WORKING PAPER

A UNIFIED MATHEMATICAL PROGRAMMING FORMULATION FOR THE DISCRIMINANT PROBLEM

Antonie Stam
Cliff T. Ragsdale

July 1989
WP-89-047

IIASA
International Institute for Applied Systems Analysis

A UNIFIED MATHEMATICAL PROGRAMMING FORMULATION FOR THE DISCRIMINANT PROBLEM

Antonie Stam
Cliff T. Ragsdale

July 1989
WP-89-047

Department of Management Sciences
and Information Technology
College of Business Administration
University of Georgia
Athens, Georgia, 30602 U.S.A.

INTERNATIONAL INSTITUTE FOR APPLIED SYSTEMS ANALYSIS
A-2361 Laxenburg, Austria

FOREWORD

The purpose of classification analysis is to predict the group membership of individuals or observations based on limited information about the group characteristics. The resulting classification or discriminant rules provide a powerful methodology in decision analysis. In fact, classification analysis has been touted as one of the most significant tools to analyze scientific and behavioral data. Applications of discriminant analysis can be found in such diverse fields as predicting bank failures, artificial intelligence, medical diagnosis, psychology, biology and credit granting. The most widely used statistical techniques are based on the assumption of multivariate normality. Frequently, this assumption is violated and nonparametric techniques are appropriate. One such technique which was recently proposed uses mathematical programming formulations of the problem.

This paper introduces a unified mathematical programming-based approach to the two-group discriminant problem which does not suffer from many of the theoretical inadequacies that have plagued previously proposed formulations. Moreover, the formulation appears to be simple, making it a promising contribution from both a theory and practice viewpoint.

Alexander B. Kurzhanski
Chairman
System and Decision Sciences Program

A Unified Mathematical Programming Formulation for the Discriminant Problem †

Antonie Stam *

Cliff T. Ragsdale *

* *Department of Management Sciences*
and Information Technology
University of Georgia
Athens, GA 30602
U.S.A.

† This paper was written in part while the first author was visiting the International Institute for Applied Systems Analysis in Austria as a participant in the 1989 Young Summer Scientists Program.

A Unified Mathematical Programming Formulation

for the Discriminant Problem

Abstract

In recent years, much research has been done on the application of mathematical programming (MP) techniques to the discriminant problem. While very promising results have been obtained, many of these techniques are plagued by a number of problems associated with the model formulation including unbounded, improper and unacceptable solutions as well as solution instability under linear transformation of the data. Some have attempted to prevent these problems by suggesting overly complex formulations which can be difficult to solve. Others have suggested formulations which solve certain problems but which create new ones. In this paper we develop a simple MP model which unifies many features of previous formulations and appears to avoid any solution problems. This approach also considers a classification gap often encountered in the related statistical techniques.

Subject Areas: Linear Programming, Linear Statistical Models, and Statistical Techniques.

A Unified Mathematical Programming Formulation for the Discriminant Problem

1. Introduction

The discriminant problem involves studying the differences between two or more groups and/or classifying new observations into one of two or more groups. This is one of the most fundamental problems of scientific inquiry and has found application in diverse fields from biology to artificial intelligence and the social and administrative sciences. For many years, well-known statistical techniques such as Fisher's linear discriminant function (LDF) (Fisher 1936) and Smith's quadratic discriminant function (QDF) (Smith 1947) have been the standard tools for attacking such problems. Recently, however, much has been written about the application of mathematical programming (MP) techniques to solve the problem in discriminant analysis. These MP technniques attempt to identify a hyperplane which can be used to distinguish between observations belonging to two different groups.

