1993

Multisurface Method of Pattern Separation

Jennifer L. Jancik '93
Illinois Wesleyan University

Multisurface Method of Pattern Separation

Jennifer L. Jancik
Illinois Wesleyan University
Bloomington, Il 61702-2900

May 12, 1993

1 Introduction

The recognition and separation of patterns is becoming increasingly important in modern applications. For example, it is currently being used at the University of Wisconsin Hospitals to aid in the diagnosis of breast cancer. Other medical applications of this technique include the diagnosis of heart disease and cystic fibrosis. Basically, the problem of recognition and separation of pattern sets can be stated as:

Given two disjoint, finite sets, $\mathcal{A}$ and $\mathcal{B}$, in n-dimensional real space $\mathbb{R}^{n\times 1}$, a discriminant function $f : \mathbb{R}^{n\times 1} \longrightarrow \mathbb{R}$, must be found that has the properties that $f(\mathcal{A}) > 0$ and $f(\mathcal{B}) \leq 0$, i.e. $f(a) > 0$ when $a \in \mathcal{A}$ while $f(b) \leq 0$ when $b \in \mathcal{B}$.

A single linear programming problem can be used to find a linear discriminant function of the form

$$f(x) = \alpha x + \beta, \text{ where } \alpha, \beta \in \mathbb{R},$$

when the convex hulls of $\mathcal{A}$ and $\mathcal{B}$ do not intersect. Some examples of this situation will be explored in the third section of this paper. In real life, however, the convex hulls often intersect, and a more complex discriminant function must be constructed. In the papers [1] and [2], Mangasarian, Setiono, and Wolberg have established a method of constructing discriminant functions for the more complex situations where the convex hulls of $\mathcal{A}$ and $\mathcal{B}$ do intersect. Their method uses linear programming techniques. The purpose of this paper is to reformulate the results of Mangasarian, Setiono, and Wolberg into the format used in our IWU linear programming course and many of our linear algebra classes in hopes that future IWU linear programming students can quickly master the concepts and algorithm. It is anticipated that future student projects may consist of some of the following:

(i) a software implementation of this algorithm in our NeXT Lab.

(ii) applications using our NeXT implementation to local pattern recognition problems.

2 Basic Notation and Concepts

Our notation is consistent with that found in Jeter [3] and Cottle and Stone [4]. Throughout this paper, a vector of ones will be denoted by e, vectors will be denoted by column vectors, and a zero vector will be denoted by 0.

Definition 2.1 *Unsymmetric primal linear programming problem: An unsymmetric primal linear programming problem is one of the form*

$$\textit{minimize } c^T x \textit{ subject to } Ax = b \textit{ and } x \geq 0,$$
$$\textit{where } x = [x_1 \ldots x_n]^T \in \mathbb{R}^{n\times 1}, c = [c_1 \ldots c_n]^T \in \mathbb{R}^{n\times 1}, \textit{ and } A = [a_{ij}] \in \mathbb{R}^{m\times n}.$$

Definition 2.2 *Dual problem: The dual problem is defined to be*

$$\textit{maximize } b^T w$$
$$\textit{subject to}$$
$$A^T w \leq c, \textit{ where } w \in \mathbb{R}^{m\times 1}.$$

The Fundamental Duality Theorem of Linear Programming, [3], states that:

Theorem 2.1 *Fundamental Duality Theorem: If either the primal or the dual problem has a finite optimal solution, then the other problem has a finite optimal solution as well. Moreover, the optimal values of the two objective functions are equal, i.e.,*

$$\min \{c^T x : Ax = b, x \geq 0\} = \max \{b^T w : A^T w \leq c\}.$$

An immediate corollary to the Fundamental Duality Theorem of Linear Programming is the well known Farkas Lemma:

Lemma 2.2 ***Farkas Lemma:*** *There exists an $x \geq 0$ such that $Ax = b$ if and only if $A^T w \geq 0$ implies that $b^T w \geq 0$.*

The Farkas Lemma is often stated in the following form:

Lemma 2.3 ***(Farkas)*** *One and only one of the following systems is consistent:*

$$Ax = b \text{ where } x \geq 0$$

or

$$A^T y \geq 0 \text{ where } b^T y < 0.$$

Geometrically, the Farkas Lemma says that either b is in the convex cone generated by the columns of A or b is separated from the convex cone generated by the columns of A by a hyperplane.

Definition 2.3 ***Symmetric primal linear programming problem:*** *A symmetric primal linear programming problem is one of the form*

$$\text{minimize } c^T x \text{ subject to } Ax \geq b \text{ and } x \geq 0.$$

Definition 2.4 ***Dual of the symmetric primal problem:*** *The dual of the symmetric primal problem is one of the form*

$$\text{maximize } b^T w \text{ subject to } A^T w \leq c \text{ and } w \geq 0.$$

When both the symmetric primal and dual problems are feasible, then there exists optimal solutions x and w to each respective problem such that $c^T x = b^T w$. Other basic notation used in this paper includes the following:

$$A_{\cdot j} := [a_{1j} \dots a_{nj}]^T, i.e., A_{\cdot j} \text{ denotes the jth column of matrix A. (Here } A = [a_{ij}] \in \mathbb{R}^{n \times m}).$$

($A_{i\cdot}$ will denote the ith row of matrix A). In particular, we define two collections $\mathcal{A}$ and $\mathcal{B}$ of points in $\mathbb{R}^{n \times 1}$ as follows:

$$A_{\cdot 1} = [a_{11} \dots a_{n1}]^T \in \mathbb{R}^{n \times 1}$$

$$A_{\cdot 2} = [a_{12} \dots a_{n2}]^T \in \mathbb{R}^{n \times 1}$$

$$\vdots$$

$$A_{\cdot m} = [a_{1m} \dots a_{nm}]^T \in \mathbb{R}^{n \times 1}$$

Then $\mathcal{A} = \{A_{\cdot 1}, \dots, A_{\cdot m}\} \subseteq \mathbb{R}^{n \times 1}$, i.e., $\mathcal{A}$ is a set of points in $\mathbb{R}^{n \times 1}$. Also define

