Acta Informatica (2008) 45:467–477
DOI 10.1007/s00236-008-0078-z

ORIGINAL ARTICLE

Regular autodense languages

Chen-Ming Fan · C. C. Huang · H. J. Shyr

Received: 20 December 2007 / Accepted: 20 September 2008 / Published online: 8 October 2008

Abstract A regular component is either autodense or anti-autodense. Characterizations of a regular component being a pure autodense language and being a pure autodense code are obtained. A relationship between intercodes and anti-autodense languages is that for an intercode L of index m, L^n is an anti-autodense language for every $n > m$.

1 Introduction

The theory of formal languages plays a considerable role in the field of computer science. Both regular languages and disjunctive languages are especially important applications. It is known in [11] that every dense languages contains a disjunctive subset. There are many researches for investigating the dense languages. For the definition and properties of dense languages, one can refer to [7,12,13,16] and [17]. We have studied algebraic properties of autodense languages and anti-autodense languages in [1]. Furthermore, we investigate characteristics of a regular component being an autodense language. The relationship between intercodes and autodense languages are studied in this paper.

This paper is organized into several sections. The first section introduces the overview of this paper. In the second section, we display some well-known definitions and properties applied in this paper. In the third section, a regular component having autodense or anti-autodense properties is shown. That is, a regular component is either autodense or

C.-M. Fan
Department of Information Management,
National Chin-Yi University of Technology, Taiping, Taichung 411, Taiwan

C. C. Huang (✉)
Department of Management Information System,
Chung Shan Medical University, Taichung 402, Taiwan
e-mail: cchuang@csmu.edu.tw

H. J. Shyr
Department of Applied Mathematics,
National Chung-Hsing University, Taichung 402, Taiwan

anti-autodense. For a regular component being autodense, a characterization concerning a regular component which is pure autodense is given. Furthermore, a characterization is studied for a regular component which is a pure autodense code. In the meanwhile, we also study some examples of autodense regular components. If a regular component is anti-autodense, then it is a code. But a regular component being a code may not imply that it is an anti-autodense language. In the final section, we investigate the relationship among bifix codes, autodense languages, and anti-autodense languages. Comma-codes, Comma-free codes, and intercodes with index greater than one are bifix codes. We construct a pure autodense comma-code in the section. Every autodense intercode being pure autodense is obtained. We also show that a skew anti-autodense language is an intercode. Furthermore, for an intercode L of index m, it is derivative that L^n is an anti-autodense language for every $n > m$.

2 Definitions and preliminaries

In this paper the alphabet X containing more than one letter is assumed. Let X^* be the free monoid generated by X. Every element of X^* is a *word* and every subset of X^* is a *language*. Let 1 denote the empty word, and $X^+ = X^* \backslash \{1\}$. A language $L \subseteq X^*$ is *dense* if for any $w \in X^*$, there exist $x, y \in X^*$ such that $xwy \in L$. That is, for every $w \in X^*$, $X^*wX^* \cap L \neq \emptyset$. A *primitive word* is a word which is not a power of any other word. Let Q be the set of all primitive words over X. Every word $u \in X^+$ can be expressed as a power of a primitive word in a unique way, that is, for any $u \in X^+$, $u = f^n$ for a unique $f \in Q$ and $n \geq 1$. In this case, f is the primitive root of u and denoted by $f = \sqrt{u}$. For a language L, let $\sqrt{L} = \{\sqrt{u} \mid u \in L\}$. A language L is a global (coglobal) language if $\sqrt{L} = Q$ ($\sqrt{L} = Q \backslash F$, where F is a finite language).

Moreover, some classes of codes in this paper are defined as follows. A language $L \subseteq X^+$ is called a *code* if $x_1x_2 \ldots x_m = y_1y_2 \ldots y_n$ and $x_i, y_j \in X$ for all $1 \leq i \leq m$, $i \leq j \leq n$ imply that $m = n$ and $x_i = y_i$ for all $1 \leq i \leq n$. For any two words $u, v \in X^*$, $u \leq_{\mathrm{p}} v$ ($u \leq_{\mathrm{s}} v$) if and only if $v = ux$ ($v = xu$) for some $x \in X^*$. $u <_{\mathrm{p}} v$ ($u <_{\mathrm{s}} v$) denotes that $u \leq_{\mathrm{p}} v$ ($u \leq_{\mathrm{s}} v$) and $u \neq v$ for $u, v \in X^*$. A language $L \subset X^+$ is an *infix code* if for all $x, y, u \in X^*$, $u, xuy \in L$ together imply $x = y = 1$. A language $L \subset X^+$ is a *prefix code* (*suffix code*) if $L \cap LX^+ = \emptyset$ ($L \cap X^+L = \emptyset$). A language is a *bifix code* if it is both a prefix code and a suffix code. It follows immediately that an infix code is a bifix code. A language $L \subset X^+$ is an *intercode of index m* if $L^{m+1} \cap X^+L^mX^+ = \emptyset$, for $m \geq 1$. The family of intercodes is an important subfamily of bifix codes [18]. A *comma-free code* is an intercode of index 1. The algebraic properties of intercodes and comma-free codes can be found in [6] or [18]. In the following, we give the definitions concerning the property of being dense.

Definition 2.1 Let $L \subseteq X^+$. A language L is autodense if for any $w \in L$, there exist $x, y \in X^+$ such that $xwy \in L$.

Definition 2.2 A language is *pure autodense* if it is autodense but not dense.

If the set $L \subseteq X^+$ is not an autodense language, then there exists $w \in L$ such that $xwy \notin L$ for all $x, y \in X^+$. Such a language is called a *non-autodense language*. Moreover, we consider a stronger version of this language type as follow.