After their introduction by Hand (1981) and Freed and Glover (1981a, 1981b), various MP techniques have been shown to rival or outperform Fisher's LDF when the assumptions underlying the LDF are seriously violated (usually encountered when the data depart from multivariate normality). The most widely proposed MP techniques are the MSD formulation (Freed and Glover 1981b), which minimizes the sum of absolute exterior deviations from the classification hyperplane, the MMD formulation (Freed and Glover 1981a), which minimizes the maximum exterior deviation from this hyperplane, and hybrid methods which attempt to both minimize the exterior deviations and maximize the interior deviations (Freed and Glover 1986; Glover, Keene and Duea 1988; Glover 1988). A deviation is said to be external if its associated observation is misclassified (ie. falls on the wrong side of the hyperplane). Internal deviations refer to the extent to which an observation is correctly classified. Thus, external deviaitions are undersirable while internal deviaitons are desirable (Glover, Keene and Duea 1988; Glover 1988).

For experimental evaluations of the MP formulations' performance on simulated and real-world data, the reader is referred to studies by Bajgier and Hill (1982), Glorfeld and Olson (1982), Freed and Glover (1986b), Markowski and Markowski (1987), Joachimsthaler and Stam (1988), Koehler and Erenguc (1989), Rubin (1989), and those studies referred to in these papers.

While these MP techniques have provided very promising results in overcoming problems encountered by the standard statistical techniques, they are not without problems of their own. Three problems, in particular, have been known to plague the various MP formulations. These problems are: 1) unbounded solutions, 2) unacceptable solutions, and 3) improper solutions. A solution is unbounded if the objective function can be increased or decreased without limit. Obviously, in such a case no meaningful discriminant rule will result. Following Koehler (1989) we call a solution to the discriminant problem unacceptable "...if it generates a discriminant function of zeros, in which case all observations will be classified in the same group" (p. 241). An improper solution occurs if all observations in both groups fall on the classification hyperplane. In such a case, the objective of zero misclassification has been achieved but the resulting classification rule is meaningless and has no discriminatory power. Koehler (1989) also notes that none of the existing MP formulations overcome all of the above problems without creating new ones. This emphasizes the need for a simplified formulation which does not suffer from these problems.

In section 2 we will develop a simple MP formulation for the discriminant problem that unifies many features of the previous formulations. In section 3 we will show that this new formulation is not plagued by the previously mentioned problems. In section 4 we discuss methods of classification using this formulation. Specifically, we indicate how to deal with the issue of the so-called classification gap. Finally, in section 5 we will recommend directions for further research in this area.

2. A Unified MP Formulation

Consider the problem of two-group discriminant analysis. Suppose we have n observations on k

independent variables where n_1 of the observations belong to group 1 and n_2 belong to group 2. Let β_1 represent the $(n_1 \times k)$ matrix made up of the observations from group 1 and define β_2 similarly for group 2. Then following Freed and Glover (1981b) it would appear that the formulation in problem (I) can be used to determine a reasonable discriminant function for these two groups.

$$\text{(I)} \quad \text{MIN} \quad z = 1'd_1 + 1'd_2 \tag{1}$$

$$\text{S.T.} \quad \beta_1 x - \mathrm{I}d_1 \leq c1 \tag{2}$$

$$\beta_2 x + \mathrm{I}d_2 > c1 \tag{3}$$

$$d_1,\ d_2 \geq 0 \tag{4}$$

$$x,\ c \text{ unrestricted.} \tag{6}$$

In this formulation (also known as the MSD formulation) 1 represents a column vector of ones of conformable dimension, d_1 and d_2 are, respectively, $(n_1 \times 1)$ and $(n_2 \times 1)$ vectors of deviational variables, I represents an appropriately dimensioned identity matrix, and c is a real-valued variable. In (4), 0 represents an appropriately dimensioned column vector of zeros.