$$A := [A_{\cdot 1} \dots A_{\cdot m}] = [a_{ij}] \in \mathbb{R}^{n \times m}$$

Next let

$$B_{\cdot 1} = [b_{11} \dots b_{n1}]^T \in \mathbb{R}^{n\times 1}$$

$$B_{\cdot 2} = [b_{12} \dots b_{n2}]^T \in \mathbb{R}^{n\times 1}$$

$$\vdots$$

$$B_{\cdot k} = [b_{1k} \dots b_{nk}]^T \in \mathbb{R}^{n\times 1}$$

Then $\mathcal{B} = \{B_{\cdot 1} \dots B_{\cdot k}\} \subseteq \mathbb{R}^{n\times 1}$ and $B := [B_{\cdot 1} \dots B_{\cdot k}] = [b_{ij}] \in \mathbb{R}^{n\times k}$. Notice that $\mathcal{A} = \{A_{1\cdot}^T, \dots, A_{m\cdot}^T\}$and $\mathcal{B} = \{B_{1\cdot}^T, \dots, B_{k\cdot}^T\}$. Extensive use will be made of the sets $\mathcal{A}$ and $\mathcal{B}$ and the matrices A and B in this paper. Also, $\mathcal{A}$ and $\mathcal{B}$ are assumed to be disjoint throughout this paper. For any $x \in \mathbb{R}^{n\times 1}$, the ∞ - norm of x, denoted by $\| x \|_\infty$, is defined by

$$\| x \|_\infty = \max \{| x_1 |, | x_2 |, \dots, | x_n |\}.$$

A point x is an extreme point of a convex set S if and only if $x = \alpha y + (1-\alpha)z$ where $y, z \in S$ and $\alpha \in [0,1]$ implies that $\alpha = 0$ or $\alpha = 1$, i.e., $\nexists y, z \in S$ and $\alpha \in (0,1)$ for which $y \neq z$ and $x = \alpha y + (1-\alpha)z$. In particular, x will be an extreme point of conv($\mathcal{A}$) if and only if $x = Au, u \geq 0$ and $e^T u = 1$ implies that u has the form $[0 \dots 0\ 1\ 0 \dots 0]^T$ where u has only one nozero coordinate, a one. Finally, for any $S \subseteq \mathbb{R}^{n\times 1}$ the convex hull of S (denoted by conv(S)) and the conical hull of S (denoted by coni(S)) are defined as follows:

$$\text{conv(S)} = \{\sum_{i=1}^{k} \alpha_i x_i : \text{ each } x_i \in S, \text{ each } \alpha_i \in \mathbb{R}, \text{ each } \alpha_i \geq 0, \sum_{i=1}^{k} \alpha_i = 1, \text{ and } k \in N\}$$

and

$$\text{coni(S)} = \{\sum_{i=1}^{k} \alpha_i x_i : \text{ each } x_i \in S, \text{ each } \alpha_i \in \mathbb{R}, \text{ each } \alpha_i \geq 0, \text{ and } k \in N\}.$$

3 Basic Results

We begin by considering the relationships between three basic problems:

When are the convex sets, conv($\mathcal{A}$) and conv($\mathcal{B}$) disjoint? When do they intersect?

When is the system

(I)

$$\begin{array}{rcrcl} Au & - & Bv & = & 0 \\ -e^T u & + & e^T v & = & 0 \end{array}$$

$$0 \neq [u^T v^T]^T \geq 0$$

consistent, where $A \in \mathbb{R}^{n\times m}, B \in \mathbb{R}^{n\times k}, u \in \mathbb{R}^{m\times 1}$, and $v \in \mathbb{R}^{k\times 1}$. When is system (I) inconsistent?

When does the linear programming problem

LP1

$$\text{maximize } 0^T u + 0^T v - e^T r - e^T s$$

$$\text{subject to}$$

$$\begin{array}{rcrcrcrcl} Au & - & Bv & + & r & - & s & = & 0 \\ -e^T u & & & & & & & = & -1 \\ & & e^T v & & & & & = & 1 \end{array}$$

$$u \geq 0, v \geq 0, r \geq 0, s \geq 0$$

have a zero maximum? When does LP1 have a negative minimum?

Our first proposition explores the relationship between the first two problems. It forms the basis from which methods of determining whether or not the sets $\mathcal{A}$ and $\mathcal{B}$ are separable are based.

Proposition 3.1

$Conv(\mathcal{A}) \cap conv(\mathcal{B}) \neq \emptyset$ iff System I is consistent.

PROOF:

Suppose that u^* and v^* is a solution of (I). Then

$$\sum_{i=1}^{m} u_i^* A_{\cdot i} - \sum_{j=1}^{k} v_j^* B_{\cdot j} = Au^* - Bv^* = 0$$

Therefore

$$\sum_{i=1}^{m} u_i^* A_{\cdot i} = \sum_{j=1}^{k} v_j^* B_{\cdot j}.$$

But

$$0 \neq \begin{pmatrix} u^* \\ v^* \end{pmatrix} \geq 0$$

means that each $u_i^* \geq 0$ and $v_j^* \geq 0$ and at least one u_i^* or $v_j^* \neq 0$. So,

$$\sum_{i=1}^{m} u_i^* A_{\cdot i} \in \text{coni}(\mathcal{A}) \text{ and } \sum_{j=1}^{k} v_j^* B_{\cdot j} \in \text{coni}(\mathcal{B}).$$

But, letting

$$M := \sum_{i=1}^{m} u_i^* = \sum_{j=1}^{k} v_j^* > 0,$$

we have that

$$\frac{1}{M} \sum_{i=1}^{m} u_i^* A_{\cdot i} = \sum_{i=1}^{m} \left(\frac{1}{M} u_i^*\right) A_{\cdot i} \in \text{conv}(\mathcal{A}),$$

since each

$$\frac{1}{M} u_i^* \geq 0, \text{ and } \sum_{i=1}^{m} \frac{1}{M} u_i^* = \frac{1}{M} \sum_{i=1}^{m} u_i^* = \frac{1}{M} M = 1.$$

Likewise,

$$\frac{1}{M}\sum_{j=1}^{k} v_j^* B_{\cdot j} \in \text{conv}(\mathcal{B}).$$

But

$$\frac{1}{M}\sum_{i=1}^{m} u_i^* A_{\cdot i} = \frac{1}{M}\sum_{j=1}^{k} v_j^* B_{\cdot j}$$

Therefore

$$\text{conv}(\mathcal{A}) \cap \text{conv}(\mathcal{B}) \neq \emptyset.$$

Now suppose that

$$\text{conv}(\mathcal{A}) \cap \text{conv}(\mathcal{B}) \neq \emptyset.$$

Then $\exists$ u^* and v^* so that

$$\sum_{i=1}^{m} u_i^* A_{\cdot i} = \sum_{j=1}^{k} v_j^* B_{\cdot j}, \text{ i.e., } Au^* - Bv^* = 0,$$

where

$$u_i^* \geq 0, \ \sum_{i=1}^{m} u_i^* = 1, \ v_j^* \geq 0, \text{ and } \sum_{j=1}^{k} v_j^* = 1.$$

