DEPARTMENT OF MATHEMATICS
UNIVERSITY OF HOUSTON HOUSTON, TEXAS

NASA CR- 141624

(NASA-CR-141624) SOME RESULTS ON THE DIVERGENCE CRITERION USED TO SOLVE THE FEATURE SELECTION PROBLEM (Houston Univ.) 19 p HC $3.25 CSCL 12A

N75-17138

Unclas
G3/64 10220

SOME RESULTS ON THE DIVERGENCE
CRITERION USED TO SOLVE THE
FEATURE SELECTION PROBLEM
BY D. R. BROWN & M. J. O'MALLY
REPORT #40 NOVEMBER 1974

PREPARED FOR
EARTH OBSERVATION DIVISION, JSC
UNDER
CONTRACT NAS-9-12777

3801 CULLEN BLVD.
HOUSTON, TEXAS 77004

SOME RESULTS ON THE DIVERGENCE CRITERION USED TO SOLVE THE FEATURE SELECTION PROBLEM

by

Dennison R. Brown

and

Matthew J. O'Malley

Mathematics Department

University of Houston

November, 1974

Report #40

INTRODUCTION.

In this paper we present several results aimed at a solution to the feature selection problem. While we do not claim to have obtained a "better solution" to the feature selection problem than that presently used, we believe our results are potentially applicable and, we hope, will offer more insight into the feature selection problem.

SECTION 1 - REDUCTION OF THE NUMBER OF VARIABLES IN "BEST B'S".

The technique presently used to solve the feature selection problem (that is, determine a $k \times n$ matrix A which maximizes the average interclass divergence D_{B_k} for the k-case) essentially treats each of the $k \cdot n$ entries of A as a variable which must be determined. Here, by using the fact that $D_{PA} = D_A$ for any $k \times k$ invertible matrix P, we make a few simple observations to show that the number of variables in A which must be determined can always be taken to be $\leq k[(n-k) + (k+1)/2]$ and perhaps, from a probabilistic point of view, can be assumed to be $\leq k[n-k]$. For the algebraic definitions and results which we make use of here, we refer the reader to D. Finkbeiner [2; Chapter 6].

We begin with two definitions.

DEFINITION: Let A and B be $\ell \times m$ matrices. We say that A is row equivalent to B if and only if there exists an invertible $\ell \times \ell$ matrix P such that $A = PB$.

<u>DEFINITION</u>. We say the $\ell \times m$ matrix E is in <u>reduced echelon form</u> if and only if the following three conditions are satisfied:

(i) The first p rows of E are nonzero; the other rows are zero. (where $p = \text{rank } E$).

(ii) The first nonzero element in each nonzero row is 1 and it appears in a column to the right of the first nonzero element of any preceding row.

(iii) The first nonzero element in each nonzero row is the only nonzero element in its column.

Thus, if $E = (e_{ij})$ is a $\ell \times m$ matrix, $\ell \le m$, and if E is in reduced echelon form, then E is "upper triangular", where by "upper triangular" we mean that $e_{ij} = 0$ if $i > j$, $j = 1, \ldots, \ell-1$. Observe that if $\ell = m$, then E is invertible if and only if $E = I_m$.

<u>THEOREM 1</u>. Any $\ell \times m$ matrix T of rank p is row equivalent to a $\ell \times m$ matrix in reduced echelon form with p nonzero rows. In particular, if $\ell = m = p$ (that is, T is an invertible $m \times m$ matrix), then the reduced echelon matrix to which T is row equivalent is I_m.

<u>Proof:</u> See [2; p. 124].

In particular, let $B = (B_1, B_2)$ be a $k \times n$ matrix, $k < n$, where B_1 and B_2 are $k \times k$ and $k \times (n-k)$ matrices, respectively. Suppose further that rank $B = k$. By Theorem 1, B_1 is row equivalent to a $k \times k$ matrix C_1 in reduced echelon form. Thus, there exists a $k \times k$ invertible matrix P such that $C_1 = PB_1$, and in particular, $C = PB = (PB_1, PB_2) = (C_1, C_2)$, where C_1 is a $k \times k$ matrix in reduced echelon form. Since C_1 is in reduced echelon form, then C_1 is upper triangular, and hence, has at most $1 + 2 + \ldots + k = (1/2)k(k+1)$ nonzero entries (in fact, if C_1 has $(1/2)k(k+1)$ nonzero entries, then all diagonal elements of C_1 are nonzero and hence C_1 is invertible. Thus, $C_1 = I_k$ and C_1 has exactly k nonzero entries). Therefore, unless $k = 1$, we can say that C_1 has $< (1/2)k(k+1)$ nonzero entries.

In general, even though $B = (B_1, B_2)$ has rank k, and hence the row rank of B = the column rank of $B = k$, it is not necessarily true that the first k columns of B are linearly independent, and hence B_1 need not be invertible. However, if B_1 is invertible, then by Theorem 1, $C_1 = I_k$ and hence $C = (I_k, C_2)$ has $k(n-k)$ "unknown" elements, namely the elements of C_2.

