LINEAR AND NONLINEAR SEPARATION OF PATTERNS BY LINEAR PROGRAMMING

O. L. Mangasarian

Shell Development Company, Emeryville, California

(Received September, 1964)

A pattern separation problem is basically a problem of obtaining a criterion for distinguishing between the elements of two disjoint sets of patterns. The patterns are usually represented by points in a Euclidean space. One way to achieve separation is to construct a plane or a nonlinear surface such that one set of patterns lies on one side of the plane or the surface, and the other set of patterns on the other side. Recently, it has been shown that linear and ellipsoidal separation may be achieved by *nonlinear* programming. In this work it is shown that *both* linear and nonlinear separation may be achieved by *linear* programming.

A BASIC problem of pattern separation is this: Given two sets of patterns A and B, the set A consisting of m patterns, the set B of k patterns, where each pattern consists of n scalar observations, find a means of 'separating' the sets A and B, i.e., describing quantitatively whether a pattern belongs to the set A or the set B. An implementable and efficient solution of this problem is the key to the construction of pattern recognizing 'machines.' If the patterns are represented by points in an n-dimensional Euclidean space the separation problem then is to find a surface in this n-dimensional space such that all points representing patterns belonging to the set A be on one side of this surface, and all points representing patterns belonging to B lie on the other side of the surface. We shall not, for brevity, distinguish between patterns and the points in the n-dimensional space representing them.

One way of separating the patterns A and B is to pass a *plane* in the n-dimensional space such that the m patterns A lie on one side of this plane and the k patterns B lie on the other side. Recently,[5] it has been shown that a necessary and sufficient condition for such linear separation is that a certain quadratic programming problem have a solution. In the present work we show that linear separation is equivalent to solving a *linear* programming problem (Theorem 1). Also, since arbitrary patterns are in general *not* linearly separable, it is important to recognize linear inseparability immediately. This can be easily achieved by invoking Theorem 3, which seems to have been first given by HIGHLEYMAN,[3] and which states that a necessary and sufficient condition for linear inseparability obtains whenever nonnegative solutions of a set of equalities can be found. This again is equivalent to finding a feasible solution of a linear program.

When the patterns are not linearly separable, one may resort to quadratic or nonlinear separation. In reference 5 it was shown that a certain unique ellipsoidal separation may be achieved by nonlinear programming. In the third section we show that a (nonunique) quadratic separation or a more general (nonunique) nonlinear separation may be achieved by *linear* programming.

LINEAR SEPARABILITY AND INSEPARABILITY OF SETS OF PATTERNS

WE DEFINE a *set of patterns* by a nonempty *pattern matrix* of real numbers, each row of which defines a *single pattern* (or simply *pattern*) and is called a *pattern vector*. A single pattern is composed of n real numbers that are called *observations* and each of which is represented by a *pattern element* of the row vectors of the pattern matrix. We shall denote a pattern matrix by a capital letter A, single patterns by the row vectors A_i, and the jth observation of the ith pattern by the scalar A_{ij}. The basic problem confronting us now may be stated thus. Given two sets of patterns, defined by the $m\times n$ pattern matrix A and the $k\times n$ pattern matrix B, determine a plane in the n-dimensional Euclidean space E^n, such that if the m rows of A and the k rows of B are taken as points in this space, then they must fall on opposite sides of this plane. Let x be an n-dimensional row vector representing a point in this n-dimensional Euclidean space. The problem then is to determine a *single* plane

$$xd-\gamma=0, \tag{1}$$

where d is an n-dimensional column vector of constants and γ is a scalar constant such that

$$Ad-e\gamma>0, \tag{2}$$

and

$$Bd-l\gamma<0, \tag{3}$$

where e and l are respectively m- and k-dimensional column vectors of ones. The two sets of patterns A and B are *linearly separable* if and only if there exist some d,γ such that (2) and (3) are satisfied.* If no such d,γ exist, then A and B are *linearly inseparable.* It is convenient now to establish the following

LEMMA 1. *The two sets of patterns A and B are linearly separable if and only if there exists an n-dimensional column vector of constants c and constant*

* This definition of linear separability is equivalent to the more commonly used one,[3-5] which states that the sets of patterns A and B are linearly separable if the convex hulls of A and B in E^n do not intersect. *Also see* remark under Theorem 3.

scalars α and β such that

$$Ac - e\alpha \geqq 0, \tag{4}$$

$$-Bc + l\beta \geqq 0, \tag{5}$$

$$\alpha - \beta > 0, \tag{6}$$

$$f \geqq c \geqq -f, \tag{7}$$

where f is an n-dimensional column vector of ones.