Intuitively, this formulation has considerable appeal as its solution, (x, c), identifies a hyperplane in R^k that either completely separates the two groups (if $z=0$) or minimizes the amount of misclassification if separation is not possible. Unfortunately, since the strict inequality in (3) cannot be directly enforced by the simplex method, this formulation always has an unacceptable (trivial) solution. Notice this trivial solution ($x=0$, $c=0$) may be produced even if an alternate optimal solution producing perfect separation with $x \neq 0$ exists. Various remedies to this problem have been suggested such as adding a linear equality constraint (ie. $a'x=1$) or non-convex constraints (such as $x'x=1$ or $\|x\|=1$) to prevent these solutions (Markowski and Markowski 1985; Freed and Glover 1986a; Koehler 1989). The linear constraint $a'x = 1$ implicitly eliminates any possible solution x, satisfying $a'x = 0$ from consideration and is therefore too restrictive. The non-convex constraints, on the other hand, make the resulting problem much more difficult to solve. Additionally, neither alternative does anything to prevent improper solutions from occurring. Thus, these alternatives are not appealling.

Consider the revised formulation of problem (I) given in (II) below:

$$\text{(II)} \quad \text{MIN} \quad z = 1'd_1 + 1'd_2 \tag{7}$$

$$\text{S.T.} \quad \beta_1 x - \mathrm{I}d_1 \leq c_1 1 \tag{8}$$

$$\beta_2 x + \mathrm{I}d_2 \geq c_2 1 \tag{9}$$

$$c_1 \leq c_2 - \epsilon \tag{10}$$

$$d_1, d_2 \geq 0 \tag{11}$$

$$x, c_1, c_2 \text{ unrestricted.} \tag{12}$$

The term ϵ in (10) represents an arbitrarily small positive number. The optimal solution to this problem (x^*, c_1^*, c_2^*) defines two hyperplanes ($\beta x^* \leq c_1^*$ and $\beta x^* \geq c_2^*$, where β is a $(1 \times k)$ vector of variables representing possible observations on our independent variables) that will be used for discrimination and/or classification purposes. A "gap" of size ϵ separates these hyperplanes.

It is easily seen that the objective in (7) will be minimized by taking c_2 as small as possible and c_1 as large as possible. Thus, (10) will be satisfied as a strict equality in the optimal solution. Therefore, we may substitute $c_2 - \epsilon$ for c_1 in (8) and let the scalar $x_0 = -c_2 + \epsilon$ to re-write (II) as follows:

$$\text{(IIa)} \quad \text{MIN} \quad z = 1'd_1 + 1'd_2 \tag{13}$$

$$\text{S.T.} \quad x_0 1 + \beta_1 x - \mathrm{I}d_1 \leq 0 \tag{14}$$

$$x_0 1 + \beta_2 x + \mathrm{I}d_2 \geq \epsilon 1 \tag{15}$$

$$d_1, d_2 \geq 0 \tag{16}$$

$$x_0, x \text{ unrestricted.} \tag{17}$$

It remains to select a value for the constant ϵ. Given that we have a general linear model it seems intuitively appealling to set $\epsilon = 1$. This is consistent with the related treatment of binary choice models in regression analysis in which the dependent variable is analogous to our right-hand-side values in (IIa) (Neter, Wasserman and Kutner 1989). In fact, it can be shown that for the two group case, regression of the independent variables on a binary dependent variable (coded for group membership) provides results equivalent to Fisher's LDF (specifically, the parameter estimates for the independent variables are proportional to Fisher's LDF). Additionally, it can be shown that the use of any other right-hand-side values in (II), say constants $\tilde{c}_1$ and $\tilde{c}_2$, produces an equivalent formulation provided $\tilde{c}_1 < \tilde{c}_2$.