Using the fact that the average interclass divergence for the k-case is invariant for row equivalent matrices (that is, $D_{PB} = D_B$, where B is a $k \times n$ matrix and P is a $k \times k$ invertible matrix), we have proven the following theorem.

THEOREM 2. Let k and n be positive integers, $1 \leq k \leq n$. There exists a "best C" which maximizes the average interclass divergence D_{B_k} for the k-case of the form: $C = (C_1, C_2)$, where C_1 is a $k \times k$ upper triangular matrix in reduced echelon form and C_2 is a $k \times (n-k)$ matrix. Moreover, if there exists a "best B" (for the k-case) of the form $B = (B_1, B_2)$, where B_1 is a $k \times k$ invertible matrix, then we may take $C_1 = I_k$, and hence there are at most $k \cdot (n-k)$ unknown entries in C - namely, the entries of C_2.

With regard to the assumption that B_1 is invertible, examples have been given to show that there exist "best B's" for which B_1 is not invertible.

However, because of the example given in [1], this does not guarantee that there does not exist a "best B" for which B_1 is invertible. Indeed, from a probabilistic point of view, it would seem reasonable to assume that B_1 is invertible. We state two questions:

<u>Question 1</u>: Let k and n be positive integers, $1 \le k < n$. Does there exist a $k \times n$ matrix $B = (B_1, B_2)$, where B_1 and B_2 are $k \times k$ and $k \times (n-k)$ matrices, respectively, which maximizes D_{B_k} for the k-case and for which B_1 is invertible?

<u>Question 2</u>: Given $\epsilon > 0$, does there exist a $k \times n$ matrix $B = (B_1, B_2)$, where B_1 is a $k \times k$ invertible matrix, for which $|D_{B_k} - D_B| < \epsilon$, where D_{B_k} denotes the maximum value of the average interclass divergence for the k-case?

Finally, we observe that Homer Walker has recently shown that if U is any $n \times n$ unitary matrix and k is any positive integer, $1 \le k < n$, then there exist $\ell \le \min \{k, n-k\}$ Householder matrices $H_1, \ldots, H_\ell$ such that $D_{(I_k|Z)U} = D_{(I_k|Z)H_\ell \cdots H_1}$. In particular, if U is such that $D_{(I_k|Z)U}$ is the maximum value of divergence for the k-case, then Walker's result effects a reduction in the number of unknowns to be determined - from $k \cdot n$ (the entries of $(I_k|Z)U$) to $n \cdot \ell$, since each Householder matrix is completely determined by a column n-vector ξ of norm 1. (In fact, if $H = I_n - 2\xi\xi^T$ is a Householder matrix, where $\xi = \begin{pmatrix} \xi_1 \\ \vdots \\ \xi_n \end{pmatrix}$, $\xi_1^2 + \ldots + \xi_n^2 = 1$, then $\xi_1 = \pm\sqrt{1 - (\xi_2^2 + \ldots + \xi_n^2)}$, and thus, to absolute value, H is determined by $n-1$ variables, and it may be argued that Walker's result reduces the number of unknowns to $(n-1)\ell$.)

However, if D_B is the maximum value of divergence for the k-case, and if $B = (I_k|Z)U = (B_1, B_2)$, where B_1 is a $k \times k$ invertible matrix, then we have shown that we may take $B_1 = I_k$, and hence the number of unknowns in B is $k(n-k)$. Thus, since $k(n-k) \leq (n-1)\ell$ always (where $\ell = \min\{k, n-k\}$), and since $k(n-k) = (n-1)\ell$ if and only if $k = 1$ or $k = n-1$, the result presented here effects a further reduction in the number of unknowns, if $1 < k < n-1$. (In actual practice, if one were to make use of Walker's result in attempting to determine a "best B", it appears that it would be necessary to take $\ell = \min\{k,n-k\}$.)

SECTION 2 - ITERATIVE SELECTION OF H_i'S.

It has been established that there exists a $k \times n$ matrix B of the form $(I_k|Z)U$, where U is an $n \times n$ unitary matrix, which maximizes the average interclass divergence D_{B_k} for the k-case. Moreover, as indicated in Section 1, it has been shown that there exist $\ell \leq \min\{k,n-k\}$ Householder matrices $H_1, \ldots, H_\ell$ for which $D_{(I_k|Z)H_\ell \ldots H_1}$ is a maximum.

Recently, it has been proposed that one might obtain a "best B" by successively maximizing over the set $\mathcal{H}$ of $n \times n$ Householder matrices the function:

$$D_{(I_k|Z)X\,H_i \ldots H_1} : \mathcal{H} \to \mathcal{R}^{\#} \;(\mathcal{R}^{\#} \text{ is the set of real numbers}),$$

where X is a variable with domain $\mathcal{H}$ and, for each $j = 1, \ldots, i$, $H_j \in \mathcal{H}$ is such that $D_{(I_k|Z)H_j \ldots H_1}$ is the maximum value of the function:

$$D_{(I_k|Z)X\,H_{j-1} \ldots H_1} : \mathcal{H} \to \mathcal{R}^{\#}.$$

A more detailed development of this, and in particular, a proof of the fact that $D_{(I_k|Z)X\ H_1 \cdots H_1}$ actually obtains a maximum on $\mathcal{H}$ will appear elsewhere.