Notice that $0 \neq u^* \geq 0$ and $0 \neq v^* \geq 0$. Therefore,

$$0 \neq \begin{pmatrix} u^* \\ v^* \end{pmatrix} \geq 0.$$

Also,

$$\sum_{i=1}^{m} u_i^* = 1 = \sum_{j=1}^{k} v_j^*$$

i.e. , $e^T u^* = e^T v^*$. So, $-e^T u^* + e^T v^* = 0$. Therefore u^* and v^* solves (I). ∎

Proposition 3.1 can be restated as follows:

System (I) is inconsistent if and only if $\text{conv}(\mathcal{A}) \cap \text{conv}(\mathcal{B}) = \emptyset$.

Next we study the relationship of LP1 and the question of whether $\text{conv}(\mathcal{A})$ and $\text{conv}(\mathcal{B})$ are disjoint. The maximal value of the objective function in LP1 is always less than or equal to 0. Hence, when system (I) has a solution u and v, then u, v and $r = s = 0$ solves LP1 with a maximal objective value of 0. Likewise, when LP1 has a solution u, v, r, and s in which the maximal value of the objective function is zero, then $r = s = 0$ and (I) has a solution u and v. Hence,

Proposition 3.2 *System (I) is consistent if and only if LP1 has a solution in which the maximal value of the objective function is 0.*

Combining Proposition 1 and 2 yields:

Proposition 3.3 *$Conv(\mathcal{A}) \cap conv(\mathcal{B}) \neq \emptyset$ if and only if LP1 has a solution in which the maximal value of the objective function is 0.*

Proposition 3.4 *$Conv(\mathcal{A}) \cap conv(\mathcal{B}) = \emptyset$ if and only if LP1 has a solution in which the maximal value of the objective function is negative.*

The constraints for LP1 can be written as follows:

$$\begin{bmatrix} A & -B & I & -I \\ -e^T & 0 & 0 & 0 \\ 0 & e^T & 0 & 0 \end{bmatrix} \begin{bmatrix} u \\ v \\ r \\ s \end{bmatrix} = \begin{bmatrix} 0 \\ -1 \\ 1 \end{bmatrix}$$

Hence, the dual of LP1 can be expressed in the following forms:

$$\min\ 0^T c - \alpha + \beta$$

$$\text{subject to}$$

$$\begin{bmatrix} A & -B & I & -I \\ -e^T & 0 & 0 & 0 \\ 0 & e^T & 0 & 0 \end{bmatrix}^T \begin{bmatrix} c \\ \alpha \\ \beta \end{bmatrix} \geq \begin{bmatrix} 0 \\ 0 \\ -e \\ -e \end{bmatrix}$$

or

$$\min\ 0^T c - \alpha + \beta$$

$$\text{subject to}$$

$$\begin{bmatrix} A^T & -e & 0 \\ -B^T & 0 & e \\ I & 0 & 0 \\ -I & 0 & 0 \end{bmatrix} \begin{bmatrix} c \\ \alpha \\ \beta \end{bmatrix} \geq \begin{bmatrix} 0 \\ 0 \\ -e \\ -e \end{bmatrix}$$

or

<u>**LP2**</u>

$$\min\ 0^T c - \alpha + \beta$$

$$\text{subject to}$$

$$\begin{array}{rcrcr} A^T c & - & \alpha e & \geq & 0 \\ -B^T c & + & \beta e & \geq & 0 \\ Ic & & & \geq & -e \\ -Ic & & & \geq & -e \end{array}$$

where $A \in \mathbb{R}^{n \times m}, B \in \mathbb{R}^{n \times k}, c \in \mathbb{R}^{n \times 1}, \alpha \in \mathbb{R}, \beta \in \mathbb{R}$.

The Fundamental Duality Theorem of Linear Programming yields the following linear separability criterion for $\mathcal{A}$ and $\mathcal{B}$.

Proposition 3.5 *$Conv(\mathcal{A}) \cap conv(\mathcal{B}) = \emptyset$ if and only if LP2 has an optimal solution in which the minimum of the objective function is negative.*

Notice that in such an optimal soution of LP2, $\alpha > \beta$ since the minimal value of $0^T c - \alpha + \beta < 0$. Hence,

$$A^T c \geq \alpha e = \frac{2\alpha}{\alpha} e > \frac{\alpha + \beta}{2} e.$$

Likewise,

$$B^T c \leq \beta e = \frac{2\beta}{2} e < \frac{\alpha + \beta}{2} e.$$

Hence, the plane $\{x \in \mathbb{R}^{n\times 1} : x^T c = \frac{\alpha+\beta}{2}\}$ separates the sets $\mathcal{A}$ and $\mathcal{B}$.

The following linear programming problem also leads to separability criterion for sets $\mathcal{A}$ and $\mathcal{B}$:

LP3

$$\max e^T u + e^T v$$

subject to

$$\begin{array}{rcrcr} Au & - & Bv & = & 0 \\ -e^T u & + & e^T v & = & 0 \\ -u & & & \geq & -e \\ & & -v & \geq & -e \\ u & & & \geq & 0 \\ & & v & \geq & 0 \end{array}$$

where $A \in \mathbb{R}^{n\times m}, B \in \mathbb{R}^{n\times k}, u \in \mathbb{R}^{m\times 1}, v \in \mathbb{R}^{k\times 1}$.