3. Properties of the formulation

Before we consider the properties of the formulation in (IIa) with $\epsilon = 1$, let us reparameterize the model as shown in (III) so that $\overline{\beta}_1$ is a $(n_1 \times (k+1))$ matrix and includes the $\mathit{1}$ vector as its first column. Define $\overline{\beta}_2$ similarly. Thus, x will now be a $((k+1) \times 1)$ vector and includes the intercept term, x_0.

$$\text{(III)} \quad \text{MIN} \quad z = \mathit{1}'d_1 + \mathit{1}'d_2 \tag{18}$$

$$\text{S.T.} \quad \overline{\beta}_1 x - \mathrm{I}d_1 \leq \mathit{0} \tag{19}$$

$$\overline{\beta}_2 x + \mathrm{I}d_2 \geq \mathit{1} \tag{20}$$

$$d_1, d_2 \geq \mathit{0} \tag{21}$$

$$x \text{ unrestricted.} \tag{22}$$

Clearly this formulation cannot be unbounded since by (21), $z \geq 0$, and when perfect discrimination (or separation) is possible $z = 0$. If this occurs we are guaranteed an acceptable solution, for by (20) we cannot have $d_2 = \mathit{0}$ and $x = \mathit{0}$ simultaneously. In fact, unacceptable solutions ($x = \mathit{0}$) cannot occur using (III) unless $\overline{x}_1 = \overline{x}_2$ and $n_1 = n_2$ (in which case no discrimination is possible using linear methods). This is proven in the following theorem.

Theorem 1: If an unacceptable solution ($x = 0$) occurs using formulation (III) then the group centroids are equal and the sample sizes are equal (ie. $\overline{x}_1 = \overline{x}_2$ and $n_1 = n_2$).

Proof: Consider the dual of (III) given in (IIIa) below:

$$\text{(IIIa) MAX} \quad z_d = 0' w_1 + 1' w_2 \tag{23}$$

$$\text{S.T.} \quad -\overline{\beta}'_1 w_1 + \overline{\beta}'_2 w_2 = 0 \tag{24}$$

$$\mathrm{I} w_1 \leq 1 \tag{25}$$

$$\mathrm{I} w_2 \leq 1 \tag{26}$$

$$w_1, w_2 \geq 0 \tag{27}$$

By contradiction, suppose that (III) has an unacceptable solution ($x = 0$). Then by (20) and (18), $d_2 = 1$. Thus, $z = n_2$ in (18) and by duality theory $z_d = n_2$ in (23). Now since w_2 is of dimension $(n_2 \times 1)$, it follows by (26) that $w_2 = 1$. So from (24) we have:

$$\overline{\beta}'_1 w_1 = \overline{\beta}'_2 1. \tag{28}$$

The right-hand-side of this equation is a $((k+1) \times 1)$ vector composed of the $k+1$ sums of the columns of $\overline{\beta}_2$. Let $\overline{\beta}'_{\alpha_{ij}}$ denote the element of the i^{th} row and j^{th} column of $\overline{\beta}'_\alpha$ and w_{α_j} denote the j^{th} element in w_α. Considering the first row of (28) we have:

$$\sum_{j=1}^{n_1} \overline{\beta}'_{1_{1j}} w_{1_j} = \sum_{j=1}^{n_1} w_{1_j} = n_2. \tag{29}$$

Now if $n_1 \neq n_2$ we may define our groups such that $n_2 > n_1$. In this case, (29) cannot be satisfied since by (25), $w_1 \leq 1$. Thus, when $n_2 > n_1$, $x = 0$ is not a minimizing solution to (III). (In this case $x_0 = 1$, $x_j = 0$, $j = 1, ..., k$, is better than the trivial solution). When $n_1 = n_2$, (29) can only be satisfied by setting $w_1 = 1$ which by (28) implies $\overline{x}_1 = \overline{x}_2$. □

The proof of this theorem highlights two seemingly troublesome characteristics of the formulation in (III). First, if $\overline{x}_1 = \overline{x}_2$ we might really like to obtain the unacceptable solution of $x = 0$ to highlight the fact that no discrimination is possible using this linear method. However, if $n_1 = n_2$ and alternate optima exist we could potentially get a non-trivial solution and, indeed, will get a non-trivial solution if $n_1 \neq n_2$. These characteristics, however, are not indicative of a flawed formulation, but an erroneous application. Preliminary exploratory data analysis should include a comparison of group centroids to reveal if any difference really exists in the groups prior to the use of any discriminant procedure. If this analysis indicates $\overline{x}_1 \simeq \overline{x}_2$ or the centroids are not significantly different, linear methods should be abandoned and an analysis of the applicability of non-linear methods should ensue.

