CELLULAR NEURAL NETWORKS: 
LOCAL PATTERNS FOR GENERAL TEMPLATES

CHENG-HSIUNG HSU* 
Department of Mathematics, National Central University, 
Chung-Li 32054, Taiwan

JONQ JUANG† and SONG-SUN LIN† 
Department of Applied Mathematics, National Chiao-Tung University, 
Hsin-Chu 30050, Taiwan

WEN-WEI LIN† 
Department of Mathematics, National Tsing-Hua University, 
Hsin-Chu 30050, Taiwan

Received March 12, 1999; Revised October 20, 1999

This work investigates the mosaic local patterns for cellular neural networks with general templates. Our results demonstrate that the set of templates can be divided into many finite regions. In each region, the same family of local patterns can be generated. Conversely, our results further demonstrate that some templates can realize a family of local patterns which can be linearly separated by a hyperplane in the configuration space. This study also proposes algorithms for verifying the linear separability for a given family of local patterns and, when it is separable, for obtaining the associated template.

1. Introduction

This work investigates Cellular Neural Networks (CNN) [Chua & Yang, 1988a, 1988b] of the following form:

\[
\frac{dx_{i,j}}{dt} = -x_{i,j} + z + \sum_{|k| \leq d, |\ell| \leq d} a_{k,\ell} f(x_{i+k,j+\ell}) \quad (i, j) \in \mathbb{Z}^2 ,
\]

\[x_{i,j}(0) = x_{i,j}^0.\]  

Herein, the nonlinearity \( f \) is a piecewise-linear function of the form

\[f(x) = \frac{1}{2}(|x + 1| - |x - 1|).\]

The numbers \( a_{k,\ell}, |k| \leq d, |\ell| \leq d, \) and \( k, \ell \in \mathbb{Z} \) are arranged in a \((2d+1) \times (2d+1)\) matrix form, \( d \in \mathbb{Z}^+ \), which is called a space-invariant \( A \)-template

\[
A = [a_{k,\ell}] = \begin{bmatrix}
a_{-d,d} & \cdots & a_{d,d} \\
\vdots & \ddots & \vdots \\
a_{-d,-d} & \cdots & a_{d,-d}
\end{bmatrix}
\]

The quantities \( x_{i,j} \) denote the state of a cell \( C_{i,j} \). If \( x_{i,j} > 1 \) (resp. \( x_{i,j} < -1 \)), then its corresponding cell \( C_{i,j} \) is called positively (resp. negatively) saturated and the state is called + (resp. −) state. A situation in which \( |x_{i,j}| = 1 \) (resp. \( |x_{i,j}| < 1 \))
is called a transitional or marginal state (resp. or linear state). The output of a cell \( C_{i,j} \), defined as \( y_{i,j} = f(x_{i,j}) \), equals 1, -1 and \( x_{i,j} \) when \( x_{i,j} \) is a +, - and linear, respectively. The quantity \( z \) is called threshold or a bias term which is related to an independent voltage source in an electrical circuit.

CNN systems (1) and (2), occasionally referred to as CY-CNN, were first proposed by Chua and Yang [1988a, 1988b]. Subsequent works have largely focused on the electrical engineering field, such as in [Chua, 1998]. Lattices also play a profound role in many scientific models, such as in modeling underlying spatial structures. Notable examples can be found in chemical reactions [Erneux, 1986; Chua, 1998], and pattern recognition [Chua & Yang, 1988b].

As generally known, stationary solutions \( \overline{x} = (\overline{x}_{i,j}) \) of (1) are essential for understanding CNN systems, in which their outputs \( \overline{y} = (f(\overline{x}_{i,j})) \) are called patterns. This work largely focuses on how patterns can be formed for each template \( A \) and threshold \( z \). The related problems can be studied by following two types of stationary solutions: mosaic and transitional. A mosaic solution \( \overline{x} \) satisfies \( |\overline{x}_{i,j}| > 1 \) for all \( (i, j) \in \mathbb{Z}^2 \) and a transitional solution satisfies \( |\overline{x}_{i,j}| \geq 1 \) for all \( (i, j) \in \mathbb{Z}^2 \) and equality holds for some \( (i, j) \). Their corresponding patterns \( \overline{y} \) can thus be called a mosaic and a transitional pattern, respectively. In addition to mosaic and transitional solutions, two other types of stationary solutions exist: defect and linear. A defect solution \( \overline{x} \) satisfies \( |\overline{x}_{i,j}| > 1 \) for \( (i, j) \in \mathbb{Z}^2 \), \( D \) and \( |\overline{x}_{k,l}| < 1 \) for \( (k, l) \in D \), where \( D \neq \emptyset \) and \( D \neq \mathbb{Z}^2 \). \( \overline{x} \) is a linear solution if \( |\overline{x}_{i,j}| < 1 \) for all \( (i, j) \in \mathbb{Z}^2 \).

This work first introduces the notion of local patterns, as proposed by Juang and Lin [1979a]. For a given mosaic solution \( \overline{x} \), the state at cell \( C_{i,j} \) is +, i.e. \( \overline{x}_{i,j} > 1 \), if and only if

\[
\sum_{|k|,|\ell| \leq d, \ (k, \ell) \neq (0,0)} a_{k,\ell} \overline{y}_{i+k,j+\ell} + a + z - 1 > 0, \tag{5}
\]

where \( a = a_{0,0} \). Similarly, if the state at cell \( C_{i,j} \) is -, i.e. \( \overline{x}_{i,j} < -1 \), if and only if

\[
\sum_{|k|,|\ell| \leq d, \ (k, \ell) \neq (0,0)} a_{k,\ell} \overline{y}_{i+k,j+\ell} - a + z + 1 < 0. \tag{6}
\]

(5) and (6) can be written in a much more compact form by introducing the following notations.

Denote by \( n = (2d+1)^2 - 1 \). Then, the \( n \) indices \( \{(k, \ell) \in \mathbb{Z}^2 : |k| \leq d, |\ell| \leq d, (k, \ell) \neq (0,0)\} \) can be arranged into \( \{m \in \mathbb{Z}^1 : 1 \leq m \leq n\} \). Furthermore, denote \( \alpha = (a_1, \ldots, a_n) \) and \( V = (v_1, \ldots, v_n) \)

\[
\begin{bmatrix}
    a_{4d^2+2d} & a_{2d^2+d+1} & a_1 \\
    \vdots & \ddots & \ddots \\
    a_n & a_{2d^2+3d} & a_{2d^2+1}
\end{bmatrix}
\]

and

\[
\begin{bmatrix}
    v_{4d^2+2d} & v_{2d^2+d+1} & v_1 \\
    \vdots & \ddots & \ddots \\
    v_n & v_{2d^2+3d} & v_{2d^2+1}
\end{bmatrix}
\]

Clearly, vector \( \alpha \) represents \( A' \); the surrounding template of \( A \) without center and is denoted by \( \alpha \cong A' \), where vector \( V \) represents the surrounding outputs at cell \( C_{i,j} \). Therefore, \( V \) is called a local pattern associated with \( (A, z) \).

Denote the configuration space by

\[
X^n = \{ V \in \mathbb{R}^n : V = (v_1, \ldots, v_n) \quad \text{and} \quad |v_m| = 1 \quad \text{for all} \quad m = 1, \ldots, n \}, \tag{7}
\]

i.e. the set of all possible local patterns. Then, for a given pair of template \( A \) and threshold \( z \), the family of feasible local patterns with \( + \) state in the center is defined by

\[
\mathcal{F}(+, A, z) = \{ V \in X^n : \alpha \cdot V + a + z - 1 > 0 \}.
\]

Similarly, the set of feasible local patterns with \( - \) state in the center is defined by

\[
\mathcal{F}(-, A, z) = \{ V \in X^n : \alpha \cdot V - a + z + 1 < 0 \}.
\]
Finally, the set of feasible local patterns of \((A, z)\) is denoted by

\[ F(A, z) = (F(+, A, z), F(-, A, z)) \,.
\]

Therefore, studying the (global) mosaic patterns on \(S^2\) for a given \((A, z)\) requires an initial examination of feasible local patterns \(F(A, z)\).

