w)\})$ is minimal showing that $\{aa, bbb, ababa\} \cup \lambda(X)$ is not a reduced unavoidable set. The proposition is proved.

Using Proposition 5.4 we generate a sequence of reduced unavoidable sets X, X_1, X_2,... by $X_k = \{aa, bbb, ababa\} \cup \lambda(X_{k-1})$, $k = 1, 2, ...$ starting from an arbitrary reduced set $X = X_0$. Denote $N_k = |X_k|$ and $H_k = \max\{|x| : x \in X_k\}$, we have $N_k = N_{k-1} + 3$ by injectivity of λ and $H_k \geq 3H_{k-1}$ as $|\lambda(x)| \geq 3|x|$ for all $x \in A^*$. Summing up: $H_k \geq 3^k H_0$ and $N_k = 3k + N_0$.

This construction shows that H_N must be bounded below by at least an exponential function in N.

We thank C. Choffrut for assistance and concerns.

REFERENCES

[1] M.-P. Schützenberger, On the Synchronizing Properties of Certain Prefix Codes. *Inform. and Control* **7** (1964) 23–36.

[2] A.V. Aho and M.J. Corasick, Efficient String Machines, an Aid to Bibliographic Research. *Comm. ACM* **18** (1975) 333–340.

[3] A. Ehrenfeucht, D. Haussler and G. Rozengerg, On Regularity of Context-free Languages. *Theoret. Comput. Sci.* **27** (1983) 311–332.

[4] Ch. Choffrut and K. Culik II, On Extendibility of Unavoidable Sets. *Discrete Appl. Math.* **9** (1984) 125–137.

[5] L. Rosaz, *Inventories of Unavoidable Languages and the Word-Extension Conjecture,* to appear.

Communicated by C. Choffrut.
Received June, 1997; accepted July 15, 1999.