Published in final edited form as:

Nonlinear Anal Theory Methods Appl. 2001 August ; 47(3): 1857–1867. doi:10.1016/

S0362-546X(01)00316-9

Set Separation Problems and Global Optimization

James E. Falk, Yelena Dandurova, and Lana Yeganova

School of Engineering and Applied Science, The George Washington University, Washington DC 20052

Abstract

Given a pair of finite, disjoint sets A and B in R^n , a fundamental problem with numerous applications is to find a simple function $f(x)$ defined over R^n which separates the sets in the sense that $f(a) > 0$ for all $a \in A$ and $f(b) < 0$ for all $b \in B$. This can always be done (e.g., with the piecewise linear function defined by the Voronoi partition implied by the points in $A \cup B$). However typically one seeks a linear (or possibly a quadratic) function f if possible, in which case we say that A and B are linearly (quadratically) separable. If A and B are separable in a linear or quadratic sense, there are generally many such functions which separate. In this case we seek a ‘robust’ separator, one that is best in a sense to be defined. When A and B are not separable in a linear or quadratic sense we seek a function which comes as close as possible to separating, according to some well defined criterion.

In this paper we examine the optimization problems associated with the set separation problem, characterize them (convex or non-convex) and suggest algorithms for their solutions.

AMS 1991 Subject Classification

15A45; 49N15; 90C25; 65K10

1 Introduction

Let $A = \{a_1, \dots, a_s\}$ and $B = \{b_1, \dots, b_t\}$ be two disjoint finite sets of points in R^n and let F denote a class of functions (e.g., linear, quadratic) each defined over R^n . We wish to *efficiently* determine a function $f \in F$ over R^n which separates these sets in the sense that

$$f(a) > 0 \text{ for all } a \in A \text{ and } f(b) < 0 \text{ for all } b \in B$$

if possible, or which ‘nearly’ separates them when no $f \in F$ separates the sets. In the event that the sets *are* separable over F , there are generally an infinite number of functions which separate. In this case, we wish to determine a separator which is robust in the sense that it is of maximal ‘distance’ from the points $A \cup B$

Such problems typically arise in, e.g., medical diagnosis. In one well-know study the points in $A \cup B$ are 9 vectors whose components correspond to attributes related to breast cancer (clump thickness, uniformity of cell size, ...). Each vector corresponds to a tissue sample of a patient who discovers a lump on her breast and each component measures a pathologist’s

judgement as to the severity of that attribute. The set A is the set of all patients in $A \cup B$ who's lumps were cancerous, and B is the complement of that set.

After a separating function f is identified, it can be used to predict the set A or B to which a new patient $c \notin A \cup B$ belongs, according to whether $f(c) > 0$ or $f(c) < 0$. Ties are not addressed here.

Since A and B are disjoint, they can always be separated by a 'Voronoi' separator (See e.g., [8]). For any finite set $C \subset R^n$, let

$$V(c_i) = \{x \in R^n : \|x - c_i\| \leq \|x - c_j\| \text{ for all } j \neq i\}$$

for all $c_i \in C$. The set $V(c_i)$ is termed the Voronoi Cell associated with c_i and consists of all points at least as close (in the Euclidean norm) to c_i as to any other point $c_j \in C$. The collection $\mathfrak{v} = \{V(c_i) : c_i \in C\}$ forms a partition of R^n in the sense that

$$\text{int} V(c_i) \cap \text{int} V(c_j) = \emptyset \text{ for all } i \neq j \text{ and } \bigcup_i V(c_i) = R^n.$$

Let $V(a_i)$ and $V(b_j)$ denote the Voronoi cells associated with the sets A and B respectively. The two sets $\mathfrak{v}(A) = \bigcup_i V(a_i)$ and $\mathfrak{v}(B) = \bigcup_j V(b_j)$ together cover R^n and one can associate any new point $c \notin A \cup B$ with either A or B according to whether or not $c \in \mathfrak{v}(A)$ or $c \in \mathfrak{v}(B)$. It is not difficult to associate a non-trivial, piecewise linear continuous function f which is positive over the interior of $\mathfrak{v}(A)$, and negative over the interior of $\mathfrak{v}(B)$. One such (somewhat complicated) function is

$$f(x) = \min \left\{ \|x - a_i\|^2 : a_i \in A \right\} - \min \left\{ \|x - b_j\|^2 : b_j \in B \right\}.$$

In this paper however, we are interested in determining either *linear* or *quadratic* functions which separate.