However, two different pairs \((A, z)\) and \((\tilde{A}, \tilde{z})\) may generate the same feasible local patterns, i.e. \(F(A, z) = F(\tilde{A}, \tilde{z})\). Hence, the following problem is studied, which is called

**Direct problem**

Partition the parameters space

\[ P^{n+2} = \{(A, z) : A \text{ is a } (2d + 1) \times (2d + 1) \text{ real matrix in (4) and } z \in \mathbb{R}^1\}, \]

into subregions such that each subregion (1) has the same mosaic patterns.

Obviously, the Direct Problem can be solved as follows:

**Theorem 1.1.** There is a positive integer \(K = K(n)\) and a unique set of open subregions \(\{P_k\}_{k=1}^K\) of \(P^{n+2}\) satisfying

(i) \(P^{n+2} = \bigcup_{k=1}^K P_k\),

(ii) \(P_k \cap P_\ell = \emptyset\) if \(k \neq \ell\),

(iii) \((A, z)\) and \((\tilde{A}, \tilde{z})\) in \(P_k\) for some \(k\) if and only if \(F(A, \tilde{z}) = F(\tilde{A}, z)\).

Here \(P\) is the closure of \(P\) in \(P^{n+2}\).

Conversely, we may ask the following “Inverse problem” or so called “Learning problem”:

Given that \((U^+, U^-) \subset X^n \times X^n\), can we find \((A, z) \in P^{n+2}\) such that

\[(U^+, U^-) = F(A, z) \,.
\]

The Learning problem is closely related to the following theorem in the classical theory of convex set, see [Lay, 1992].

**Theorem 1.2** (Linearly Separating Theorem). Two bounded sets \(U\) and \(W\) in \(\mathbb{R}^n\) can be separated by a hyperplane in \(\mathbb{R}^n\) if and only if

\[ \operatorname{conv}(U) \cap \operatorname{conv}(W) = \emptyset, \]

where \(\operatorname{conv}(S)\) is the convex hull of set \(S\) in \(\mathbb{R}^n\).

Given \(U \subset X^n\), denote by \(U^c\) the complement of \(U\) in \(X^n\). Then, according to Theorem 1.2, \(U\) and \(U^c\) can be separated by a hyperplane if and only if

\[ \operatorname{conv}(U) \cap \operatorname{conv}(U^c) = \emptyset. \quad (8) \]

A set \(U \subset X^n\) which satisfies (8) is called linearly separable since there exists hyperplanes in \(\mathbb{R}^n\) which separate \(U\) and \(U^c\). Furthermore, let \((a, b)\) be the coefficients of such hyperplanes, then \(U = F(+, A, z)\) where \((a, z)\) satisfies \(a + z - 1 = b\).

Since the number of elements in \(X^n\) is equal to \(2^n\), the number of linearly separable sets \(U\) in \(X^n\) is expected to be quite large; the number of elements in \(U\) is generally large as well. It is highly desired to describe or represent such a linearly separable set \(U\). Indeed, for such \(U\), we may choose some appropriate hyperplanes (extreme supporting planes which are defined in Sec. 3) in \(\mathbb{R}^n\), denoted by \(H^{(i)}(U)\) for \(i = 0, 1, \ldots, n-1\), such that \(U\) can be characterized by these hyperplanes. The results are as follows:

**Theorem 1.3.** Let \(H^{(n-1)}\) be an \((n-1)\)-dimensional extreme supporting hyperplane of a linearly separable set \(U \subseteq X^n\) in \(\mathbb{R}^n\). Then, there is a sequence of supporting hyperplanes \(\{H^{(i)}\}_{i=0}^{n-1}\) such that

(i) \(H^{(i)}\) is an \(i\)-dimensional supporting hyperplane in \(\mathbb{R}^{i+1}\) which separates \(\operatorname{conv}(H^{(i+1)} \cap U)\) and \(\operatorname{conv}(H^{(i+1)} \cap U^c)\) from \(i = 0\) to \(n - 2\). Moreover,

\[ U = \bigcup_{i=0}^{n-1} (H^{(i)}, + \cap X^n), \]

and

\[ U^c = \bigcup_{i=1}^{n-1} (H^{(i)}, - \cap X^n) \bigcup H^{(0)}, - \,.
\]

(ii) Conversely, if \(\{H^{(i)}\}_{i=0}^{n-1}\) is a finite sequence of hyperplanes such that \(H^{(i)} \cap X^n\) is \(i\)-dimensional and satisfies

(a) \(H^{(j)} \cap X^n \subset H^{(j+1)} \cap X^n\), for \(j = 0, \ldots, n-2\),

(b) \(\operatorname{conv}(H^{(j)}, + \cap X^n) \cap \operatorname{conv}(H^{(j)}, - \cap X^n) = \emptyset\),

for \(j = 0, \ldots, n-1\).

Define

\[ U = \bigcup_{i=0}^{n-1} (H^{(i)}, + \cap X^n) \quad \text{and} \]

\[ W = \bigcup_{i=1}^{n-1} (H^{(i)}, - \cap X^n) \bigcup H^{(0)}, - , \]

where \(H^{(0)} = X^n\).
then
\[ \text{conv}(\mathcal{U}) \cap \text{conv}(\mathcal{W}) = \emptyset \quad \text{and} \quad \mathcal{W} = X^n - \mathcal{U}. \]

Notably, Sec. 3 provides definitions of \( H^{(i),+} \) and \( H^{(i),-} \).

Therefore, studying a linearly separable set is equivalent to studying the finite sequence of (extreme) hyperplanes that satisfy condition (ii) in Theorem 1.3. By using this theorem, we propose an algorithm to write down a complete list of all nonequivalent local patterns for all symmetric \( 3 \times 3 \) templates, see Appendix. In general, the following two-phased method of linear programming can be applied to verify a given set \( \mathcal{U} \) is linearly separable or not.

**Theorem 1.4** (Separation Algorithm). For given subsets \( \mathcal{U} = \{U_i\}_{i=1}^n \) and \( \mathcal{U}' = \{U_i\}_{i=r+1}^N \) in \( X \subset \mathbb{R}^n \), consider the following linear programming problem:

\[
\begin{aligned}
\min & \sum_{i=1}^{n+2} \beta_i, \\
\text{s.t.} & \quad M \cdot B + \tilde{B} = C, \\
& \quad B \geq 0 \quad \text{and} \quad \tilde{B} \geq 0.
\end{aligned}
\] (9)

Here
\[ B = (\beta_1, \ldots, \beta_N)^t, \quad \tilde{B} = (\beta_1, \ldots, \beta_{n+2})^t \in \mathbb{R}^n, \]
\[ C = C_{(n+2) \times 1} = (0, \ldots, 0, 1, 1)^t, \]
and \( M \) is a \((n + 2) \times N\) matrix by
\[ M = \begin{bmatrix}
U_1 & \cdots & U_t & -U_{r+1}^t & \cdots & -U_N^t \\
1 & \cdots & 1 & 0 & \cdots & 0 \\
0 & \cdots & 0 & 1 & \cdots & 1
\end{bmatrix}. \]

Let \((B^*, \tilde{B}^*)\) be an optimal solution of (9), then we have
\begin{enumerate}
\item if \( \tilde{B}^* \neq 0 \), then \( \mathcal{U} \) is a linearly separable set.
\item if \( B^* = 0 \), then \( \mathcal{U} \) is not linearly separable.
\end{enumerate}

After establishing the algorithm to test the linear separability, we propose an algorithm to obtain some \((A, z) \in \mathcal{P}^{n+2}\) such that \( \mathcal{U} = \mathcal{F}(+, A, z) \). The result is as follows.

