PATTERN RECOGNITION VIA LINEAR PROGRAMMING : THEORY AND APPLICATION TO MEDICAL DIAGNOSIS

O.L. MANGASARIAN†, R. SETIONO † AND W.H. WOLBERG ‡

REVISED MAY 18, 1990

Abstract. A decision problem associated with a fundamental nonconvex model for linearly inseparable pattern sets is shown to be NP-complete. Another nonconvex model that employs an $\infty-$ norm instead of the 2-norm, can be solved in polynomial time by solving $2n$ linear programs, where n is the (usually small) dimensionality of the pattern space. An effective LP-based finite algorithm is proposed for solving the latter model. The algorithm is employed to obtain a nonconvex piecewise-linear function for separating points representing measurements made on fine needle aspirates taken from benign and malignant human breasts. A computer program trained on 369 samples has correctly diagnosed each of 45 new samples encountered and is currently in use at the University of Wisconsin Hospitals.

1. Introduction. The fundamental problem we wish to address is that of distinguishing between elements of two distinct pattern sets. Mathematically we can formulate the problem as follows. Given two disjoint finite points sets $\mathcal{A}$ and $\mathcal{B}$ in the n-dimensional real space R^n, construct a discriminant function f, from R^n into the real line R, such that $f(\mathcal{A}) > 0$ and $f(\mathcal{B}) \leq 0$. When the convex hulls of the two point sets $\mathcal{A}$ and $\mathcal{B}$ do not intersect, a single linear program [6,7,9,2,3] can be used to obtain a linear discriminant function of the following type

$$f(x) := cx + \gamma \tag{1.1}$$

where c is in R^n and γ is in R. Unfortunately in many real-life problems the convex hulls of the sets $\mathcal{A}$ and $\mathcal{B}$ intersect and one must resort to a more complex discriminant function, such as a piecewise-linear function which is usually nonconvex.

†Computer Sciences Department, University of Wisconsin, 1210 West Dayton Street, Madison, WI 53706. Research of these authors supported by National Science Foundation Grants DCR-8521228 and CCR-8723091 and Air Force Office of Scientific Research Grants AFOSR-86-0172 and AFOSR 89-0410.

‡Departments of Surgery and Human Oncology, Clinical Sciences Center, University of Wisconsin, Madison, WI 53792.

In the multisurface method of [7] it was shown how to construct such a function by solving a sequence of nonlinear programs, each containing a single nonconvex constraint. A gradient approach for solving the same problem is given in Takayima [10]. One purpose of this paper is to show that a decision problem associated with the nonlinear program of the multisurface method is NP-complete (Theorem 3.1). Another purpose is to show that by a change from a $2-$norm to an ∞-norm, the nonconvex program can be solved in polynomial time by solving $2n$ linear programs (Theorem 3.2). Solution of a sequence of these ∞-norm nonconvex programs leads to an efficient algorithm for obtaining a piecewise-linear discriminant function that will separate two disjoint point sets, regardless of whether their convex hulls intersect or not. When applied to the diagnosis of breast cancer the method completely separated 369 points in R^9 into 201 points belonging to a benign category and 168 points belonging to a malignant category (Section 4). Other methods [11] had failed to achieve such complete separation on the same set of samples. By contrast, our discriminant function which consists of 4 pairs of parallel planes defined on R^9, not only achieved complete separation of these sample sets, but also correctly classified each of 45 new points subsequently obtained.

It is worthwhile to point out here, that for the linearly inseparable case (i.e. the case of intersecting convex hulls), solution of any of the single linear programs proposed in [6], [9] or [2,3] may not provide any useful information. We will demonstrate this by means of small examples and by citing computational experience with the medical diagnosis problem.

A brief word about the notation employed. For vectors x and y in the n-dimensional real space R^n, xy will denote the scalar product, $\|x\|_p$ will denote the p-norm, $\left(\sum_{i=1}^n |x_i|^p\right)^{\frac{1}{p}}$ for positive p with $\|x\|_\infty := max_{1\le i\le n} |x_i|$. For an $m \times n$ real matrix signified by $A \in R^{m\times n}$, A_i denotes the i^{th} row, while A^t will denote the transpose. A vector of ones of any dimension will be denoted by e. For an optimization problem $min_{x\in X} f(x)$, the set of its solutions will be denoted by $arg\, min_{x\in X} f(x)$. Cardinality of a set denotes the number of elements in it.