Proposition 3.6 *Conv($\mathcal{A}$) $\cap$ conv($\mathcal{B}$) $\neq \emptyset$ if and only if LP3 has an optimal solution in which the maximal value of the objective function is positive.*

PROOF:

Suppose that conv($\mathcal{A}$) $\cap$ conv($\mathcal{B}$) $\neq \emptyset$. By Proposition 3.1 $\exists\, u, v \in \mathbb{R}^{n\times 1}$ so that

$$\begin{array}{rcrcr} Au & - & Bv & = & 0 \\ -e^T u & + & e^T v & = & 0 \end{array}$$

$$0 \leq u \neq 0$$

$$0 \leq v \neq 0.$$

Notice that u and v each have at least one positive component. Therefore,

$$e^T u = \sum_{i=1}^{m} u_i > 0 \text{ and } e^T v = \sum_{i=1}^{k} v_i > 0.$$

Therefore,

$$e^T u + e^T v > 0.$$

We will show that a feasible solution for LP3 can be constructed from u and v. Let

$$\hat{u} = \frac{1}{e^T u} u \text{ and } \hat{v} = \frac{1}{e^T v} v.$$

Then

$$\hat{u} \geq 0 \text{ and } \hat{v} \geq 0.$$

Notice that

$$\frac{1}{e^T u} u_i = \frac{u_i}{u_1 + \ldots + u_m} \leq 1 \text{ for all i.}$$

Therefore,

$$\hat{u} = \frac{1}{e^T u} u \leq e,$$

and

$$-\hat{u} \geq -e.$$

Also, notice that

$$\frac{1}{e^T v} v_i = \frac{v_i}{v_1 + \ldots + v_k} \leq 1 \text{ for all i.}$$

Therefore

$$\hat{v} = \frac{1}{e^T v} v \leq e,$$

and

$$-\hat{v} \geq -e.$$

Since $Au - Bv = 0$, then

$$\frac{1}{e^T u}(Au - Bv) = \frac{1}{e^T u} 0,$$

$$A \frac{1}{e^T u} u - B \frac{1}{e^T v} v = 0,$$

$$A\hat{u} - B\hat{v} = 0,$$

and

$$\frac{1}{e^T u}(-e^T u + e^T v) = 0(\frac{1}{e^T u}).$$

Therefore,

$$-e^T \hat{u} + e^T \hat{v} = 0.$$

Thus, $\hat{u}$ and $\hat{v}$ together form a feasible solution for LP3. Moreover, $0 < e^T \hat{u} + e^T \hat{v} \leq 2n$. Since LP3 is feasible and the objective function is bounded from above, it follows that LP3 has an optimal solution for which the objective function is positive.

Next, if LP3 has an optimal solution u and v for which the objective function is positive, then

$$\begin{array}{rcrcl} Au & - & Bv & = & 0 \\ -e^T u & + & e^T v & = & 0 \end{array}$$

$$u \geq 0, v \geq 0.$$

Since $e^T u + e^T v > 0$, then either u or v has at least one positive entry. But $e^T u = e^T v$ implies that each u and v must have at least one positive entry. Therefore,

$$0 \leq u \neq 0, \text{ and } 0 \leq v \neq 0.$$

Thus system I has a solution. So, $\text{conv}(\mathcal{A}) \cap \text{conv}(\mathcal{B}) \neq \emptyset$ by Proposition 3.1. ∎

Now consider LP3 in the following format

$$\max e^T u + e^T v$$

subject to

$$\begin{array}{rcrcl} Au & - & Bv & = & 0 \\ -e^T u & + & e^T v & = & 0 \\ -u & & & \geq & -e \\ & & -v & \geq & -e \end{array}$$

$$u \geq 0, v \geq 0.$$

We shall now express LP3 in the symmetric dual form. This requires several rewrites of LP3 as follows:

$$\max e^T u + e^T v$$

subject to

$$\begin{array}{rcrcl} Au & - & Bv & \leq & 0 \\ -Au & + & Bv & \leq & 0 \\ -e^T u & + & e^T v & \leq & 0 \\ e^T u & - & e^T v & \leq & 0 \\ u & & & \leq & e \\ & & v & \leq & e \end{array}$$

$$u \geq 0, v \geq 0$$

$$\max e^T u + e^T v$$

subject to

$$\begin{bmatrix} A & -B \\ -A & B \\ -e^T & e^T \\ e^T & -e^T \\ I & 0 \\ 0 & I \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix} \leq \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ e \\ e \end{bmatrix}$$

$$u \geq 0, v \geq 0$$

Hence, the dual of LP3 has the symmetric primal form

$$\min e^T y + e^T w$$

subject to

$$\begin{bmatrix} A^T & -A^T & -e & e & I & 0 \\ -B^T & B^T & e & -e & 0 & I \end{bmatrix} \begin{bmatrix} x \\ z \\ \alpha \\ \beta \\ y \\ w \end{bmatrix} \geq \begin{bmatrix} e \\ e \end{bmatrix}$$

$$x \geq 0, y \geq 0, \alpha \geq 0, \beta \geq 0, z \geq 0, w \geq 0$$

or

$$\min e^T y + e^T w$$

subject to

$$\begin{array}{rcrcccl} A^T(x-z) & - & (\alpha-\beta)e & + & y & & \geq e \\ -B^T(x-z) & + & (\alpha-\beta)e & + & & w & \geq e \end{array}$$

$$x \geq 0, y \geq 0, \alpha \geq 0, \beta \geq 0, z \geq 0, w \geq 0$$

and finally, letting $c = x - z$ and $\gamma = \alpha - \beta$

LP4

$$\min e^T y + e^T w$$

subject to

$$\begin{array}{rcccccl} A^T c & - & \gamma e & + & y & & \geq e \\ -B^T c & + & \gamma e & & & + \ w & \geq e \end{array}$$

$$y \geq 0, w \geq 0$$

We then have the following proposition:

Proposition 3.7 *Conv($\mathcal{A}$) $\cap$ conv($\mathcal{B}$) $= \emptyset$ if and only if LP4 has an optimal solution for which the objective function has value 0.*

PROOF:

It follows from the Fundamental Duality Theorem of Linear Programming and from Proposition 3.6 that: conv($\mathcal{A}$) $\cap$ conv($\mathcal{B}$) $\neq \emptyset$ if and only if LP4 has an optimal solution for which the objective function is positive. (Notice that for any feasible solution $c, \gamma, y,$ and w of LP4 it follows that $e^T y + e^T w \geq 0$.) Thus, conv($\mathcal{A}$) $\cap$ conv($\mathcal{B}$) $= \emptyset$ if and only if LP4 has an optimal solution for which the objective function has value 0. ∎

In addition, the plane $H(c, \gamma) := \{x : c^T x = \gamma\}$ separates the sets $\mathcal{A}$ and $\mathcal{B}$ where (c, γ, y, z) is any solution of LP4. To see this, define

$$H(c, \gamma) = \{x = [x_1 \ldots x_n]^T : c^T x = \gamma\}$$

Note that

$$A^T c \geq e + \gamma e - y \text{ and } B^T c \leq -e + \gamma e + w.$$

But

$$y = w = 0 \text{ if } e^T y + e^T w = 0.$$

Therefore,

$$A^T c \geq e + \gamma e > \gamma e \text{ and } B^T c \leq -e + \gamma e < \gamma e.$$

Hence, the plane $x^T c = \gamma$ separates the sets $\mathcal{A}$ and $\mathcal{B}$. In fact it strictly separates the sets $\mathcal{A}$ and $\mathcal{B}$.