**Theorem 1.5.** For given subsets \( \mathcal{U} = \{U_i\}_{i=1}^r \) and \( \mathcal{U}' = \{U_i\}_{i=r+1}^N \) in some configuration space, define the perception matrix \( Q = Q(\mathcal{U}) \) by
\[ Q = \begin{bmatrix}
-U_1 & -1 \\
\vdots & \vdots \\
-U_r & -1 \\
U_{r+1} & 1 \\
\vdots & \vdots \\
U_N & 1
\end{bmatrix}, \]

then \( \mathcal{U} \) is a linearly separable set and \( \mathcal{U} = \mathcal{F}(+, A, z) \) if and only if
\[ Q \begin{pmatrix} \alpha^t \\ a + z - 1 \end{pmatrix} < 0. \]

Notably, all arguments presented here can be applied to any rectangular template, not only a square template given by (4). Therefore, the theorems also hold for general systems of CNN. A recent textbook reported on similar results about separability for an uncoupled system [Chua, 1998].

The rest of this paper is organized as follows. Section 2 describes the direct problem, demonstrating that the parameters space can be divided into finitely many subregions such that in each region, (1) has the same mosaic patterns. Section 3 addresses the learning problem and characterizes a linearly separable set by a finite sequence of extreme hyperplanes. Section 4 presents the separation algorithm to test the linear separability. Finally, the Appendix lists some extreme hyperplanes for certain \( 3 \times 3 \) symmetric templates.

## 2. Partitioning of the Parameters Space

In this section, we first partition the parameters space \( \mathcal{P}^{n+2} \) into finite subregions such that each region (1) has the same mosaic pattern. This is the first step to study the more difficult problem: Can two different pairs \((A, z)\) and \((\tilde{A}, \tilde{z})\) determine “the same” dynamics of (1) and (2)?

Next, we prove Theorem 1.1.

**Proof of Theorem 1.1.** Clearly, the cardinal number \( N \) of \( X^n \) is \( 2^n \). Therefore, the cardinal number \( N^* \) of
\[ \mathcal{F} = \{ \mathcal{U} \subset X^n : \mathcal{U} \text{ satisfies (8)} \} \] (10)
is strictly less than $2^N$. For any $\mathcal{U} \subset X^n$, define

$$\mathcal{A}^+(\mathcal{U}) = \{(\alpha, b) \in \mathbb{R}^{n+1} : \alpha \cdot V + b > 0 \text{ for any } V \in \mathcal{U}\},$$

$$\mathcal{A}^-(\mathcal{U}) = \{(\alpha, c) \in \mathbb{R}^{n+1} : \alpha \cdot V' + c < 0 \text{ for any } V' \in \mathcal{U}^c\}.$$  \hspace{1cm} (11)

Then, $\mathcal{A}^+(\mathcal{U}) \cap \mathcal{A}^-(\mathcal{U}) \neq \emptyset$ if and only if $\mathcal{U}$ satisfies (8). In this case, $\mathcal{A}^+(\mathcal{U})$ is an open convex cone with vertex $(0, 0)$ in $\mathbb{R}^{n+1}$. The boundary $\partial \mathcal{A}^+(\mathcal{U})$ of $\mathcal{A}^+(\mathcal{U})$ in $\mathbb{R}^{n+1}$ consists of $(A, z)$ such that the equality holds in (5), i.e. there is a local transitional state for these $(A, z)$.

Define

$$\hat{\mathcal{F}}(+, \alpha, b) = \{V \in X^n : \alpha \cdot V + b > 0\},$$

and

$$\hat{\mathcal{F}}(-, \alpha, c) = \{V \in X^n : \alpha \cdot V + c < 0\},$$

then $\mathbb{R}^{n+1} = \{(\alpha, b) : \alpha \in \mathbb{R}^n, b \in \mathbb{R}^1\}$ is divided into $N^*$-many regions $\{\hat{\mathcal{A}}^+(\mathcal{U}) : \mathcal{U} \in \mathcal{F}\}$. For each $\mathcal{A}^+(\mathcal{U})$, $\hat{\mathcal{F}}(+, \alpha, b) = \mathcal{U}$. Similarly, $\mathbb{R}^{n+1} = \{(\alpha, c) : \alpha \in \mathbb{R}^n, c \in \mathbb{R}^1\}$ is also divided by $N^*$-many regions of $\{\hat{\mathcal{A}}^-(\mathcal{W}) : \mathcal{W} \in \mathcal{F}\}$. For each $\mathcal{A}^-(\mathcal{W})$, $\hat{\mathcal{F}}(-, \alpha, c) = \mathcal{W}$.

Therefore, for any pair $(\mathcal{U}, \mathcal{W}) \in \mathcal{F} \times \mathcal{F}$, if there are $(\alpha, b) \in \mathcal{A}^+(\mathcal{U})$ and $(\alpha, c) \in \mathcal{A}^-(\mathcal{W})$ for some $\alpha \in \mathbb{R}^n$, $b$ and $c \in \mathbb{R}^1$, define

$$a = 1 + \frac{1}{2}(b - c), \quad z = \frac{1}{2}(b + c),$$  \hspace{1cm} (12)

and

$$\mathcal{A}' \cong \alpha.$$  \hspace{1cm} (13)

Then, we have $\mathcal{F}(+, A, z) = \hat{\mathcal{F}}(+, \alpha, b) = \mathcal{U}$ and $\mathcal{F}(-, A, z) = \hat{\mathcal{F}}(-, \alpha, c) = \mathcal{W}$. The converse is also true. Hence

$$\mathbb{P}^{n+2} = \bigcup_{k=1}^{K} \hat{\mathcal{F}}_k,$$

with $K \leq (N^*)^2$. The proof is complete.  \hspace{1cm} ■

Closely examining the local patterns reveals that the numerical range of $\alpha \in \mathbb{R}^n$ acting on $X^n$ play an important role in partitioning the parameters space. Obviously, given an $\alpha \in \mathbb{R}^n$, the $z - \alpha$ plane is partitioned into finite subregions such that each subregion contains the same local patterns.

The elements of $X^n$ are first labeled by one-dimensional indices running from $i = 1$ to $i = N$, i.e. $X^n = \{V_i\}_{i=1}^N$. The following notion is then introduced.

**Definition 2.1.** Let $\alpha$ and $\tilde{\alpha}$ be in $\mathbb{R}^n$. $\alpha$ and $\tilde{\alpha}$ are called equivalent if there is a one-to-one function $\varphi$ from $X^n$ onto itself such that

$$\alpha \cdot V_i > \alpha \cdot V_j$$  \hspace{1cm} (14)

if and only if

$$\tilde{\alpha} \cdot \varphi(V_i) > \tilde{\alpha} \cdot \varphi(V_j),$$  \hspace{1cm} (15)

for any $V_i$ and $V_j$ in $X^n$. Moreover, $\alpha$ and $\tilde{\alpha}$ are called strictly equivalent when $\varphi$ is the identity map.

We will see any two equivalent $\alpha$ and $\tilde{\alpha}$ determine the same feasible local patterns under $\varphi$. Indeed, we have the following result:

**Theorem 2.2.** If $\alpha$ and $\tilde{\alpha}$ are equivalent, then for given $b$ and $c \in \mathbb{R}$, there exists $\tilde{b}$ and $\tilde{c} \in \mathbb{R}$ such that $\mathcal{F}(+, \alpha, b) = \mathcal{F}(+, \tilde{\alpha}, \tilde{b})$ and $\mathcal{F}(-, \alpha, c) = \mathcal{F}(-, \tilde{\alpha}, \tilde{c})$.