2. Linearly separable pattern sets. Our primary concern is how to discriminate between two disjoint point sets $\mathcal{A}$ and $\mathcal{B}$ in R^n. We shall represent the sets $\mathcal{A}$ and $\mathcal{B}$ by the matrices $A \in R^{m\times n}$ and $B \in R^{k\times n}$ and we begin with the following evident lemma.

LEMMA 2.1. *The convex hulls of the sets $\mathcal{A}$ and $\mathcal{B}$ are disjoint if and only if there exists* **no** *$u \in R^m$ and $v \in R^k$ such that*

$$uA - vB = 0, \quad -ue + ve = 0, \quad 0 \ne \begin{pmatrix} u \\ v \end{pmatrix} \ge 0 \tag{2.1}$$

The dual of a linear program that attempts to solve the system (2.1) generates a plane that separates the sets $\mathcal{A}$ and $\mathcal{B}$ when their convex hulls do not intersect. For example, the system (2.1) having a solution is equivalent to the linear program

$$\max_{u,v,r,s} \left\{ -(r+s)e \mid uA - vB + r - s = 0, -ue = -1, ve = 1, (u,v,r,s) \ge 0 \right\} \tag{2.2}$$

having a zero maximum.

The dual of (2.2) leads to the following linear separability criterion for the pattern sets $\mathcal{A}$ and $\mathcal{B}$ proposed in [6].

THEOREM 2.1. *[6] The convex hulls of the sets $\mathcal{A}$ and $\mathcal{B}$ are disjoint if and only if the linear program*

$$\min_{c,\alpha,\beta}\Big\{-\alpha+\beta \;\Big|\; Ac-e\alpha\geq 0, -Bc+e\beta\geq 0, e\geq c\geq -e\Big\} \tag{2.3}$$

has a negative minimum in which case the plane $xc=\frac{\alpha+\beta}{2}$ separates the sets $\mathcal{A}$ and $\mathcal{B}$, where (c,α,β) is any solution of (2.3). Thus

$$\begin{cases} Ac\geq e\alpha > e\frac{(\alpha+\beta)}{2} \\ Bc\leq e\beta < e\frac{(\alpha+\beta)}{2} \end{cases} \tag{2.4}$$

By considering the dual of another linear program that attempts to solve (2.1), we are led to the linear separability criterion of Smith [9]. Again, it is easy to see that (2.1) having a solution is equivalent to the linear program

$$\max_{u,v}\Big\{ue+ve \;\Big|\; uA-vB=0, -ue+ve=0, e\geq u\geq 0, e\geq v\geq 0\Big\} \tag{2.5}$$

having a positive maximum. By considering the equivalence of (2.1) having no solution and the dual of (2.5) having a zero minimum, we obtain the following.

THEOREM 2.2. *[9] The convex hulls of the sets $\mathcal{A}$ and $\mathcal{B}$ are disjoint if and only if the linear program*

$$\min_{c,\gamma,y,z}\Big\{ey+ez \;\Big|\; Ac-e\gamma+y\geq e, -Bc+e\gamma+z\geq e, (y,z)\geq 0\Big\} \tag{2.6}$$

has a zero minimum in which case the plane $xc=\gamma$ separates the sets $\mathcal{A}$ and $\mathcal{B}$, where (c,γ,y,z) is any solution of (2.6).

Finally consider the following less obvious linear program that attempts to solve (2.1):

$$\max_{u,v,\eta}\left\{\eta(m+k) \;\middle|\; \begin{array}{c} uA-vB+\eta(eA-eB)=0, -ue+ve+\eta(-m+k)=0, \\ -ue-ve=-1, (u,v)\geq 0 \end{array}\right\} \tag{2.7}$$

Since $(\eta=0,u,v)$ is feasible for (2.7) for some $u\geq 0, v\geq 0$ such that $eu=ev=\frac{1}{2}$, when the convex hulls of $\mathcal{A}$ and $\mathcal{B}$ intersect, it follows that the maximum of (2.7) is nonnegative for this case. Conversely, if the maximum of (2.7) is nonnegative, then $\left(\frac{u+e\eta}{eu+m\eta}A\right)=\left(\frac{v+e\eta}{ev+k\eta}B\right)$ is a point in the intersection of the convex hulls of $\mathcal{A}$ and $\mathcal{B}$. Employing this equivalence of the nonnegativity of the maximum of (2.7) and the nonemptiness of the intersection of the convex hulls of $\mathcal{A}$ and $\mathcal{B}$, we obtain Grinold's [2,3] separability criterion when we consider (2.8) below, the dual of (2.7).