Definition 2.3 A language L is called anti-autodense if $L \cap X^+LX^+ = \emptyset$.

Furthermore, there are some results used in the rest of this paper as follow.

Lemma 2.1 ([10]) *Let $u, v \in X^+$. Then $uv = vu$ implies that u and v are powers of a common word.*

Lemma 2.2 ([9]) *If $uv = vz$, $u, v, z \in X^*$ and $u \neq 1$, then $u = (pq)^i$, $v = (pq)^j p$, $z = (qp)^i$ for some $p, q \in X^*$, $i \geq 1$, $j \geq 0$ and $pq, qp \in Q$.*

Lemma 2.3 ([4]) *Let $x, y \in X^+$. Then $(xy)^*x = x(yx)^*$ is a code if and only if $\{x, y\}$ is a code.*

Lemma 2.4 ([15]) *Let $x, y \in X^+$. Then $xy \neq yx$ if and only if $\{x, y\}$ is a code.*

Lemma 2.5 ([19]) *Let X be an alphabet having at least two letters and let $L \subseteq X^+$ be an intercode of index n with $n \geq 1$. Then for every m, $m \geq n$, L is an intercode of index m.*

3 Regular component related to autodense and anti-autodense languages

An infinite regular language contains at least an infinite subset of the form uv^+w, where $u, w \in X^*$, $v \in X^+$. Such an infinite subset is called a *regular component*. First, we discuss the relationship between regular components and autodense or anti-autodense languages.

Proposition 3.1 *A regular component is either autodense or anti-autodense.*

Proof Suppose that $L = uv^+w$, $u, w \in X^*$, $v \in X^+$. Assume that L is not anti-autodense. Then we will show that L is autodense. From the assumption, there exist $i_0 \geq 1$ and $x, y \in X^+$ such that $uv^{i_0}w \in L$ and $xuv^{i_0}wy = uv^jw \in L$ for some $j > i_0$. From $xuv^{i_0}wy = uv^jw$, it follows that $xuv = uvv^kv_1$, $wy = v_2v^{j-k-2}w$, where $0 \leq k \leq j-2$ and $v_1, v_2 \in X^*$ with $v = v_1v_2$. First, we consider $xuv = uvv^kv_1$. By Lemma 2.2, there exist $p, q \in X^*$ with $pq \in Q$ and $k_1 \geq 0$, $j_1 \geq 1$ such that

$$x = (pq)^{j_1}, \quad uv = (pq)^{k_1}p, \quad v^kv_1 = (qp)^{j_1}. \tag{1}$$

Next, we consider $wy = v_2v^{j-k-2}w$. By Lemma 2.2 again, there exist $r, s \in X^*$ with $rs \in Q$ and $k_2 \geq 0$, $j_2 \geq 1$ such that

$$v_2v^{j-k-2} = (rs)^{j_2}, \quad w = (rs)^{k_2}r, \quad y = (sr)^{j_2}. \tag{2}$$

Since $xuv = uvv^kv_1$ for some $k \geq 0$, this in conjunction with $\lg(v) = \lg(v_1v_2) = \lg(v_2v_1)$ yields that $v_1v_2 = v_2v_1$. By Lemma 2.1, it implies that there exists $g \in Q$ such that $v_1 = g^{m_1}$, $v_2 = g^{m_2}$, where $m_1, m_2 \geq 0$. Hence $v = g^{m_1+m_2}$. This yields that $(qp)^{j_1} = v^kv_1 = g^{k(m_1+m_2)+m_1}$ from $v^kv_1 = (qp)^{j_1}$ in Eq. (1). Since $qp, g \in Q$, $g = qp$. Moreover, we have that $(rs)^{j_2} = v_2v^{j-k-2} = g^{m_2+(j-k-2)(m_1+m_2)}$ from $v_2v^{j-k-2} = (rs)^{j_2}$ in Eq. (2). Since $rs, g \in Q$, $g = rs$. Therefore $qp = g = rs$. This implies that $v = (qp)^{m_1+m_2}$ and $w = (rs)^{k_2}r = (qp)^{k_2}r$. Recall that $uv = (pq)^{k_1}p$ in Eq. (1). Now for any $i \geq 1$,

$$\begin{aligned} uv^iw &= uvv^{i-1}w \\ &= (pq)^{k_1}p(qp)^{(i-1)(m_1+m_2)}(qp)^{k_2}r \\ &= (pq)^{k_1+(i-1)(m_1+m_2)+k_2}pr. \end{aligned}$$

Let $x_1 = (pq)^{m_1+m_2},\ y_1 = (sr)^{m_1+m_2}$. Then,

$$\begin{aligned} x_1uv^iwy_1 &= (pq)^{m_1+m_2}(pq)^{k_1+(i-1)(m_1+m_2)+k_2}pr(sr)^{m_1+m_2} \\ &= (pq)^{k_1}p(qp)^{i(m_1+m_2)}(qp)^{k_2}(rs)^{m_1+m_2}r \\ &= (pq)^{k_1}p(qp)^{i(m_1+m_2)}(qp)^{k_2}(qp)^{m_1+m_2}r \\ &= (pq)^{k_1}p(qp)^{(i+1)(m_1+m_2)}(qp)^{k_2}r \\ &= uvv^{i+1}w = uv^{i+2}w \in L. \end{aligned}$$

Hence L is autodense. □

A regular component of the form f^+, where f is a primitive word is called a *tree* [3]. Any proper infinite subset of a tree is called a *half tree*. It is clear that trees and half trees are all pure autodense languages. If a regular component uv^+w is autodense, then uv^+w is a pure autodense language because no regular component can be dense. We will study a characterization for a regular component which is pure autodense in the following proposition.