2 Definitions and Preview

Let A and B be two disjoint, finite sets in R^n , and let f be a function defined over R^n which is not identically zero. We say that f *separates* A and B if

$$f(a) > 0 \text{ for all } a \in A$$

$$f(b) < 0 \text{ for all } b \in B$$

and we term the function f a *separator* of A and B . (Some authors refer to this as 'strict' or 'strong' separation – see [4]).

If F is a class of functions (e.g., linear or quadratic), we say that A and B are (strongly) separable over F if there is a non-trivial function $f \in F$ which (strongly) separates A and B . Otherwise the sets are *inseparable* over F .

When A and B are separable by some function $f \in F$, the set

$$S_f = \{x: f(x) = 0\}$$

is termed a *separating set* of A and B .

If A and B are separable over F , there are generally an infinite number of non-trivial functions which separate. Let F^S denote the set of all non-trivial functions in F which separate. In this case, we seek a function $\bar{f} \in F^S$ whose separating set $S_{\bar{f}}$ is of maximal ‘distance’ from $A \cup B$ where distance is measured by some given function $D: F^S \times (A \cup B) \rightarrow R$. Such a function $\bar{f}$ is termed a *robust separator with respect to D*.

In this paper, we focus on two types of separators:

- Linear functions:

$$L = \{f(x) = x^T \omega - \gamma\}$$

where $\omega \in R^n$, and $\gamma \in R$. Non-trivial linear functions have $\omega \neq 0$.

- Quadratic functions:

$$Q = \{f(x) = (1/2) x^T M x + x^T \omega - \gamma\}$$

where M is a symmetric $n \times n$ matrix, $\omega \in R^n$, and $\gamma \in R$. Non-trivial quadratic functions have at least one non zero element of M or ω .

Also in this paper we will focus on two ‘distance’ measures:

- ‘Nearest Euclidean Distance’: Given a function $f(x)$ which separates A and B , the distance between f and $A \cup B$ is the minimum of the distances between points in the set $A \cup B$ and the separating set $S_f = \{x: f(x) = 0\}$.

A robust separator $\bar{f}$ is one whose separating set $S_{\bar{f}}$ is of maximal distance away from the closest point(s) in $A \cup B$.

- ‘Weighted Euclidean Distance’ (Linear separators only)

One such measure will be shown to be:

$$e^{\gamma} \left(\frac{1}{s} \right) \sum_{i=1}^s e^{-a_i \cdot \omega} + e^{-\gamma} \left(\frac{1}{t} \right) \sum_{j=1}^t e^{-b_j \cdot \omega}$$

and a robust linear separator $\bar{f}$ is one which minimizes this over the set $\{\omega \neq 0\}$. This measure takes into account all of the points of $A \cup B$, not only the closest ones.

In this paper we will summarize results obtained for the cases:

- Linear Separators
 - Nearest Euclidean Distance – A and B are either separable or not.
 - Weighted Euclidean Distance – A and B are either separable or not.
- Quadratic Separators
 - Nearest Euclidean Distance – A and B are separable over Q

In what follows, the symbol e denotes a vector of appropriate size, all of whose components are 1. Also the symbols A and B will stand for sets and also (with some abuse of notation) matrices whose rows are the vectors of the sets A and B .

3 Brief Literature Survey

Nearly all of the literature on the 2 set separation problem focuses on the linear separation problem with the notion of ‘robustness’ only tangentially discussed. O. L. Mangasarian was one of the earliest and most influential contributors to the linear separation problem starting in 1965 (see [6]). In the same year, J. B. Rosen [9] described a (non-robust) quadratic model.

During the ‘80’s and ‘90’, an entire sequence of papers by and/or referencing Freed and Glover (starting with [5]) appeared in the journal *Decision Sciences*. All of this work is essentially independent of the aforementioned references, and is directed at establishing linear programming models whose solutions will separate (or nearly separate) a pair of sets A and B . Some of the models work only under very special conditions.

The earliest model seeking robust linear separators is due to Cavalier et al, [1] who suggested a heuristic method for its solution when A and B are separable. Falk and Cardona [4] later showed that the suggested heuristic was, in fact, convergent and improved on the rate of convergence by introducing a method to accurately predict which constraints would be binding. They also addressed the case where A and B are not separable, and introduced a Branch and Bound method for its solution.

In [3] Falk and Karlov looked at the robust quadratic problem when A and B are separable, and presented an algorithm for its solution.

The weighted Euclidean distance measure is believed to be new although Mangasarian did address a related model using a weighted (non-Euclidean) distance measure (see [7]).