Several questions have been raised in conjunction with this suggestion and most still remain unanswered. In this brief section we present two observations and ask five questions related to this proposal.

We first note that the $n \times n$ Householder matrices are precisely those matrices of the form: $I_n - 2\xi\xi^T$, where ξ is a column n-unit vector. In particular, every Householder matrix H is a symmetric unitary matrix and hence $H = H^{-1}$. Of prime importance, Henry Decell has shown that:

$$D_{(I_k|Z)H_p \cdots H_1} \le D_{(I_k|Z)H_{p+1}H_p \cdots H_1},$$

where, for each $i = 1, \ldots, p+1$, H_i denotes a Householder matrix which maximizes (over $\mathcal{H}$) $D_{(I_k|Z)X\ H_{i-1} \cdots H_1}$. For our first observation, we make use of the fact that $\mathcal{H}$ is invariant under the group of inner automorphisms of the group $\mathcal{U}$ of unitary matrices. In particular, $H_1\ H\ H_1^{-1} = H_1\ H\ H_1 \in \mathcal{H}$, whenever $H_1, H \in \mathcal{H}$.

<u>PROPOSITION</u>. Let $G_1, \ldots, G_\ell \in \mathcal{H}$. (That is, $G_1, \ldots, G_\ell$ are fixed but arbitrary Householder matrices.) Let $G_{\ell+1}$ maximize the function:

$$D_{(I_k|Z)X\ G_\ell \cdots G_1} : \mathcal{H} \to \mathcal{R}^{\#}.$$

Then, for any $G \in \mathcal{H}$,

$$D_{(I_k|Z)G_\ell \cdots G_1 G} \le D_{(I_k|Z)G_{\ell+1}G_\ell \cdots G_1}.$$

Henry Decell has pointed out to the authors the following generalization of the Proposition.

RESULT. Under the same hypothesis as for the Proposition, if G is any element of $\mathcal{H}$, then, for any $p = 0, 1, \ldots, \ell-1$,

$$D(I_k|Z)G_\ell \cdots G_{\ell-p} G G_{\ell-(p+1)} \cdots G_1 \le D(I_k|Z)G_{\ell+1}G_\ell \cdots G_1 .$$

Proof: Observe that $G_\ell \cdots G_{\ell-p} G\, G_{\ell-p} \cdots G_\ell = H \in \mathcal{H}$ and hence,

$$D(I_k|Z)HG_\ell \cdots G_1 \le D(I_k|Z)G_{\ell+1}G_\ell \cdots G_1 .$$

But $HG_\ell \cdots G_1 = HG_\ell \cdots G_{\ell-p}G_{\ell-(p+1)} \cdots G_1 =$

$[G_\ell \cdots G_{\ell-p} G\, G_{\ell-p} \cdots G_\ell](G_\ell \cdots G_{\ell-p})(G_{\ell-(p+1)} \cdots G_1) =$

$G_\ell \cdots G_{\ell-p} G\, G_{\ell-(p+1)} \cdots G_1$, since $(G_{\ell-p} \cdots G_\ell)(G_\ell \cdots G_{\ell-p}) = I_n$.

As a consequence of the Proposition or Result, observe that if, for each $i > 0$, H_i denotes a Householder matrix which maximizes the function $D(I_k|Z)X\, H_{i-1} \cdots H_1$, $X \in \mathcal{H}$, then, for any $p > 1$,

$$\begin{aligned}
D(I_k|Z)H_p \cdots H_1 &\le D(I_k|Z)T_{p+1}H_p \cdots H_2 \\
&\le D(I_k|Z)T_{p+2}T_{p+1}H_p \cdots H_3 \\
&\le \\
&\ \vdots \\
&\le D(I_k|Z)T_{2p} \cdots T_{p+1} \\
&\le \\
&\ \vdots
\end{aligned}$$

where $T_{p+j} \in \mathcal{H}$ for each $j \in \omega$, and T_{p+j} maximizes the function $D(I_k|Z)X\, T_{p+j-1}\cdots T_{p+1}H_p\cdots H_j$, $X \in \mathcal{H}$, if $j < p$. (If $j \geq p$, then no H_i's appear and the form of the function maximized is clear.)

This points out the known fact that the H_i's are not unique, and perhaps indicates that one must be "selective" in choosing the H_i's which maximize at each step.

We list five questions below; the first one seems to continually arise in any attempt to show that a "best B" can be reached by this successive maximization over the set $\mathcal{H}$.