Proposition 3.2 *Let $L = uv^+w,\ u, w \in X^*,\ v \in X^+$ be a regular component. Then L is pure autodense if and only if there exists $z \in X^*$ such that $Lz \subseteq f^+$ for some $f \in Q$. Moreover, L is pure autodense if and only if there exist $x \in X^*, y \in X^+$ such that $L \subseteq (xy)^+x$.*

Proof Let $L = uv^+w,\ u, w \in X^*,\ v \in X^+$ be a regular component. Assume that there exists $z \in X^*$ such that $Lz \subseteq f^+$ for some $f \in Q$. If $z = 1$, then $uv^+w \subseteq f^+$; hence $L = uv^+w$ is autodense. This clearly implies that L is pure autodense. Now assume that $z \in X^+$. Since $Lz \subseteq f^+,\ f \in Q$, it follows that $uvwz = f^{j_1}$ and $uv^2wz = f^{j_2}$, for some $j_1, j_2 \geq 1$. Note that $j_1 < j_2$. Consider $uvwz = f^{j_1}$. Then

$$uv = f^{k_1}f_1, \quad wz = f_2f^{k_2}, \tag{3}$$

where $f_1 \in X^*,\ f_2 \in X^+$ with $f = f_1f_2$ and $k_1, k_2 \geq 0$. Now we consider $uv^2wz = f^{j_2}$. Then

$$uv = f^{k_3}f_3, \quad vwz = f_4f^{k_4}, \tag{4}$$

where $f_3 \in X^*,\ f_4 \in X^+$ with $f = f_3f_4$ and $k_3, k_4 \geq 0$.

From Eqs. (3) and (4), it follows that $f^{k_1}f_1 = uv = f^{k_3}f_3$. This implies that $k_3 = k_1$ and $f_3 = f_1$. Since $f_1f_2 = f = f_3f_4$, we have that $f_4 = f_2$. Consider $wz = f_2f^{k_2}$ in Eq. (3) and $vwz = f_4f^{k_4} = f_2f^{k_4}$ in Eq. (4). Since $\lg(vwz) > \lg(wz)$, it shows that $k_4 > k_2$ and $\lg(v) = \lg(f^{k_4-k_2})$. This in conjunction with $uv = f^{k_1}f_1$ and $vwz = f_2f^{k_4}$ yields that $v = (f_2f_1)^{k_4-k_2}$. Consider $wz = f_2f^{k_2}$ in Eq. (3), there exist $f_5 \in X^*,\ f_6 \in X^*$ with $f = f_5f_6$ and $0 \leq m \leq k_2$ such that $z = f_6f^m$. There are the following cases:

(1) $m = k_2$. Then $f_2 = wf_6$. This implies that $f = f_1f_2 = f_1wf_6$. For any $i \geq 1$ such that $uv^iw \in L$, let $x = f^{k_4-k_2},\ y = (f_6f_1w)^{k_4-k_2}$. It follows that

$$\begin{aligned} xuv^iwy &= f^{k_4-k_2}uvv^{i-1}w(f_6f_1w)^{k_4-k_2} \\ &= f^{k_4-k_2}f^{k_1}f_1(f_2f_1)^{(i-1)(k_4-k_2)}(wf_6f_1)^{k_4-k_2}w \\ &= f^{k_1}f_1(f_2f_1)^{i(k_4-k_2)}(f_2f_1)^{k_4-k_2}w \\ &= uvv^{i+1}w = uv^{i+2}w. \end{aligned}$$

Hence $xuv^iwy = uv^{i+2}w \in L$.

(2) $m < k_2$. Then there is $n \geq 0$ such that $w = f_2 f^n f_5,\ z = f_6 f^m$. For any $i \geq 1$ such that $uv^i w \in L$, let $x = f^{k_4-k_2},\ y = (f_6 f_5)^{k_4-k_2}$. It follows that

$$\begin{aligned} xuv^i wy &= f^{k_4-k_2} uvv^{i-1} w (f_6 f_5)^{k_4-k_2} \\ &= f^{k_4-k_2} f^{k_1} f_1 (f_2 f_1)^{(i-1)(k_4-k_2)} f_2 f^n f_5 (f_6 f_5)^{k_4-k_2} \\ &= f^{k_1} f_1 (f_2 f_1)^{i(k_4-k_2)} (f_2 f_1)^{k_4-k_2} f_2 f^n f_5 \\ &= uvv^{i+1} w = uv^{i+2} w. \end{aligned}$$

Hence $xuv^i wy = uv^{i+2} w \in L$.

In both cases, i is chosen arbitrarily, thus L is autodense. This clearly implies that L is pure autodense.

Now, assume that L is autodense. Consider $uvw \in L$, there exist $x, y \in X^+$ such that $xuvwy = uv^i w$, for some $i \geq 2$. Then $xuv = uvv^k v_1,\ wy = v_2 v^{i-k-2} w$, where $0 \leq k \leq i-2$ and $v_1, v_2 \in X^*$ with $v = v_1 v_2$. From $xuv = uvv^k v_1$, by Lemma 2.2, there exist $p, q \in X^*$ with $pq \in Q$ and $k_1 \geq 0,\ j_1 \geq 1$ such that

$$x = (pq)^{j_1}, \quad uv = (pq)^{k_1} p, \quad v^k v_1 = (qp)^{j_1}. \tag{5}$$

From $wy = v_2 v^{i-k-2} w$, by Lemma 2.2 again, there exist $r, s \in X^*$ with $rs \in Q$ and $k_2 \geq 0,\ j_2 \geq 1$ such that

$$v_2 v^{i-k-2} = (rs)^{j_2}, \quad w = (rs)^{k_2} r, \quad y = (sr)^{j_2}. \tag{6}$$