THEOREM 2.3. *[2,3] The convex hulls of the sets $\mathcal{A}$ and $\mathcal{B}$ are disjoint if and only if the linear program*

$$\min_{c,\gamma,\rho}\left\{-\rho \;\middle|\; \begin{array}{c} Ac - e\gamma - e\rho \geq 0, -Bc + e\gamma - e\rho \geq 0, \\ (eA - eB)c + (-m+k)\gamma = m + k \end{array}\right\} \tag{2.8}$$

has a negative minimum, in which case the plane $xc = \gamma$ separates the sets $\mathcal{A}$ and $\mathcal{B}$, where (c, γ, ρ) is any solution of (2.8)

Our purpose in describing three different linear programs, (2.3), (2.6) and (2.8), each of which generates a separating plane for sets with disjoint convex hulls, is to point out the fact that contrary to previous claims, none of these linear programs can be guaranteed to generate by itself a useful plane for the case when the convex hulls of the sets $\mathcal{A}$ and $\mathcal{B}$ intersect. We note first that for the linearly inseparable case, the linear program (2.3) is solved by $c = 0, \alpha = 0, \beta = 0$. This provides no useful information for a plane that can possibly be utilized to separate subsets of $\mathcal{A}$ and $\mathcal{B}$ from each other as required for example in the multisurface method [7]. In [9], Smith claimed for the linear program (2.6), "the capability to compute weights for the nonseparable as well as separable pattern sets". This unfortunately may not always be true as can be seen from the simple example

$$A = \begin{bmatrix} 1 \\ 2 \end{bmatrix}, B = \begin{bmatrix} -1 \\ 0 \\ 4 \end{bmatrix} \tag{2.9}$$

for which Smith's LP (2.6) is solved by $c = 0, \gamma = 1, y_1 = y_2 = 2$ and $z_1 = z_2 = z_3 = 0$. The solution $c = 0$ is unique for this problem, and hence no useful plane is generated that minimizes some error criterion.

Similarly, Grinold stated [2, Theorem (iii)] that if for a solution (c, γ, ρ) of the linear program (2.8), $\rho \leq 0$, then (c, γ) "defines a hyperplane that minimizes the maximum error". Again this is not true in general as can be seen from the example

$$A = \begin{bmatrix} 1 \\ 3 \end{bmatrix}, B = \begin{bmatrix} \frac{1}{2} \\ 2 \\ 4 \end{bmatrix} \tag{2.10}$$

for which Grinold's LP (2.8) is solved by $c = 0, \gamma = 5, \rho = -5$. Again the solution component $c = 0$ is unique, and hence it cannot provide a hyperplane that minimizes a maximum error.

It is therefore important to have a scheme capable of generating a plane which, in the least, provides partial separation for the linearly inseparable case. This enables us to produce an algorithm for the construction of a discriminant function for this case. It seems essential then to introduce some condition which ensures that $c \neq 0$, as was done in [7]. This leads to a problem with a single nonconvex constraint which we shall discuss in the next section. Nonconvexity appears to be the inevitable price paid for a method that is guaranteed to handle the linearly inseparable case.

