CR-134158

JSC - 08412
JSC INTERNAL TECHNICAL NOTE

ON MINIMIZING THE PROBABILITY OF MISCLASSIFICATION FOR LINEAR FEATURE SELECTION

by

L.F. Guseman, Jr.*
Department of Mathematics
Texas A & M University
College Station, Texas 77843

and

Homer F. Walker**
Department of Mathematics
Texas Tech University
Lubbock, Texas 79409

For

Earth Observations Division

* NASA-ASEE Summer Faculty Fellow

** Supported under NASA Contract NAS-9-12777 with University of Houston

(NASA-CR-134058) ON MINIMIZING THE PROBABILITY OF MISCLASSIFICATION FOR LINEAR FEATURE SELECTION (Houston Univ.) 42 p HC $4.25 CSCL 12A

N73-32486

Unclas

G3/19 18203

National Aeronautics and Space Administration

LYNDON B. JOHNSON SPACE CENTER

Houston, Texas

August 1973

Distribution:
FA2/E. Davis
FM8/H. DeVezin
A. Actkinson
W. Coberly
A. Wylie
FS5/J. Parker
T. Price
GA/R. Piland
S. Whitley
HA/C. Charlesworth
G. Smith
F. Ravet
J. Mitchell
JM2/(3 copies)
JM6/R. Shirkey (4)
TA/A. Calio
TF/R. MacDonald
TF2/R. Erb
TF3/A. Potter
SR&T File (2)
TF4/J. Overton
TF5/A. Patteson
TF6/T. Kloves
TF7/J. Dragg
K. Baker (3)
A. Feiveson
F. Hall
F. Johnson
M. Trichel
G. Lendaris
LEC/C. Clouse
C. Chang
E. Kan
T. Minter
S. Yao
W. Eppler
Philco/G. Austin
TRW/J. Quirein
IBM/G. Meyers
M. Felix
D. Ingram
NASA MSFC, C. Messer, S&D-DIR
NASA JPL, F. Billingsley, 168-427
NASA GSFC, W. Alford, 563.0

Dr. Henry Decell (2)
Department of Mathematics
University of Houston
Houston, TX 77004

Dr. Tom Boullion
Dr. Ben Duran
Department of Mathematics
Texas Tech University
Post Office Box 4319
Lubbock, TX 79409

Dr. Patrick L. Odell
Dr. Raj Chhikara
University of Texas at Dallas
Post Office Box 688
Richardson, TX 75080

Dr. M. Stuart Lynn
Dr. David Van Rooy
Institute for Computer Services
and Applications
Rice University
Houston, TX 77001

Dr. R. J. P. de Figueiredo
Mathematical Sciences Department
Rice University
Houston, TX 77001

Dr. Richard Northouse
Department of Elec. Eng.
and Computer Science
University of Wisconsin
Milwaukee, WI 53201

Dr. James A. Smith
Department of Watershed Sciences
Colorado State University
Fort Collins, CO 80521

Dr. Robert M. Haralick
Remote Sensing Laboratory
Center for Research in Engineering
Science
University of Kansas
Lawrence, KS 66044

Dr. David Landgrebe
Dr. Phillip H. Swain
Mr. Terry L. Phillips
Laboratory for the Application of
Remote Sensing
Purdue University
1220 Potter Drive
W. Lafayette, IN 47906

Dr. Ralph K. Cavin, III
Dr. Darrell R. Gimlin
Department of Electrical Engineering
Texas A&M University
College Station, TX 77843

Dr. Lawrence F. Guseman (2)
Department of Mathematics
Texas A&M University
College Station, TX 77843

Dr. Jon Erickson
Mr. Robert Crane
Mr. Robert Marshall
Environmental Research Institute of Michigan
Post Office Box 618
Ann Arbor, MI 48107

Dr. Dennis L. Tebbe
Department of Electrical Engineering
University of Missouri
Colombia, MO 65201

Dr. Larry J. Stephens
Mathematics Department
Kearney State College
Kearney, NB 68847

Dr. Ricahrd E. Haskell
School of Engineering
Oakland University
Rochester, MI 48063

Dr. Homer F. Walker (2)
Mathematics Department
University of Denver
University Park
Denver, CO 80210

Dr. Cecil Hallum
Department of Mathematical Sciences
Loyola University
6363 St. Charles Avenue
New Orleans, LA 70118

ON MINIMIZING THE PROBABILITY OF

MISCLASSIFICATION FOR LINEAR FEATURE SELECTION

Prepared By:

L. F. Guseman, Jr.

L. F. Guseman, Jr., NASA-ASSE

Homer F. Walker

Homer F. Walker, NAS-9-12777, University of Houston

Approved by:

James L. Dragg

Deputy Chief, Applications Analysis Branch

Approved by:

R. B. MacDonald

Chief, Earth Observations Division

ABSTRACT

The use of techniques for feature selection allows one to treat classification problems in spaces of reduced dimensions. This note considers a method of linear feature selection for n-dimensional observation vectors which belong to one of two populations, where each population is described by a known multivariate normal density function. More specifically, we consider the problem of finding a $1 \times n$ transformation matrix B for which the probability of misclassification with respect to the one-dimensional transformed density functions is minimized. Theoretical results are presented which give rise to a numerically tractable expression for the variation in the probability of misclassification with respect to B. Using this expression we discuss a computational procedure for obtaining a B which minimizes the probability of misclassification. Preliminary numerical results are discussed.

TABLE OF CONTENTS

Preface iv

1. Introduction 1

2. Differentiating the Probability of Misclassification 5

3. Computational Procedure 15

4. Preliminary Numerical Results 24

5. Concluding Remarks 34

References 35

PREFACE

Multispectral Scanners have been developed to remotely collect data (from aircraft and spacecraft) from which earth resources information can be extracted. The analysis and interpretation of this data requires sophisticated mathematical techniques. The ability to use the data for specific applications depends to a large extent upon the accuracy and speed of the analytical techniques together with the ability to compute complicated mathematical expressions.

The specific problem addressed in this report is that of combining a given set of spectral features, to obtain a smaller set of spectral features which retain "to the greatest possible degree" the information inherent in the original measurements. Combining spectral features while retaining inherent information, is a preclassification technique used to reduce prohibitive data storage and computation time requirements encountered in the classification technique itself. A technique is developed in this report for combining spectral features in such a way that the optimal retention of information is accomplished by minimizing the probability of incorrectly classifying observations.

The mathematical expression for computing the probability that an observation will be incorrectly classified is complicated. However, it is generally agreed that it is both theoretically and empirically the best criterion by which one can measure information degradation when attempting to combine spectral feature.

ON MINIMIZING THE PROBABILITY OF

MISCLASSIFICATION FOR LINEAR FEATURE SELECTION

1. Introduction

Consider two populations Π_1 and Π_2 with associated multivariate normal density functions defined for $x = (x_1,\ldots,x_n)^T \in E^n$ by

$$p_i(x) = (2\pi)^{-n/2}|\Sigma_i|^{-1/2}\exp(-\frac{1}{2}(x-\mu_i)^T\Sigma_i^{-1}(x-\mu_i)), \quad i = 1,2.$$

If $B = (b_1,\ldots,b_n)$ is a nonzero $1\times n$ vector and $x \in E^n$, then $Bx \in E^1$ and the populations Π_1 and Π_2 have transformed normal density functions defined for $y \in E^1$ by

$$p_i(y,B) = (2\pi)^{-1/2}(B\Sigma_i B^T)^{-1/2}\exp(-\frac{(y-B\mu_i)^2}{2B\Sigma_i B^T}), \quad i = 1,2.$$

The linear feature selection problem considered in the sequel is to choose a B which minimizes the probability of misclassification of a transformed observation in E^1 using a Bayes optimal (maximum likelihood) classification scheme. If the a priori probabilities that an observation comes from either Π_1 or Π_2 are equal, then the transformed probability of misclassification in E^1, as a function of B, denoted by g, is given [1] by

$$g(B) = \frac{1}{2}\int_{R_1(B)} p_2\,(y,B)\,dy + \frac{1}{2}\int_{R_2(B)} p_1(y,B)\,dy,$$

where

$$R_1(B) = \{y \in E^1 : p_1(y,B) \geq p_2(y,B)\}$$

and

$$R_2(B) = \{y \in E^1 : p_1(y,B) < p_2(y,B)\}.$$

(If $p_1(y,B) \equiv p_2(y,B)$, we define $g(B) = \frac{1}{2}$). If g is to be minimized as a function of B, it is natural to ask whether g is a differentiable function of the elements of B. If such be the case, then the minimum value of g will occur only if these derivatives all vanish.

In the sequel it is shown that g is a differentiable function of the elements of B, and a formula for its derivatives is given. A method for obtaining numerically a B which minimizes g is discussed. Results obtained, using statistics from C-1 Flight Line Data as population parameters, are presented.

We remark that a more general result is given in [2] concerning the differentiability of g when B is a $k\times n$ matrix of rank k. Unfortunately, when k is greater than 1, the result of [2] does not always guarantee the differentiability of g. Moreover, when g is differentiable, the formulas for its derivatives are not numerically tractable.