Proof. If c, α, β exist satisfying (4) to (7) then $d=c$, $\gamma=\frac{1}{2}(\alpha+\beta)$ satisfy (2) and (3), and hence A and B are linearly separable. Conversely if A and B are linearly separable then there exist d,γ satisfying (2) and (3). Now at least one component of d must be different from zero, otherwise (2) and (3) lead to $-\gamma>0$ and $-\gamma<0$. Denote the largest absolute value of any component of d by δ. Dividing (2) and (3) by δ gives

$$(Ad/\delta) - e\gamma/\delta > 0, \tag{8}$$

$$(Bd/\delta) - l\gamma/\delta < 0. \tag{9}$$

Observe that

$$f \geqq d/\delta \geqq -f. \tag{10}$$

Now define

$$\alpha \equiv \min_{i\epsilon 1,\ldots,m} \sum_{j=1}^{j=n} (A_{ij}d_j/\delta), \tag{11}$$

$$\beta \equiv \max_{i\epsilon 1,\ldots,k} \sum_{j=1}^{j=n} (B_{ij}d_j/\delta), \tag{12}$$

$$c \equiv d/\delta. \tag{13}$$

It follows from (11), (12), (8), and (9) that

$$\alpha - \beta > 0, \tag{14}$$

and from (8) through (13) that

$$f \geqq c \geqq -f, \tag{15}$$

$$Ac - e\alpha = (Ad/\delta) - e\min_{i\epsilon 1,\ldots,m} \sum_{j=1}^{j=n} (A_{ij}d_j/\delta) \geqq 0, \tag{16}$$

and

$$Bc - l\beta = (Bd/\delta) - l\max_{i\epsilon 1,\ldots,k} \sum_{j=1}^{j=n} (B_{ij}d_j/\delta) \leqq 0. \tag{17}$$

Therefore, conditions (16), (17), (14), and (15) are precisely conditions (4) through (7).

Observe that the presence of a strict inequality (6) among the inequalities (4) to (7) prevents a routine application of the linear programming algorithm[2] in order to obtain a feasible solution c, α, β to the system (4) to (7). However by considering $(\alpha-\beta)$ as the objective function of a linear programming problem with constraints (4), (5), and (7), the following linear-programming criterion for linear separability may be obtained.

THEOREM 1 (*Linear Separability*). *A necessary and sufficient condition for linear separability of the pattern sets A and B is*

$$\theta(A, B) > 0, \tag{18}$$

where $\theta(A, B)$ is the solution of the linear programming problem

$$\theta(A,\ B) = \max_{c,\alpha,\beta}\{(\alpha-\beta) \mid Ac - e\alpha \geqq 0,\ \ -Bc + l\beta \geqq 0,\ f \geqq c \geqq -f\}. \tag{19}$$

Proof. If $\theta(A,\ B) > 0$, then (4) to (7) are obviously satisfied by the solution of (19) and hence A and B are linearly separable. Conversely, if A and B are linearly separable, then some c, α, β must satisfy (4) to (7). The same c, α, β is a feasible point for the linear programming problem and renders $\alpha - \beta > 0$. Hence $\theta(A, B) > 0$.

COROLLARY 1. (*Linear Inseparability*). *A necessary and sufficient condition for linear inseparability of the pattern sets A and B is that $\theta(A, B) = 0$.*

Proof. The proof follows immediately from Theorem 1 by observing that $c = 0$, $\alpha = 0$, $\beta = 0$ is a feasible point and hence $\theta(A, B) \geqq 0$.

It should be remarked that the linear programming problem (19) is a very well behaved problem because of the two following aspects: (1) It always has a feasible solution, $c = 0$, $\alpha = 0$, $\beta = 0$. (2) Its solution is bounded from above by

$$\{\max_{i\in 1,\ \ldots,\ m} \textstyle\sum_{j=1}^{j=n} |A_{ij}| + \max_{i\in 1,\ \ldots,\ k} \textstyle\sum_{j=1}^{j=n} |B_{ij}|\}. \tag{20}$$