Since $xuv = uvv^k v_1$ and $\lg(v) = \lg(v_1 v_2) = \lg(v_2 v_1)$, this yields that $v_1 v_2 = v_2 v_1$. By Lemma 2.1, there exists $g \in Q$ such that $v_1 = g^{m_1},\ v_2 = g^{m_2}$, where $m_1, m_2 \geq 0$. Then $v = g^{m_1+m_2}$. Now, consider $v^k v_1 = (qp)^{j_1}$ in Eq. (5). We have that $(qp)^{j_1} = v^k v_1 = g^{k(m_1+m_2)+m_1}$. Since $qp, g \in Q$, we have that $g = qp$. Next, consider $v_2 v^{i-k-2} = (rs)^{j_2}$ in Eq. (6). We have that $(rs)^{j_2} = v_2 v^{i-k-2} = g^{m_2+(i-k-2)(m_1+m_2)}$. Since $rs, g \in Q$, we have that $g = rs$. This yields that $qp = g = rs$. By replacing rs by qp, it follows have that $v = (qp)^{m_1+m_2}$ and $w = (rs)^{k_2} r = (qp)^{k_2} r$. Recall that $uv = (pq)^{k_1} p$ in Eq. (5). Then for any $i \geq 1$,

$$\begin{aligned} uv^i w &= uvv^{i-1} w \\ &= (pq)^{k_1} p (qp)^{(i-1)(m_1+m_2)} (qp)^{k_2} r \\ &= (pq)^{k_1+(i-1)(m_1+m_2)+k_2} pr. \end{aligned}$$

Let $z = sq$. We have that

$$\begin{aligned} uv^i wz &= (pq)^{k_1+(i-1)(m_1+m_2)+k_2} prsq \\ &= (pq)^{k_1+(i-1)(m_1+m_2)+k_2+2}. \end{aligned}$$

Since i is chosen arbitrarily, we have that $Lz \subseteq (pq)^+ = f^+$, where $f = pq \in Q$.

Moreover, $Lz \subseteq f^+$ where $z = yf^i$ with $y \leq_s f$ for some $i \geq 0$ if and only if there exists $x \in X^*$ with $f = xy$ such that $L \subseteq (xy)^+ x$. □

Furthermore, we study a characterization for a regular component which is a pure autodense code.

Proposition 3.3 *Let $L = uv^+ w,\ u, w \in X^*,\ v \in X^+$ be a regular component. Then L is a pure autodense code if and only if there exist $x, y \in X^+$ with $\sqrt{x} \neq \sqrt{y}$ such that $L \subseteq (xy)^* x$.*

Proof Let $L = uv^+w,\ u, w \in X^*,\ v \in X^+$. Then by Proposition 3.2, L is a pure autodense language if and only if there exist $x \in X^*, y \in X^+$ such that $L \subseteq (xy)^+x$. In order to show our result, we only show that $L = (xy)^+x$ is a code if and only if $\sqrt{x} \neq \sqrt{y}$. First, assume that L is not a code. Then $(xy)^*x$ is not a code. By Lemma 2.3, $\{x, y\}$ is not a code. Moreover, by Lemma 2.4, $xy = yx$. This in conjunction with Lemma 2.1 yields that x and y are powers of a common word. This implies that $\sqrt{x} = \sqrt{y}$, a contradiction. Conversely, if $\sqrt{x} \neq \sqrt{y}$, i.e., x, y are not powers of a common word, then by Lemmas 2.1 and 2.4, $\{x, y\}$ is a code. By Lemma 2.3 again, $(xy)^*x$ is a code and hence L is a code. □

The set $(xy)^*x$, where $x, y \in X^*$ is called a δ-language in [4], which is neither a prefix code nor a suffix code. If a δ-language is a code, then it is called a δ-code. We rewrite Propositions 3.2 and 3.3 as a remark:

Remark 3.1 Let $L = uv^+w,\ u, w \in X^*,\ v \in X^+$ be a regular component. Then the following two statements are true:

(1) L is a pure autodense language if and only if L is a δ-language.
(2) L is a pure autodense code if and only if L is a δ-code.

Proof Let $L = uv^+w,\ u, w \in X^*,\ v \in X^+$ be a regular component. Then the results are immediately from Propositions 3.2 and 3.3. □

A regular component uv^+w, where $u, v, w \in X^+$ is called a *real regular component* if $v \not\leq_s u,\ v \not\leq_p w$. Let $L = uv^+w$, where $u, v, w \in X^+$ be a real regular component. Then for $i \geq 1$, $z_i = uv^iw \in L$. Note that if L is an autodense language and for some $x, y \in X^+$ such that $uv^jw = z_j = xuv^iwy \in L$, where $j \geq 1$, then $j > i$.