We have found it convenient to work with the Gateaux differential of g at B with increment C, denoted by $\delta g(B;C)$, and defined (if the limit exists) by

$$\delta g(B;C) = \lim_{s\to 0} \frac{g(B+sC) - g(B)}{s}$$

for a $1\times n$ vector C. If for a given B the above limit exists for each $1\times n$ vector C, then g is said to be Gateaux differentiable at B [3, p. 171]. If g is Gateaux differentiable at B, then the derivative of g with respect to, say, the j^{th} component of B is given by $\delta g(B;C_j)$, where C_j is the $1\times n$ vector with a 1 in the j^{th} slot and zeros elsewhere. Similarly, if B is a nonzero $1\times n$ vector, and C is a $1\times n$ vector, we define

$$\delta p_i(y,B;C) = \lim_{s\to 0} \frac{p_i(y,B+sC) - p_i(y,B)}{s}, \qquad i = 1,2.$$

We also define

$$F(y,B) = \log \frac{p_1(y,B)}{p_2(y,B)}$$

$$= \frac{1}{2} \log \frac{B\Sigma_2 B^T}{B\Sigma_1 B^T} - \frac{(y-B\mu_1)^2}{2\, B\Sigma_1 B^T} + \frac{(y-B\mu_2)^2}{2\, B\Sigma_2 B^T}.$$

Then

$$R_1(B) = \{y \in E^1 : F(y,B) \geq 0\},$$

$$R_2(B) = \{y \in E^1 : F(y,B) < 0\},$$

and we let

$$S(B) = \{y \in E^1 : F(y,B) = 0\}.$$

Note that since $F(y,B)$ is a quadratic function of y, $S(B)$ consists of at most two points.

2. Differentiating The Probability of Misclassification

In this section we show that for a nonzero $1\times n$ vector B, $\delta g(B;C)$ exists for each $1\times n$ vector C. We also obtain a formula for $\delta g(B;C)$ which is numerically tractable.

THEOREM 1. Let B be a nonzero $1\times n$ vector. Then $\delta g(B;C)$ exists for each $1\times n$ vector C and is given by

$$\delta g(B;C) = \begin{cases} 0, & \text{if } p_1(y,B) \equiv p_2(y,B) \\ \frac{1}{2}\int\limits_{R_1(B)} \delta p_2(y,B;C)\,dy + \frac{1}{2}\int\limits_{R_2(B)} \delta p_1(y,B;C)\,dy, & \text{if } p_1(y,B) \not\equiv p_2(y,B) \end{cases}$$

Proof: If $p_1(y,B) \equiv p_2(y,B)$, then g has attained its maximum (namely ...) as a function of B. Applying the techniques developed in the remainder of the proof, one can show that g is Gateaux differentiable for this B. Therefore $\delta g(B;C) = 0$ for all C.

If $p_1(y,B) \not\equiv p_2(y,B)$, then for small $s \neq 0$, we have

$$\begin{aligned} \frac{g(B+sC) - g(B)}{s} &= \frac{1}{2s}\left\{ \int\limits_{R_1(B+sC)} p_2(y,B+sC)\,dy + \int\limits_{R_2(B+sC)} p_1(y,B+sC)\,dy - \int\limits_{R_1(B)} p_2(y,B)\,dy - \int\limits_{R_2(B)} p_1(y,B)\,dy \right\} \\ &= \frac{1}{2}\int\limits_{R_1(B+sC)} \frac{p_2(y,B+sC) - p_2(y,B)}{s}\,dy + \frac{1}{2}\int\limits_{R_2(B+sC)} \frac{p_1(y,B+sC) - p_1(y,B)}{s}\,dy \\ &+ \frac{1}{2}\int\limits_{R_1(B+sC)\sim R_1(B)} \frac{p_2(y,B) - p_1(y,B)}{s}\,dy + \frac{1}{2}\int\limits_{R_1(B)\sim R_1(B+sC)} \frac{p_1(y,B) - p_2(y,B)}{s}\,dy. \end{aligned}$$

It is readily verified that

$$\lim_{s\to 0}\left\{\frac{1}{2}\int_{R_1(B+sC)}\frac{p_2(y,B+sC)-p_2(y,B)}{s}\,dy+\frac{1}{2}\int_{R_2(B+sC)}\frac{p_1(y,B+sC)-p_1(y,B)}{s}\,dy\right\}$$

$$=\frac{1}{2}\int_{R_1(B)}\delta p_2(y,B;C)\,dy+\frac{1}{2}\int_{R_2(B)}\delta p_1(y,B;C)\,dy.$$

Then the theorem will be proved if it can be shown that

$$\lim_{s\to o}\int_{R_1(B)\sim R_1(B+sC)}\frac{p_1(y,B)-p_2(y,B)}{s}\,dy=\lim_{s\to o}\int_{R_1(B+sC)\sim R_1(B)}\frac{p_2(y,B)-p_1(y,B)}{s}\,dy=0.$$

We show that the first limit exists and is equal to zero. The second limit is handled similarly.

First note that $S(B)$ is the set of solutions of the equation

$$(*)\quad F(y,B)=\alpha(B)y^2+2\beta(B)y+\gamma(B)=0,$$

where

$$\alpha(B)=B(\Sigma_1-\Sigma_2)B^T,$$

$$\beta(B)=-B(\Sigma_1B^TB\mu_2-\Sigma_2B^TB\mu_1),$$

and

$$\gamma(B) = B\Sigma_1 B^T (B\mu_2)^2 - B\Sigma_2 B^T (B\mu_1)^2 + (B\Sigma_1 B^T)(B\Sigma_2 B^T)\log \frac{B\Sigma_2 B^T}{B\Sigma_1 B^T} .$$

We make the following observation: Since $p_1(y,B) \not\equiv p_2(y,B)$, $F(y,B)$ must take on both positive and negative valves. (Otherwise, we would have either $p_1(y,B) \geq p_2(y,B)$ or $p_1(y,B) \leq p_2(y,B)$ for all y. Since both $p_1(y,B)$ and $p_2(y,B)$ are continuous and have integral 1, either eventuality would imply $p_1(y,B) \equiv p_2(y,B)$.) From this observation, we see that if $\alpha(B) \neq 0$, then (*) has two distinct real solutions, and if $\alpha(B) = 0$, then $\beta(B) \neq 0$ and (*) has exactly one solution.

If $\alpha(B) \neq 0$, let $y_+(B)$ and $y_-(B)$ denote the two distinct real solutions of (*). Since $F(y,B)$ is quadratic in y and (*) has distinct solutions, it follows that $\frac{\partial}{\partial y} F(y_+(B),B) \neq 0$ and $\frac{\partial}{\partial y} F(y_-(B),B) \neq 0$. Then, by the Implicit Function Theorem, there exist unique functions $y_+(s)$ and $y_-(s)$, defined and continuously differentiable for small s, which satisfy $y_+(0) = y_+(B)$, $y_-(0) = y_-(B)$, and $F(y_+(s), B+sC) = F(y_-(s), B+sC) = 0$ for small s. This is to say that the points of $S(B + sC)$ vary in a continuously differentiable way for small s, and it follows that, for small s, $R_1(B) \sim R_1(B + sC)$ is an interval or pair of intervals not exceeding $K|s|$ in length (for an appropriate constant K independent of s). Then, for small s, there exists a constant K' for which

$$\left| \int_{R_1(B)\sim R_1(B+sC)} \frac{p_1(y,B) - p_2(y,B)}{s} \right| \leq K' \max_{y \in R_1(B)\sim R_1(B+sC)} \left| p_1(y,B) - p_2(y,B) \right| .$$

Since $R_1(B) \sim R_1(B+sC)$ is composed of intervals of length not exceeding $K|s|$ and having either $y_+(B)$ or $y_-(B)$ as an endpoint, and since $p_1(y_+(B),B) - p_2(y_+(B),B) = p_1(y_-(B),B) - p_2(y_-(B),B) = 0$, we have

$$\lim_{s\to 0} \int_{R_1(B)\sim R_1(B+sC)} \frac{p_1(y,B) - p_2(y,B)}{s} = 0$$

as desired.