When the sets A and B are linearly inseparable, $\theta(A, B) = 0$ and the maximizing c will vanish in general. If such is the case it may be desirable to determine which patterns of either of the sets A and B are precluding a linear separation. These points cannot be determined from the solution $\theta(A,\ B) = 0$ of (19). One way to determine them is the following. Augment the constraints of (19) by the constraint $c_1 = 1$,* and denote the solution of (19) with the augmented constraint by $\bar{c}$, $\bar{\alpha}$, $\bar{\beta}$. The deletion of the points in the pattern set A (i.e., the rows of A) that satisfy

$$\textstyle\sum_{j=1}^{j=n} A_{ij}\, \bar{c}_j \leqq \bar{\beta}$$

or the deletion of the points in the pattern set B (i.e., the rows of B) that satisfy

$$\textstyle\sum_{j=1}^{j=n} B_{ij}\, \bar{c}_j \geqq \bar{\alpha}$$

will render the pattern sets A and B linearly separable. If one is interested in deleting the least number of such points, one may solve $2n$ linear programming problems (19) each with a different component of c_i ($i = 1, 2, \cdots, n$) set equal to plus and minus one. The least number of points to be deleted

* The choice of this constraint is arbitrary. Any other constraint of the type $c_i = 1$ or $c_i = -1$ would work to prevent the identical vanishing of all the components of c.

is obtained from the solution with the least number of points satisfying either of the last two relations.

A variation of the procedure described in the previous paragraph may also be used to separate two linearly inseparable sets of patterns by a finite number of planes. This is done as follows: After $\bar{c}$, $\bar{\alpha}$, $\bar{\beta}$ are obtained, by setting some c_i equal to 1 or -1 and solving (19), we obtain two planes, $x\bar{c}=\bar{\alpha}$ and $x\bar{c}=\bar{\beta}$, which leave parts of the pattern sets A and B unseparated. We take the unseparated parts only of the sets A and B and obtain two other planes that will again leave parts of A and B unseparated. This process is repeated until the remaining unseparated parts can be separated by one plane only. It is easy to see that this process is finite (i.e., it will terminate after a finite number of linear programs), and that it will separate the originally linearly inseparable pattern sets A and B into two sets that are separable by a finite number of planes.

By invoking the duality principles of linear programs, reference 2, pp. 71–74, it is possible to obtain immediately the following dual theorem to Theorem 1.

THEOREM 2 (*Dual Linear Separability*). *A necessary and sufficient condition for the linear separability of the pattern sets A and B is that*

$$\varphi(A, B)>0, \tag{21}$$

where $\varphi(A, B)$ *is the solution of the linear programming problem*

$$\begin{aligned}\varphi(A, B)=\min_{u,v,p}\{f'p \mid e'u=1,\ l'v=1,\ -A'u+B'v+p\geqq 0,\\ A'u-B'v+p\geqq 0,\ u\geqq 0,\ v\geqq 0\},\end{aligned} \tag{22}$$

*where u, v, and p are m-, k-, and n-dimensional column vectors, and the prime denotes the transpose.**

COROLLARY 2 (*Dual Linear Inseparability*). *A necessary and sufficient condition for the linear inseparability of the pattern sets A and B is that* $\varphi(A, B)=0$. Corollary 2 follows from Theorem 2 by observing that from the constraints of (22), $p\geqq 0$, hence $f'p\geqq 0$ and $\varphi(A, B)\geqq 0$.

We are now in a position to derive the final main result of this section, a necessary and sufficient condition for linear inseparability that is equivalent to finding a nonnegative solution of linear equalities. This condition is similar to a condition of Highleyman[3] and NILSSON.[4]

* Problem (22) is equivalent to

$$\begin{aligned}\psi(A,B)=\min_{u,v,r,s}\{f'(r+s) \mid A'u-B'v+r-s=0,\quad e'u=1,\quad l'v=1,\\ u\geqq 0,\ v\geqq 0,\ r\geqq 0,\ s\geqq 0\},\end{aligned} \tag{22*}$$

where r and s are n-dimensional vectors. Problem (22*) is in the exact format required by the primal simplex algorithm, and hence is the best candidate for solution by that method. The shadow prices associated with the first three sets of constraints of (22*) are precisely c, $-\alpha$, β.

THEOREM 3 (*Dual Inseparability Criterion*). *A necessary and sufficient condition that the sets of patterns A and B be linearly inseparable is that the system*

$$A'u-B'v=0, \tag{23}$$

$$e'u=1, \tag{24}$$

$$l'v=1, \tag{25}$$

$$u\geqq 0, \tag{26}$$

$$v\geqq 0, \tag{27}$$

*has a solution.**

REMARK. Conditions (23) to (27) are nothing more than a mathematical statement of the fact that the convex hulls of A and B intersect.