Secondly, although we have shown that by defining our groups so that $n_2 > n_1$ when the sample sizes are unequal we technically do prevent the trivial solution, this may seem to be of little real comfort if we instead obtain the near trivial solution $x_0 = 1$, $x_j = 0$, $j=1, ..., k$. It is important to remember, however, that the real issue is not just avoiding a trivial solution but having a formulation that avoids the trivial solution when a non-trivial solution provides as good, or better, discrimination. If a solution to (III) exists which provides perfect discrimination then $z = 0$ in (18). The trivial or near trivial solutions discussed above produce objective function values in (18) of $z > 0$. Thus, if perfect linear separation is possible, the corresponding solution will be selected over the trivial one by the solution procedure. In the same way, it is possible for formulation (III) to generate an improper solution of say, $\overline{\beta}_1 x = \epsilon 1$, $\overline{\beta}_2 x = \epsilon 1$. However, for any value of ϵ it is easy to show that $z > 0$ in (18). So again, if perfect linear separation is also possible, the corresponding non-trivial solution will be selected.

Another disturbing property of the traditional MSD and MMD formulations is that differing results may be obtained depending on where the data is located with respect to the origin. Suppose that we transform the values of the sample data on the independent variables in each group by equal amounts via $\tilde{\beta}_i = a\beta_i + 1b$, $i = 1, 2$. Here a is a non-zero scalar and b is a $(1 \times k)$ vector in which the j^{th} element is a constant indicating the amount by which we are shifting the values of the observations on the j^{th} independent variable. Intuitively it seems that this should not impact our ability to discriminate between the groups since we are transforming the values of the same variables in each group by equal amounts. Markowski and Markowski (1985) however, show that while such transformations (with $a = 1$) have no impact on the discrimination ability of Fisher's LDF they can have a significant impact on the discriminatory power of the traditional MSD and MMD formulations.

Transforming the data in this way, however, is not a problem with the formulation in (IIa). To see this, suppose we have an optimal solution to (IIa) given by (x_0^*, x^*) with an objective function value of z^*. Now suppose we replace β_i in (IIa) with $a\beta_i + 1b$, $(a \neq 0)$ and substitute y_0 and y for x_0 and x, respectively. After some simple algebra we have the following problem:

$$\text{(IIb)} \quad \text{MIN} \quad z = 1'd_1 + 1'd_2 \tag{30}$$

$$(y_0 + by)1 + a\beta_1 y - \mathrm{I}d_1 \leq 0 \tag{31}$$

$$(y_0 + by)1 + a\beta_2 y + \mathrm{I}d_2 \geq \epsilon 1 \tag{32}$$

$$d_1, d_2 \geq 0 \tag{33}$$

$$y_0, y \text{ unrestricted.} \tag{34}$$

If we let $\tilde{y}_0 = y_0 + by$ and $\tilde{y} = ay$ it is easy to see that the formulation in (IIb) is equivalent to the original formulation in (IIa). Thus, the optimal solution to the transformed problem in (IIb) is a linear function of the solution to problem (IIa) and is given by $(y_0, y) = (x_0^* - \frac{1}{a}bx^*, \frac{1}{a}x^*)$ with an objective function value of $z = z^*$. Since our formulation in (III) is equivalent to that in (IIa) (with $\epsilon = 1$), it is also insensitive to linear transformations of the data. To distinguish our model in (III) from the traditional MSD formulation we shall henceforth refer to it as the unified MSD or UMSD.