If $\alpha(B) = 0$, then (*) has exactly one solution $y(B) = \frac{B\mu_1 + B\mu_2}{2}$. Since $\alpha(B+sC)$ is a polynomial in s, either $\alpha(B+sc) \equiv 0$ for all s or $\alpha(B+sC) \neq 0$ in some deleted neighborhood of $s = 0$. If $\alpha(B+sC) \equiv 0$ for all s, then $y(s) = \frac{(B+sC)\mu_1 + (B+sC)\mu_2}{2}$ is the unique continuously differentiable function of s satisfying $y(0) = y(B)$ and $F(y(s), B+sC) = 0$. Reasoning as in the paragraph above, we obtain

$$\lim_{s\to 0} \int_{R_1(B)\sim R_1(B+sC)} \frac{p_1(y,B) - p_2(y,B)}{s} = 0.$$

If $\alpha(B + sC) \neq 0$ in some deleted neighborhood of $s = 0$, then, for small $s \neq 0$, the roots of (*) are

$$y_\pm(s) = \frac{-\beta(B+sC) \pm \sqrt{\beta(B+sC)^2 - \alpha(B+sC)\gamma(B+sC)}}{\alpha(B + sC)}.$$

Without loss of generality, we assume that $-\beta(B+sC) > 0$ for small s. Then, for small $s \neq 0$, we have that $|y_{\pm}(s)| \geq \frac{K}{|s|}$ for an appropriate constant K independent of s. Furthermore, we have that

$$y_{-}(s) = \frac{-\beta(B+sC) - \sqrt{\beta(B+sC)^2 - \alpha(B+sC)\,\gamma(B+sC)}}{\alpha(B+sC)}$$

$$= \frac{\gamma(B+sC)}{2\sqrt{\beta(B)^2}} + O(s)$$

$$= \frac{B\mu_2 + 3\mu_1}{2} + O(s) = y(B) + O(s)$$

for small s. Consequently, we can find a constant K' for which

$$\left| \int_{R_1(B)\sim R_1(B+sC)} \frac{p_1(y,B) - p_2(y,B)}{s} \right| \leq K' \max_{|y-y(B)| \leq K'|s|} \left| p_1(y,B) - p_2(y,B) \right| +$$

$$+ \frac{1}{|s|} \int_{|y| \geq \frac{K}{|s|}} |p_1(y,B) - p_2(y,B)|$$

for small $s \neq 0$. Since $p_1(y(B),B) - p_2(y(B),B) = 0$ and $|p_1(y,B) - p_2(y,B)|$ approaches zero exponentially as $|y|$ becomes large, it follows that

$$\lim_{s \to 0} \int_{R_1(B)\sim R_1(B+sC)} \frac{p_1(y,B) - p_2(y,B)}{s} = 0.$$

This completes the proof of Theorem 1.

According to Theorem 1, for nonzero B the Gateaux differential $\delta g(B;C)$ exists and can be calculated using the Gateaux differentials $\delta p_1(y,B;C)$ and $\delta p_2(y,B;C)$ of the transformed density functions. In the lemma below, we calculate $\delta p_i(y,B;C)$. For convenience, we omit subscripts.

Lemma. Let B be a nonzero $1\times n$ vector. Then

$$\delta p(y,B;C) = -p(y,B)\left\{\frac{C\Sigma B^T}{B\Sigma B^T} - \frac{C\mu}{B\Sigma B^T}(y-B\mu) - \frac{C\Sigma B^T}{(B\Sigma B^T)^2}(y-B\mu)^2\right\}$$

for each $1\times n$ vector C.

Proof: We have

$$p(y,B) = (2\pi)^{-1/2} f_o(B)^{-1/2}\exp(-\tfrac{1}{2} f_1(B)),$$

where $f_o(B) = B\Sigma B^T$ and $f_1(B) = \dfrac{(y - B\mu)^2}{B\,\Sigma\,B^T}$. It is easily verified that

$$\delta p(y,B;C) = (2\pi)^{-1/2}\{-\tfrac{1}{2} f_o(B)^{-3/2}\delta f_o(B;C)\exp(-\tfrac{1}{2}f_1(B)) -$$

$$-\tfrac{1}{2}f_o(B)^{-1/2}\exp(-\tfrac{1}{2}f_1(B))\delta f_1(B;C)\}$$

$$= -\tfrac{1}{2}p(y,B)\{f_o(B)^{-1}\delta f_o(B;C) + \delta f_1(B;C)\}$$

and that $\delta f_o(B;C) = 2C\Sigma B^T$. Now $f_1(B) = \dfrac{(y - B\mu)^2}{f_o(B)}$, and one sees without difficulty that

$$\delta f_1(B;C) = \frac{-2C\mu}{f_o(B)}\ (y-B\mu) - \frac{(y - B\mu)^2}{f_o(B)^2}\ \delta f_o(B;C)$$

$$= \frac{-2C\mu}{B\Sigma B^T}(y-B\mu) - \frac{2C\Sigma B^T}{(B\Sigma B^T)^2}\ (y - B\mu)^2.$$

After a brief calculation, the lemma follows.

Theorem 1 and the above lemma provide an explicit formula for $\delta g(B;C)$. Unfortunately, this formula does not lend itself to easy computation because of the integrals that appear. Remarkably enough, a short calculation yields the formula of Theorem 2 below, in which no integrals appear. In order to summarize the results of this section, we incorporate some of the statement of Theorem 1 in the statement of Theorem 2. We also recall that, if $p_1(y,B) \neq p_2(y,B)$, then either $S(B) = \{a\}$ or $S(B) = \{a_-,a_+\}$ for some a,a_-,a_+ in E^1 (with $a_- < a_+$). In the statement of the theorem, we use the notation

$$f(y)\Big|_{S(B)} = \begin{cases} f(a) & \text{if } S(B) = \{a\} \\ f(a_+) - f(a_-) & \text{if } S(B) = \{a_+,a_-\} \end{cases}$$

for a function $f(y)$ on E^1.

<u>Theorem 2:</u> Let B be a non-zero $1\times n$ vector. Then $\delta g(B;C)$ exists for each $1\times n$ vector C and is given by

$$\delta g(B;C) = \begin{cases} 0 \quad \text{if} \quad p_1(y,B) \equiv p_2(y,B) \\ \pm\, p_1(y,B)[C(\mu_2-\mu_1) + \dfrac{C\Sigma_2B^T}{B\Sigma_2B^T}(y-B\mu_2) - \dfrac{C\Sigma_1B^T}{B\Sigma_1B^T}(y-B\mu_1)]\Big|_{S(B)} \\ \text{if} \quad p_1(y,B) \not\equiv p_2(y,B), \quad \text{where the sign is taken to be} \\ \lim\limits_{y\to\infty} [\text{sign} \log \dfrac{p_1(y,B)}{p_2(y,B)}]. \end{cases}$$

Proof: We already know from Theorem 1 that, for nonzero B, $\delta g(B;C)$ exists for all C and $\delta g(B;C) = 0$ if $p_1(y,B) \equiv p_2(y,B)$. If $p_1(y,B) \not\equiv p_2(y,B)$, then, according to Theorem 1,

$$\delta g(B;C) = \int_{R_1(B)} \delta p_2(y,B;C) + \int_{R_2(B)} \delta p_1(y,B;C).$$

It must be shown that this expression is the same as the corresponding expression in the statement of Theorem 2. We show below that

$$\int_{R_1(B)} \delta p_2(y,B;C) = \pm p_1(y,B)[C\mu_2 + \frac{C\Sigma_2B^T}{B\Sigma_2B^T}(y-B\mu_2)]\Big|_{S(B)}$$

where the sign is taken to be $\lim\limits_{y\to\infty} [\text{sign} \log \dfrac{p_1(y,B)}{p_2(y,B)}]$. For simplicity, we assume that the equation $F(y,B) = 0$ has only one root, i.e., $S(B) = \{a\}$

for some $a \in E^1$, and that $R_1(B) = (-\infty, a]$. The remaining cases and expressions are dealt with similarly.

From the preceding lemma, we obtain

$$\int_{R_1(B)} \delta p_2(y,B;C)\,dy = -\frac{C\Sigma_2 B^T}{B\Sigma_2 B^T} \int_{-\infty}^{a} p_2(y,B)\,dy + \frac{C\mu_2}{B\Sigma_2 B^T} \int_{-\infty}^{a} (y-B\mu_2)\,p_2(y,B)\,dy$$

$$+ \frac{C\Sigma_2 B^T}{(B\Sigma_2 B^T)^2} \int_{-\infty}^{a} (y-B\mu_2)^2\,p_2(y,B)\,dy.$$

Integrating by parts, we obtain

$$\frac{C\Sigma_2 B^T}{(B\Sigma_2 B^T)^2} \int_{-\infty}^{a} (y-B\mu_2)^2\,p_2(y,B)\,dy = -\frac{C\Sigma_2 B^T}{B\Sigma_2 B^T}(y-B\mu_2)\,p_2(y,B)\Big|_{y=a} +$$

$$+ \frac{C\Sigma_2 B^T}{B\Sigma_2 B^T} \int_{-\infty}^{a} p_2(y,B)\,dy$$

and

$$\frac{C\mu_2}{B\Sigma_2 B^T} \int_{-\infty}^{a} (y-B\mu_2)\,p_2(y,B)\,dy = C\mu_2\,p_2(y,B)\Big|_{y=a}.$$

Since $p_2(a,B) = p_1(a,B)$, substitution gives the desired result

$$\int_{R_1(B)} \delta p_2(y,B;C)\,dy = -p_1(y,B)\left[C\mu_2 + \frac{C\Sigma_2 B^T}{B\Sigma_2 B^T}(y-B\mu_2)\right]\Bigg|_{S(B)},$$

and the proof of Theorem 2 is complete.