4 The Linear Separation Problem

We are given two disjoint sets A and B in R^n and wish to first determine if *there is* a linear (affine) function

$$f(x) = x^T \omega - \gamma$$

which (strongly) separates them. If so, there must be a solution to the system

$$a_i^T \omega - \gamma > 0 \text{ for all } i=1, \dots, s$$

$$b_j^T \omega - \gamma > 0 \text{ for all } j=1, \dots, t.$$

Mangasarian [6] seems to have been the first to point out the

Theorem 1

Consider the linear program:

$$V^+ = \begin{cases} \max & s - t \\ \text{s.t.} & s^T e \leq A\omega, t^T e \geq B\omega, -e \leq \omega \leq e \end{cases}$$

The sets A and B are separable if and only if the solution value V^+ is positive. The optimal value (ω^+, γ^+) gives a separator (albeit generally not robust).

Given that a pair of sets A and B are separable, we can seek to find a *robust separator*. Geometrically, we are seeking a separator which is of maximal distance to the closest points in A and B . It turns out (after eliminating γ) that the appropriate optimization problem (see [4]) is:

$$\max_{\|\omega\|=1} \left\{ \min_{a \in A} \{a \cdot \omega\} - \max_{b \in B} \{b \cdot \omega\} \right\}. \quad (1)$$

The objective function of this problem is piecewise linear and concave, but the problem is not convex due to the restriction $\|\omega\| = 1$.

The optimal solution of this problem is unique when A and B are separable.

In the case that the sets A and B are not separable, we can seek a *linear hyperplane which minimizes the distance of the farthest mis-classified point(s)*. Interestingly, this problem leads to precisely the same optimization problem (1) above! In this case however, the optimal solution of the problem need not be unique.

However, appropriate solution methods are very different for the two cases.

In the separable case, it turns out that the constraint $\|\omega\| = 1$ can be replaced by the constraint $\|\omega\| \leq 1$.

Theorem 2

(Falk, Lopez-Cardona): In the case that the sets A and B are separable, the solution of the convex problem

$$\max_{\|\omega\| \leq 1} \left\{ \min_{a \in A} \{a \cdot \omega\} - \max_{b \in B} \{b \cdot \omega\} \right\}$$

gives the unique optimal weight vector ω^* of the robust separator. The optimal threshold value is

$$\gamma^* = \left(\min_{a \in A} \{a \cdot \omega\} + \min_{b \in B} \{b \cdot \omega\} \right) / 2.$$

In [4] a cutting plane procedure (previously suggested as a heuristic in [1]) is shown to converge to the solution of this problem. The first problem solved is exactly Mangasarian's linear program which determines the separability of A and B . In addition, an acceleration scheme based on predicting the points of A and B which determine ω^* is incorporated.

In the non-separable case, the constraint $\|\omega\| = 1$ can be replaced by the constraint $\|\omega\| \leq 1$ which describes a non-convex region.

Theorem 3

(Falk, Lopez-Cardona): In the case that the sets A and B are not separable, the solution of the non-convex problem

$$\max_{\|\omega\| \geq 1} \left\{ \min_{a \in A} \{a \cdot \omega\} - \max_{b \in B} \{b \cdot \omega\} \right\}$$

gives an optimal weight vector ω^* of a robust separator. A corresponding optimal threshold value is

$$\gamma^* = \left(\min_{a \in A} \{a \cdot \omega\} + \max_{b \in B} \{b \cdot \omega\} \right) / 2.$$

In [4] a Branch and Bound procedure is developed to solve this problem. Again, the first problem solved is exactly Mangasarian's linear program which determines the separability of A and B .

5 The Linear Separability Problem: Weighted Distance

The material summarized in this section is extracted from [2] and is available through the authors.

One of the criticisms of the robust separator defined above is that it is only determined by a small number of points in $A \cup B$. In an attempt to address this issue, we can introduce a new

separator $H(\omega, \gamma)$ which takes into account the distance between it and the set $A \cup B$ but is determined by *all* of the points in this set. This is done by ‘weighting’ points relative to $A \cup B$ according to their distance to $H(\omega, \gamma)$ with the closer points getting higher weights, but still taking into account the points distant from $H(\omega, \gamma)$. We chose the negative exponential function for this purpose.