The following two propositions in this section form the foundation of a degeneracy procedure which is used in the algorithm to follow in the next section. They are taken from Mangasarian [5].

Proposition 3.8 *If conv($\mathcal{A}$) $\cap$ conv($\mathcal{B}$) $\neq \emptyset$, then at least one of the systems*

$$\begin{array}{rcl} Au & = & B_{\cdot j} \text{ , for some } j \in \{1, \ldots, k\} \\ e^T u & = & 1 \\ u & \geq & 0 \end{array}$$

or

$$\begin{array}{rcl} Bv & = & A_{\cdot j} \text{ , for some } j \in \{1, \ldots, m\} \\ e^T v & = & 1 \\ v & \geq & 0 \end{array}$$

is inconsistent.

PROOF:

Suppose that all of the above systems are always consistent. If the first system is always consistent then

$$B_{\cdot j} \in \text{ conv}(\mathcal{A}) \ \forall\, j = 1, \ldots, k.$$

Hence,

$$\mathcal{B} \subseteq \text{conv}(\mathcal{A}).$$

Likewise, if the second system is consistent $\forall j = 1, \ldots, m$, then

$$\mathcal{A} \subseteq \text{conv}(\mathcal{B}).$$

But this means that

$$\text{conv}(\mathcal{B}) \subseteq \text{conv}(\mathcal{A}) \text{ and } \text{conv}(\mathcal{A}) \subseteq \text{conv}(\mathcal{B}), i.e.,$$

$\text{conv}(\mathcal{A}) = \text{ conv}(\mathcal{B})$. Pick an $A_{\cdot j} \in \mathcal{A}$ that is an extreme point of $\text{conv}(\mathcal{A})$. Since $A_{\cdot j} = Bv$, $e^T v = 1$ for some $v \geq 0$, and each $B_{\cdot i} \in \text{conv}(\mathcal{A})$, it follows that $A_{\cdot j} = B_{\cdot k}$ for some k, i.e., $v = [0 \ldots 0\ 1\ 0 \ldots 0]^T$ where the nonzero coordinate of v is in the k-th coordinate. This is a contradiction since $\mathcal{A} \cap \mathcal{B} = \emptyset$ in this paper. Hence, at least one of the systems in inconsistent for some j. ∎

Proposition 3.9 *If conv($\mathcal{A}$) ∩ conv($\mathcal{B}$) ≠ ∅, then one of the following systems*

$$\begin{array}{rcrcl} \alpha & - & \beta & > & 0 \\ A^T c & - & \alpha e & \geq & 0 \\ -B^T{}_{\cdot j} c & + & \beta & \geq & 0 \text{ for some } j \end{array}$$
$$-e \leq c \leq e$$

or

$$\begin{array}{rcrcl} \alpha & - & \beta & > & 0 \\ A^T{}_{\cdot j} c & - & \alpha & \geq & 0 \text{ for some } j \\ -B^T c & + & \beta e & \geq & 0 \end{array}$$
$$-e \leq c \leq e$$

is consistent.

PROOF:

Without loss of generality we can assume from Proposition 3.8 that a system of equations of the form

$$\begin{array}{rcl} Au & = & B_{\cdot j} \\ e^T u & = & 1 \end{array}$$
$$u \geq 0$$

which may be written

$$\begin{bmatrix} A \\ e^T \end{bmatrix} u = \begin{bmatrix} B_{\cdot j} \\ 1 \end{bmatrix} \text{ and } u \geq 0$$

has no solution.

From the Farkas Lemma, we know that

$$\begin{bmatrix} A \\ e^T \end{bmatrix}^T y \geq 0, \begin{bmatrix} B_{\cdot j} \\ 1 \end{bmatrix}^T y < 0 \text{ has a solution.}$$

This may be written as

$$\begin{bmatrix} A^T & e \end{bmatrix} \begin{bmatrix} z \\ -\lambda \end{bmatrix} \geq 0, \begin{bmatrix} B^T._j & 1 \end{bmatrix} \begin{bmatrix} z \\ -\lambda \end{bmatrix} < 0, \text{ when } y = \begin{bmatrix} z \\ -\lambda \end{bmatrix}$$

or

$$(1)\ A^T z - \lambda e \geq 0,\ B^T._j z - \lambda < 0 \text{ has a solution.}$$

We know $z \neq 0$, since if $z = 0$, then

$$-\lambda e \geq 0 \text{ and } -\lambda < 0$$

or

$$\lambda e \leq 0 \text{ and } \lambda > 0.$$

But e is a vector of ones, so we obtain

$$\lambda \leq 0 \text{ and } \lambda > 0, \text{ which is a contradiction.}$$

Now set

$$c = \frac{1}{\max \mid z_i \mid} z,\ \alpha = \frac{\lambda}{\max \mid z_i \mid},\ \beta = B^T._j c$$

After dividing each equation in (1) by max$\mid z_i \mid$ and substituting c, and α into (1), we obtain

$$\begin{array}{rcrcl} A^T c & - & \alpha e & \geq & 0 \\ -B^T._j c & + & \alpha & > & 0 \text{ for some j} \end{array}$$

In addition, substituting β for $B^T._j c$ in the last inequality yields $\alpha - \beta > 0$.

Also, note that $\max \mid z_i \mid > 0$ and $\max \mid z_i \mid \geq z_j$. Therefore

$$-e \leq c \leq e.$$

So

$$\begin{array}{rcrcl} \alpha & - & \beta & > & 0 \\ A^T c & - & \alpha e & \geq & 0 \\ -B^T._j c & + & \beta & \geq & 0 \text{ for some j} \end{array}$$
$$-e \leq c \leq e$$

has a solution if (1) has a solution. In a similar fashion we can prove that if

$$\begin{array}{rcl} Bv & = & A._j \text{ for some j} \\ e^T v & = & 1 \\ v & \geq & 0 \end{array}$$

has no solution,

then

$$\begin{array}{rcrcl} \alpha & - & \beta & > & 0 \\ A^T._j c & - & \alpha & \geq & 0 \text{ for some j} \\ -B^T c & + & \beta e & \geq & 0 \end{array}$$
$$-e \leq c \leq e$$

has a solution.