It is also interesting to study the properties of this formulation with an MMD objective. This unified MMD (UMMD) formulation is given as follows:

$$\text{(IV)} \quad \text{MIN} \quad z = d \tag{35}$$

$$\text{S.T.} \quad \overline{\beta}_1 x - d1 \leq 0 \tag{36}$$

$$\overline{\beta}_2 x + d1 \geq 1 \tag{37}$$

$$d \geq 0 \tag{38}$$

$$x \text{ unrestricted} \tag{39}$$

Most other MMD formulations have taken d in (38) to be unrestricted. However, if perfect separation is possible then there exists a solution x such that for any positive scalar $\alpha \geq 1$ we have:

$$\alpha \overline{\beta}_1 x \leq \overline{\beta}_1 x \leq 0 \tag{40}$$

$$\alpha \overline{\beta}_2 x \geq \overline{\beta}_2 x \geq 1 \tag{41}$$

So if left unrestricted, d can be made arbitrarily small as α increases resulting in an unbounded solution. This is prevented in (38) by restricting d to be non-negative.

To see that an unacceptable soluiton cannot occur using UMMD consider the dual of (IV) given by (IVa) below:

$$\text{(IVa)} \quad \text{MAX } z_d = 0' w_1 + 1' w_2 \tag{42}$$

$$\text{S.T.} \quad \overline{\beta}_1' w_1 - \overline{\beta}_2' w_2 = 0 \tag{43}$$

$$1' w_1 + 1' w_2 \leq 1 \tag{44}$$

$$w_1, w_2 \geq 0 \tag{45}$$

Now suppose there is an unacceptable solution to (IV) (ie. $x = 0$). Then by (35) and (37), $z = d = 1$ in (IV). Therefore, $z_d = 1'w_2 = 1$ in (IVa). Hence, by (44) $w_1 = 0$. However, there is no solution to (43) with $w_1 = 0$ and $1'w_2 = 1$ since the first row of $\overline{\beta}_2'$ is a (1 x n_2) vector of ones. Thus, an unacceptable solution to (IV) cannot occur. Also, if perfect separation is possible in (IV) then $z = 0$. This solution would obviously be selected over the trivial solution with $z=1$.

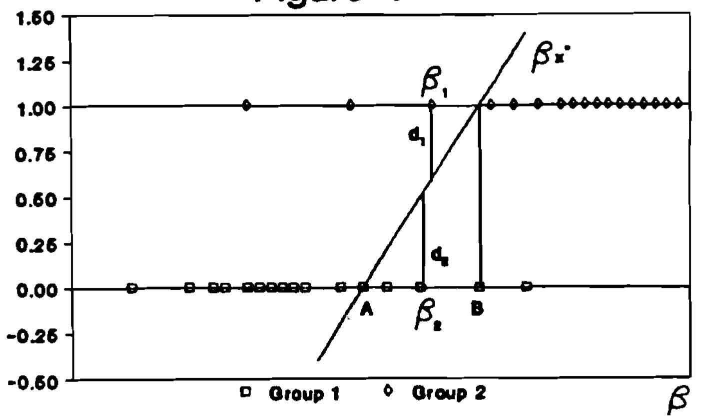
Similarly, it is easy to see that if an improper solution to (IV) occurs, so that $\overline{\beta}_1 x = \epsilon 1$ and $\overline{\beta}_2 x = \epsilon 1$, then $z = d > 0$. So if separation is also possible, the associated non-trivial solution with $z=0$ would be selected. Thus, our UMMD problem formulation does not suffer from the problems of unboundedness, improper or unacceptable solutions which were encountered in previous formulations (Markowski and Markowski 1985; Koehler 1989). The UMMD formulation can also be shown to be insensitive to linear transformations of the data in the same way as discussed above for the UMSD formulation.

4. The Classification Gap and Alternative Classification Methods

Previous methods for the discriminant problem have generally devised rules for classifying a new observation β as follows: If $\beta x^* \leq c^*$ classify the new observation as coming from population (group) 1, otherwise classify the new observation as coming from population (group) 2, where x^* and c^* are determined by the optimization procedure. Using the formulation in (III), classification is not as straightforward due to the gap created by setting $\epsilon=1$. This gap leaves us with an infinite number of possible cut-off values in the interval (0, 1] to use for classification purposes.