3. Linearly inseparable pattern sets. When the convex hulls of the sets $\mathcal{A}$ and $\mathcal{B}$ intersect none of the linear programs of Section 2 are guaranteed to generate a plane that partially separates $\mathcal{A}$ from $\mathcal{B}$. To ensure the generation of such a plane, we impose a nonzeroness condition on the vector c, the normal to the separating plane. In [7] this was done in conjunction with the linear program (2.3) by imposing the condition $\|c\|_2^2 \geq \delta$ for some $\delta \in (0, n]$. This condition ensured the generation of 2 planes for the linearly inseparable case, $xc = \alpha$ and $xc = \beta$, with minimum distance apart, and such that the intersection of the convex hulls of $\mathcal{A}$ and $\mathcal{B}$ is contained in the closed set between the 2 parallel planes. By repeating this procedure for the subsets of $\mathcal{A}$ and $\mathcal{B}$ which are contained between and on the 2 planes, a piecewise-linear discriminant function is obtained for separating the sets $\mathcal{A}$ and $\mathcal{B}$ [7]. Because a decision problem associated with the linear program (2.3) augmented by a nonconvex constraint $\|c\|_2^2 \geq n$ is NP-complete (Theorem 3.1 below), we shall introduce another means of imposing nonzeroness on c, namely by means of the nonconvex constraint $\|c\|_\infty = 1$. We will show that such a problem can be solved in polynomial time by solving $2n$ linear programs (Theorem 3.2 below). Because in many applications, n is relatively small compared to m and k, this is a viable formulation for generating a piecewise-linear discriminant function (Algorithm 3.3 below).

We begin by showing the NP-completeness of a decision problem associated with the linear program (2.3) augmented by the 2-norm condition $\|c\|_2^2 \geq n$ as follows

$$\min_{c,\alpha,\beta}\left\{-\alpha + \beta \mid Ac - e\alpha \geq 0, -Bc + e\beta \geq 0, e \geq c \geq -e, \|c\|_2^2 \geq n\right\} \tag{3.1}$$

THEOREM 3.1. *The following decision problem associated with the nonconvex program (3.1) with rational entries for A and B is NP-complete:*

$$\text{(3.2)}\quad \textit{Is}\ \min_{c,\alpha,\beta}\left\{-\alpha + \beta \mid Ac - e\alpha \geq 0, -Bc + e\beta \geq 0, e \geq c \geq -e, \|c\|_2^2 \geq n\right\} \leq 0?$$

Proof. We first note that the constraints $\|c\|_\infty \leq 1$ and $\|c\|_2^2 \geq n$ of (3.2) are equivalent to c being one of the vertices of the cube $\{c \mid \|c\|_\infty = 1\}$. Therefore problem (3.2) is in NP, because a correct guess of a vertex of the cube that gives the minimum value of $-\alpha + \beta$ will answer the question of (3.2) in polynomial time by evaluating $\alpha - \beta$, where

$$\alpha = \min_{1 \leq i \leq m} A_i c, \quad \beta = \max_{1 \leq i \leq k} B_i c$$

and c is an optimum vertex. We now show that (3.2) is NP-hard by reducing to it the partition problem [1] :

$$\text{(3.3)} \qquad \textit{Is } dx = 0 \textit{ for some } x \in R^n \textit{ such that } e \geq x \geq -e, \|x\|_2^2 \geq n?$$

where $(d_1, d_2, \ldots, d_n)$, the components of d, are given positive integers. The partition problem (3.3) is the following instance of (3.2) :

$$\textit{Is } \min_{\alpha,\beta,x}\left\{-\alpha+\beta \mid \begin{pmatrix} d \\ -d \end{pmatrix} x \geq \begin{pmatrix} \alpha \\ \alpha \end{pmatrix}, \begin{pmatrix} -d \\ d \end{pmatrix} x \leq \begin{pmatrix} \beta \\ \beta \end{pmatrix}, e \geq x \geq -e, \|x\|_2^2 \geq n, \right\} \leq 0?$$

If the minimum is less than or equal to zero, then since $\alpha \leq 0$ and $\beta \geq 0$ it follows that $\alpha = 0$, $\beta = 0$ and thus $dx = 0$. Hence (3.2) is NP-hard. Since it is also in NP, it is NP-complete. ■

We show now that if the nonzeroness condition $\|c\|_\infty = 1$ is imposed on the linear program (2.3) instead of $\|c\|_2^2 \geq n$, then the resulting nonconvex program

$$\text{(3.4)} \qquad \min_{c,\alpha,\beta}\left\{-\alpha+\beta \mid Ac - e\alpha \geq 0, -Bc + e\beta \geq 0, \|c\|_\infty = 1\right\}$$

can be solved in polynomial time.