∎

4 The Main Results

The previous linear programs are not guaranteed to form a plane that will partially separate $\mathcal{A}$ from $\mathcal{B}$ when their convex hulls intersect. By imposing a nonzeroness condition on the normal to the separating plane, c, the partial separation of $\mathcal{A}$ and $\mathcal{B}$ can be assured. The following nonconvex linear program can be solved in polynomial time.

<u>LP5</u>

$$\min \; -\alpha + \beta$$

$$\text{subject to}$$

$$\begin{array}{rcrcr} A^T c & - & \alpha e & \geq & 0 \\ -B^T c & + & \beta e & \geq & 0 \\ \| c \|_\infty & & & = & 1 \end{array}$$

Theorem 4.1 *Problem LP5, with rational entries for A and B can be solved in polynomial time by solving the 2n linear programming problems, each of the form found in LP6, for $i = 1, 2, \ldots, n$ and taking the solution with the least $-\alpha + \beta$ among the 2n solutions of LP6.*

<u>LP6</u>

$$\min \; 0^T c - \alpha + \beta$$

$$\textit{subject to}$$

$$\begin{array}{rcrcr} A^T c & - & \alpha e & \geq & 0 \\ -B^T c & + & \beta e & \geq & 0 \\ c & & & \geq & -e \\ c & & & \leq & e \\ c_i & & & = & \pm 1 \end{array}$$

PROOF:

For each $i = 1, \ldots, n$ solve the two linear programming problems
LP6.1

$$\min \; 0^T c - \alpha + \beta$$

$$\text{subject to}$$

$$\begin{array}{rcrcr} A^T c & - & \alpha e & \geq & 0 \\ -B^T c & + & \beta e & \geq & 0 \\ c & & & \geq & -e \\ c & & & \leq & e \\ c_i & & & = & 1 \end{array}$$

and
LP6.2

$$\min \; 0^T c - \alpha + \beta$$

$$\text{subject to}$$

$$\begin{array}{rcrcr} A^T c & - & \alpha e & \geq & 0 \\ -B^T c & + & \beta e & \geq & 0 \\ c & & & \geq & -e \\ c & & & \leq & e \\ c_i & & & = & -1 \end{array}$$

For the two solutions compute the corresponding value of the objective function, $0^T c - \alpha + \beta$. Let c^i, α^i, and β^i denote the solution that yields the smallest value of $-\alpha+\beta$. Next let $\bar{c}, \bar{\alpha}$, and $\bar{\beta}$ be any solution of LP5. Since $\| c^i \|_\infty = 1$, it follows that each c^i, α^i, and β^i is a feasible solution of LP5 and

(1)

$$-\bar{\alpha} + \bar{\beta} \leq \min \{-\alpha^l + \beta^l : l = 1, \ldots, n\}.$$

Since $\| \bar{c} \|_{infty} = 1$, we have that $\exists\, l \in \{1, \ldots, n\}$ such that $\bar{c}_l = \pm 1$ while $-1 \leq \bar{c}_j \leq 1$ for all $j = 1, \ldots, n$. Thus, $\bar{c}, \bar{\alpha}$, and $\bar{\beta}$ is a feasible solution of one of the linear programming problems LP6. In particular,

$$-\alpha^l + \beta^l \leq -\bar{\alpha} + \bar{\beta}.$$

Thus,

(2)

$$\min \{-\alpha^l + \beta^l : l = 1, \ldots, n\} \leq -\bar{\alpha} + \bar{\beta}.$$

Combining (1) and (2) we have that

$$-\bar{\alpha} + \bar{\beta} = \min \{-\alpha^l + \beta^l : l = 1, \ldots, n\}.$$

Therefore, when $\min \{-\alpha^l + \beta^l : l = 1, \ldots, n\} = -\alpha^h + \beta^h$, then c^h, α^h, and β^h solves LP5. Since 2n LP's are needed to compute (c^h, α^h, β^h), and each LP is solvable in polynomial time, then LP5 is solvable in polynomial time by solving the 2n linear programs of LP6. We can now outline an algorithm for discriminating between two disjoint point sets $\mathcal{A}$ and $\mathcal{B}$ represented by the matrices $A \in \mathbb{R}^{m \times n}$ and $B \in \mathbb{R}^{k \times n}$.

<u>**Algorithm 1**</u>

1. Set $j = 0, \mathcal{A}^{0'} = \mathcal{A}, A^0 = A, \mathcal{B}^0 = \mathcal{B}, B^0 = B$, and input an integer jmax. Solve LP2. If the min of LP2 is negative, stop, the plane

$$x^T c = \frac{\alpha + \beta}{2}$$

separates $\mathcal{A}$ and $\mathcal{B}$.

2. For each $i = 1, \ldots, n$ solve LP6.1 and LP6.2. Let c^i, α^i, and β^i denote corresponding solutions of LP6.1, and c^{-i}, α^{-i}, and β^{-i} denote corresponding solutions of LP6.2. For each i compute

$$(i) \text{ cardinality } \{r : A^T{}_{\cdot r} c^i \leq \beta^i\} + \text{cardinality } \{s : B^T{}_{\cdot s} c^i \geq \alpha^i\}$$

Likewise for each i compute

$$(ii) \text{ cardinality } \{r : A^T{}_{\cdot r} c^{-i} \leq \beta^{-i}\} + \text{cardinality } \{s : B^T{}_{\cdot s} c^{-i} \geq \alpha^{-i}\}.$$

Define $i(j)$ to be the minimum of the 2n values in (i) and (ii), and let $c^{i(j)}, \alpha^{i(j)}$, and $\beta^{i(j)}$ be a solution of the corresponding LP6 that yields $i(j)$.

<u>Comment</u> This step picks the LP6.1 or LP6.2 for which the closed set between the parallel planes $c^{i(j)^T} x = \alpha^{i(j)}$ and $c^{i(j)^T} x = \beta^{i(j)}$ contains the least number of points from $\mathcal{A}^j$ and $\mathcal{B}^j$, while the open half spaces outside this region separate the remaining portions of A^j and B^j.

3. Let

$$\mathcal{A}^{j+1} = \{A_{\cdot r} \in \mathcal{A}^j \mid A^T{}_{\cdot r}c^{i(j)} \leq \beta^{i(j)}\}$$

and

$$\mathcal{B}^{j+1} = \{B_{\cdot s} \in \mathcal{B}^j \mid B^T{}_{\cdot s}c^{i(j)} \geq \alpha^{i(j)}\}$$

If $\mathcal{A}^{j+1} \neq \mathcal{A}^j$ or $\mathcal{B}^{j+1} \neq \mathcal{B}^j$ go to (5).