**Proof.** Denote by $b_j = -\alpha \cdot V_j$. Without loss of generality, we may assume that $b_i \leq b_{i+1}$, for $i = 1, \ldots, N - 1$. Therefore, if $b_k < b < b_{k+1}$ for some $k$, then

$$V_i \in \mathcal{F}(+, \alpha, b) \text{ if and only if } i \leq k.$$  

Similarly, if $b_k < c < b_{k+1}$ then

$$V_j \in \mathcal{F}(-, \alpha, c) \text{ if and only if } j \geq \ell + 1.$$  

Since $\tilde{\alpha}$ is equivalent to $\alpha$ by $\varphi$, if we denote by $\tilde{b}_j = -\tilde{\alpha} \cdot \varphi(V_j)$,

then (15) implies

$$\tilde{b}_i \leq \tilde{b}_{i+1}$$

for $i = 1, \ldots, N - 1$. Denote

$$\tilde{V}_i = \varphi(V_i).$$

Then the results follow. The proof is complete.  \hspace{1cm} ■

If we express $a$ and $z$ in terms of $b$ and $c$, then we have the bifurcation diagrams in $(z, a)$ plane.
which is a generalization of the previous results in [Juang & Lin, 1997a].

Indeed, for a given $\alpha \in \mathbb{R}^n$, we have two sets of parallel straight lines \( \{ \ell_j^+ \}_{j=1}^N \) and \( \{ \ell_j^- \}_{j=1}^N \) in \((z, a)\) plane, here
\[
\mathcal{L}_j^+(z, a) = a + z - 1 + \alpha \cdot V_j, \quad (16)
\]
\[
\ell_j^+ = \{ (z, a) : \mathcal{L}_j^+(z, a) = 0 \},
\]
\[
\mathcal{L}_j^-(z, a) = -a + z + 1 + \alpha \cdot V_j, \quad (17)
\]
and
\[
\ell_j^- = \{ (z, a) : \mathcal{L}_j^-(z, a) = 0 \},
\]
for \( j = 1, \ldots, N \). In general, it may happen that \( \alpha \cdot V_i = \alpha \cdot V_j \) for some \( i \neq j \), and in this case \( \ell_i^+ = \ell_j^+ \) and \( \ell_i^- = \ell_j^- \). For simplicity, we only treat the following simple cases, and the others can be treated analogously. We need the following definition.

**Definition 2.3.** \( \alpha \in \mathbb{R}^n \) is called simple if
\[
\alpha \cdot V_i \neq \alpha \cdot V_j \quad (18)
\]
for any \( i \neq j, 1 \leq i, j \leq N \).

**Example 2.4**

(i) For any \( r > 1 \), denote by \( \alpha_r = (a_1, \ldots, a_n) \) with \( a_j = r^j \), for \( j = 1, \ldots, N \). Then \( \alpha_r \) is simple and equivalent to each other for all \( r > 2 \).

(ii) Denote by \( I[k, \ell] = \{ i \in \mathbb{Z}^1 : k \leq i \leq \ell \} \), the set of all integers that are no greater then \( \ell \) and no smaller than \( k \). A function \( \tau \) defined on \( I[k, \ell] \) is called a sign function if
\[
|\tau(i)| = 1
\]
for all \( i \in I[k, \ell] \). Now, for any permutation \( \sigma \) and sign function \( \tau \) defined on \( I[1, n] \), and \( \alpha = (a_1, \ldots, a_n) \in \mathbb{R}^n \), define
\[
\alpha_{\sigma, \tau} = (\tau(1)a_{\sigma(1)}, \ldots, \tau(n)a_{\sigma(n)}) \quad (19)
\]
and
\[
\varphi_{\sigma, \tau}(V) = (\tau(1)V_{\sigma(1)}, \ldots, \tau(n)V_{\sigma(n)}).
\]
Then \( \alpha_{\sigma, \tau} \) is equivalent to \( \alpha \) by \( \varphi_{\sigma, \tau} \).

**Theorem 2.5.** Assume \( \alpha \) is simple and
\[
\alpha \cdot V_j > \alpha \cdot V_{j+1}, \quad (20)
\]
then
\[
D_{j, k} = \{ (z, a) : \mathcal{L}_j^+(z, a) > 0 > \mathcal{L}_{j+1}^+(z, a) \} \quad (21)
\]
for \( j = 1, \ldots, N - 1 \). Denote the region
\[
V_p \in \mathcal{F}(+, A, z) \text{ if and only if } p \leq j,
\]
and
\[
V_q \in \mathcal{F}(-, A, z) \text{ if and only if } q \geq k + 1.
\]

**Proof.** For the first part, if \( p \leq j \), then by (20) we have
\[
\alpha \cdot V_p + a + z - 1 \geq \alpha \cdot V_j + a + z - 1 > 0,
\]
hence \( V_p \in \mathcal{F}(+, A, z) \) on \( D_{j, k} \). Similarly, if \( p > j \) then
\[
\alpha \cdot V_p + a + z - 1 < \alpha \cdot V_{j+1} + a + z - 1 < 0,
\]
and \( V_p \notin \mathcal{F}(+, A, z) \) on \( D_{j, k} \). The results of the second part can also be obtained similarly. The proof is complete. ☐

**Remark 2.6.** The other undefined regions in Theorem 2.5 will be denoted by \( D_{0, k}, D_{j, 0}, D_{N, k} \) and \( D_{j, N} \), respectively. See Fig. 3.

To illustrate the previous notions and theorems, we give the following example of one-dimensional CNN.

**Example 2.7.** Consider the one-dimensional CNN with \( A = [r, p, s] \) and denote the eight local patterns and lines by
Then in each region of \((r, s)\) plane as Fig. 2, the \((z, p - 1)\) plane can be partitioned up to 25 subregions. For example, when \(s > r > 0\), we have the bifurcation diagram as in Fig. 3.

Fig. 2. Partition of \((r, s)\) plane.

If \(\alpha\) is non-simple, then some of \(\ell^+_{ij}\) and \(\ell^-_{ij}\) collide with each other, and \((z, a)\) plane is reduced to fewer subregions than simple \(\alpha\). For example, when \(d = 1\), \(A\) is square-cross, e.g.

\[
A = \begin{bmatrix} 0 & a\varepsilon & 0 \\ a\varepsilon & a & a\varepsilon \\ 0 & a\varepsilon & 0 \end{bmatrix},
\]

then there are five straight lines for both \(\ell^+_{ij}\) and \(\ell^-_{ij}\). Details can be found in [Juang & Lin, 1997a]. Furthermore, as in many applications, the templates with certain symmetry frequently play important roles, e.g. [Chua, 1998]. In this case, \(\alpha\) is obviously non-simple. In the following, we introduce several templates used from time to time.

**Definition 2.8.** Template \(A = [a_{ij}]\) is called symmetric if

\[
a_{-i,-j} = a_{i,j}
\]  

(22)

for all \((i, j)\), and is called skew-symmetric (or anti-symmetric) if

\[
a_{-i,-j} = -a_{i,j}
\]  

(23)

for all \((i, j) \neq (0, 0)\). \(A\) is called isotropic if

\[
a_{i,j} = a_{|i|,|j|} \quad \text{for} \ i \cdot j \neq 0,
\]

and

\[
a_{i,j} = a_{|i|,|j|} = a_{j,i} \quad \text{for} \ i \cdot j = 0.
\]  

(24)

Denote by \(P^n_{s+1}\), \(P^n_{sk}\) and \(P^n_{iso}\) the parameter spaces of all symmetric, skew-symmetric, and isotropic templates in \(P^{n+1}\), respectively.

Clearly, symmetry reduces the number of parameters. Indeed, we have the following result.