Proposition 3.4 *Let $L = uv^+w$, where $u, w \in X^+$, $v \in Q$ be a real regular component. If L is an autodense language, then u is a proper suffix of v and w is a proper prefix of v.*

Proof Let $L = uv^+w$ be a real regular component, where $u, w \in X^+, v \in Q$. Assume that L is an autodense language. Consider $uvw \in L$. Since L is an autodense language, there exist $x, y \in X^+$ such that $xuvwy \in L$. If $xuvwy = uv^2w$, then $\lg(v) = \lg(x) + \lg(y)$. Since $x, y \in X^+$, there exist $v_1, v_2 \in X^+$ with $v = v_1v_2$ such that $xu = uv_1,\ v = v_2v_1,\ wy = v_2w$. From $v = v_2v_1$, we have that $v_1v_2 = v_2v_1$. By Lemma 2.1, $v \notin Q$, a contradiction. Assume that $xuvwy = uv^jw$ for some $j \geq 3$. We consider the following cases:

(1) $\lg(x) < \lg(v)$. Then $xu = uv_1$ and $vwy = v_2v^{j-1}w$ for some $v_1, v_2 \in X^+$ with $v = v_1v_2$. From $vwy = v_2v^{j-1}w$, this implies that $v_1v_2 = v_2v_1$. By Lemma 2.1, $v \notin Q$, a contradiction.
(2) $\lg(v^i) < \lg(x) < \lg(v^{i+1})$ for $1 \leq i < j$. Then $xu = uv^iv_1$ and $vwy = v_2v^{j-i-1}w$ for some $v_1, v_2 \in X^+$ with $v = v_1v_2$. From $vwy = v_2v^{j-1}w$, this implies that $v_1v_2 = v_2v_1$. By Lemma 2.1, $v \notin Q$, a contradiction.

Therefore $xu = uv^i$, $v^kw = wy$ with $i + k = j - 1,\ i, k \geq 1$. This in conjunction with the definition of real regular component, $v \not\leq_s u$ and $v \not\leq_p w$, yields that u is a proper suffix of v and w is a proper prefix of v. □

Proposition 3.5 *Let $L = uv^+w$ be a general infinite regular component, where $u, v, w \in X^+$. If u is a suffix of v and w is a prefix of v, then L is an autodense language.*

Proof Let $L = uv^+w$, where $u, v, w \in X^+$ with u is a suffix of v and w is a prefix of v. Then $v = g_1u,\ v = wg_2$ for some $g_1, g_2 \in X^*$. Let $x = ug_1,\ y = g_2w$. This yields that $xu = uv$, $wy = vw$. Then for any word $uv^iw \in L$, $xuv^iwy = uv^{i+2}w \in L$. This shows that L is an autodense language. □

The following two corollaries are direct consequences of the above proposition. Moreover, we also study some examples of regular components which are autodense.

Corollary 3.1 *Let $L = uv^+w$ be a real regular component with $v \in Q$. If $\{u, v\}$ is a suffix code or $\{w, v\}$ is a prefix code, then L is not an autodense language.*

Corollary 3.2 *Let $L = uv^+w$ be a regular component with $\{u, v, w\}$ a bifix code contained in Q. Then L is not an autodense language.*

Example 1 $L_1 = a(ab)^+b$ is not an autodense language.

Example 2 $L_2 = u(v)^+w = b(ab)^+a$ is an autodense language. Here $u = b,\ v = ab$, $w = a$. L_2 is in fact a half tree, that is, $L_2 = ba(ba)^+ = (ba)^+ \setminus \{ba\}$. Since a half tree is a pure autodense language, L_2 is a pure autodense language.

Example 3 $L_3 = b(a^2b^2)^+a$ is a pure autodense language. This is true, with a given word $z = (b)(a^2b^2)^n(a) \in L,\ n \geq 1$. By taking $x = ba^2b,\ y = ab^2a$, then $xzy \in L$.

Example 4 Let $L_4 = a^+,\ L_5 = b^+$. Both L_4 and L_5 are pure autodense languages. The catenation of L_4 and L_5 is $L_4L_5 = a^+b^+$ which is a pure autodense language.

Lemma 3.1 ([1]) *Every coglobal language is dense and hence is not pure autodense.*

Every regular global (or coglobal) language L contains infinitely many trees [5]. That is, L contains infinitely many f^+, where f is a primitive word. The similar results for the case of pure autodense languages are interesting. By Lemma 3.1, every pure autodense language is not coglobal. But a regular pure autodense language may contain infinitely many trees. For example: Let $ab, ba \in X^+$, where $a \neq b \in X$. Then $L = \{ab, ba\}^+$ is a pure autodense language. If L is an autodense code, then certainly L is not an infix code.

Example 5 $L = \{a^nb^n \mid n \geq 1\}$ is a pure autodense code.

Example 6 $F_1^1, F_2^1, F_1^0, F_2^0$ (see [2]) are all pure autodense codes.

Example 7 Let $X = \{a, b\}$ and D be the *Dyck language* over X, i.e., $D = \{w \in X^+ \mid w_a = w_b,\ z_b \leq z_a$ for all $z \leq_{\mathrm{p}} w\}$ where w_a is the number of the letter a in w. It is known that D is a dense language. The subset aDb is a bifix code which has been called the *prime Dyck code*. The aDb is a dense language; hence it is an autodense language.

A characterization of an anti-autodense regular component is considered in the following.

Proposition 3.6 *Let L be a regular component. If L is an anti-autodense language, then L is a code.*

Proof Let L be a regular component. Then $L = uv^+w$ for some $u, w \in X^*,\ v \in X^+$. Assume that L is anti-autodense, i.e., $L \cap X^+LX^+ = \emptyset$. Now suppose that L is not a code. Then there exist $m, n \geq 1$ and $x_1, x_2 \ldots, x_m, y_1, y_2, \ldots, y_n \in L$ such that $x_1x_2 \ldots x_m = y_1y_2 \ldots y_n$, where $x_1 \neq y_1, x_m \neq y_n$. Let $x_i = uv^{k_{1i}}w,\ y_j = uv^{k_{2j}}w, k_{1i}, k_{2j} \geq 1$, where $1 \leq i \leq m,\ 1 \leq j \leq n$. Since $x_1 \neq y_1$, it follows that $k_{11} \neq k_{21}$. We consider the following two cases:

(1) $m = 1$. This in conjunction with $x_1 \neq y_1$ yields that $n \geq 2$. Then $uv^{k_{11}}w = uv^{k_{21}}w \cdots uv^{k_{2n}}w$. If $n \geq 3$, then it is true that $L \cap X^+uv^{k_{22}}wX^+ \neq \emptyset$; hence $L \cap X^+LX^+ \neq \emptyset$, a contradiction. Thus $n = 2$. We have that $v^{k_{11}-k_{21}-k_{22}} = wu$. If $k_{11} - k_{21} - k_{22} = 0$, then $wu = 1$ and $L = v^+$. This implies that L is an autodense language, a contradiction. Hence $k_{11} - k_{21} - k_{22} \geq 1$. Since $v \in X^+$, we have that $wu \in X^+$ and then $uw \in X^+$. Thus $u(wu)w = uv^{k_{11}-k_{21}-k_{22}}w \in L$. This implies that $uw(uv^{k_{11}-k_{21}-k_{22}}w)uw = uw(u(wu)w)uw = u(wu)^3w = uv^{3(k_{11}-k_{21}-k_{22})}w \in L$. Since $uw \in X^+$ and $(uv^{k_{11}-k_{21}-k_{22}}w) \in L$, $L \cap X^+LX^+ \neq \emptyset$, a contradiction.

(2) $m \geq 2$. Then $uv^{k_{11}}wuv^{k_{12}}w \cdots uv^{k_{1m}}w = uv^{k_{21}}wuv^{k_{22}}w \cdots uv^{k_{2n}}w$. Since $x_1 \neq y_1$, $k_{11} \neq k_{21}$. Without loss of generality, let $k_{11} > k_{21}$. Thus $v^{k_{11}-k_{21}}wuv \leq_p wuv^{k_{22}}w \cdots uv^{k_{2n}}w$. This implies that there exists $z \in X^+$ such that $v^{k_{11}-k_{21}}wuv = wuvz$. By Lemma 2.2, there exist $p, q \in X^*$ with $pq \in Q$ and $k \geq 0$, $i_0 \geq 1$ such that $v^{k_{11}-k_{21}} = (pq)^{i_0}$, $wuv = (pq)^k p$, $z = (qp)^{i_0}$. From $v^{k_{11}-k_{21}} = (pq)^{i_0}$ and $pq \in Q$, $v \in X^+$, there exists $k_1 \geq 1$ such that $v = (pq)^{k_1}$ and then $\lg(pq) = \lg(qp) \leq \lg(v)$. This implies that $pq \leq_s v$. But from $wuv = (pq)^k p$ and $\lg(qp) \leq \lg(v)$, this yields that $qp \leq_s v$. Thus $pq = qp$. By Lemma 2.1, p, q are powers of a common word. This implies that $pq \notin Q$, a contradiction.

Both cases imply that L must be a code. □

The converse of the above proposition is not true, that is, a regular component being a code may not imply that it is an anti-autodense language. For example, let $L = (ab)^+a$, where $a \neq b \in X$. By Proposition 3.2, L is autodense. But since $a \neq b \in X$, by Lemma 2.3, L is a code. Therefore, we study a characterization for a regular component that is a code in following proposition.

Proposition 3.7 *A regular component is a code if and only if it is not a subset of a tree.*

Proof The sufficient condition is clear. Now, let $L = uv^+w$, $u, w \in X^*$, $v \in X^+$. Assume that L is not a subset of a tree. That is, $L \not\subseteq f^+$, for every $f \in Q$. We will show that L is a code. By Proposition 3.1, L is either anti-autodense or autodense. If L is anti-autodense, then by Proposition 3.6, L is a code. If L is autodense, then by Proposition 3.2, there exist $x, y \in X^*$ such that $L \subseteq (xy)^+x$. If L is not a code, then $(xy)^+x$ is not a code. By Lemma 2.3, $\{x, y\}$ is not a code and then $xy = yx$. Moreover, by Lemma 2.1, this implies that x, y are powers of a common word. Let $x = f^i$, $y = f^j$, where $i, j \geq 1$ and $f \in Q$. Thus $L \subseteq (xy)^+x \subseteq f^+$. That is, L is a subset of a tree, a contradiction. Therefore L is a code. □

4 Intercodes related to autodense languages and anti-autodense languages

It is known that every intercode with index greater than or equal to one is a bifix code [18]. An intercode of index 1 is a comma-free code. There exist bifix codes which are not intercodes of any index. Such a bifix code has been called a comma-code [8]. Recall that a comma-free code is an infix code and an infix code can never be an autodense language. Hence, there are no autodense comma-free codes. But pure autodense intercodes of index greater than one exist. In fact, it has been pointed out that the context-free language $L = \{a^n b^n \mid n \geq 1\}$ is an intercode of index 2 which is a pure autodense language. Moreover, we will study comma-codes with autodense property. The following example is a bifix code which is a comma-code, in the sense that it is not an intercode of any index.

Example 1 (see [14]) Let $X = \{a, b\}$ and $X^+ = \{x_1 = a \leq x_2 = b \leq x_3 = aa \leq x_4 = ab \leq \cdots\}$, where $\leq$ is the lexicographic order on X^*. For any $x_n \in X^+,\ n \geq 1$, the word x'_n is defined as

$$\begin{aligned}
x'_n &= b^n x_n b^n \quad \text{if } x_n = ay = za,\\
x'_n &= b^n x_n a^n \quad \text{if } x_n = ay = zb,\\
x'_n &= a^n x_n b^n \quad \text{if } x_n = by = za,\\
x'_n &= a^n x_n a^n \quad \text{if } x_n = by = zb,
\end{aligned}$$

where $y, z \in X^*$. Let $R = \{x'_n \mid x_n \in X^+,\ n \geq 1\}$. Then R being an autodense language is immediate. Since R is dense, it is not pure autodense.