4. Degeneracy Procedure: Find a column $A_{\cdot r}$ of A^j (case a) or column $B_{\cdot s}$ or B^j (case b) such that when LP2 is solved with $A = A_{\cdot r}$ and $B = B^j$ (case a) or $A = A^j$ and $B = B_{\cdot s}$ (case b) the minimum of LP2 is negative. In either case denote the solution of the LP by $(\bar{c}, \bar{\alpha}, \bar{\beta})$.

   Case a: Define $c^{i(j)} = \bar{c}$, $\alpha^{i(j)} = -\infty$, $\beta^{i(j)} = \bar{\beta}$, $\mathcal{B}^{j+i} = \mathcal{B}^j$,
   $\mathcal{A}^{j+1} = \{A_i \mid A_i \in \mathcal{A}^j,\ A_i^T\bar{c} \leq \bar{\beta}\}$,

   Case b: Define $c^{i(j)} = \bar{c}$, $\alpha^{i(j)} = \bar{\alpha}$, $\beta^{i(j)} = \infty, \mathcal{A}^{j+1} = \mathcal{A}^j, \mathcal{B}^{j+1} = \{B_i \mid B_i \in \mathcal{B}^j, B_i^T c \geq \bar{\alpha}\}$

   Comment: This degeneracy procedure eliminates at least one point from $\mathcal{A}^j$ or $\mathcal{B}^j$ and thus ensures that either $\mathcal{A}^{j+1} \neq \mathcal{A}^j$ or $\mathcal{B}^{j+1} \neq \mathcal{B}^j$. It is based on Propositions 3.8 and 3.9.

5. Save the planes

$$x^T c^{i(j)} = \alpha^{i(j)} \text{ and } x^T c^{i(j)} = \beta^{i(j)}$$

6. If $\mathcal{A}^{j+1} = \mathcal{B}^{j+1} = \emptyset$, replace jmax by j and stop. If $j = jmax$ stop, else increment j by 1 and go to (2).

When $\mathcal{A}$ and $\mathcal{B}$ are not linearly separable the previous algorithm constructs a sequence of parallel planes:

$$x^T c^{i(j)} = \beta^{i(j)}, x^T c^{i(j)} = \alpha^{i(j)}, j = 0, \ldots, jmax$$

such that if jmax is sufficiently large, the sets $\mathcal{A}$ and $\mathcal{B}$ are separated by the following procedure.

PROCEDURE: Set $j = 0$, input jmax and a given pattern $x^T \in \mathbb{R}^n$.

1. If $j = jmax$, go to (4).
2. If $x^T c^{i(j)} > \beta^{i(j)}$, then $x \in \mathcal{A}$, stop.
   If $x^T c^{i(j)} < \alpha^{i(j)}$, then $x \in \mathcal{B}$, stop.
3. Increment j by 1 and go to (1).
4. If $x^T c^{i(j)} \geq \frac{\alpha^{i(j)}+\beta^{i(j)}}{2}$, the $x \in \mathcal{A}$, stop.
   If $x^T c^{i(j)} < \frac{\alpha^{i(j)}+\beta^{i(j)}}{2}$, then $x \in \mathcal{B}$, stop.

The following sets were separated, with the help of the linear programming software package Lindo, using the previously described algorithm.

$$\mathcal{A} = \{[1\ 3]^T, [2\ 7]^T, [3\ 4]^T, [3\ 10]^T, [6\ 5]^T\}$$

Therefore,

$$A = \begin{bmatrix} 1 & 2 & 3 & 3 & 6 \\ 3 & 7 & 4 & 10 & 5 \end{bmatrix}$$

$$\mathcal{B} = \{[4\ 7]^T, [5\ 10]^T, [9\ 8]^T\}$$

Therefore,

$$B = \begin{bmatrix} 4 & 5 & 9 \\ 7 & 10 & 8 \end{bmatrix}$$

Let

$$\hat{A} = \{A^T{}_{\cdot r} \mid A^T{}_{\cdot r} c^{\pm i} \leq \beta^{\pm i}\}$$

and

$$\hat{B} = \{B^T{}_{\cdot s} \mid B^T{}_{\cdot s} c^{\pm i} \geq \alpha^{\pm i}\}$$

(2) $c_1 = 1$

$\alpha = -.33333$
$\beta = 6.33333$ $\quad \hat{A} = \{A_{1\cdot}, A_{2\cdot}, A_{3\cdot}, A_{4\cdot}, A_{5\cdot}\}$
$c_1 = 1$ $\quad \hat{B} = \{B_{1\cdot}, B_{2\cdot}, B_{3\cdot}\}$
$c_2 = -.33333$

$c_1 = -1$

$\alpha = -8$
$\beta = -7.5$ $\quad \hat{A} = \{A_{2\cdot}, A_{3\cdot}, A_{4\cdot}, A_{5\cdot}\}$
$c_1 = -1$ $\quad \hat{B} = \emptyset$
$c_2 = -.5$

$c_2 = 1$

$\alpha = 2.8$
$\beta = 9$ $\quad \hat{A} = \{A_{1\cdot}, A_{2\cdot}, A_{3\cdot}, A_{5\cdot}\}$
$c_1 = -.2$ $\quad \hat{B} = \{B_{1\cdot}, B_{2\cdot}, B_{3\cdot}\}$
$c_2 = 1$

$c_2 = -1$

$\alpha = -13$
$\beta = -11$ $\quad \hat{A} = \{A_{4\cdot}, A_{5\cdot}\}$
$c_1 = -1$ $\quad \hat{B} = \{B_{1\cdot}\}$
$c_2 = -1$

Therefore, $(c^{i(j)}, \alpha^{i(j)}, \beta^{i(j)}) = (\begin{bmatrix} -1 \\ -1 \end{bmatrix}, -13, -11)$

(3) Now

$$A = \begin{bmatrix} 3 & 6 \\ 10 & 5 \end{bmatrix} \text{ and } B = \begin{bmatrix} 4 \\ 7 \end{bmatrix}$$