**Proposition 2.9.** The number of independent parameters of \(P^n_{s+1}\) and \(P^n_{sk}\) is

\[
n_2 = 2d^2 + 2d
\]  

(25)

and of \(P^n_{iso}\) is

\[
n_4 = d^2 + d.
\]  

(26)
Furthermore, the set of all possible local patterns \(X^n\) surrounding a cell is equivalent to \(X_{n2.3}^{n+1}\) for \(P_{sk}^{n+1}\) and \(P_{sk}^{n+1}\) for \(P_{iso}^{n+1}\), where

\[X^m,2\ell+1 = \{v = (v_1, \ldots, v_m): v_j \in I[-\ell, \ell]\},\]

for positive integers \(m\) and \(\ell\). In particular, when \(d = 1\), we have \(n = 8\), \(n_2 = 4\), \(n_4 = 2\), and \(X^{4,3}\) and \(X^{2,5}\) for \(P_{sk}^9\) and \(P_{iso}^9\), respectively.

**Proof.** We only treat \(d = 1\), the general cases can be treated analogously. If \(A \in P_{sk}^9\) is given by

\[A = \begin{bmatrix} e_4 & e_3 & e_2 \\ e_1 & a & e_1 \\ e_2 & e_3 & e_4 \end{bmatrix},\]

then (5) is equivalent to

\[2e_1v_1 + 2e_2v_2 + 2e_3v_3 + 2e_4v_4 + a + z - 1 > 0,\]

where \(v_i \in I[-1, 1]\) for \(i = 1\) to 4. Therefore, the set of all possible local patterns surrounding a cell + is \(X^{4,3}\). Similar results hold for \(P_{sk}^9\) and \(P_{iso}^9\). □

**Remark 2.10**

(i) If

\[|a_{-i,-j}| = |a_{i,j}| \quad (27)\]

for all \((i, j)\) with \((22)\) and \((23)\) holds for some indices \((i, j) \neq (0, 0)\), \(A\) is called partial skew-symmetric. In this case, the set of all possible local patterns is still \(X^{n2.3}\). Similar results hold for partial isotropic template \(A = [a_{ij}]\) which satisfies

\[|a_{i,j}| = |a_{|i|,|j|}| \quad (28)\]

for all \((i, j) \neq (0, 0)\).

(ii) If \(A\) enjoys certain symmetry but is not isotropic, then the configuration space can be written as a product of some \(X^{k,\ell}\). For example, if \(d = 2\) and

\[A = \begin{bmatrix} e_4 & e_7 & e_2 & e_6 & e_4 \\ e_8 & e_3 & e_1 & e_3 & e_5 \\ e_2 & e_1 & a & e_1 & e_2 \\ e_5 & e_3 & e_1 & e_3 & e_8 \\ e_4 & e_6 & e_2 & e_7 & e_4 \end{bmatrix},\]

then the configuration space \(X\) is \(X^{4,3} \times X^{4,5}\).

### 3. Learning Problems

As mentioned in Sec. 1, the learning problem is a basic problem in CNN. Resolving this problem requires a more thorough understanding of linearly separable sets in a given configuration space. In this section, we apply the classical theory of convex set to study the property of linearly separable sets. In particular, supporting hyperplanes are used to characterize the linear separability of \(U\) and its complement \(U^c\) in configuration space \(X\).

We first recall some notions from the theory of convex sets. Details can be found in [Lay, 1992].

**Definition 3.1**

(i) Let \(S\) be a convex set in \(\mathbb{R}^n\). A point \(x\) in \(S\) is called an extreme point of \(S\) if there exists no non-degenerate line segment in \(S\) that contains \(x\) in its relative interior.

(ii) Let \(S\) be a set in \(\mathbb{R}^n\) and \(H\) be a hyperplane in \(\mathbb{R}^n\). Then \(H\) is said to be a supporting plane at \(x \in S\) if \(x \in H\) and there exists a linear function \(\ell: \mathbb{R}^n \rightarrow \mathbb{R}^1\), such that

\[\ell(H) = 0, \quad (29)\]

and

\[\ell(S) \geq 0 \text{ or } \ell(S) \leq 0. \quad (30)\]

(iii) Let \(S\) be a compact convex subset in \(\mathbb{R}^n\). We call a subset \(F\) of \(S\) an \(m\)-face if there is a supporting hyperplane \(H\) of \(S\) such that \(F = S \cap H\) which is \(m\)-dimensional. In particular, \(0\)-face of \(S\) is a vertex of \(S\), \((n - 1)\)-face of \(S\) is called a facet of \(S\).

(iv) A polytope \(P\) in \(\mathbb{R}^n\) is a convex hull of some finite set \(\{x_1, \ldots, x_k\}\), i.e.,

\[P = \text{conv}\{x_1, \ldots, x_k\}. \quad (31)\]

The set \(\{x_1, \ldots, x_k\}\) is called a minimal representation of the polytope \(P\) if (31) holds and for each \(i = 1, \ldots, k\), \(x_i \notin \text{conv}\{\bigcup_{j \neq i}\{x_j\}\}.

**Remark 3.2.** Clearly, if \(\{x_i\}_{i=1}^k\) is a minimal representation of the polytope \(P\), each \(x_i\) is an extreme point of \(P\) and a vertex of \(P\) for \(i = 1, 2, \ldots, k\).

A set \(H\) in \(\mathbb{R}^n\) is a hyperplane if there exists \((\alpha, b) \in \mathbb{R}^{n+1}\) with \(\alpha \neq 0\) such that

\[H = H_{\alpha, b} \equiv \{x \in \mathbb{R}^n: \alpha \cdot x + b = 0\}.\]
Then $H$ separates $\mathbb{R}^n$ into two half-spaces:

$$H_{a,b}^+ = \{ x \in \mathbb{R}^n : \alpha \cdot x + b > 0 \}$$

and

$$H_{a,b}^- = \{ x \in \mathbb{R}^n : \alpha \cdot x + b < 0 \}.$$ 

Hereinafter, assume that $X \subseteq \mathbb{R}^n$ is an $n$-dimensional finite set which does not necessarily equal to $X^{n,t}$, a finite rectangular lattice, and for each $U \subseteq X$ denote by

$$U^c = X - U.$$ 

$U$ is called a linearly separable set (in $X$) if (8) holds. In this case, the positive convex cone $A^+(U)$ is nonempty. Clearly, any $(A, z) \in A^+(U)$ will solve the learning problem. Then, we denote $\mathcal{H}(U)$ by

$$\mathcal{H}(U) = \{ \text{the set of all hyperplanes separating $U$ and $U^c$} \}. \quad (32)$$

It is clear that for any $\gamma > 0$, $(\gamma \alpha, \gamma b) \in A^+(U)$ if and only if $(\alpha, b) \in A^+(U)$. It is also clear that

$$H_{\gamma \alpha, \gamma b} = H_{\alpha, b}$$

for any $\gamma \neq 0$. Hence, each element in $\mathcal{H}(U)$ can be represented by a ray $\{ (\gamma \alpha, \gamma b) : \gamma > 0 \}$ for some $(\alpha, b) \in A^+(U)$ and vice versa.

Recall that $\partial A^+(U)$, in which the boundary of $A^+(U)$, then for each $(\overline{\alpha}, \overline{b}) \in \partial A^+(U)$ it is easy to see that the associate hyperplane $H_{\overline{\alpha}, \overline{b}}$ is a supporting hyperplane of $U$ or $U^c$ or both.

**Definition 3.3.** If $(\alpha^*, b^*) \in \partial A^+(U)$ and $H_{\alpha^*, b^*} \cap U$ is $(n-1)$-dimensional, then $H_{\alpha^*, b^*}$ is called an extreme supporting hyperplane.

For any linearly separable set $U$, the set of extreme supporting hyperplanes $\mathcal{H}^+(U)$ of $U$ is nonempty and finite since $X$ is finite.

In the following, we prove Theorem 1.3 such that for a given linearly separable set $U$ and an extreme supporting hyperplane in $\mathcal{H}^+(U)$, by some reduction procedure, we can obtain a finite sequence of supporting hyperplanes in $\partial A^+(U)$ which separate $U$ and $U^c$.