3. Computational Procedure

In this section we present a method for computing a $1\times n$ vector B which minimizes the probability of misclassification g. It is well-known [3, p. 178] that when B is an extremum of g, then $\delta g(B;C) = 0$ for each $1\times n$ vector C. It follows that if B minimizes g, then B must satisfy the vector equation

$$\frac{\partial g}{\partial B} \equiv \begin{pmatrix} \delta g(B;C_1) \\ \vdots \\ \delta g(B;C_n) \end{pmatrix} = \begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix},$$

where C_j, $1 \le j \le n$, is a $1\times n$ vector with a one in the $j^{\underline{th}}$ slot and zeros elsewhere. Our method consists of using a numerically tractable formula for $\frac{\partial g}{\partial B}$, which we obtain from Theorem 2, in a Davidon-Fletcher-Powell iterative procedure, SUBROUTINE DAVIDN, for finding a local minimum of g.

Assuming that both the $n\times 1$ mean vectors μ_1 and μ_2 and the $n\times n$ covariance matrices Σ_1 and Σ_2 are known, we describe below the way in which the necessary functions are computed for SUBROUTINE DAVIDN.

To compute the error function

$$\Phi(a) = (2\pi)^{-1/2} \int_{-\infty}^{a} \exp(-\tfrac{1}{2} t^2)dt,$$

a double precision function, DPHI, is used.

For a given $1 \times n$ vector B, let

$$D_i(a,B) = \int_{-\infty}^{a} p_i(y,B)\,dy, \quad i = 1,2, \quad -\infty < a < \infty .$$

In computing the values of $D_i(a,B)$, one uses the relationship

$$D_i(a,B) = \Phi\left(\frac{a - B\mu_i}{(B\Sigma_i B^T)^{1/2}}\right), \quad i = 1,2.$$

After computing the scalars $B\mu_1$, $B\mu_2$, $B\Sigma_1 B^T$, and $B\Sigma_2 B^T$, we solve the quadratic equation

$$\alpha(B)y^2 + 2\beta(B)y + \gamma(B) = 0,$$

where

$$\alpha(B) = B\Sigma_1 B^T - B\Sigma_2 B^T$$

$$\beta(B) = (B\Sigma_2 B^T)B\mu_1 - (B\Sigma_1 B^T)B\mu_2,$$

and

$$\gamma(B) = (B\Sigma_1 B^T)(B\mu_2)^2 - (B\Sigma_2 B^T)(B\mu_1)^2 + (B\Sigma_1 B^T)(B\Sigma_2 B^T)\log\frac{B\Sigma_2 B^T}{B\Sigma_1 B^T} .$$

As noted in the proof of Theorem 1, this quadratic has either a single root or else two distinct real roots. We treat these cases separately.

Single Root Case

The quadratic equation has a single root precisely when $\alpha(B) = 0$, that is, when the transformed covariances are equal. If a denotes the single root, then it is easily verified that

$$a = \frac{B\mu_1 + B\mu_2}{2},$$

$$g(B) = \begin{cases} \frac{1}{2} - \frac{1}{2}[D_1(a,B) - D_2(a,B)], & \text{if } B\mu_1 < a \\ \frac{1}{2} + \frac{1}{2}[D_1(a,B) - \Gamma\ (a,B)], & \text{if } B\mu_2 < a, \end{cases}$$

and

$$\frac{\partial g}{\partial B} = \begin{cases} \mu_1 - \mu_2 - \frac{(\Sigma_1 + \Sigma_2)B^T}{B\Sigma_1 B^T + B\Sigma_2 B^T}(B\mu_1 - B\mu_2), & \text{if } B\mu_1 < a \\ \mu_2 - \mu_1 + \frac{(\Sigma_1 + \Sigma_2)B^T}{B\Sigma_1 B^T + B\Sigma_2 B^T}(B\mu_1 - B\mu_2), & \text{if } B\mu_2 < a. \end{cases}$$

Two Root Case

Let a_1, a_2 denote the two distinct roots of the quadratic equation arranged so that $a_1 < a_2$. Then one can easily verify that

$$g(B) = \begin{cases} \frac{1}{2} - K, & \text{if } R_1(B) = [a_1, a_2] \\ \frac{1}{2} + K, & \text{if } R_2(B) = (a_1, a_2) \end{cases}$$

where

$$K = \frac{1}{2}\{[D_1(a_2,B) - D_1(a_1,B)] - [D_2(a_2,B) - D_2(a_1,B]\},$$

and

$$\frac{\partial g}{\partial B} = \begin{cases} K_1\mu_1 - K_2\mu_2 + K_3 \dfrac{\Sigma_1 B^T}{B\Sigma_1 B^T} - K_4 \dfrac{\Sigma_2 B^T}{B\Sigma_2 B^T}, & \text{if} \quad R_1(B) = [a_1,a_2] \\ K_2\mu_2 - K_1\mu_1 + K_4 \dfrac{\Sigma_2 B^T}{B\Sigma_2 B^T} - K_3 \dfrac{\Sigma_1 B^T}{B\Sigma_1 B^T}, & \text{if} \quad R_2(B) = (a_1,a_2) \end{cases}$$

where

$$K_1 = p_1(a_2,B) - p_1(a_1,B)$$

$$K_2 = p_2(a_2,B) - p_2(a_1,B)$$

$$K_3 = (a_2 - B\mu_1)p_1(a_2,B) - (a_1 - B\mu_1)p_1(a_1,B)$$

$$K_4 = (a_2 - B\mu_2)p_2(a_2,B) - (a_1 - B\mu_2)p_2(a_1,B).$$

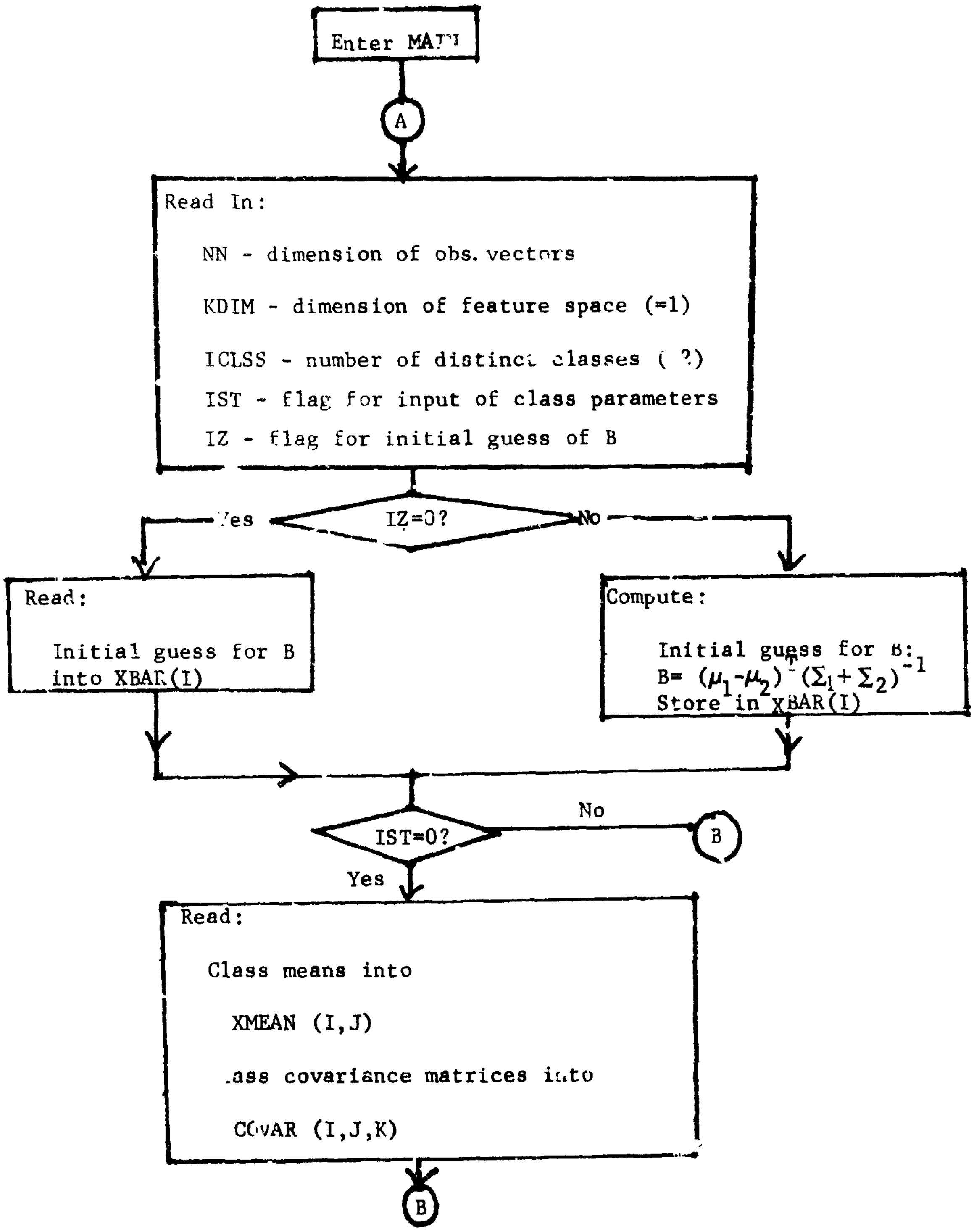
It should be noted that for the case of equal $n\times n$ covariance matrices, we always have the single root case, and one can verify that