THEOREM 3.2. *The nonconvex program (3.4) with rational entries for A and B can be solved in polynomial time by solving the $2n$ linear programs for $i = 1, 2, \ldots, n$:*

$$\text{(3.5)} \qquad \min_{c,\alpha,\beta}\left\{-\alpha+\beta \mid Ac - e\alpha \geq 0, -Bc + e\beta \geq 0, e \geq c \geq -e, c_i = \pm 1\right\}$$

and taking the solution with the least $-\alpha+\beta$ among the $2n$ solutions of (3.5) with $i = 1, 2, \ldots, n$.

Proof. Let

$$\text{(3.6)} \qquad \begin{array}{l} (c^i, \alpha^i, \beta^i) \in \\ \qquad \arg\min_{c,\alpha,\beta}\left\{-\alpha+\beta \mid Ac - e\alpha \geq 0, -Bc + e\beta \geq 0, e \geq c \geq -e, |c_i| = 1\right\} \end{array}$$

Note that for a fixed i, (c^i, α^i, β^i) can be obtained by solving the two LP's of (3.5) and taking the solution with the lower value of $-\alpha+\beta$. Let $(\bar{c}, \bar{\alpha}, \bar{\beta})$ be a solution of (3.4). Since $\|c^i\|_\infty = 1$ it follows that

$$\text{(3.7)} \qquad -\bar{\alpha} + \bar{\beta} \leq \min_{1 \leq i \leq n} -\alpha^i + \beta^i$$

Since $|\bar{c}_l| = 1$ for some l and $-1 \leq \bar{c}_{i \neq l} \leq 1$, it follows that

$$\text{(3.8)} \qquad -\alpha^l + \beta^l \leq -\bar{\alpha} + \bar{\beta}$$

Combining (3.7) and (3.8) gives

$$-\bar{\alpha}+\bar{\beta} \leq \min_{1\leq i\leq n} -\alpha^i+\beta^i \leq -\alpha^l+\beta^l \leq -\bar{\alpha}+\bar{\beta} \tag{3.9}$$

Hence (c^l, α^l, β^l) solves (3.4). Since $2n$ LP's are needed to compute (c^l, α^l, β^l), and each LP is solvable in polynomial time [5,4], it follows that (3.4) is solvable in polynomial time by solving the $2n$ linear programs of (3.5). ■

We outline now an algorithm based on Theorem (3.2) for discriminating between two disjoint point sets $\mathcal{A}$ and $\mathcal{B}$ represented by the matrices $A \in R^{m\times n}$ and $B \in R^{k\times n}$.

ALGORITHM 3.3.

(0) *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 the linear program (2.3). If the minimum of (2.3) is negative, stop, the plane $xc=\frac{\alpha+\beta}{2}$ separates $\mathcal{A}$ and $\mathcal{B}$.*

(i) *Solve the $2n$ LP's of (3.5) with $A=A^j$ and $B=B^j$. Let $(c^{\pm i}, \alpha^{\pm i}, \beta^{\pm i})$ denote the solution of each LP corresponding to $c_i=\pm 1$ and define*

$$i(j) = arg \min_{1\leq i\leq n} \Big(cardinality\big\{ r \;\big|\; A_r c^{\pm i} \leq \beta^{\pm i} \big\} \\ + cardinality\big\{ s \;\big|\; B_s c^{\pm i} \geq \alpha^{\pm i} \big\} \Big)$$

and let $(c^{i(j)}, \alpha^{i(j)}, \beta^{i(j)})$ be a solution of one of the pair of LP's (3.5) corresponding to $i(j)$.

COMMENT. This step picks that LP of (3.5) for which the closed set between the parallel planes $xc^{i(j)}=\alpha^{i(j)}$ and $xc^{i(j)}=\beta^{i(j)}$ contains the least number of points from both $\mathcal{A}^j$ and $\mathcal{B}^j$, while the remaining open half spaces outside this closed set contain the remaining separated portions of $\mathcal{A}^j$ and $\mathcal{B}^j$.