The distance between a point x and a hyperplane $H(\omega, \gamma)$ is

$$d(x, H) = \frac{|x \cdot \omega - \gamma|}{\|\omega\|}$$

or simply

$$d(x, H) = |x \cdot \omega - \gamma|$$

if we restrict ω to have norm 1. Points $a \in A$ should, if possible, have a distance $d(a, H) = a \cdot \omega - \gamma \geq 0$ and points in B should, if possible have a distance $d(b, H) = \gamma - b \cdot \omega \geq 0$.

The function

$$e^{-(x \cdot \omega - \gamma)}$$

will be smaller for those properly classified points in A farthest from H than for those properly classified points in A closest to H . It also will be larger for any mis-classified point in A than for any properly classified point in A .

Likewise the function

$$e^{-(\gamma - x \cdot \omega)}$$

will be largest for mis-classified points in B , and smallest for properly classified points in B with the farthest points in B from H having the smallest values.

In order to account for possible differences in the number of points in A and B (s and t), we will weight the terms $\sum_{i=1}^s e^{-(a_i \cdot \omega - \gamma)}$ and $\sum_{j=1}^t e^{-(\gamma - b_j \cdot \omega)}$ with the coefficients $(1/s)$ and $(1/t)$ respectively.

Thus the function

$$f(\omega, \gamma) = \left(\frac{1}{s}\right) \sum_{i=1}^s e^{-(a_i \cdot \omega - \gamma)} + \left(\frac{1}{t}\right) \sum_{j=1}^t e^{-(\gamma - b_j \cdot \omega)} = e^{\gamma} \left(\frac{1}{s}\right) \sum_{i=1}^s e^{-a_i \cdot \omega} + e^{-\gamma} \left(\frac{1}{t}\right) \sum_{j=1}^t e^{b_j \cdot \omega}$$

will tend to be minimized over the set $S = \{(\omega, \gamma) : \|\omega\| = 1\}$ at a point which will separate A and B when they are separable, or at least give a ‘nearly’ separating hyperplane.

Thus we are led to study the problem

$$\min_{\|\omega\|=1} f(\omega, \gamma) \quad (2)$$

First we can eliminate γ by fixing ω and solving the problem

$$\min T e^{\gamma} + R e^{-\gamma}$$

where $T = \left(\frac{1}{s}\right) \sum_{i=1}^s e^{-a_i \cdot \omega}$ and $R = \left(\frac{1}{t}\right) \sum_{j=1}^t e^{b_j \cdot \omega}$. We obtain

$$\gamma = \left(\frac{1}{2}\right) \ln \left(\frac{s \sum_{j=1}^t e^{b_j \cdot \omega}}{t \sum_{i=1}^s e^{-a_i \cdot \omega}} \right).$$

and the above problem becomes

$$\min 2 \sqrt{TR}$$

which is equivalent to the problem

$$\min TR$$

or

$$(1/st) \min_{\|\omega\|=1} \left(\sum_{i=1}^s e^{-a_i \cdot \omega} \right) \left(\sum_{j=1}^t e^{b_j \cdot \omega} \right).$$

This seems somewhat cumbersome, but since the terms in brackets are positive, one can take logs and address the problem

$$\min_{\|\omega\|=1} \ln \left(\sum_{i=1}^s e^{-a_i \cdot \omega} \right) + \ln \left(\sum_{j=1}^t e^{b_j \cdot \omega} \right). \quad (3)$$

In this form, we can characterize the objective as convex.

Lemma 4

The function

$$g(s) = \ln \left(\sum_{i=1}^n e^{s_i} \right)$$

is convex in $s = (s_1, \dots, s_m)$.

The proof involves showing that $g''(s)$ is positive semi-definite.

Nevertheless, problem (3) is non-convex due to the constraint $\|\omega\| = 1$. Let

$$G(\omega) = \ln \left(\sum_{i=1}^s e^{-a_i \cdot \omega} \right) + \ln \left(\sum_{j=1}^t e^{b_j \cdot \omega} \right)$$

denote the objective function of (3). We get the

Theorem 5

The level sets $G(\omega)$ are unbounded if and only if the sets A and B are separable.

Thus when A and B are separable, the non-convex constraint $\|\omega\| = 1$ can be replaced by the convex constraint $\|\omega\| \leq 1$. However this is not so in the non-separable case and Branch and Bound must be used.

To close this section, we note that the use of the exponential function e^{-t} is arbitrary and, in fact, can be replaced by the more general function $e^{-\lambda t}$ where λ is a positive parameter. It can be shown that the hyperplane produced with a large λ approximates the hyperplane produced by the Falk-Lopez model.