Proof. If the sets A and B are inseparable, then by Corollary 2, $\varphi(A,B)=0$ and hence $p=0$. This immediately implies that the solution of (22) satisfies (23) through (27). Conversely, if (23) through (27) are satisfied by u, v, then this implies that u, v, $p=0$, satisfy the constraints of (22) and hence $\varphi(A,B)=0$. By Corollary 2, then we conclude that A and B are inseparable.

NONLINEAR SEPARATION OF SETS OF PATTERNS BY LINEAR PROGRAMMING

SOMETIMES it may not be possible to separate two sets of patterns by a plane. One may then resort to a separating surface that is nonlinear. In reference 5 it was shown that given one set of patterns it may be possible to determine a *unique* ellipsoid enclosing the set of patterns by solving a nonlinear programming problem. In the present work we drop the uniqueness requirement, and show that a quadratic separation or a more general nonlinear separation may be achieved by linear programming. For the sake of simplicity we will confine ourselves to quadratic separation here with the understanding that a more complicated nonlinear separation can be achieved analogously by the linear programming technique of this section.

Again let x be an n-dimensional row vector representing a point in an n-dimensional Euclidean space. The *quadratic separation* problem of the pattern sets A and B consists of determining a *single quadratic surface*

$$x\,Ex'+xd-\gamma=0, \tag{28}$$

* The conditions (4) to (7) of the original Lemma 1 can be recovered directly from the conditions (23) to (27) of Theorem 3 by invoking Theorem 1 of GALE.[1] Conversely, Theorem 3 may be derived directly from the Lemma by invoking the same Theorem.

where E is an $n\times n$ matrix of constants, d an n-dimensional column vector of constants, and γ a scalar constant, such that

$$A_iEA_i'+A_id-\gamma>0, \qquad (i=1, \cdots, m) \quad (29)$$

and

$$B_jEB_j'+B_jd-\gamma<0. \qquad (j=1, \cdots, k) \quad (30)$$

The two sets of patterns A and B are *quadratically separable* if and only if there exist some E, d, γ such that (29) and (30) are satisfied. If no such E, d, γ exist, then A and B are *quadratically inseparable*. It is obvious that linear separability implies quadratic separability and that quadratic inseparability implies linear inseparability.

It is possible now to establish results for nonlinear separability analogous to those of the preceding section for linear separability. This is possible because of the linearity of (29) and (30) in E, d, and γ. We will confine ourselves here to a mere statement of the simpler results for quadratic separability, the proofs of which are essentially identical to the corresponding theorems of the preceding section.

LEMMA 1A. *The two sets of patterns A and B are quadratically separable if and only if there exists an $n\times n$ matrix D of constants, an n-dimensional column vector c of constants, and constant scalars α and β such that*

$$A_iDA_i'+A_ic-\alpha\geqq 0, \qquad (i=1, \cdots, m) \quad (31)$$

$$-B_jDB_j'-B_jc+\beta\geqq 0, \qquad (j=1, \cdots, k) \quad (32)$$

$$\alpha-\beta>0, \quad (33)$$

$$f\geqq c\geqq -f, \quad (34)$$

$$K\geqq D\geqq -K, \quad (35)$$

where K is an $n\times n$ matrix of ones.

THEOREM 1A (*Quadratic Separability*). *A necessary and sufficient condition for the quadratic separability of the pattern sets A and B is that*

$$\psi(A, B)>0 \quad (36)$$

where $\psi(A,B)$ is the solution of the linear programming problem

$$\psi(A,B)=\max_{D,c,\alpha,\beta}\{(\alpha-\beta)|A_iDA_i'+A_ic-\alpha\geqq 0,\ i=1, \cdots, m,\ -B_jDB_j'-B_jc +\beta\geqq 0,\ j=1, \cdots, k,\ f\geqq c\geqq -f,\ K\geqq D\geqq -K\}. \quad (37)$$

COROLLARY 1A (*Quadratic Inseparability*). *A necessary and sufficient condition for quadratic inseparability of the pattern sets A and B is that $\psi(A,B)=0$.*