**Proof of Theorem 1.3.** The first part of the theorem can be proved by induction on $i$, $i = 2$ to $n-1$. The detail is omitted.

As for the converse part of the theorem, if $\text{conv}(U) \cap \text{conv}(W) \neq \emptyset$, then there is $j$ such that $\text{conv}(H^{(j)} \cap X) \cap \text{conv}(H^{(j)}, X) \neq \emptyset$, a contradiction to assumption (b). The proof is complete. $\blacksquare$

By Theorem 1.3, a linearly separable set $U$ can be represented by a finite sequence of hyperplanes. Among these representations, under suitable normalization, $U$ can be uniquely represented by a finite set of extreme supporting hyperplanes as follows:

**Theorem 3.4.** Given any linearly separable set $U$ of $X$ in $\mathbb{R}^n$, there is a unique family of finite many extreme supporting hyperplanes $\{H_j\}_{j=0}^m$ such that

$$U = \bigcap_{j=0}^m (H_j^+ \cap X)$$

and

$$U^c = \bigcup_{j=0}^m (H_j^- \cap X).$$

Conversely, if $\{H_j\}_{j=0}^m$ is a set of hyperplanes in $\mathbb{R}^n$, define

$$U = \bigcap_{j=0}^m (H_j^+ \cap X), \quad (33)$$

and

$$W = \bigcup_{j=0}^m (H_j^- \cap X). \quad (34)$$

If

$$\bigcap_{j=0}^m H_j = \emptyset \text{ and } U \cup W = X, \quad (35)$$

then $\text{conv}(U) \cap \text{conv}(W) = \emptyset$, i.e. $U$ is linearly separable and $W = U^c$.

**Proof.** According to Theorem 1.3, $U$ can be determined by a finite sequence of supporting hyperplanes. The first part of the theorem can be proved by induction, i.e. the $(n-i)$-face can be determined by $(n-i+1)$-face. Without a loss of generality, we prove only that each $(n-2)$-face is determined by $(n-1)$-faces that contains it. Assume that $H$ is an $(n-2)$-face of $U$ that separates $\text{conv}(U)$ and $\text{conv}(U^c)$. Let $H_1$ and $H_2$ be $(n-1)$-dimensional supporting hyperplanes.
that contain $H$. If $h$, $h_1$, $h_2$ be a linear function corresponding to $H$, $H_1$ and $H_2$, for $x \in (H_1 \cap X) \setminus (H \cap X)$, $h(x) > 0$, $y \in (H_2 \cap X) \setminus (H \cap X)$, $h(y) > 0$, $x \in (H_1 \cap X) \setminus (H \cap X)$, $h_2(x) > 0$, $h_1(y) > 0$, $h_1(x) = 0$, $h_2(y) = 0$. Define $G(t) = h(t) - (h(y)/h_1(y))h_1(t) - (h(x)/h_2(x))h_2(t)$. Then $G$ is linear and $G(t) = 0$ for all $t \in (H_1 \cup H_2) \cap X$. Hence $G(0) = 0$, i.e.$$
abla(\mathcal{U}) = \frac{h(y)}{h_1(y)}h_1(t) + \frac{h(x)}{h_2(x)}h_2(t),$$thus, $H$ is determined by $H_1$ and $H_2$. By induction, we prove the first part. Conversely, suppose that $\text{conv}(\mathcal{U}) \cap \text{conv}(\mathcal{V}) \neq \emptyset$, then by the definition of $\mathcal{U}$ and $\mathcal{V}$ there exists $x_0 \in (\bigcap_{j=1}^m H_j) \cap (\bigcap_{j=1}^m H_j')$, i.e. $x_0 \in \bigcap_{j=1}^m H_j \cap H_j'$, hence, $x_0 \in H_j \cap H_j'$ for all $i$, but $H_j \cap H_j' = \emptyset$ so $x_0 \in H_j$ for all $j$, i.e. $x_0 \in \bigcap_{j=1}^m H_j = \emptyset$ contradicts the assumption. The proof is complete.  

Remark 3.5

(i) If the space $X$ is a finitely rectangular lattice, the extreme supporting hyperplanes for a linearly separable set $\mathcal{U}$ in $X$ can be obtained with the aid of a computer. Therefore, to obtain all linearly separable sets $X$, it suffices to find all nonequivalent extreme hyperplanes $\{H_j\}_{j=0}^m$ which satisfy (33), (34) and (35). When the number of elements in $X$ is small, the complete list of all linearly separable sets $\mathcal{F}(X)$ can be written down explicitly by a computer program. Nevertheless, $\mathcal{F}(X)$ may still be extremely large. For example, if $X = X^{4,3}$, there are 308 nonequivalent extreme supporting hyperplanes. Details can be found in the Appendix.

(ii) The extreme supporting hyperplanes $H^*(\mathcal{U})$ of a linearly separable set $\mathcal{U}$ act as a vertex of polytopes of suitable cross-sections of $\mathcal{T}(\mathcal{U})$. This duality can be constructed as follows:

For each $H_{\alpha,b} \in H^*(\mathcal{U})$, it can be normalized by $|(|\alpha, b|) = 1$, and denoted by $H^*(\mathcal{U}) = \{H_{\alpha,b_1}, \ldots, H_{\alpha_m,b_m}\}$. Let $$\bar{Q} = \frac{1}{m} \sum_{i=1}^m (\alpha_i, b_i)$$ and $$Q = \frac{\bar{Q}}{|Q|}.$$ The cross-section $\mathcal{C}(\mathcal{U}, Q)$ of $\mathcal{A}(\mathcal{U})$ is defined by $$\mathcal{C}(\mathcal{U}, Q) = \{x \in \mathbb{R}^{n+1} : (x - Q) \cdot Q = 0\}.$$ The polytope $P(\mathcal{U}) = P(\mathcal{U}, Q)$ is then defined by $$P(\mathcal{U}) = \mathcal{C}(\mathcal{U}, Q) \cap \mathcal{P}(\mathcal{U}).$$ It is clear that $$P(\mathcal{U}) = \text{conv}\{(\alpha^*_1, b^*_1), \ldots, (\alpha^*_m, b^*_m)\}$$ with $$H_{\alpha^*_i,b^*_i} \in H^*(\mathcal{U}).$$

Finally, Theorem 1.3 can be applied to solve the following problems, which is closely related to the Learning Problem:

Given $\mathcal{U}_0 \subset X$, a set of local patterns which is required to be linearly separable, find $\mathcal{U} \in \mathcal{F}(X)$ and $\mathcal{U}_0 \subset \mathcal{U}$ such that $$z(\mathcal{U} - \mathcal{U}_0) = \min\{z(\mathcal{U} - \mathcal{U}_0) : \mathcal{U}_0 \subset \mathcal{U} \text{ and } \mathcal{U} \in \mathcal{F}(X)\}.$$ (36)

**Theorem 3.6.** Given $\mathcal{U}_0 \subset X$ in $\mathbb{R}^n$, then there exists $\mathcal{U} \in \mathcal{F}(X)$ such that $\mathcal{U}_0 \subset \mathcal{U}$ and (36) holds. Furthermore, the set $\mathcal{U}$ is not unique in general.