<u>Question 1</u>: If $D(I_k|Z)H \leq D(I_k|Z)G$, is $D(I_k|Z)H_1H \leq D(I_k|Z)H_1G$, where $H, G, H_1 \in \mathcal{H}$?

<u>Question 2</u>: Let $H_1,\ldots,H_j, G_1,\ldots,G_j \in \mathcal{H}$ and suppose that $D(I_k|Z)G_j\cdots G_1 \leq D(I_k|Z)H_j\cdots H_1$. Let G_{j+1} and H_{j+1} maximize $D(I_k|Z)XG_j\cdots G_1$ and $D(I_k|Z)X\,H_j\cdots H_1$, respectively. Is $D(I_k|Z)G_{j+1}G_j\cdots G_1 \leq D(I_k|Z)H_{j+1}H_j\cdots H_1$?

<u>Question 3</u>: If the answer to Question 2 is "No, in general", can we choose $G_1',\ldots,G_j'$ and $H_1',\ldots,H_j'$ so that $D(I_k|Z)G_i'\cdots G_1' = D(I_k|Z)G_i\cdots G_1$ and $D(I_k|Z)H_i'\cdots H_1' = D(I_k|Z)H_i\cdots H_1$, for each $i=1,\ldots,j$, and for which there exist G'_{j+1} and H'_{j+1} which maximize $D(I_k|Z)X\,G_j'\cdots G_1'$ and $D(I_k|Z)X\,H_j'\cdots H_1'$, respectively, and are such that $D(I_k|Z)G'_{j+1}G'_j\cdots G'_1 \leq D(I_k|Z)H'_{j+1}H'_j\cdots H'_1$?

<u>Question 4:</u> Replace $\leq$ in Question 1 by =. That is, let $H, G \in \mathcal{H}$ and suppose that $D(I_k|Z)H = D(I_k|Z)G$. Is $D(I_k|Z)H_1H = D(I_k|Z)H_1G$, where H_1

represents an arbitrary element of $\mathcal{H}$?

Question 5: In Question 4, replace $H_1 \in \mathcal{H}$ with H_1 maximizes $D_{(I_k|Z)X\,H}$ and G_1 maximizes $D_{(I_k|Z)X\,G}$. Is $D_{(I_k|Z)H_1H} = D_{(I_k|Z)G_1G}$?

SECTION 3 - A REPRESENTATION THEOREM USING H_i'S.

As observed in Section 1, Homer Walker has recently shown that if $B = (I_k|Z)U$, where U is an $n \times n$ unitary matrix, then there exists $\ell \leq \min\{k,n-k\}$ Householder matrices $H_1,\ldots,H_\ell$ such that $D_B = D_{(I_k|Z)H_\ell\ldots H_1}$. In particular, this fact follows directly from the following two results.

(i) There exists $1 \leq v \leq k$ Householder matrices $G_1,\ldots,G_v$ such that $(I_k|Z)U = (I_k|Z)G_v\ldots G_1$.

(ii) There exists $1 \leq j \leq n-k$ Householder matrices $T_1,\ldots,T_j$ such that $(I_k|Z)U = A(I_k|Z)T_j\ldots T_1$, where A is a $k \times k$ invertible matrix.

In particular, if $k = n-1$, then $n-k = 1$ and thus, by (ii), there exists a Householder matrix H and an invertible $(n-1)\times(n-1)$ matrix A such that $(I_{n-1}|Z)U = A(I_{n-1}|Z)H$. In this section we make repeated use of this result, together with the fact that the nth row of H can be chosen equal to the nth row of U, to derive a result which has application in attempting to maximize divergence by a procedure closely related to that suggested in Section 2.

Throughout this section $\mathcal{U}_\ell$ and $\mathcal{H}_\ell$ will denote the sets of $\ell \times \ell$ unitary and Householder matrices, respectively. Observe that $\mathcal{H}_\ell \subseteq \mathcal{U}_\ell$ for all positive integers ℓ. The main result of this section is the following theorem.

THEOREM 3. Let p and n be fixed positive integers, $p < n$, and let $U \in \mathcal{U}_n$. Then there exist p Householder matrices $\hat{H}_1, \ldots, \hat{H}_p$ having the following property (*): given any integer ℓ, $1 \le \ell \le p$, there exists $C_\ell \in \mathcal{U}_{n-\ell}$ such that

$$(I_{n-\ell}|Z)U = C_\ell (I_{n-\ell}|Z)\hat{H}_\ell \ldots \hat{H}_1.$$

Moreover, for each $j = 1, \ldots, p$, $\hat{H}_j = \left(\begin{array}{c|c} H_j & Z \\ \hline Z & I_{j-1} \end{array}\right)$ where $H_j \in \mathcal{H}_{n-(j-1)}$ and Z denotes the zero matrix of appropriate dimension. Therefore, for any integer ℓ, $1 \le \ell \le p$, $D_{(I_{n-\ell}|Z)U} = D_{(I_{n-\ell}|Z)\hat{H}_\ell \ldots \hat{H}_1}$.