(4) Save the planes

$$-x_1 - x_2 = -13 \text{ and } -x_1 - x_2 = -11$$

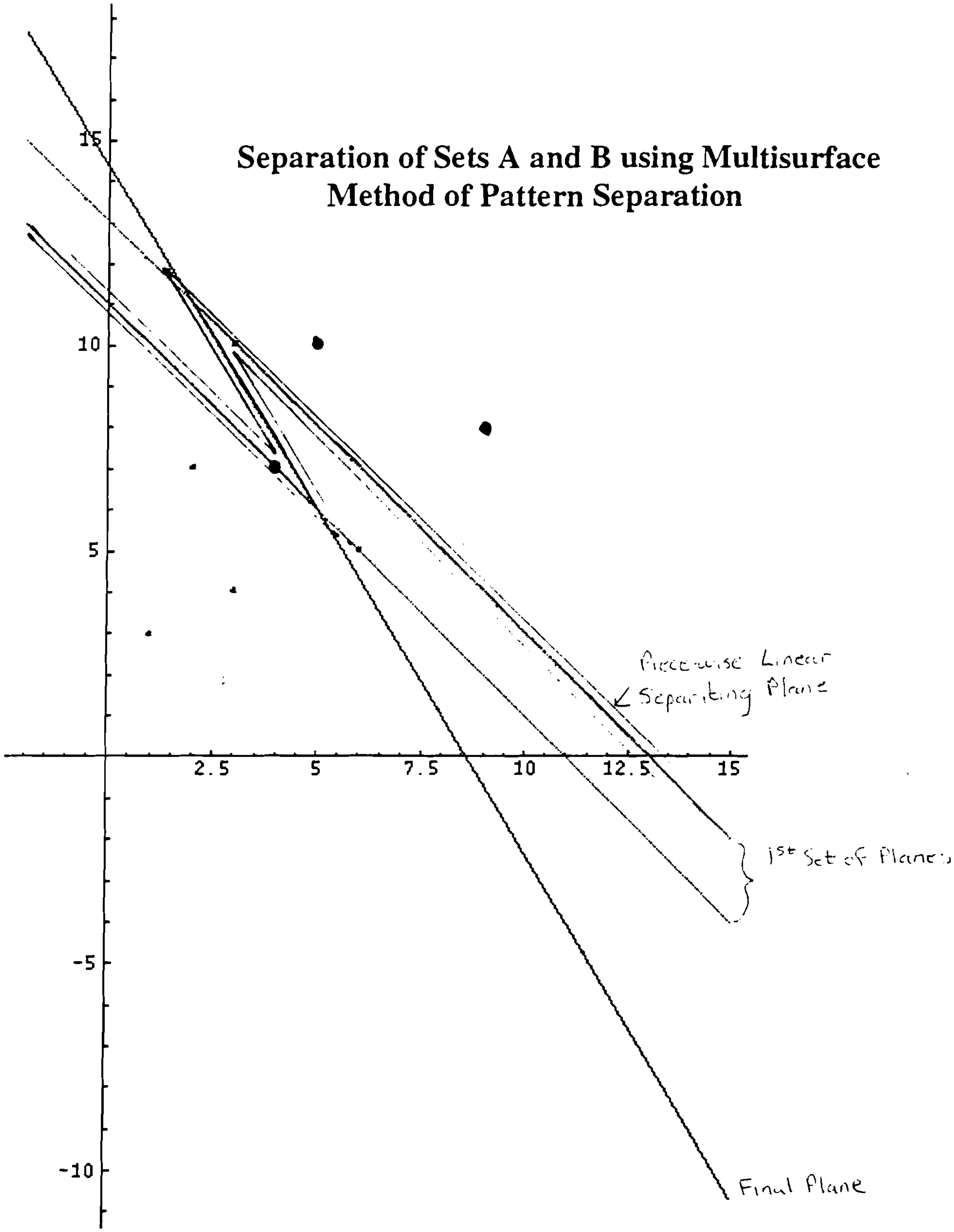
(1)

$$\text{min of LP2} = -.8 \text{ with}$$

$$\begin{aligned} \alpha &= 9 \\ \beta &= 8.2 \\ c_1 &= 1 \\ c_2 &= .6 \end{aligned}$$

Therefore, the plane $x_1 + .6x_2 = 8.6$ separates the previous A and B.

Separation of Sets A and B using Multisurface Method of Pattern Separation



5 Applications

This method of pattern separation is currently in use at the University of Wisconsin Hospitals to aid in the diagnosis of breast cancer. A fine needle aspirate is taken from a patient and nine measurements are made. These nine measurements are: clump thickness, size uniformity, marginal adhesion, cell size, shape uniformity, bare nuclei, bland chromatin, normal nucleoli and mitosis. Each of the nine measurements is designated by an integer between 1 and 10, with larger numbers indicating a greater likelihood of malignancy. The discriminant function constructed, which is based on algorithm 1, has been applied to 369 points, 168 of which come from patients with confirmed malignancy. Algorithm 1 generated four pairs of parallel planes which completely separated the benign samples from the malignant ones. The resulting discriminant function can instantly classify any sample point given to it in $\mathbb{R}^9$. So far, 45 new sample points have been encountered since the construction of the discriminant function, and all were classified correctly [1].

Another possible application of this method is in the acceptance of students to college. Schools must accept many more students than they have space for since many of those students will attend school elsewhere. The difficulty lies in determining the yield of students. Students decide which college to attend based on many factors, including: location, academic reputation, cost, financial aid and size. Each of these variables, and several others, could be assigned a value and a discriminant function could be constructed which separates students based upon the likelihood they would attend a particular college. This would aid the admissions office in deciding who, and how many students to accept.

References

[1] O.L. Mangasarian, R. Setiono, and W. H. Wolberg, *Pattern recognition via Linear Programming: Theory and Application to Medical Diagnosis*, Large - Scale Numerical Optimization, (1990), pp. 22-30.

[2] O.L. Mangasarian, *Linear and Nonlinear separation of Patterns by Linear Programming*, Operations Research, 13, (1965), pp. 444-452.

[3] M.W. Jeter, *Mathematical Programming: An Introduction to Optimization*, (1986), Marcel Dekker, Inc..

[4] R.W. Cottle and R.E. Stone, *On the Uniqueness of Solutions to Linear Complementarity Problems*, Mathematical Programming, vol. 27, no. 2, (1983), pp. 191-213, North-Holland, Amsterdam.

[5] O.L. Mangasarian, *Multisurface Method of Pattern Separation*, IEEE Transactions on Information Theory, IT-14, 6, (1968), pp. 801-807.

This research was funded by a grant from the National Science Foundation. In addition, I would like to thank Dr. Faiz Al-Khayyal (Georgia Institute of Technology) and Dr. Melvyn Jeter (Illinois Wesleyan University) for their help with this research.