Contiguously, we construct a subset R' of R which is a pure autodense language. Start with the first element $z_1 = bab \in R$. Put $z_1 \in R'$. With the word $z_1 = bab$ in R', there is only one word z_2 in $X^+(z_1)X^+ \cap R$. Put that z_2 in R'. Now suppose we have number n element z_n in R', then put the element z_{n+1} which is the only element in $X^+(z_n)X^+ \cap R$. By continuing the process, we got the subset R' of R. R' is an autodense language which is not dense because the word $a^2ba^2 \notin R'$ and so $X^+(a^2ba^2)X^+ \cap R' = \emptyset$. Thus R' is not dense.

Remark 4.1 Since both the bifix codes R and R' are not infix codes, they cannot be comma-free codes. It can be shown that they are also not intercodes of index greater than one, and hence are comma-codes.

In the following propositions, we investigate some combinatorics of autodense languages and anti-autodense languages which are intercodes of index greater than one. First, we consider the case of autodense languages. The language L is called a skew language if no proper prefix of a word is a suffix of any word in L and no proper suffix of a word is a prefix of any word in L. It is known in [9] that every skew language is a bifix code with $\mathrm{Pref}(L) \cap \mathrm{Suff}(L) = L$ where $\mathrm{Pref}(L) = \{u \in X^+ \mid u \leq_{\mathrm{p}} v$ for all $v \in L\}$ and $\mathrm{Suff}(L) = \{u \in X^+ \mid u \leq_{\mathrm{s}} v$ for all $v \in L\}$. A skew language may be an autodense language. For example: Let $a \neq b \in X$. Then $L = \{a^n b^n \mid n \geq 1\}$ is a skew language. Recall that L is an autodense language. We now show that if L is an autodense intercode, then L is pure autodense.

Lemma 4.1 ([18]) *No intercode is dense.*

Proposition 4.1 *Every autodense intercode is pure autodense.*

Proof Immediate from Lemma 4.1. □

We investigate the relationship between anti-autodense language and intercodes in the following proposition.

Proposition 4.2 *Let L be an anti-autodense language. If L is a skew language, then L is an intercode of index m for some $m \geq 1$.*

Proof Let L be an anti-autodense language. Then $L \cap X^+LX^+ = \emptyset$. Assume that L is a skew language. Recall that L is a bifix code. Suppose that L is not an intercode of index m for any $m \geq 1$. Then there exist $x_1, x_2, \ldots, x_{m+1}, y_1, y_2, \ldots, y_m \in L$ and $x, y \in X^+$ such that $x_1x_2 \ldots x_{m+1} = xy_1y_2 \ldots y_m y$. There are the following subcases:

(1) $\lg(x_1) = \lg(x)$. Then $x_1 = x$ and $x_2 \ldots x_{m+1} = y_1 \ldots y_m y$. If $y_j = x_{j+1}$ for every $j = 1, \ldots, m-1$, then $x_{m+1} = y_m y$. This contradicts that L is a bifix code. Hence $y_j \neq x_{j+1}$, for some $j = 1, \ldots, m-1$. Thus there exists $z \in X^+$ such that either $y_j = x_{j+1}z$ or $x_{j+1} = y_j z$. Since $x_{j+1}, y_j \in L$, $z \in X^+$, both cases imply that L is not a bifix code, a contradiction.

(2) $\lg(x) < \lg(x_1)$. There exists $z \in X^+$ such that $x_1 = xz$ and $zx_2 \ldots x_{m+1} = y_1 \ldots y_m y$. If $\lg(z) \leq \lg(y_1)$, then z is a proper suffix of x_1 and is also a prefix y_1. This in conjunction with $x_1, y_1 \in L$ yields that L is not skew, a contradiction. If $\lg(z) > \lg(y_1)$, then there exists $t \in X^+$ such that $z = y_1 t$. This implies that $x_1 = xy_1 t$. This contradicts that $L \cap X^+ L X^+ = \emptyset$.

(3) $\lg(x_1) < \lg(x)$. There exists $x' \in X^+$ such that $x = x_1 x'$ and $x_2 \ldots x_{m+1} = x' y_1 \ldots y_m y$. If $\lg(x_2) = \lg(x')$, then $x_3 \ldots x_{m+1} = y_1 \ldots y_m y$. The proof of this case is similar to case(1). If $\lg(x_2) < \lg(x')$, then there exist $x_{k1} \in X^+, x_{k2} \in X^*$ such that $x_2 \ldots x_{k-1} x_{k1} = x'$ and $x_{k2} x_{k+1} \ldots x_{m+1} = y_1 \ldots y_m y$ with $x_k = x_{k1} x_{k2}$ for some $k = 3, \ldots, m$. As $x_{k2} = 1$, we have $x_{k+1} \ldots x_{m+1} = y_1 \ldots y_m y$. The proof of this case is also similar to case(1); hence $x_{k2} \in X^+$. For $x_{k2} x_{k+1} \ldots x_{m+1} = y_1 \ldots y_m y$, the proof of this case is similar to case(2). Then we consider $\lg(x_2) > \lg(x')$. There exist $y_{k'1} \in X^*, y_{k'2} \in X^+$ such that $x_2 = x' y_1 \ldots y_{k'-1} y_{k'1}$ and $x_3 \ldots x_{m+1} = y_{k'2} y_{k'+1} \ldots y_m y$ with $y'_k = y_{k'1} y_{k'2}$ for some $k' = 1, \ldots, m-1$. As $y_{k'1} = 1$, we have $x_2 = x' y_1 \ldots y_{k'-1}$. This contradicts that L is a bifix code. Hence $y_{k'1} \in X^+$. This implies that $x_2 = x' y_1 t$ for some $t \in X^+$. This contradicts that $L \cap X^+ L X^+ = \emptyset$.