(ii) *Let*

$$\mathcal{A}^{j+1} = \{A_r \in \mathcal{A}^j | A_r c^{i(j)} \leq \beta^{i(j)}\}$$
$$\mathcal{B}^{j+1} = \{B_s \in \mathcal{B}^j | B_s c^{i(j)} \geq \alpha^{i(j)}\}$$

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

(iii) *Degeneracy Procedure : Find a row A_r of A^j (Case a) or row B_s of B^j (case b) such that when the LP (2.3) is solved with $A=A_r$ and $B=B^j$ (case a) or $A=A^j$ and $B=B_s$ (case b), the minimum of the LP is negative. In either case denote the solution of the LP by $(\bar{c}, \bar{\alpha}, \bar{\beta})$.*

Case a : Define $c^{i(j)}=\bar{c}, \alpha^{i(j)}=-\infty, \beta^{i(j)}=\bar{\beta}$

$$\mathcal{A}^{j+1} := \big\{ A_i \;\big|\; A_i \in \mathcal{A}^j, A_i\bar{c} \leq \bar{\beta} \big\}, \quad \mathcal{B}^{j+1} := \mathcal{B}^j$$

Case b : Define $c^{i(j)}=\bar{c}, \alpha^{i(j)}=\bar{\alpha}, \beta^{i(j)}=\infty$

$$\mathcal{A}^{j+1} := \mathcal{A}^j, \quad \mathcal{B}^{j+1} := \big\{ B_i \;\big|\; B_i \in \mathcal{B}^j, B_i\bar{c} \geq \bar{\alpha} \big\}$$

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 [7, Lemma 2.11].

(iv) Save the planes

$$xc^{i(j)} = \alpha^{i(j)} \text{ and } xc^{i(j)} = \beta^{i(j)}$$

(v) 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 (i). ■

When $\mathcal{A}$ and $\mathcal{B}$ are not linearly separable, Algorithm 3.3 constructs a sequence of parallel planes :

$$xc^{i(j)} = \beta^{i(j)}, \quad xc^{i(j)} = \alpha^{i(j)}, \qquad j = 0, \ldots, jmax$$

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

PROCEDURE 3.4. *Set $j = 0$, input $jmax$ and a given pattern $x \in R^n$.*

(i) If $j = jmax$ go to (iv).

(ii) If $xc^{i(j)} > \beta^{i(j)}$ then $x \in \mathcal{A}$, stop.
If $xc^{i(j)} < \alpha^{i(j)}$ then $x \in \mathcal{B}$, stop.

(iii) Increment j by 1 and go to (i).

(iv) If $xc^{i(j)} \geq \frac{\alpha^{i(j)}+\beta^{i(j)}}{2}$ then $x \in \mathcal{A}$, stop
If $xc^{i(j)} < \frac{\alpha^{i(j)}+\beta^{i(j)}}{2}$ then $x \in \mathcal{B}$, stop

In most real problems that have been run using Algorithm 3.3, the degeneracy procedure was not required. In fact if we make the following assumption, then the Degeneracy Procedure (iii) in Algorithm 3.3 is not needed.

ASSUMPTION 3.5. *If the disjoint sets $\mathcal{A}$ and $\mathcal{B}$ are linearly inseparable, then for at least one of the $2n$ linear programs of (3.5), $i = 1, 2, \ldots, n$ whose solution is denoted by (c^i, α^i, β^i),*

$$\{A_l \mid A_l c^i > \beta^i\} \cup \{B_l \mid B_l c^i < \alpha^i\} \neq \emptyset$$

It is possible however to concoct examples that violate this assumption. Gary Schultz provided one such example. The set $\mathcal{A}$ consists of the 4 vertices of the unit square in R^2 and the set $\mathcal{B}$ consists of 8 points on the edges of the square, symmetrically situated, each at distance 0.1 from a vertex.

4. An application to medical diagnosis. We describe now how the ideas of the previous section were applied to generate a discriminant function, for the diagnosis of breast cancer [11], which is currently in use at the University of Wisconsin Hospitals. The discriminant function constructed is based on Algorithm 3.3 applied to 369 points in the nine-dimensional real space R^9. Each point represents nine measurements made on a fine needle aspirate (fna) taken from a patient's breast. These nine measurements are : clump thickness, size uniformity, shape uniformity, marginal adhesion, cell size, 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. Of the 369 points, 201 came from patients with no breast malignancy, while 168 came from patients with confirmed malignancy. By using Algorithm 3.3 we were able to construct four pairs of parallel planes which completely separated the benign samples from the malignant ones. The first pair of planes classified all but 88 of the samples, the second pair left 45 unclassified points, the third pair left 16 unclassified points and the fourth pair classified all the remaining points. The resulting discriminant function has been placed on a personal computer diskette and can instantly classify any sample point given to it in R^9. All 45 new sample points encountered after the construction of the discriminant function were classified correctly, with 44 of the points being classified by the first pair of parallel planes and the 45th point by the second pair of parallel planes.