Before proving Theorem 3, we derive two preliminary results.

LEMMA 1. Let p and n be positive integers $p < n$, let A be a $p \times p$ matrix, and let $\hat{A} = \left(\begin{array}{c|c} A & Z \\ \hline Z & I_{n-p} \end{array}\right)$, where Z denotes the zero matrix of appropriate dimension. Then the following are true.

(i) $A \in \mathcal{U}_p$ if and only if $\hat{A} \in \mathcal{U}_n$.

(ii) $A \in \mathcal{H}_p$ if and only if $\hat{A} \in \mathcal{H}_n$.

Moreover, $A(I_p|Z_{p\times(n-p)}) = (I_p|Z_{p\times(n-p)})\hat{A}$.

Proof: Observe that $(\hat{A})^T = \left(\begin{array}{c|c} A^T & Z \\ \hline Z & I_{n-p} \end{array}\right)$ and therefore, $\hat{A}\cdot(\hat{A})^T = \left(\begin{array}{c|c} AA^T & Z \\ \hline Z & I_{n-p} \end{array}\right)$. Thus, (i) $\hat{A} \in \mathcal{U}_n \longleftrightarrow \hat{A}\cdot(\hat{A})^T = I_n \longleftrightarrow AA^T = I_p \longleftrightarrow A \in \mathcal{U}_p$. For (ii) we have: $\hat{A} \in \mathcal{H}_n \longleftrightarrow \hat{A} = I_n - 2\xi\xi^T$ for a unit n vector

$$\xi = \begin{bmatrix} \xi_1 \\ \vdots \\ \xi_p \\ 0 \\ \vdots \\ 0 \end{bmatrix} \longleftrightarrow A = I_p - 2\bar{\xi}\bar{\xi}^T, \text{ where } \bar{\xi} = \begin{pmatrix} \xi_1 \\ \vdots \\ \xi_p \end{pmatrix} \text{ is a unit p-vector.}$$

The last statement of Lemma 1 follows directly from matrix multiplication.

Observe that if $k \in \omega$ and if $H \in \mathcal{H}_k$, $U \in \mathcal{U}_k$, then $U = (UH^{-1})H = (UH)H$, where $UH \in \mathcal{U}_k$. Moreover, this representation is unique; that is, if $U = AH$, then $A = UH \in \mathcal{U}_k$.

LEMMA 2. For any positive integer k, let $H \in \mathcal{H}_k$, $U \in \mathcal{U}_k$, where the kth row of H is the kth row of U. If A is a $(k-1)\times(k-1)$ matrix such that $A\left(I_{k-1} \;\middle|\; \begin{matrix}0\\ \vdots\\ 0\end{matrix}\right) H = \left(I_{k-1} \;\middle|\; \begin{matrix}0\\ \vdots\\ 0\end{matrix}\right) U$, then $A \in \mathcal{U}_{k-1}$.

Proof: Let $H = \begin{pmatrix} h_1 \\ \vdots \\ h_{k-1} \\ u_k \end{pmatrix}$, $U = \begin{pmatrix} u_1 \\ \vdots \\ u_{k-1} \\ u_k \end{pmatrix}$, and

$$\hat{A} = \left(\begin{array}{c|c} & 0 \\ A & \vdots \\ & 0 \\ \hline 0\ldots0 & 1 \end{array}\right). \text{ Then } \hat{A}H = \left(\begin{array}{c} A\left(I_{k-1} \;\middle|\; \begin{matrix}0\\ \vdots\\ 0\end{matrix}\right) H \\ \hline u_k \end{array}\right) = \left(\begin{array}{c} \left(I_{k-1} \;\middle|\; \begin{matrix}0\\ \vdots\\ 0\end{matrix}\right) U \\ \hline u_k \end{array}\right) = U.$$

Thus, by the observation preceding Lemma 2, $\hat{A} = UH \in \mathcal{U}_k$, and therefore, by Lemma 1, $A \in \mathcal{U}_{k-1}$.

Proof of Theorem 3.

Fix $n > 1$ and let $U \in \mathcal{U}_n$. We proceed by induction on p. If $p = 1$, then by result (ii) there exists an $(n-1)\times(n-1)$ invertible matrix C_1 and $\hat{H}_1 \in \mathcal{H}_n$ such that $\left(I_{n-1} \;\middle|\; \begin{matrix}0\\ \vdots\\ 0\end{matrix}\right) U = C_1 \left(I_{n-1} \;\middle|\; \begin{matrix}0\\ \vdots\\ 0\end{matrix}\right) \hat{H}_1$. Moreover, the nth row of $\hat{H}_1$ can be chosen to be the same as the nth row of U, and hence, by Lemma 2, $C_1 \in \mathcal{U}_{n-1}$.