Therefore L is an intercode of index for $m \geq 1$. □

Lemma 4.2 *Let $L \subseteq X^+$ and let $m \geq 2$. Then $X^+ L^m X^+ \subseteq X^+ L^n X^+$ for all $1 \leq n < m$.*

Proof Let L be a language and let $1 \leq n < m$. Suppose that $z = xu_1 u_2 \ldots u_m y \in X^+ L^m X^+$, where $x, y \in X^+$ and $u_1, u_2, \ldots, u_m \in L$. Then $z = (x) u_1 u_2 \cdot u_n (u_{n+1} \ldots u_m y)$ and $z \in X^+ L^n X^+$. □

Proposition 4.3 *For any anti-autodense language $L \subseteq X^+$, $L \cap X^+ L^n X^+ = \emptyset$ for any $n \geq 2$.*

Proof Let L be an anti-autodense language, i.e. $L \cap X^+ L X^+ = \emptyset$. Since $X^+ L^2 X^+ \subseteq X^+ L X^+$ by Lemma 4.2, $L \cap X^+ L^2 X^+ \subseteq L \cap X^+ L X^+ = \emptyset$. By similar argument, for any $n \geq 3$, $L \cap X^+ L^n X^+ = \emptyset$. □

Proposition 4.4 *For any $m \geq 1$, if L is an intercode of index m, then L^{m+1} is an anti-autodense language.*

Proof Suppose L is an intercode of index m, $m \geq 1$. Then by definition $L^{m+1} \cap X^+ L^m X^+ = \emptyset$. By Lemma 4.2, $X^+ L^{m+1} X^+ \subset X^+ L^m X^+$ and so $L^{m+1} \cap X^+ L^{m+1} X^+ \subset L^{m+1} \cap X^+ L^m X^+ = \emptyset$. It follows immediately that $L^{m+1} \cap X^+ L^{m+1} X^+ = \emptyset$, and by definition, L^{m+1} is an anti-autodense language. □

By Lemma [19], every intercode of index m is an intercode of index $m+1$, the following is immediate:

Corollary 4.1 *For any $m \geq 1$, if L is an intercode of index m, then for all $n > m$, L^n is an anti-autodense language.*

References

1. Chen, K.-H., Fan, C.-M., Huang, C.C., Shyr, H.J.: Autodense languages and anti-autodense languages. Int. J. Comp. Math. (submitted)
2. Fan, C.-M., Shyr, H.J.: Some properties of fibonacci languages. Tamkang J. Math. **27**(2), 165–182 (1998)
3. Fan, C.-M., Shyr, H.J.: Catenation Closed Pairs and Forest Languages. Words, Semigroups, & Transductions, pp. 115–127. World Scientific, Singapore (2000)
4. Fan, C.-M., Shyr, H.J.: δ-Codes and δ-languages. J. Discrete Math. Sci. Cryptogr. **8**(3), 381–394 (2005)
5. Chang, R.K., Shyr, H.J.: Global and coglobal languages. Algebra Coll. **2**, 11–22 (1995)
6. Hsieh, C.Y., Hsu, S.C., Shyr, H.J.: Some algebraic properties of comma-free codes, vol. 697, pp. 57–65. RIMS Kenkyuroku, Japan (1989)
7. Lallement, G.: Semigroups and combinatorial applications. Wiley, New York (1979)
8. Li, Z.-Z., Shyr, H.J., Tsai, Y.S.: Annihilators of bifix codes. Int. J. Comp. Math. **83**(1), 81–99 (2006)
9. Lin, Y.Y.: Properties of words and the related homomorphisms. M.S. Thesis, Institute of Applied Mathematics, Chung-Yuan University (1995)
10. Lyndon, R.C., Schützenberger, M.P.: On the equation $a^M = b^N c^P$ in a free group. Mich. Math. J. **9**, 289–298 (1962)
11. Shyr, H.J.: Disjunctive languages on a free monoid. Inform. Cont. **34**(2), 123–129 (1977)
12. Shyr, H.J.: A characterization of dense languages. Semigroup Forum **30**, 237–240 (1984)
13. Shyr, H.J.: Characterization of right dense languages. Semigroup Forum **33**, 23–30 (1986)
14. Shyr, H.J.: Free monoids and languages, 3rd edn. Hon Min Book Company, Taichung (2001)
15. Shyr, H.J., Thierrin, G.: Codes and binary relations. Lecture Notes in Mathematics Springer, Séminare d'Algèbre, Paul Dubreil, Paris, vol. 586, pp. 180–188 (1975–1976)
16. Shyr, H.J., Tsai, Y.S.: Note on languages which are dense subsemigroups. Soochow J. Math. **11**, 117–122 (1988)
17. Shyr, H.J., Tseng, D.C.: Some properties of dense languages. Soochow J. Math. **10**, 127–131 (1984)
18. Shyr, H.J., Yu, S.S.: Intercodes and some related properties. Soochow J. Math. **16**(1), 95–107 (1990)
19. Yu, S.S.: A characterization of intercodes. Int. J. Comp. Math. **36**, 39–48 (1990)