$$B = (\mu_1 - \mu_2)^T(\Sigma_1 + \Sigma_2)^{-1}$$

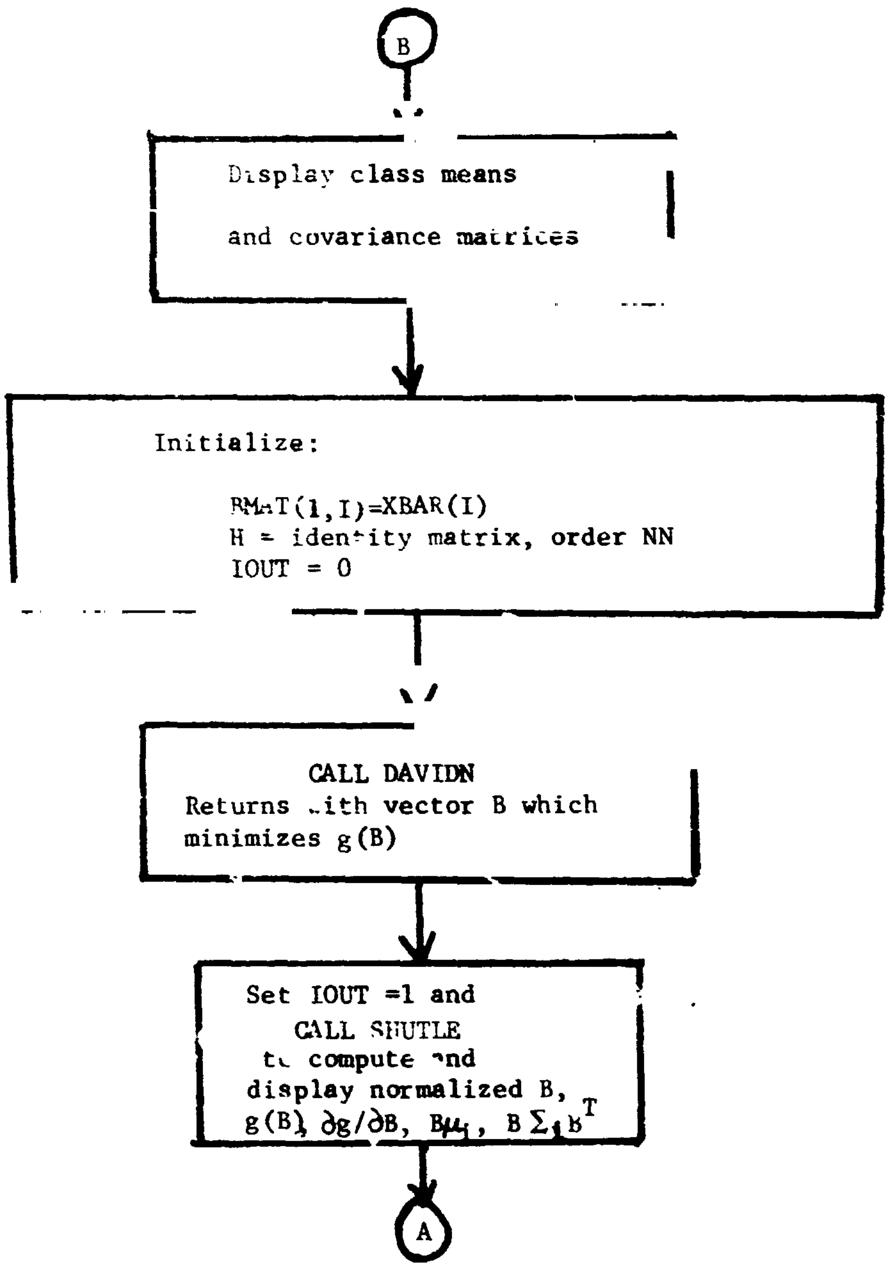
satisfies $\frac{\partial g}{\partial B} = 0$. This suggests that one should start the iterative procedure using this B as an initial guess, even if $\Sigma_1 \neq \Sigma_2$.

A flow chart for the preliminary version of our procedure is presented in the remainder of this section.

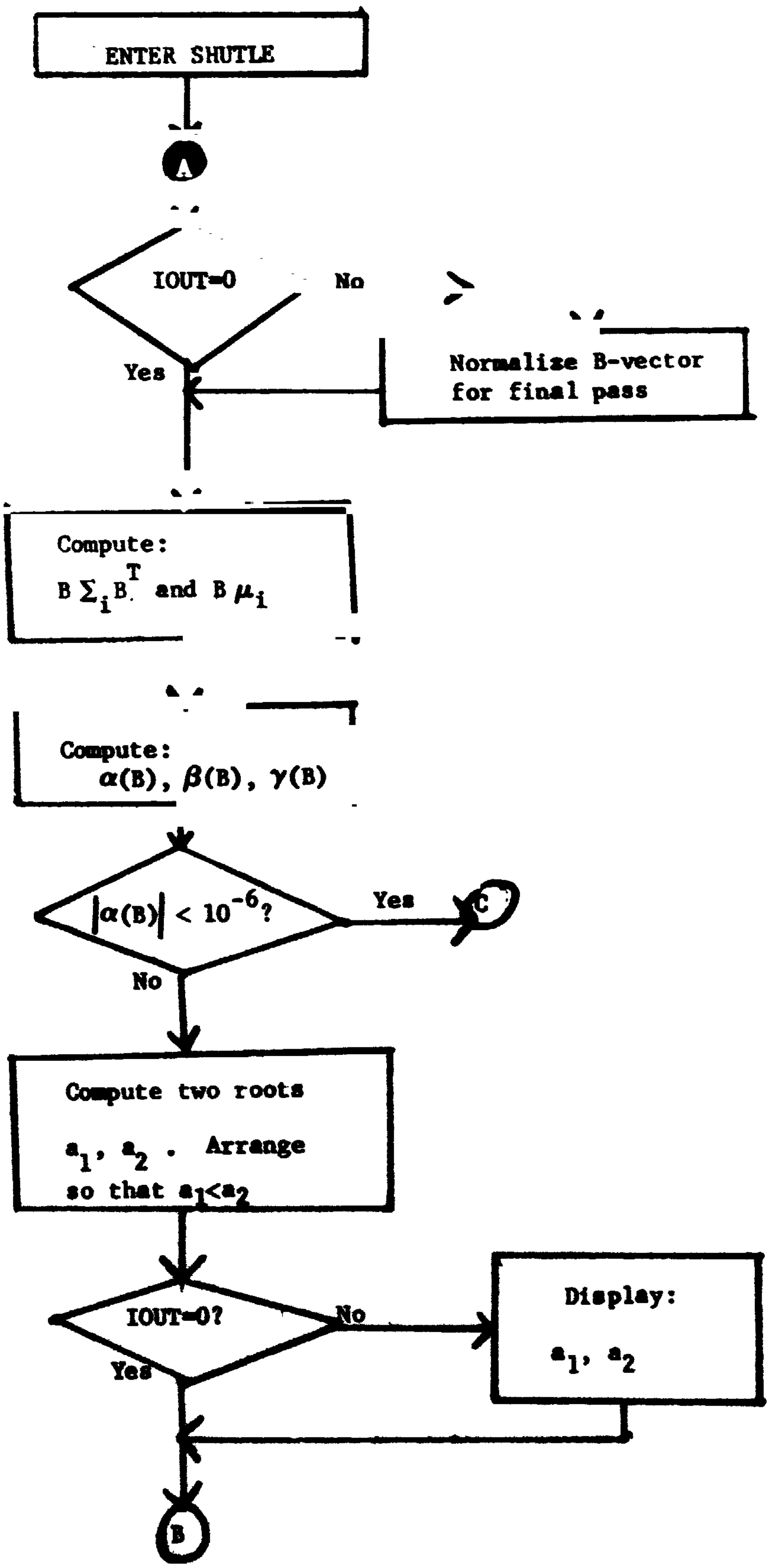
MAIN-1


Enter MAIN
A
Read In:
NN - dimension of obs. vectors
KDIM - dimension of feature space (=1)
ICLSS - number of distinct classes (2)
IST - flag for input of class parameters
IZ - flag for initial guess of B
IZ=0?
Yes
No
Read:
Initial guess for B into XBAR(I)
Compute:
Initial guess for B:
B= $(\mu_1-\mu_2)^T(\Sigma_1+\Sigma_2)^{-1}$
Store in XBAR(I)
IST=0?
No
B
Yes
Read:
Class means into
XMEAN (I,J)
Class covariance matrices into
COVAR (I,J,K)
B