For the sake of clarity we will go through the case for $p = 2$. By result (ii), applied to $\left(I_{n-2} \middle| \begin{matrix} 0 \\ \vdots \\ 0 \end{matrix}\right) C_1$, there exists an $(n-2)\times(n-2)$ invertible matrix C_2 and $H_2 \in \mathcal{H}_{n-1}$ such that $\left(I_{n-2} \middle| \begin{matrix} 0 \\ \vdots \\ 0 \end{matrix}\right) C_1 = C_2 \left(I_{n-2} \middle| \begin{matrix} 0 \\ \vdots \\ 0 \end{matrix}\right) H_2$, where the $(n-1)$st row of H_2 is chosen equal to the $(n-1)$st row of C_1. Thus, by Lemma 2, $C_2 \in \mathcal{U}_{n-2}$. Therefore,

$$\begin{aligned}(I_{n-2}|Z_{(n-2)\times 2})U &= \left(I_{n-2} \middle| \begin{matrix} 0 \\ \vdots \\ 0 \end{matrix}\right)\left(I_{n-1} \middle| \begin{matrix} 0 \\ \vdots \\ 0 \end{matrix}\right)U \\ &= \left(I_{n-2} \middle| \begin{matrix} 0 \\ \vdots \\ 0 \end{matrix}\right) C_1 \left(I_{n-1} \middle| \begin{matrix} 0 \\ \vdots \\ 0 \end{matrix}\right)\hat{H}_1 \\ &= C_2 \left(I_{n-2} \middle| \begin{matrix} 0 \\ \vdots \\ 0 \end{matrix}\right) H_2 \left(I_{n-1} \middle| \begin{matrix} 0 \\ \vdots \\ 0 \end{matrix}\right)\hat{H}_1\end{aligned}$$

$$= C_2 \left(I_{n-2} \middle| \begin{matrix} 0 \\ \vdots \\ 0 \end{matrix}\right)\left(I_{n-1} \middle| \begin{matrix} 0 \\ \vdots \\ 0 \end{matrix}\right)\hat{H}_2\hat{H}_1 \quad \text{by Lemma 1, and therefore}$$

$$= C_2\ (I_{n-2}|Z_{(n-2)\times 2})\hat{H}_2\hat{H}_1,$$

where $\hat{H}_2 = \left(\begin{array}{c|c} H_2 & \begin{matrix} 0 \\ \vdots \\ 0 \end{matrix} \\ \hline 0\ldots 0 & 1 \end{array}\right) \in \mathcal{H}_n$.

Suppose now that the theorem is true for p; we show it is true for $p+1$. (That is, suppose we have constructed p $n \times n$ Householder matrices $\hat{H}_1,\ldots,\hat{H}_p$ having property (*). We construct $\hat{H}_{p+1}$, having the desired form, and such that $(I_{n-(p+1)}|Z)U = C_{p+1}(I_{n-(p+1)}|Z)\hat{H}_{p+1}\hat{H}_p \ldots \hat{H}_1$, where $C_{p+1} \in \mathcal{U}_{n-(p+1)}$.)

Let $(I_{n-p}|Z)U = C_p(I_{n-p}|Z)\hat{H}_p \ldots \hat{H}_1$, where $C_p \in \mathcal{U}_{n-p}$ and, for each $i = 1,\ldots,p$, $\hat{H}_i$ has property (*). Then by result (ii), applied to $\left(I_{n-(p+1)} \middle| \begin{matrix} 0 \\ \vdots \\ 0 \end{matrix}\right) C_p$, and by Lemma 2 there exist $C_{p+1} \in \mathcal{U}_{n-(p+1)}$ and $H_{p+1} \in \mathcal{H}_{n-p}$ such that $\left(I_{n-(p+1)} \middle| \begin{matrix} 0 \\ \vdots \\ 0 \end{matrix}\right) C_p = C_{p+1}\left(I_{n-(p+1)} \middle| \begin{matrix} 0 \\ \vdots \\ 0 \end{matrix}\right) H_{p+1}$. Thus, if $\hat{H}_{p+1} = \left(\begin{array}{c|c} H_{p+1} & Z \\ \hline Z & I_{n-(p+1)} \end{array}\right)$, then $\hat{H}_{p+1} \in \mathcal{H}_n$ and

$$(I_{n-(p+1)}|Z)U = \left(I_{n-(p+1)} \middle| \begin{matrix} 0 \\ \vdots \\ 0 \end{matrix}\right) (I_{n-p}|Z_{(n-p)\times p})U$$

$$= \left(I_{n-(p+1)} \middle| \begin{matrix} 0 \\ \vdots \\ 0 \end{matrix}\right) C_p \; (I_{n-p}|Z_{(n-p)\times p})\hat{H}_p \ldots \hat{H}_1$$

$$= C_{p+1} \left(I_{n-(p+1)} \middle| \begin{matrix} 0 \\ \vdots \\ 0 \end{matrix}\right) H_{p+1} \; (I_{n-p}|Z_{(n-p)\times p})\hat{H}_p \ldots \hat{H}_1$$

$$= C_{p+1}\left(I_{n-(p+1)} \middle| \begin{matrix} 0 \\ \vdots \\ 0 \end{matrix}\right) (I_{n-p}|Z_{(n-p)\times p})\hat{H}_{p+1}\hat{H}_p \ldots \hat{H}_1$$

$$= C_{p+1}(I_{n-(p+1)}|Z)\hat{H}_{p+1}\hat{H}_p \ldots \hat{H}_1.$$

Thus, by induction, the proof is complete.