We observe again here that the linear programming problem (37) always has a feasible solution $D=0$, $c=0$, $\alpha=0$, $\beta=0$, and is bounded from above by

$$\{\max_{i\epsilon 1,\ldots,m}[(\textstyle\sum_{j=1}^{j=n}|A_{ij}|)^2+\sum_{j=1}^{j=n}|A_{ij}|] \\ +\max_{i\epsilon 1,\ldots,k}[(\textstyle\sum_{j=1}^{j=n}|B_{ij}|)^2+\sum_{j=1}^{j=n}|B_{ij}|]\}. \tag{38}$$

Further results analogous to Theorems 2, and 3, and Corollary 2 may also be stated for quadratic and more general nonlinear separability problems. They are somewhat more complicated and hence will not be given here. Suffice it to say that a nonlinear separation problem may be viewed as a linear programming problem.

REMARKS

THE MOST widely used method for nonparametric pattern separation is ROSENBLATT's error correction procedure[6, 7] for linear separation. Recently, GREENBERG AND KONHEIM[8] have extended this to nonlinear separation. Rosenblatt's method and various modifications thereof[9] are based on a very simple iterative procedure. In the present work we have presented a linear and nonlinear separation method based on linear programming. It is difficult to assess the relative advantages of the two methods without an extensive series of test problems. However, the obvious advantage of the error correction procedure is its simplicity. Its main disadvantage seems to be its inability to determine inseparability of pattern sets when it occurs. This is a consequence of the fact that the error correction procedure converges only when the pattern sets are indeed separable, a fact that is not known a priori. In the linear programming method however, inseparability is immediately detected by either $\theta(A,B)=0$ or $\psi(A,B)=0$. Since it is possible to construct some simple examples for which the error correction procedure converges very slowly, the problem of distinguishing between slow convergence and nonconvergence may be a difficult one. Another advantage of the linear programming method is that it can be readily extended in order to separate two sets by more than one plane or surface as described in the second section.

SINGLETON[10]* has described a linear programming method similar to that of MINNICK[11] that, however, differs from the present one in that it involves the use of a relatively large matrix essentially made up of A, $-B$, $-A$ and B. More recently, CHARNES[12] has described a linear programming method for *linear* separability that is equivalent to (22*). Our manner of derivation is different here and extends directly to nonlinear separability.

* I am indebted to a referee for this reference.

REFERENCES*

1. D. Gale, "The Basic Theorems of Real Linear Equations, Inequalities, Linear Programming and Game Theory," *Nav. Res. Log. Quart.* **3,** 193–200 (1956).
2. S. I. Gass, *Linear Programming*, McGraw-Hill, New York, 1958.
3. W. Highleyman, "A Note on Linear Separability," IRE *Trans. on Electronic Computers*, Correspondence Section, 777–778 (December 1961).
4. N. J. Nilsson, "Learning Machines," Class Notes EE 353, pp. 83–84, Summer 1962, Stanford University.
5. J. B. Rosen, "Pattern Separation by Convex Programming," Technical Report No. 30, Contract Nonr-225(37), Stanford University, June 28, 1963.
6. F. Rosenblatt, *Principles of Neurodynamics*, Spartan Books, Washington, D. C., 1961.
7. ———, "On the Convergence of Reinforcement Procedures in Simple Perceptrons," Cornell Aeronautical Laboratory Report No. VG-1196-G-4, Buffalo, New York, 1960.
8. H. J. Greenberg and A. G. Konheim, "Linear and Nonlinear Methods in Pattern Classification," *IBM J. Res. and Devel.* **8,** 299–307 (1964).
9. N. J. Nilsson, "Introduction to Theory of Trainable Pattern Classifying Machines," Stanford Research Institute, Menlo Park, California, 1964.
10. R. C. Singleton, "A Test for Linear Separability as Applied to Self-Organizing Machines," in *Conference on Self-Organizing Systems—1962*," M. C. Yovits, G. T. Jacobi, and G. D. Goldstein (eds.), pp. 503–524, Spartan Books, Washington, D. C., 1962.
11. R. C. Minnick, "Linear-Input Logic," *IRE Trans. Electronic Computers*, V.EC **10,** 6–16 (March 1961).
12. A. Charnes, "Some Fundamental Theorems of Perceptron Theory and Their Geometry," in *Computer and Information Sciences*, J. T. Tou and R. H. Wilcox (eds.) pp. 67–74, Spartan Books, Washington, D. C., 1964.

* References 8 through 12 came to the author's attention after finishing the work.