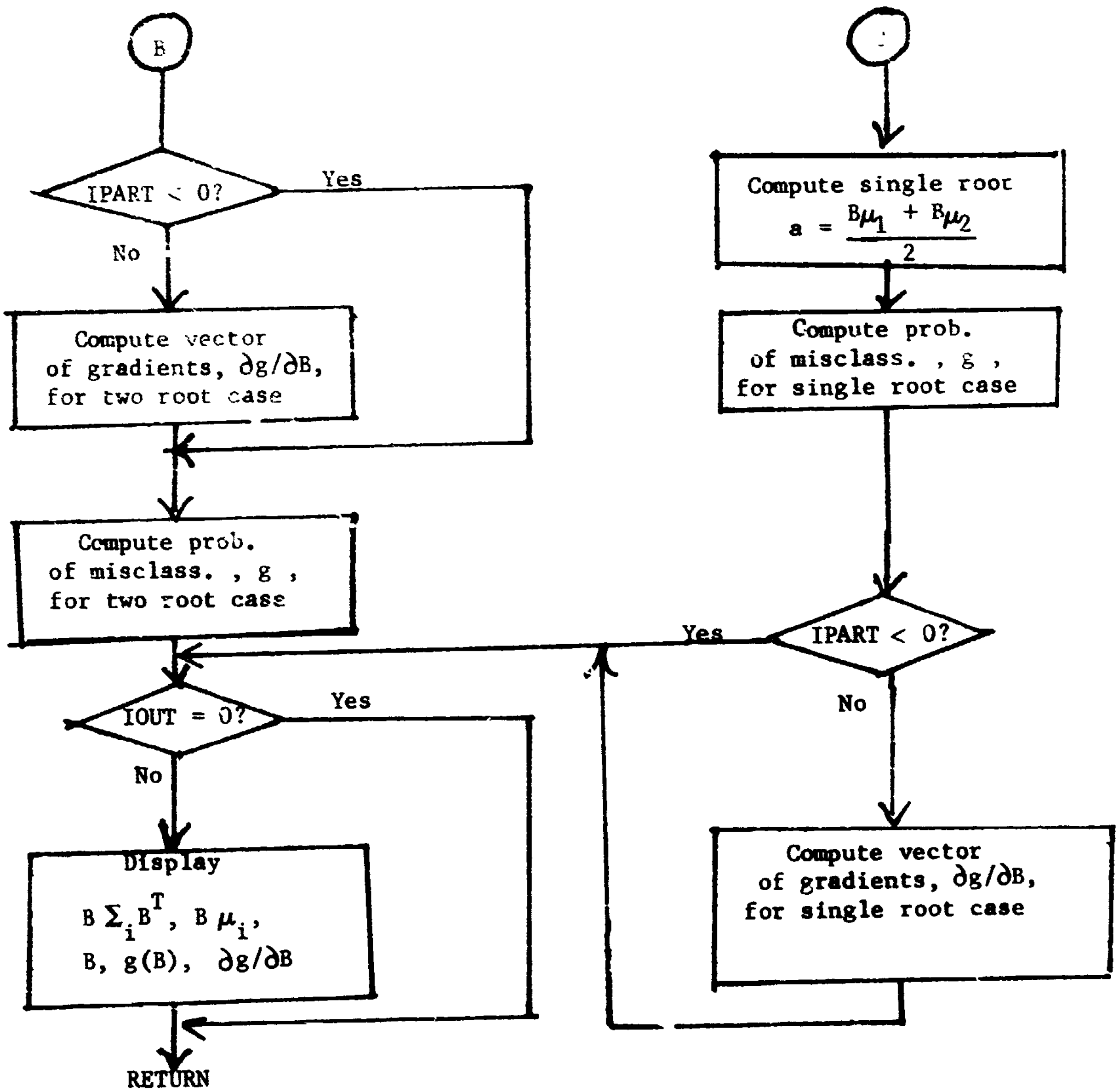
MAIN-2


B
Display class means
and covariance matrices
Initialize:
BMAT(1,I)=XBAR(I)
H = identity matrix, order NN
IOUT = 0
CALL DAVIDN
Returns with vector B which
minimizes g(B)
Set IOUT =1 and
CALL SHUTLE
to compute and
display normalized B,
$g(B)$, $\partial g/\partial B$, $B\mu_i$, $B\Sigma_i B^T$
A

SHUTLE-1


ENTER SHUTLE
A
IOUT=0
No
Yes
Normalize B-vector for final pass
Compute:
$B \Sigma_i B^T$ and $B \mu_i$
Compute:
$\alpha(B)$, $\beta(B)$, $\gamma(B)$
$|\alpha(B)| < 10^{-6}$?
Yes
C
No
Compute two roots a_1, a_2 . Arrange so that $a_1 < a_2$
IOUT=0?
No
Yes
Display:
a_1, a_2
B

SHUTLE-2


B
IPART < 0?
Yes
No
Compute vector of gradients, ∂g/∂B, for two root case
Compute prob. of misclass. , g , for two root case
IOUT = 0?
Yes
No
Display
B Σi B^T, B μi, B, g(B), ∂g/∂B
RETURN
Compute single root
a = (Bμ1 + Bμ2)/2
Compute prob. of misclass. , g , for single root case
IPART < 0?
Yes
No
Compute vector of gradients, ∂g/∂B, for single root case

4. Preliminary Numerical Results

In this section we present some preliminary results of the computational procedure discussed in the previous section. Mean vectors and covariance matrices for several pairs of classes from C-1 Flight Line Data were used as population parameters. These parameters are given on pages 29 - 33 .

In each of the Cases 1-4, presented in the following pages, the formula

$$(**) \qquad B_o = (\mu_1 - \mu_2)^T(\Sigma_1 + \Sigma_2)^{-1}$$

is used to compute the initial vector, B_o, for starting the iterative procedure. The final vector B, which minimizes the probability of misclassification, was determined using the computational procedure discussed in the previous section. The values of the transformed means, transformed covariances, and probability of misclassification for B_o and B are given in each case. We have also given the number of iterations (computations of g) needed to determine B for each of the cases.

Case 1

Population Parameters: $\left.\begin{matrix}\mu_1 \\ \Sigma_1\end{matrix}\right\}$ = Class 1 ; $\left.\begin{matrix}\mu_2 \\ \Sigma_2\end{matrix}\right\}$ = Class 2

Initial vector, B_o, computed using (**)	Final vector, B, which minimizes g
.152381	.149740
.103120	.101319
-.210958	-.207372
-.271028	-.266399
-.049500	-.047121
.485537	.477211
.375000	.368575
.003148	.003040
-.203273	-.199847
-.673343	-.661844
.054181	.052954
.082849	.081234
$g(B_o)$ = .017290986	$g(B)$ = .017290920
	Number of iterations: 8
$B_o\Sigma_1B_o^T$ = 4.517525	$B\Sigma_1B^T$ = 4.361629
$B_o\Sigma_2B_o^T$ = 4.413498	$B\Sigma_2B^T$ = 4.257698
$B_o\mu_1$ = -34.301838	$B\mu_1$ = -33.855418
$B_o\mu_2$ = -43.232862	$B\mu_2$ = -42.629203

Case 2

Population Parameters: $\left.\begin{matrix}\mu_1\\ \Sigma_1\end{matrix}\right\}$ = Class 3 ; $\left.\begin{matrix}\mu_2\\ \Sigma_2\end{matrix}\right\}$ = Class 4

Initial vector, B_o, computed using (**)	Final vector, B, which minimizes g
-.213967	-.196691
-.469183	-.438855
.177938	.179309
-.001497	-.018567
.188612	.140462
-.103755	-.274768
-.097442	-.156862
.009932	.081037
.601157	.729173
-.086085	.009754
-.084609	-.145805
-.145998	-.238270
$g(B_o)$ = .003018366	$g(B)$ = .002715326
	Number of iterations: 49
$B_o\Sigma_1B_o^T$ = 10.706851	$B\Sigma_1B^T$ = 25.185144
$B_o\Sigma_2B_o^T$ = 2.607698	$B\Sigma_2B^T$ = 45.544406
$B_o\mu_1$ = -35.144301	$B\mu_1$ = -48.456476
$B_o\mu_2$ = -48.458851	$B\mu_2$ = -68.128863

Case 3

Population Parameter : $\left.\begin{matrix}\mu_1 \\ \Sigma_1\end{matrix}\right\}$ = Class 5 ; $\left.\begin{matrix}\mu_2 \\ \Sigma_2\end{matrix}\right\}$ = Class 6

Initial vector, B_o, computed using (**)	Final vector, B, which minimizes g
.808819	.577773
.231155	.167764
-.290136	-.205380
.193884	.140068
.118042	.088422
.755009	.540627
.164943	.118282
-.362365	-.253048
-.447546	-.314009
.440773	-.310622
.014834	.011237
-.083220	-.053831
$g(B_o)$ = .024515472	g(B) = .024407769
	Number of iterations: 19
$B_o\Sigma_1B_o^T$ = 4.523550	$B\Sigma_1B^T$ = 2.443916
$B_o\Sigma_2B_o^T$ = 3.145724	$B\Sigma_2B^T$ = 1.586862
$B_o\mu_1$ = 111.417584	$B\mu_1$ = 85.707183
$B_o\mu_2$ = 103.748308	$B\mu_2$ = 80.153584