Since some of the justification for using the value $\epsilon=1$ is derived from the related statistical techniques, we might also look to these techniques for assistance in developing our classification rule. Graphically, the classification gap and the problem formulation in (III) can be represented as in Figure 1 for the case of one independent variable β.

Figure 1



When estimating the optimal solution x^* to (III), d_1 and d_2 represent the external deviations of the observations β_1 and β_2, respectively. The UMSD objective is to minimize the sum of such external deviations for all observations. Note that any observation i in group 2 with $\beta_i x^* \geq B$ or any observation i in group 1 with $\beta_i x^* \leq A$ have zero external deviation and do not fall in the classification gap. Any observation i with $0 < \beta_i x^* < 1$ falls in the region of the classification gap.

The question, therefore, is how to choose a cut-off value, c, such that an observation β_i is classified into group 1 if and only if $\beta_i x^* \leq c$. Following the statistical techniques, it is reasonable to use a cut-off value of $c = 0.5$, assuming equal costs of misclassification and equal prior probabilities. If the prior probabilities are proportional to the sample size of each group, a search procedure is recommended to find the cut-off value, c, which minimizes the number of misclassifications in the data set (Neter, Wasserman and Kutner 1989, p. 609).

Summarizing, we recommend the following methodology for establishing the classification rule:

STEP 1: Solve the UMSD model (III) (or the UMMD model (IV)) with $n_1 \leq n_2$ to find the optimal estimate x^* of x.

STEP 2: Use the appropriate criterion to determine a the cut-off value c (using either c=0.5 or the optimal cut-off value which minimizes the number of misclassified cases).

We then classify a new observation β_i as from population 1 if $\beta_i x^* \leq c$, and otherwise as from population 2. In a sense one can view the classification gap as the "fuzzy area" in which there is greater uncertainity involved in classifying observations as coming from one population versus the other.

An interesting extension of the UMSD and UMMD models described above is to consider a general

l_p-metric objective (with $1 \leq p \leq \infty$) in the optimization in STEP 1. The rationale for such an objective is that the MSD and the MMD criteria are special cases of such a general metric, with p=1 and p=∞, respectively (see Stam and Joachimsthaler 1989). The general l_p-metric objective for UMSD is defined as follows:

$$\text{MIN} \quad z_p = \left[\sum_{i=1}^{n_1} (d_{1_i})^p + \sum_{i=1}^{n_2} (d_{2_i})^p \right]^{1/p} \tag{46}$$

Let us define model (IIIb) as our UMSD formulation in (III) with a new objective function given by (46). The advantage of (IIIb) is that it is flexible, allowing a variety of data conditions to be modeled effectively by successively solving (IIIb) using different values of p. The discriminant function associated with the value of p which gives the best classification (in terms of the smallest objective function value or lowest total number of misclassifications) can then be selected. It is well-known that the MSD formulation is robust with respect to outliers, whereas the MMD formulation is very sensitive to outliers (Bajgier and Hill 1982; Glorfled and Olson 1982; Stam and Joachimsthaler 1989). Any metric with $1 \leq p \leq \infty$ will place a relative emphasis in between these two extremes on outlying observations. Stam and Ragsdale (1989) have done some preliminary work which suggests the UMSD formulation with an l_p-metric objective may be very promising.

5. Conclusions and directions for future research

We have introduced a unified MP formulation for solving the classification problem in discriminant analysis that does not seem to be plagued by problems associated with other formualtions. Our formulation is quite simple and does not involve normalization vectors or complicating non-linear constraints. Therefore, it provides a valuable contribution to the methodology of the MP-based approaches to discriminant analysis.