As already remarked in Section 2, it has recently been proposed that one might obtain a "best B" by successively maximizing over the set $\mathcal{H}_n$ the average interclass divergence

$$D_{(I_k|Z)X\,H_i \ldots H_1} : \mathcal{H}_n \to \mathcal{R}^{\#},$$

where $X \in \mathcal{H}_n$. (See Section 2 for a full statement of the proposed procedure.)

Another procedure, closely related to the preceding, is to successively maximize over the set $\mathcal{H}_{n-p}$ the function:

$$D_{\alpha_{p+1}} : \mathcal{H}_{n-p} \to \mathcal{R}^{\#},$$

where $\alpha_{p+1} = \left(I_{n-(p+1)} \left| \begin{matrix} 0 \\ \vdots \\ 0 \end{matrix} \right.\right) X \left(I_{n-p} \left| \begin{matrix} 0 \\ \vdots \\ 0 \end{matrix} \right.\right) H_p \cdots \left(I_{n-1} \left| \begin{matrix} 0 \\ \vdots \\ 0 \end{matrix} \right.\right) H_1,$

and, for each $i = 1, \ldots, p$, $H_i \in \mathcal{H}_{n-(i-1)}$ maximizes the function

$$D_{\alpha_i} : \mathcal{H}_{n-(i-1)} \to \mathcal{R}^{\#},$$

where $\alpha_i = \left(I_{n-i} \left| \begin{matrix} 0 \\ \vdots \\ 0 \end{matrix} \right.\right) X \left(I_{n-(i-1)} \left| \begin{matrix} 0 \\ \vdots \\ 0 \end{matrix} \right.\right) H_{n-(i-1)} \cdots \left(I_{n-1} \left| \begin{matrix} 0 \\ \vdots \\ 0 \end{matrix} \right.\right) H_1.$

Note that the problem of obtaining $H_{p+1} \in \mathcal{H}_{n-p}$ which maximizes the above function is a "new feature selection problem". In particular, the problem of maximizing over the set $\mathcal{H}_{n-1}$ the function:

$$D_{\left(I_{n-2} \left| \begin{smallmatrix} 0 \\ \vdots \\ 0 \end{smallmatrix} \right.\right) X \left(I_{n-1} \left| \begin{smallmatrix} 0 \\ \vdots \\ 0 \end{smallmatrix} \right.\right) H_1} : \mathcal{H}_{n-1} \to \mathcal{R}^{\#},$$

is the new feature selection problem with new statistics $\left(I_{n-1} \left| \begin{matrix} 0 \\ \vdots \\ 0 \end{matrix} \right.\right) H_1 \Omega_i H_1 \left(\frac{I_{n-1}}{0 \ldots 0}\right)$ and $\left(I_{n-1} \left| \begin{matrix} 0 \\ \vdots \\ 0 \end{matrix} \right.\right) H_1 \mu_i$, $i = 1, \ldots, m$, as "new covariance matrices and means", respectively (where $\Omega_1, \ldots, \Omega_m$ and $\mu_1, \ldots, \mu_m$ denote the covariance matrices and means, respectively, for the m-classes of the original problem). More details on this procedure will appear elsewhere.

Now, $B = \left(I_{n-p} \,\middle|\, \begin{matrix} 0 \\ \vdots \\ 0 \end{matrix} \right) H_p \cdots \left(I_{n-1} \,\middle|\, \begin{matrix} 0 \\ \vdots \\ 0 \end{matrix} \right) H_1$ is a $(n-p)\times n$ matrix and hence D_B is a particular value for the average interclass divergence for the case $k = n-p$; thus $D_B \le D_{B_{n-p}}$, where $D_{B_{n-p}}$ denotes the maximum value of the average interclass divergence for the case $k = n-p$. Is it possible that $D_B = D_{B_{n-p}}$? Less hopefully, can we determine a $U \in \mathcal{U}_n$ such that $D_B = D_{(I_{n-p}|Z)U}$?