6 The Quadratic Separation Problem

In the previous two sections, we sought a *linear* robust separator. In this section we treat the problem of finding a *quadratic* robust separator. There are two reasons for this – quadratic separators may separate when there is no linear separator (see Figure 1), and will generally separate better even if A and B are linearly separable. In this section we will assume that A and B are indeed quadratically separable.

Thus we seek a function of the form

$$f(x) = (1/2) x^T M x + x^T \omega - \gamma$$

where M is an $m \times m$ matrix, ω is an m vector and γ is a scalar such that the minimal distance between the set $A \cup B$ and the set

$$S_f = \left\{ x : (1/2) x^T M x + x^T \omega - \gamma = 0 \right\}$$

is maximal. As in the linear case, we can determine if the sets A and B are quadratically separable by solving a linear program.

Theorem 6

The optimal value V^* of the linear program

$$V^* = \begin{cases} \max_{s,t,M,\omega} & s - t \\ \text{s.t.} & s^T e \leq (1/2) a^T M a + a^T \omega \quad \text{for all } a \in A \\ & t^T e \geq (1/2) b^T M b + b^T \omega \quad \text{for all } b \in B \\ & M_{ij}, \omega_i \in [-1, 1] \quad \text{for all } i, j \end{cases}$$

is positive if and only if the sets A and B are quadratically separable. The optimal solution $M^*, \omega^*, \gamma^* = (s^* + t^*)/2$ gives a separator (albeit generally not robust).

Clearly if A and B are linearly separable, they are quadratically separable. In the linear case when A and B are separable, the robust separator is unique. This is not true in the quadratic case (e.g., $A = \{(-1,0)\}$, $B = \{(1,0)\}$).

One of the more serious difficulties in generating a *robust* quadratic separator is the problem of actually finding the minimum distance between a point $c \in A \cup B$ and the set S_f above. This problem has the form

$$\min_{\text{s.t.}} \|x - c\| \\ (1/2)x^T M x + x^T \omega - \gamma = 0$$

where M , ω , and γ are given. This, in itself, is a non-convex problem! However, Falk and Karlov [3] present an algorithm which involves:

- diagonalizing M , and
- maximizing the dual of the above problem where M is diagonal.

The second step is a simple convex problem with one variable (the dual variable associated with the constraint).

Also the first step needs only to be done once per iteration (it does not change for different points $c \in A \cup B$).

The general algorithm starts by determining a ‘trial’ quadratic separator which solves the linear program of Theorem 6. With this ‘incumbant’ separator $f \leftrightarrow (M, \omega, \gamma)$, the nearest point(s) in S_f to $A \cup B$ are determined. These points are then ‘added’ to the sets A and B to form new sets A^-, B^- . With these new points, a new incumbent is generated.

Geometrically this amounts to incrementally ‘blowing up’ hyperspheres of increasingly larger radii around the crucial points $c \in A \cup B$ (see Figure 1). Convergence is established in [3].

Numerical results also can be found in the references ([4] and [3]). All of these methods were tested on the Wisconsin breast cancer data and proved to show excellent accuracy on this data.

References

1. Cavalier TM, Ignizio JP, Soyster AL. Discriminant Analysis via Mathematical Programming: Certain Problems and their Causes. *Computers and Operations Research*. 1989; 16(4):353–362.
2. Dandurova Y, Yeganova L, Falk JE. Separating Sets via Exponentials. 2000 Draft paper.
3. Falk JE, Karlov V. Quadratic Separation of Sets. *Computers and Operations Research*. 2000 to appear.
4. Falk JE, Lopez-Cardona E. The Surgical Separation of Sets. *Journal of Global Optimization*. 1997; 00:1–31.
5. Freed N, Glover F. A Linear Programming Approach to the Discriminant Problem. *Decision Sciences*. 1981; 12:68–74.
6. Mangasarian OL. Linear and Nonlinear Separation of Patterns. *Operations Research*. 1965; 13:444–452.
7. Mangasarian, OL., Setiono, R., Wolberg, WH. Pattern Recognition via Linear Programming. In: Coleman, TF., Li, Y., editors. *Large-Scale Numerical Optimization*. SIAM; 1990. p. 22-30.
8. Preparata, FP. *Computational Geometry, An Introduction*. Springer-Verlag; Paris: 1990.
9. Rosen JB. Pattern Separation by Convex Programming. *Journal of Mathematical Analysis and Applications*. 1965; 10:123–134.

Figure 1.
Fitting Data with Nonlinear Curves