**Proof.** For given $\mathcal{U}_0 \subset X$ in $\mathbb{R}^n$, let $$\mathcal{C}(\mathcal{U}_0) = \text{conv}(\mathcal{U}_0).$$ We may assume that $\mathcal{C}(\mathcal{U}_0)$ is $n$-dimensional and $\{H_{i}^{(n-1)}\}_{i=1}^{k_{n-1}}$ are the $(n-1)$-dimensional extreme supporting hyperplanes of $\mathcal{C}(\mathcal{U}_0)$ such that $H_{i}^{(n-1)} \cap X$ contains at least $n$ points of $X$, for $i = 1$ to $k_{n-1}$. Consider the subsets $$\text{conv}\left(H_{i}^{(n-1)} \cap X \cap \mathcal{C}(\mathcal{U}_0)\right) \text{ for } i = 1 \text{ to } k_{n-1},$$ then each subset is a polytope in the $(n-1)$-dimensional hyperplanes, $\{H_{i}^{(n-1)}\}_{i=1}^{k_{n-1}}$. By the same process, we can construct finitely many sequences of hyperplanes, each determining a linearly separable subset in $X$ which contains $\mathcal{U}_0$. Hence, there exist linearly separable subsets to satisfy (36). The proof is complete.

4. Algorithms

In this section, we first apply the method of linear programming to determine whether or not a given set $\mathcal{U}$ in $X \subset \mathbb{R}^n$ is linearly separable, i.e. to test whether (8) is true or not. Next, an algorithm is
developed to obtain the vertices for polytope $P(U)$ for any linearly separable set $U$. This work also studies some properties of generalized perceptron matrix (as defined later) of $U$ and $U^c$.

Denote by

$$U = \{U_i\}_{i=1}^r \quad \text{and} \quad U^c = \{U_i\}_{i=r+1}^N . \quad (37)$$

If

$$\text{conv}(U) \cap \text{conv}(U^c) \neq \emptyset , \quad (38)$$

then there exist $\beta_i \geq 0, i \in [1, N]$ such that

$$\sum_{i=1}^r \beta_i U_i = \sum_{i=r+1}^N \beta_i U_i \quad (39)$$

and

$$\sum_{i=1}^r \beta_i = 1 = \sum_{i=r+1}^N \beta_i . \quad (40)$$

That is, the matrix equation

$$MB = C \quad (41)$$

has a non-negative solution $B$, where

$$M = M_{(n+2) \times N} \equiv \begin{bmatrix} U_1 \cdots U_r & -U_{r+1} & \cdots & -U_N \\ 1 & \cdots & 1 & 0 & \cdots & 0 \\ 0 & \cdots & 0 & 1 & \cdots & 1 \end{bmatrix}$$

and

$$C = C_{(n+2) \times 1} \equiv (0, \ldots, 0, 1, 1)^t .$$

To solve (41), we shall use the "Two-phases" method, see, e.g. [Fang & Puthenpura, 1993]. We introduce an artificial variable

$$\tilde{B} \equiv (\tilde{\beta}_1, \ldots, \tilde{\beta}_{n+2})^t \in \mathbb{R}^{n+2}$$

to (41) and consider the following optimization problem:

$$\begin{cases} 
\min \sum_{i=1}^{n+2} \tilde{\beta}_i , \\
MB + \tilde{B} = C , \\
B \geq 0 \quad \text{and} \quad \tilde{B} \geq 0 .
\end{cases} \quad (42)$$

It is clear that (42) is a standard linear programming problem.

Since matrix $M$ is of full row rank, and $(0, C^t)^t$ is a basic feasible solution of (42), by the fundamental theorem of linear programming, see e.g. [Fang & Puthenpura, 1993], the optimal objective value can be attained which is denoted by

$$(B^*, \tilde{B}^*)^t \equiv (\beta_1^*, \ldots, \beta_N^*, \tilde{\beta}_1^*, \ldots, \tilde{\beta}_{n+2}^*) . \quad (43)$$

Then the linear separability problem of $U$ with its complement $U^c$ can be answered by $(B^*, \tilde{B}^*)$ as Theorem 1.4.

**Proof of Theorem 1.4**

(i) If $\tilde{B}^* \neq 0$, then we claim that (41) has no non-negative solution. Otherwise, if there was a non-negative solution $B$ of (41), then $(B, 0)$ is a basic feasible solution with zero objective value, a contradiction to $\tilde{B}^* \neq 0$ which gives a positive objective value.

(ii) If $\tilde{B}^* = 0$, then $B^*$ is a non-negative solution of (41). Hence (38) holds. The proof is complete. ■

**Remark 4.1.** If the objective function in (42) is replaced by $\sum_{i=1}^{n+2} \tilde{\beta}_i^2$, the result still holds. In this case, the square root of optimal objective value corresponds to the distance between conv($U$) and conv($U^c$).

After establishing the algorithm to test whether or not (8) holds, we present an algorithm and also recall the perception convergence theorem to find some $(a, b) \in A(U)$ whenever it is nonempty, i.e. (8) holds.

Given $U$ and $U^c$ as in (37), the perception matrix $Q = Q(U)$ defined by

$$Q = \begin{bmatrix}
-U_1 & -1 \\
\vdots & \vdots \\
-U_{r-1} & 1 \\
U_{r+1} & 1 \\
\vdots & \vdots \\
U_N & 1
\end{bmatrix} \quad (44)$$

is a $N \times (n + 1)$ matrix, see e.g. [Rosenblatt, 1992].

Therefore, if $U$ is linearly separable, $(a, b) \in A^+(U)$ if and only if

$$Q \begin{pmatrix} a^t \\ b \end{pmatrix} < 0 . \quad (45)$$
On the other hand, if there is \((\alpha, b) \in \mathbb{R}^{n+1}\) such that (45) holds, \(\mathcal{U}\) is linearly separable.

When \(\mathcal{U}\) is linearly separable, two different methods can be applied to obtain a solution \((\alpha, b)\) of (45). One is the linear-programming method and the other is the Perceptron Convergence Theorem.

**Method 1.** (45) is a linear-strictly inequality problem (LSI) in linear programming. Denote the unknown by \(X = [\alpha, b]^\top \in \mathbb{R}^{n+1}\) and introduce some artificial variables \(\underline{x}, \hat{x} \in \mathbb{R}^{n+1}\) and \(\hat{x} \in \mathbb{R}^n\) to (45). The following optimization problem to (45) as in (42) is considered as follows.

\[
\begin{align*}
\min_{\underline{x}, \hat{x}} & \sum_{i=1}^N \underline{x}_i, \\
Q(\underline{x} - \hat{x}) + \hat{x} & = 0, \\
\underline{x}, \hat{x} & \geq 0,
\end{align*}
\]

(46)

Obviously, (46) is a standard linear programming problem, we can easily obtain solutions for it.

**Method 2.** Perceptron Convergence Theorem [Rosenblatt, 1992].

A perceptron is a network which depends on the sequence of past activity state of the network. The vector \((\alpha, b)\) in (45) can be considered as connection-value vector in neural network [Rosenblatt, 1992]. To obtain the solution of (45), perceptron takes the “error-correction rule,” i.e. consider \((\alpha, b)\) as a function of discrete \(t\), \((\alpha(t), b(t))\), \(t \in \mathbb{Z}^+\). Connection-value vectors change only if the patterns are misclassified; otherwise, they keep the same value. The change procedure is as follows.

Let
\[
\begin{align*}
\mathcal{U}_i & \equiv \begin{cases}
(U_i, 1) & U_i \in \mathcal{U}, \\
(-U_i, -1) & U_i \in \mathcal{U}^c,
\end{cases}
\end{align*}
\]

and
\[
W(t) = \begin{pmatrix} \alpha(t)^T \\ b(t) \end{pmatrix}.
\]

If after \(t\) iteration, we have \(W(t) \cdot \mathcal{U}_k < 0\) (misclassified), then at \(t + 1\) iteration we change
\[
W(t + 1) = W(t) + \alpha \mathcal{U}_k,
\]
with \(W(t + 1)\) as the new connection-vector value, where \(\alpha\) is a constant and lies between 0 and 1. This is called the learning coefficient.

Throughout this process, after finite time iteration, the Perceptron Convergence Theorem can ensure that such a connection-vector value exists for linearly separable sets \(\mathcal{U}\) and \(\mathcal{U}^c\).