Theorem 3, or more precisely, the proof used in Theorem 3, provides an affirmative answer to our last question. In particular, the proof of Theorem 3 shows that

$$
\begin{aligned}
B &= \left(I_{n-p} \,\middle|\, \begin{matrix} 0 \\ \vdots \\ 0 \end{matrix} \right) H_p \cdots \left(I_{n-1} \,\middle|\, \begin{matrix} 0 \\ \vdots \\ 0 \end{matrix} \right) H_1 \\
&= \left(I_{n-p} \,\middle|\, \begin{matrix} 0 \\ \vdots \\ 0 \end{matrix} \right)\left(I_{n-(p-1)} \,\middle|\, \begin{matrix} 0 \\ \vdots \\ 0 \end{matrix} \right) \cdots \left(I_{n-1} \,\middle|\, \begin{matrix} 0 \\ \vdots \\ 0 \end{matrix} \right) \hat{H}_p \hat{H}_{p-1} \cdots \hat{H}_1 \\
&= (I_{n-p} | Z_{(n-p)\times p}) \hat{H}_p \cdots \hat{H}_1,
\end{aligned}
$$

where $\hat{H}_j = \left(\begin{array}{c|c} H_i & Z \\ \hline Z & I_{j-1} \end{array} \right) \in \mathcal{H}_n$, for $j = 2,\ldots,p$, $\hat{H}_1 = H_1$. Therefore, $D_B = D_{(I_{n-p}|Z)U}$; where $U = \hat{H}_p \cdots \hat{H}_1 \in \mathcal{U}_n$.

SECTION 4 - A CONCLUDING OBSERVATION

We conclude the paper with an observation and two questions. Homer Walker has pointed out to us that if U is any $n \times n$ matrix, then

$D_{(I_k|Z)U} \le D_{(I_{k+1}|Z)U}$. In particular, this is true if U is unitary and if $D_{(I_k|Z)U}$ is the maximum value for D_B, $B \in M_{kn} = \{A|A$ is a $k \times n$ matrix with real entries$\}$.

Using this fact, observe that if

$$D_{(I_k|Z)U_k} = D_{(I_k|Z)\begin{pmatrix} u_1^{(k)} \\ \vdots \\ u_n^{(k)} \end{pmatrix}} \quad \text{and} \quad D_{(I_{k+1}|Z)U_{k+1}} = D_{(I_{k+1}|Z)\begin{pmatrix} u_1^{(k+1)} \\ \vdots \\ u_n^{(k+1)} \end{pmatrix}}$$

are maximum values for the case k and $k+1$, respectively (where $u_i^{(j)}$ denotes the ith row of U_j), then we have

$$D_{(I_k|Z)U_{k+1}} = D_{\begin{pmatrix} u_1^{(k+1)} \\ \vdots \\ u_k^{(k+1)} \end{pmatrix}}$$

$$\le D_{(I_k|Z)U_k} = D_{\begin{pmatrix} u_1^{(k)} \\ \vdots \\ u_k^{(k)} \end{pmatrix}}$$

$$\le D_{(I_{k+1}|Z)U_k} = D_{\begin{pmatrix} u_1^{(k)} \\ \vdots \\ u_k^{(k)} \\ u_{k+1}^{(k)} \end{pmatrix}}$$

$$\le D_{(I_{k+1}|Z)U_{k+1}} = D_{\begin{pmatrix} u_1^{(k+1)} \\ \vdots \\ u_k^{(k+1)} \\ u_{k+1}^{(k+1)} \end{pmatrix}}$$

This motivates the following question.

Question 1: Let $D_{(I_j|Z)U_j} = D_{\begin{pmatrix} u_1^{(j)} \\ \vdots \\ u_j^{(j)} \end{pmatrix}}$ be the maximum value for D_B, $B \in M_{jn}$, where $u_i^{(j)}$ denotes the ith row of U_j, an $n \times n$ unitary matrix, for each $j = 1, \ldots, n$, $i = 1, \ldots, j$. Let

$$C_k = \begin{pmatrix} u_1^{(1)} \\ u_2^{(2)} \\ \vdots \\ u_k^{(k)} \end{pmatrix}$$

for each $k = 1, \ldots, n$. Does $D_{C_k} = D_{(I_k|Z)U_k}$ for any $k = 1, \ldots, n$? If not, do there exist $n \times n$ unitary matrices $\hat{U}_1, \ldots, \hat{U}_n$ for which the answer is yes?

Finally, we close with a somewhat unrelated (to what we have discussed in this paper) question. We have already observed that there existsan $n \times n$ Householder matrix H such that $D_{(I_{n-1}|Z)H}$ is the maximum value for D_B, $B \in M_{n-1,n}$. For $k < n-1$, we do not know if there always existsan $n \times n$ Householder matrix H_k for which $D_{(I_k|Z)H_k}$ is the maximum value of D_{B_k}, $B_k \in M_{kn}$. (In fact, the evidence seems to indicate the contrary, in general). However, we propose the following question.

Question 2: Suppose $D_{(I_k|Z)H_k}$ is the maximum value for D_{B_k}, $B_k \in M_{kn}$, where H_k is an $n \times n$ Householder matrix and $k < n-1$. Is $D_{(I_j|Z)H_k}$ the maximum value for D_{B_j}, $B_j \in M_{jn}$, for any $k < j \leq n-1$?

REFERENCES

1. Brown, D. and O'Malley, M., "Best B's need not be row equivalent", to appear.

2. Finkbeiner, D., Introduction to Matrices and Linear Transformations, 2nd Edition, (1966), W. H. Freeman and Company, San Francisco.