Case 4

Population Parameters: $\left.\begin{matrix}\mu_1 \\ \Sigma_1\end{matrix}\right\}$ = Class 7 ; $\left.\begin{matrix}\mu_2 \\ \Sigma_2\end{matrix}\right\}$ = Class 8

Initial vector, B_o,
computed using (**)

2.243915
.945839
.308322
1.222243
-.500153
-1.406996
-1.643716
-1.009229
.829381
-.490735
-.610768
-1.316042

$g(B_o) = .20299 \times 10^{-11}$

$B_o \Sigma_1 B_o^T = 56.866767$

$B_o \Sigma_2 B_o^T = 38.391677$

$B_o \mu_1 = -176.557149$

$B_o \mu_2 = -271.815594$

Final vector, B,
which minimizes g

Same as B_o

$g(B)$ = Same as $g(B_o)$
Number of iterations: 1

$B\Sigma_1 B^T =$
$B\Sigma_2 B^T =$
$B\mu_1 =$
$B\mu_2 =$

Same as for B_o

Class 1	MEAN	.16997769+03	.17495210+03	.19327100+03	.19252034+03	.16906102+03
	MEAN	.16705184+03	.19051837+03	.17088714+03	.18449409+03	.17286024+03
	MEAN	.16242323+03	.18200591+03			
Class 2	MEAN	.17085855+03	.17711908+03	.19527697+03	.19497829+03	.17264605+03
	MEAN	.16823223+03	.19251316+03	.17830000+03	.19304539+03	.18202960+03
	MEAN	.14818816+03	.17376513+03			
Class 3	MEAN	.17872219+03	.18174953+03	.19723617+03	.19657800+03	.17510097+03
	MEAN	.17049099+03	.19132753+03	.16950404+03	.18331013+03	.16658421+03
	MEAN	.14759664+03	.17001492+03			
Class 4	MEAN	.18260169+03	.18367145+03	.19701956+03	.19534094+03	.17455737+03
	MEAN	.17377575+03	.19161604+03	.16234224+03	.17112712+03	.15899739+03
	MEAN	.17283181+03	.18596349+03			
Class 5	MEAN	.18234438+03	.18560949+03	.20049707+03	.20072515+03	.18267511+03
	MEAN	.17498570+03	.19682326+03	.18611501+03	.19962053+03	.17815335+03
	MEAN	.11384275+03	.14711826+03			
Class 6	MEAN	.17783389+03	.18299663+03	.19925926+03	.19918069+03	.17933109+03
	MEAN	.16979461+03	.19422110+03	.18477329+03	.19993603+03	.17645118+03
	MEAN	.10342536+03	.14320426+03			
Class 7	MEAN	.17489414+03	.17520609+03	.19134082+03	.18964956+03	.16111868+03
	MEAN	.15829190+03	.18242903+03	.15211467+03	.16892622+03	.15519086+03
	MEAN	.15898717+03	.17907939+03			
Class 8	MEAN	.16494737+03	.16969132+03	.18957895+03	.18815505+03	.16247937+03
	MEAN	.16552632+03	.18764068+03	.15987767+03	.17153627+03	.16476387+03
	MEAN	.18318919+03	.19461735+03			

Mean Vectors for C-1 Flight Line Data

Class 1	COVAR	.65274073+01	.53574336+01	.77711774+01	.39174090+01	.39079335+01
	COVAR	.39718175+01	.37877161+01	.42324260+01	.28162174+01	.38085157+01
	COVAR	.74806794+01	.82846057+01	.54437276+01	.60450481+01	.13914199+02
	COVAR	.64108748+01	.73504820+01	.48119443+01	.54267190+01	.11094011+02
	COVAR	.12005847+02	.44160372+01	.48613533+01	.35376967+01	.37425676+01
	COVAR	.72447743+01	.70952387+01	.56575742+01	.71964298+01	.86472496+01
	COVAR	.56510909+01	.63811547+01	.12734403+02	.11478605+02	.77570125+01
	COVAR	.16037155+02	.55520678+01	.70039848+01	.44864996+01	.52106884+01
	COVAR	.10100492+02	.87373384+01	.62906535+01	.12599464+02	.12289525+02
	COVAR	.57565647+01	.70405760+01	.48323850+01	.53846570+01	.10931712+02
	COVAR	.10475405+02	.72931145+01	.13130637+02	.11170218+02	.14280519+02
	COVAR	.32038015+01	.23827286+01	.25201620+01	.24244123+01	.61632565+01
	COVAR	.97626821+01	.56964213+01	.35008505+01	-.16000719+00	.606772[illegible]+01
	COVAR	.37219972+02	.89507602+00	.74552272+00	.40220684+00	.95687257+00
	COVAR	.27100793+01	.50692931+01	.25373175+01	.17656046+01	-.48881714-01
	COVAR	.31544700+01	.14667230+02	.13187752+02		
Class 2	COVAR	.95184916+01	.69115220+01	.88646785+01	.44309080+01	.37431651+01
	COVAR	.34406807+01	.39567691+01	.39940287+01	.22252403+01	.3121[illegible]773+01
	COVAR	.74805143+01	.70079424+01	.42429308+01	.39982355+01	.95862911+01
	COVAR	.48544641+01	.49499441+01	.27374775+01	.29444791+01	.57372899+01
	COVAR	.58729561+01	.29416254+01	.29717698+01	.19558660+01	.1998640[illegible]+01
	COVAR	.34101902+01	.31941114+01	.32717161+01	.56928900+01	.61907175+01
	COVAR	.35284398+01	.38050691+01	.69838051+01	.57373930+01	.39782752+01
	COVAR	.95662936+01	.43204476+01	.50512071+01	.27207958+01	.3347[illegible]249+01
	COVAR	.55295736+01	.46932032+01	.36034181+01	.75018433+01	.82316443+01
	COVAR	.33267710+01	.38898232+01	.24236578+01	.26517294+01	.5111[illegible]+01
	COVAR	.48265631+01	.37793995+01	.70766951+01	.65806170+01	.8697[illegible]087+01
	COVAR	-.26494707+01	-.43871340+01	-.16018528+01	-.33224429+01	-.24777953+01
	COVAR	.14784744+00	-.78588752+00	-.70749177+01	-.86234251+01	-.37[illegible]4[illegible]226+01
	COVAR	.53809208+02	-.29094777+01	-.30246799+01	-.17406981+01	-.21301647+01
	COVAR	-.22589593+01	.14740524+00	-.87083781+00	-.40789335+01	-.45364016+01
	COVAR	-.20127919+01	.23381943+02	.20961917+02		

Covariance Matrices for C-1 Flight Line Data

Class 3	COVAR	.91161802+01	.66045035+01	.82326256+01	.47323188+01	.41854335+01
	COVAR	.39852331+01	.40835606+01	.44924960+01	.27806968+01	.39517793+01
	COVAR	.98973590+01	.94961882+01	.63655668+01	.63851358+01	.16003301+02
	COVAR	.98305172+01	.91012862+01	.54772540+01	.58472518+01	.13257359+02
	COVAR	.15131915+02	.60014969+01	.56218878+01	.38423741+01	.39368126+01
	COVAR	.81580653+01	.82277669+01	.62664118+01	.71519314+01	.86643805+01
	COVAR	.59418314+01	.67607232+01	.12994063+02	.10407219+02	.77813252+01
	COVAR	.18960338+02	.37690475+01	.65223772+01	.42861632+01	.58287180+01
	COVAR	.93032812+01	.62325851+01	.59841799+01	.16819953+02	.19493933+02
	COVAR	.43850395+01	.63043770+01	.45273624+01	.54767917+01	.92163461+01
	COVAR	.65711884+01	.61959697+01	.16203484+02	.16684998+02	.18009228+02
	COVAR	.31018034+01	-.20718610+01	-.94758997+00	-.36734321+01	-.36875191+01
	COVAR	.25831165+01	-.76354983+00	-.15302163+02	-.21300202+02	-.15697665+02
	COVAR	.53635088+02	-.18416247+01	-.42929035+01	-.31558881+01	-.41417114+01
	COVAR	-.68693179+01	-.25502386+01	-.34632219+01	-.12824065+02	-.14913833+02
	COVAR	-.11681605+02	.26076516+02	.24453757+02		
Class 4	COVAR	.78732130+01	.65937731+01	.86930263+01	.46457593+01	.[illegible]7161491+01
	COVAR	.45136421+01	.43315809+01	.49940203+01	.33801512+01	.42900776+01
	COVAR	.89083828+01	.99973336+01	.67359941+01	.69500961+01	.15971593+02
	COVAR	.83313652+01	.93587592+01	.62085633+01	.66909869+01	.13493629+02
	COVAR	.14117976+02	.58906708+01	.63134262+01	.45411082+01	.46626307+01
	COVAR	.93328693+01	.93437159+01	.75400166+01	.10912009+02	.12441285+02
	COVAR	.83905627+01	.90841533+01	.18665612+02	.18224221+02	.12909707+02
	COVAR	.27528586+02	.10326594+02	.11844141+02	.79355423+01	.88170364+01
	COVAR	.17647302+02	.17422523+02	.12573303+02	.25320458+02	.26200270+02
	COVAR	.15398830+02	.17003709+02	.11564956+02	.12488322+02	.25537658+02
	COVAR	.25390785+02	.18426265+02	.36561884+02	.35929882+02	.54530976+02
	COVAR	.14244121+02	.14743792+02	.98199905+01	.10013674+02	.21709285+02
	COVAR	.22493244+02	.16482673+02	.30686431+02	.28524981+02	.44914107+02
	COVAR	.55121075+02	.10988059+02	.11674234+02	.73607926+01	.7810[illegible]+01
	COVAR	.16864457+02	.17394939+02	.12055119+02	.23843552+02	.2224[illegible]+02
	COVAR	.34414125+02	.34789681+02	.30148699+02		