In contrast to our approach, statistically based schemes [12] failed to obtain complete separation of the 369 sample points. Similarly, the linear programming formulation of Smith (2.6) can be employed in Step (i) of Algorithm 3.3 as a replacement for the $2n$ linear programs (3.5). In this case $(c^{i(j)}, \alpha^{i(j)}, \beta^{i(j)})$ can be defined as

$$c^{i(j)} := \bar{c}, \quad \alpha^{i(j)} := \min_{A_i \in \mathcal{A}^j} A_i \bar{c}, \quad \beta^{i(j)} := \max_{B_i \in \mathcal{B}^j} B_i \bar{c}$$

where $(\bar{c}, \bar{\gamma}, \bar{y}, \bar{z})$ is a solution of (2.6). When this was done, the method failed to obtain a discriminant function because it required a degeneracy procedure like that of Step (iii) of Algorithm 3.3 which was not performed. In fact the first pair of parallel planes obtained by Smith's linear program left 105 unclassified points, while the second pair failed to separate any of these 105 points. When Grinold's linear program (2.8) was similarly employed in Step (i) of Algorithm 3.3, no degeneracy procedure was required, however seven pair of parallel planes were needed, instead of our four, to completely separate the 369 sample points. Specifically these seven pairs of planes left unclassified 82, 53, 32, 20, 19, 13 and 0 points respectively.

5. Conclusions. We have presented a fundamental model for pattern recognition that can handle linearly inseparable pattern sets. A nonconvex optimization problem (3.4) represents this model and can be solved in polynomial time by solving the $2n$ linear programs (3.5). These linear programs are used to generate a sequence of parallel planes which result in a piecewise-linear nonconvex discriminant function. This procedure has been successfully used to generate a practical computer program for breast cancer diagnosis. This program is now in use at the University of Wisconsin Hospitals, and is superior to programs based on other linear programming or statistical formulations. It is hoped that further important applications of our approach will be found.

REFERENCES

[1] M.R. Garey and D.S. Johnson, *Computers and intractability: A guide to the theory of NP-completeness*, W.H. Freeman, San Francisco, CA 1979.

[2] R.C. Grinold, *A note on pattern separation*, Operations Research, 18, (1970), pp. 187-190.

[3] ———, *Mathematical Programming methods of pattern classification*, Management Science, 19 (1972), pp. 272-289.

[4] N. Karmarkar, *A new polynomial time algorithm for linear programming*, Combinatorica, 4, (1984), pp. 373-395.

[5] L.G. Khachian, *A polynomial algorithm in linear programming*, Dokl. Akad. Nauk SSR 244, 5, (1979), pp. 1093-1096; Soviet Math. Doklady, 20, (1979), pp. 191-194.

[6] O.L. Mangasarian, *Linear and nonlinear separation of patterns by linear programming*, Operations Research, 13, (1965), pp. 444-452.

[7] ———, *Multisurface method of pattern separation*, IEEE Transactions on Information Theory, IT-14, 6 (1968), pp. 801-807.

[8] O.L. Mangasarian and T.H. Shiau, *A variable complexity norm maximization problem*, SIAM J. Alg. Disc. Meth., 7, (1986), pp. 455-461.

[9] F.W. Smith, *Pattern classifier design by linear programming*, IEEE Transaction on Computers, C-17, 4, (1968), pp. 367-372.

[10] R. Takayima, *A learning procedure for multisurface method of pattern separation*, Pattern Recognition, 12, (1980), pp. 75-82.

[11] W.H. Wolberg and O.L. Mangasarian, *Multisurface method of pattern separation for medical diagnosis applied to breast cytology* , Proceedings of the National Academy of Sciences, U.S.A., 87, (1990).

[12] W.H. Wolberg, M.A. Tanner and W.-Y. Loh, *Diagnostic schemes for fine needle aspirates of breast masses*, Analytical and Quantitative Cytology and Histology, 10, (1988), pp. 225-228.