**Theorem 4.2** (Perceptron Convergence Theorem). For linearly separable sets \(\mathcal{U}\) and \(\mathcal{U}^c\) in \(\mathbb{R}^n\), after finite time iteration of (47), there exists \(W\) to satisfy (45).

**Proof.** Details can be found in [Lunberger, 1986; Rosenblatt, 1992]. In addition, the learning rules are outlined as follows.

**Step 1.** Let \(W(0) \equiv 0\), a zero vector in \(\mathbb{R}^{n+1}\).

**Step 2.**
\[
W(t + 1) = \begin{cases}
W(t) + \alpha \mathcal{U}_k & \text{if } W(t) \mathcal{U}_k < 0 \\
W(t) & \text{if } W(t) \mathcal{U}_k > 0
\end{cases}
\]
and
\[
0 < \alpha \leq 1.
\]

**Step 3.** Returning to Step 2 until for all \(\mathcal{U}_k, k = 1\) to \(N\) satisfy
\[
W(t^\ast) \mathcal{U}_k > 0, \quad \text{for some } t^\ast \geq 0.
\]

Finally, for two given disjoint finite subsets \(\mathcal{U}\) and \(\mathcal{U}^c\) in \(\mathbb{R}^n\), we can study some properties of generalized perceptron matrix \(R\) defined as follows.

Assume that \(\mathcal{U} = \{U_i\}_{i=1}^N, \mathcal{U}^c = \{U_i\}_{i=r+1}^N\). Define
\[
\mathcal{U}_i \equiv \begin{cases}
(1, U_i), & U_i \in \mathcal{U} \\
(-1, -U_i), & U_i \in \mathcal{U}^c
\end{cases}
\]
for \(i = 1\) to \(N\) and
\[
\mathcal{U} = \{\mathcal{U}_i\}_{i=1}^N, \quad \mathcal{U}^c = \{\mathcal{U}_i\}_{i=r+1}^N.
\]

Then (40) is equivalent to there exists \(\beta_i \geq 0, i = 1\) to \(N\) such that
\[
\begin{align*}
\sum_{i=1}^N \beta_i \mathcal{U}_i & = 0 \\
\sum_{i=1}^N \beta_i & = 1 \quad \text{(by rescale)}.
\end{align*}
\]

In (48), if we let \(X = \sum_{i=1}^N \beta_i \mathcal{U}_i = 0\), then
Remark 4.4. In Theorem 1.4, if \( \text{conv}(\mathcal{U}) \cap \text{conv}(\mathcal{U}^c) \neq \emptyset \) if and only if 0 is an eigenvalue of \( R \) with eigenvector \( B = (\beta_1, \ldots, \beta_N)^T \geq 0 \), i.e., \( \beta_i \geq 0 \) for \( i = 1 \) to \( N \).

Then, the separation of \( \mathcal{U} \) and \( \mathcal{U}^c \) can be determined by the following criterion of \( R \).

**Theorem 4.3.** Assume that \( \mathcal{U} \) and \( \mathcal{U}^c \) are disjoint finite subsets in \( \mathbb{R}^n \), then

1. \( \text{conv}(\mathcal{U}) \cap \text{conv}(\mathcal{U}^c) \neq \emptyset \) if and only if 0 is an eigenvalue of \( R \) with eigenvector \( B = (\beta_1, \ldots, \beta_N)^T \geq 0 \), i.e., \( \beta_i \geq 0 \) for \( i = 1 \) to \( N \).
2. \( \text{conv}(\mathcal{U}) \cap \text{conv}(\mathcal{U}^c) = \emptyset \) if and only if there exists \( B \geq 0 \) s.t. \( R \cdot B > 0 \).

**Proof.** The proof can be verified by the Perceptron Convergence Theorem and is omitted.

**Remark 4.4.** In Theorem 1.4, if \( \text{conv}(\mathcal{U}) \cap \text{conv}(\mathcal{U}^c) \neq \emptyset \) then \( B^* = 0 \), and in fact \( B^* \) is an eigenvector of \( R \) with eigenvalue 0, since

\[
R \cdot B^* = Q \cdot Q^t B^* = 0
\]

where \( R \equiv (m_{ij})_{N \times N} \) with \( m_{ij} = \mathcal{U}_i \cdot \mathcal{U}_j \) and \( B \equiv (\beta_1, \ldots, \beta_N)^T \).

The proof can be verified by the Perceptron Convergence Theorem and is omitted.

**Acknowledgments**

The authors would like to thank the National Science Council of the Republic of China for financially supporting this research under Contract No. NSC.


Appendix

This appendix lists the nonequivalent linearly separable sets by its associated extreme hyperplanes for $X^{2,3}$ and $X^{3,3}$. Some typical linearly separable sets of $X^{4,3}$ are also given.

(I) The integer coefficients $(A, B, C)$ of all extreme supporting hyperplanes, $H : AX + BY + C = 0$, in $X^{2,3}$ is given by

\[
A : 21111 \\
B : 11100 \\
C : 11010
\]

Above data can be interpreted such that there are five nonequivalent linearly separable sets represented by five straight lines $AX + BY + C = 0$ in $X^{2,3}$, see the following diagrams. In each diagram, the vertex which is circled by $\circ$ belongs to the associated linearly separable set.
Notably, in the above diagrams, we only list $\mathcal{U}$ with $\mathfrak{U} \geq 6$. For those $\mathcal{U}$ with $\mathfrak{U} \leq 5$ can be reflected with respect to the hyperplane. For example, $\mathcal{U} = \{(1, 1)\}$ is equivalent to the complement of (ii).

(II) The integer coefficients $(A, B, C, D)$ of all extreme supporting hyperplanes, $H: AX + BY + CZ + D = 0$, in $X^{3,3}$ is given by

\begin{align*}
A : & \quad 4 \ 4 \ 4 \ 3 \ 3 \ 3 \ 3 \ 2 \ 2 \ 2 \ 2 \ 2 \ 1 \ 1 \ 1 \ 1 \ 1 \ 1 \\
B : & \quad 3 \ 2 \ 2 \ 2 \ 2 \ 1 \ 1 \ 1 \ 2 \ 1 \ 1 \ 1 \ 1 \ 1 \ 1 \ 1 \ 0 \ 0 \\
C : & \quad 2 \ 1 \ 1 \ 2 \ 1 \ 1 \ 1 \ 1 \ 1 \ 1 \ 1 \ 1 \ 0 \ 1 \ 1 \ 0 \ 0 \ 0 \\
D : & \quad 1 \ 3 \ 1 \ 1 \ 2 \ 1 \ 2 \ 1 \ 3 \ 1 \ 2 \ 1 \ 0 \ 1 \ 2 \ 1 \ 0 \ 1 \ 0 \ 0 
\end{align*}

(III) The number of integer coefficients $(A, B, C, D, E)$ of all extreme supporting hyperplanes, $H : AX + BY + CZ + DW + E = 0$ in $X^{4,3}$ is 308 and we only list some typical ones.

\begin{align*}
A : & \quad 10 \ 9 \ 8 \ 7 \ 6 \ 5 \ 4 \ 3 \ 2 \ 1 \\
B : & \quad 9 \ 8 \ 7 \ 6 \ 5 \ 4 \ 4 \ 3 \ 2 \ 1 \\
C : & \quad 4 \ 5 \ 6 \ 5 \ 4 \ 4 \ 3 \ 2 \ 2 \ 1 \\
D : & \quad 2 \ 2 \ 4 \ 4 \ 4 \ 2 \ 2 \ 2 \ 1 \ 1 \\
E : & \quad 3 \ 6 \ 5 \ 2 \ 3 \ 3 \ 5 \ 4 \ 5 \ 3 
\end{align*}

The algorithm to generate these data and the complete list can be found in http://www.math.nthu.edu.tw/lin.