Future research should explore the extension of the current UMSD and UMMD formulations to a more general class of l_p-metric models and to models which include both internal and external deviations explicitly in the objective function. It appears that the latter can be achieved with a formulation similar to the hybrid model proposed by Glover, Keene and Duea (1988). The issue of the classification gap is well-known in the statistics literature and provides another interesting research direction to explore. This would provide a better understanding of the nature and interpretation of the gap associated with our UMSD and UMMD formulations.

References

[1] Bajgier, S.M. and A.V. Hill, "An Experimental Comparison of Statistical and Linear Programming Approaches to the Discriminant Problem," *Decision Sciences*, 1982, 13, 604-618.

[2] Fisher, R.A., "The Use of Multiple Measurements in Taxonomy Problems," *Annals of Eugenics*, 1936, 7, 179-188.

[3] Freed, N. and F. Glover, "A Linear Programming Approach to the Discriminant Problem," *Decision Sciences*, 1981a, 12, 68-74.

[4] Freed, N. and F. Glover, "Simple But Powerful Goal Programming Formulations for the Discriminant Problem," *European Journal of Operational Research*, 1981b, 7, 44-60.

[5] Freed, N. and F. Glover, "Resolving Certain Difficulties and Improving the Classification Power of LP Discriminant Analysis Formulations," *Decision Sciences*, 1986a, 17, 589-595.

[6] Freed, N. and F. Glover, "Evaluating Alternative Linear Programming Models to Solve the Two-Group Discriminant Problem," *Decision Sciences*, 1986b, 17, 151-162.

[7] Glorfeld, L.W. and D.L. Olson, "Using the L-1 Metric for Robust Analysis of the Two-Group Discriminant Problem," *Proceedings of the American Institute of the Decision Sciences*, 2, San Francisco, 1982, 297-298.

[8] Glover, F., "Improved Linear and Integer Programming Models for Discriminant Analysis," Working Paper, Center of Applied Artificial Intelligence, University of Colorado, Boulder, Co., 1988.

[9] Glover, F., S. Keene and B. Duea, "A New Class of Models for the Discriminant Problem," *Decision Sciences*, 1988, 19, 269-280.

[10] Hand, D.J., *Discrimination and Classification*, 1981, Wiley, New York.

[11] Joachimsthaler, E.A. and A. Stam, "Four Approaches to the Classification Problem in Discriminant Analysis - An experimental study," *Decision Sciences*, 1988, 19, 322-333.

[12] Koehler, G.J., "Characterization of Unacceptable Solutions in LP Discriminant Analysis," *Decision Sciences*, 1989, 20, 239-257.

[13] Koehler, G.J. and S.S. Erenguc, "Minimizing Misclassifications in Linear Discriminant Analysis," Working Paper, College of Business Administration, University of Florida, 1988.

[14] Markowski, E.P. and C.A. Markowski, "Some Difficulties and Improvements in Applying Linear Programming Formulations to the Discriminant Problem," *Decision Sciences*, 1985, 16, 237-247.

[15] Markowski, C.A. and E.P. Markowski, "An Experimental Comparison of Several Approaches to the Discriminant Problem With Both Qualitative and Quantitative Variables," *European Journal of Operational Research*, 1987, 28, 74-87.

[16] Neter, J., W. Wasserman and M.H. Kutner, *Applied Linear Regression Models*, Second Edition, 1989, Irwin, Homewood, Illinois.

[17] Rubin, P.A., "Evaluating the Maximize Minimum Distance Formulation of the Linear Discriminant Problem," *European Journal of Operational Research*, 1989, forthcoming.

[18] Smith, C.A.B., "Some Examples of Discrimination," *Annals of Eugenics*, 1947, 13, 272-282.

[19] Stam, A. and E.A. Joachimsthaler, "Solving the Classification Problem in Discriminant Analysis via Linear and Nonlinear Programming Methods," *Decision Sciences*, 1989, 20, 285-293.

[20] Stam, A. and C.T. Ragsdale, "A Robust Nonparametric Procedure to Estimate Response Functions for Binary Choice Models," forthcoming in *Operations Research Letters*.