Covariance Matrices for C-1 Flight Line Data

Class 5	COVAR	.43000513+01	.27386034+01	.42186615+01	.21128412+01	.18471369+01
	COVAR	.23645883+01	.19068092+01	.20096312+01	.13551776+01	.21591267+01
	COVAR	.40020739+01	.40453467+01	.27818843+01	.27845078+01	.76100952+01
	COVAR	.40016751+01	.43084581+01	.27385798+01	.29434026+01	.68653137+01
	COVAR	.10031005+02	.26428291+01	.26298962+01	.20345934+01	.20216123+01
	COVAR	.41551566+01	.52269907+01	.41364936+01	.39005493+01	.41281670+01
	COVAR	.28342116+01	.31259097+01	.65373893+01	.64847400+01	.41971928+01
	COVAR	.84698590+01	.26567734+01	.31176464+01	.19793839+01	.24352961+01
	COVAR	.44871693+01	.40680440+01	.29523959+01	.56737098+01	.57284728+01
	COVAR	.40576895+01	.42328746+01	.28606567+01	.28217591+01	.67202036+01
	COVAR	.91874990+01	.52553382+01	.60837326+01	.33612186+01	.11564847+02
	COVAR	.49560077+01	.31544186+01	.26165750+01	-.46498452-01	.65000956+01
	COVAR	.21281886+02	.82790828+01	-.42543352+01	-.91409824+01	.28502022+02
	COVAR	.21452662+02	.34650994+01	.24480321+01	.10881223+01	.31731040+00
	COVAR	.36359748+01	.12624578+02	.46594059+01	-.17684862+01	-.47743413+01
	COVAR	.16360267+02	.11183525+03	.77383924+02		
Class 6	COVAR	.33948499+01	.25954401+01	.50617864+01	.21565959+01	.25340818+01
	COVAR	.30394507+01	.20761157+01	.28601596+01	.20531003+01	.31010202+01
	COVAR	.34786529+01	.46314531+01	.35623803+01	.37389828+01	.80846391+01
	COVAR	.17849468+01	.25869481+01	.18847690+01	.21843453+01	.39456058+01
	COVAR	.42240608+01	.18502503+01	.25434419+01	.21325010+01	.21892167+01
	COVAR	.36267141+01	.27769228+01	.34712846+01	.45274681+01	.64025065+01
	COVAR	.44936745+01	.52252815+01	.89684043+01	.53848447+01	.48782721+01
	COVAR	.13406571+02	.39994741+01	.59548405+01	.41357054+01	.49273030+01
	COVAR	.77874967+01	.47508909+01	.45006772+01	.11354020+02	.11864442+02
	COVAR	.20885106+01	.31060152+01	.23975031+01	.25172587+01	.46931487+01
	COVAR	.34219839+01	.30405818+01	.62507169+01	.55648507+01	.56771081+01
	COVAR	-.13339378+02	-.18982836+02	-.12817145+02	-.16108408+02	-.25014025+02
	COVAR	-.13251863+02	-.13568311+02	-.36761883+02	-.35367139+02	-.14273030+02
	COVAR	.15790088+03	-.81828862+01	-.11334143+02	-.84552642+01	-.92627939+01
	COVAR	-.15442987+02	-.77197972+01	-.83564496+01	-.21086223+02	-.20131738+02
	COVAR	-.87529376+01	.80553464+02	.57769464+02		

Covariance Matrice for C-1 Flight Line Data

Class 7	COVAR	.49165547+01	.35723952+01	.66244251+01	.27712584+01	.28550640+01
	COVAR	.33452266+01	.25022005+01	.32616881+01	.21989867+01	.32165788+01
	COVAR	.44058318+01	.53944606+01	.37548627+01	.39188325+01	.87499474+01
	COVAR	.30541981+01	.43250253+01	.25601463+01	.31882508+01	.51250390+01
	COVAR	.56659295+01	.25037174+01	.29187314+01	.23071127+01	.24442102+01
	COVAR	.36849948+01	.32887907+01	.36319977+01	.51782802+01	.67187226+01
	COVAR	.46591196+01	.52472817+01	.88234575+01	.73300627+01	.54515640+01
	COVAR	.14565491+02	.49408197+01	.68539010+01	.43839127+01	.54075115+01
	COVAR	.86026636+01	.74982783+01	.56424323+01	.13265289+02	.15120174+02
	COVAR	.49800909+01	.65810188+01	.45143542+01	.51375642+01	.84701067+01
	COVAR	.71898295+01	.55528820+01	.12795913+02	.13242022+02	.15011698+02
	COVAR	-.86492268+00	-.13986376+01	-.50846461+00	-.[illegible]7849683+00	-.19069832+01
	COVAR	-.50989533+00	.96199417-01	-.29073372+01	-.34271111+01	-.16644834+01
	COVAR	.17327283+02	-.55900517+00	-.94253881+00	-.69481619+00	-.88467602+00
	COVAR	-.14524477+01	-.49831313+00	-.82782820+00	-.23742799+01	-.27734318+01
	COVAR	-.16796910+01	.39055138+01	.62561319+01		
Class 8	COVAR	.32550607+01	.12615085+01	.33219616+01	.11530214+01	.69605637+00
	COVAR	.18139151+01	.62498125+00	.79294274+00	.38161643+00	.13705[illegible]22+01
	COVAR	.11976308+01	.12165587+01	.74486429+00	.76317410+00	.31644600+01
	COVAR	.10077973+01	.11926076+01	.57662318+00	.80004498+00	.13071675+01
	COVAR	.22610586+01	.78932373+00	.62739054+00	.70775228+00	.48912475+00
	COVAR	.61779188+00	.77627830+00	.15472314+01	.77417903+00	.86959226+00
	COVAR	.67491378+00	.58737077+00	.10544532+01	.12339931+01	.72692125+0[illegible]
	COVAR	.33097997+01	.46986055+00	.73557160+00	.24606388+00	.65462425+00
	COVAR	.88500848+00	.11760384+01	.58983275+00	.12109964+01	.21664174+01
	COVAR	.91205578+00	.82444388+00	.79787074+00	.77740493+00	.12629289+01
	COVAR	.14620633+01	.10777113+01	.16277674+01	.15384919+01	.36108294+01
	COVAR	.11438746+01	.82058982+00	.90170940+00	.66729566+00	.73965274+00
	COVAR	.12279202+01	.13620929+01	.12468237+01	.11263186+01	.18567028+01
	COVAR	.77661507+01	.16501724+00	.37032174+00	.39511171-01	.39132249+00
	COVAR	.73212484+00	.66891588+00	.12413628+00	.72520293+00	.84082260+00
	COVAR	.10505424+01	.40155540+00	.31083594+01		

Covariance Matrices for C-1 Flight Line Data

5. Concluding Remarks

Although the theory and subsequent computational procedure presented herein yield encouraging numerical results, there is still much work to be done. Extensive testing of the existing program is needed. Other optimization techniques should be tested as substitutes for the Davidon-Fletcher Powell routine. Theoretical results are needed which justify the use of the formula (**) of Section 4 for computing starting vectors.

The theory for the two population case can be extended to the case of m populations. The associated computational procedure is in developmental stage.

Finally, we note the possibility of developing a suboptimal method, using the computational procedure discussed herein, which determines a $k \times n$ matrix, B, one row at a time so as to minimize the probability of misclassification in k-dimensional space. Such a procedure has not been developed, even in the case of two populations.

REFERENCES

1. T. W. Anderson, An Introduction to Multivariate Statistical Analysis, John Wiley and Sons, Inc. New York, 1958.

2. L. F. Guseman, Jr. and Homer F. Walker, On Minimizing the Probability of Misclassification for Linear Feature Selection, Report #28 NAS-9-12777, University of Houston, Department of Mathematics, Aug. 1973.

3. David G. Luenberger, Optimization by Vector Space Methods, John Wiley and Sons, Inc., New York, 1969.