DEPARTMENT OF MATHEMATICS
UNIVERSITY OF HOUSTON HOUSTON, TEXAS

NASA CR- 141879

(NASA-CR-141879) HOUSEHOLDER TRANSFORMATIONS AND OPTIMAL LINEAR COMBINATIONS (Houston Univ.) 14 p HC $3.25 CSCL 12A

N75-26742

Unclas
G3/64 27275

HOUSEHOLDER TRANSFORMATIONS AND OPTIMAL LINEAR COMBINATIONS

BY HENRY P. DECELL, JR.
& WILLIAM SMILEY

JUNE 1974 REPORT # 38

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

3801 CULLEN BLVD.
HOUSTON, TEXAS 77004

Householder Transformations and Optimal Linear Combinations

June, 1974

Henry P. Decell, Jr. and W. G. Smiley, III
Department of Mathematics
University of Houston

Report # 38

Householder Transformations and Optimal Linear Combinations

Henry P. Decell, Jr. and W. G. Smiley, III

Department of Mathematics
University of Houston

I. Preliminaries

Notation. Let M_n be the set of linear transformations on R^n, and suppose that O_n denotes the group of orthogonal transformations on R^n.

Theorem 1. In the norm operator topology on M_n, O_n is a compact topological group. [1]

Theorem 2. If $A \in O_n$, then $\det(A) = \pm 1$. [7]

Definition 1. The orthogonal transformation A is a rotation, in case $\det(A) = 1$. Otherwise, A is called a reflection. Suppose that R is the subgroup of O_n consisting of all rotations, and let $\hat{R}$ denote $O_n \sim R$.

Theorem 3. In the norm operator topology on M_n, O_n consists of the components R and $\hat{R}$. Hence, R and $\hat{R}$ are compact. [1]

Definition 2. Let B denote $\{x : \|x\| = 1\}$, the set of unit vectors in R^n.

Theorem 4. In the norm topology on R^n, the set B is compact and connected.

Proof. Since B is closed and bounded, it is compact. Moreover, B is the continuous image of $R^n \sim \{\theta\}$ and hence is connected.

Definition 3. For each $x \in B$, let the Householder transformation H_x be defined by $H_x = I - 2xx^T$. Let H_n denote $\{H_x : x \in B\}$, the set of all Householder transformations. [5]

<u>Theorem 5.</u> The set H_n is a compact, connected subset of $\hat{R}$.

<u>Proof.</u> By the continuity of matrix operations, H_n is the continuous image of B under the mapping $x \to I - 2xx^T$. Thus H_n is compact and connected. Choose $x \in B$. Since $(I - 2xx^T)^T = I - 2xx^T$ and $(I - 2xx^T)^2 = I$, it follows that $I - 2xx^T \in O_n$. Also, $\det(I - 2xx^T) = -1$ implies that $I - 2xx^T \in \hat{R}$.

<u>Theorem 6.</u> (Householder) If $y \in R^n \setminus \{\theta\}$ and $x \in B$, then there exists a vector $w \in B$ such that $(I - 2ww^T)y = \| y \|x$. [4]

<u>Proof.</u> In the case where $y = \| y \|x$, choose any w satisfying $\langle w,y \rangle = 0$.

Otherwise, let w be defined by $w = \dfrac{y - \| y \|x}{\|y - \| y \|x\|}$.

<u>Corollary 7.</u> For each $x,y \in B$, there exists some $w \in B$ satisfying $H_w(y) = x$.

It has been shown by Decell in [2], for optimal selection of linear combinations (feature selection), that the search for an optimal solution ($k \times n$, rank k matrix B) may be restricted to the set of $k \times n$ matrices of the form $B = (I_k|Z)U$, where U is an $n \times n$ orthogonal matrix.
H. Walker has shown that, given an optimal linear transformation $(I_k|Z)U$, there exists a positive integer $p \leq \min\{k,n-k\}$ such that $(I_k|Z)U$ may be factored into the product $(I_k|Z)H_p \cdots H_1$, for some $H_1,\ldots,H_p \in H_n$. Note that Theorems 8 and 10, with their corollaries, in addition to establishing the existence of the $p \leq \min\{k,n-k\}$ factors $H_1,\cdots H_p$, yield Walker's result for a very particular sequence of transformations in H_n (i.e., those derived by Householder's technique for upper triangularization of U^T).

These remarks apply to all separability criteria which are invariant under nonsingular transformation (e.g., probability of misclassification, Divergence, Bhattacharyya distance, Chernoff distance, Transformed Divergence). This discussion will be summarized in Theorem 12.

Theorem 8. Let $\{e_1,\ldots,e_n\}$ denote the usual orthonormal basis for R^n. Suppose that $\{u_1,\ldots,u_n\}$ is any orthonormal set of vectors. Then there exist transformations $H_1,\ldots,H_{n-1} \in \mathcal{H}_n$ such that:

(1) if $i \leq n-1$, then $H_i \cdots H_1 u_j = e_j$ for all $j \leq i$

(2) in addition, if $i = n-1$, $H_{n-1} \cdots H_1 u_n = \pm e_n$.

Proof. If $i = 1$, by Theorem 6 there exists a transformation $H_1 \in \mathcal{H}_n$ such that $H_1 u_1 = e_1$. Let $p < n-1$. Suppose that $H_1,\ldots,H_p \in \mathcal{H}_n$ have been chosen such that if $i \leq p$, then $H_i \cdots H_1 u_j = e_j$ for all $j \leq i$. Let the vector $u = H_p \cdots H_1 u_{p+1}$, and suppose that $\hat{u}$ denotes the vector in R^{n-p} which consists of the last $n-p$ components of u. Likewise, let $\hat{e}_1$ consist of the last $n-p$ components of e_{p+1}. Since H_i is an isometry for each $i = 1,\ldots,p$, we have $\| u \| = 1$. It follows from the relations $\langle u,e_j \rangle = 0$, $j = 1,\ldots,p$, that $\hat{u}$ is a unit vector in R^{n-p}. Again using Theorem 6, choose $\hat{H} = I_{n-p} - 2x_{n-p}x_{n-p}^T \in \mathcal{H}_{n-p}$ such that $\hat{H}\hat{u} = \hat{e}_1 \in R^{n-p}$. Define

$$H_{p+1} = \left(\begin{array}{c|c} I_p & Z \\ \hline Z & \hat{H} \end{array}\right) = I - 2\left(\begin{array}{c} Z \\ \hline x_{n-p} \end{array}\right)(Z|x_{n-p}^T) \in \mathcal{H}_n ,$$

where each occurrence of Z denotes the appropriate matrix or vector of zeros. It is evident that if $i \leq p+1$, then for all $j \leq i$,

$H_{p+1}H_p\cdots H_1u_j = H_{p+1}e_j = \left(\begin{array}{c|c} I_p & Z \\ \hline Z & \hat{H} \end{array}\right) e_j = e_j$. Therefore, by induction, (1) holds for all $i = 1,\ldots,n-1$. Finally, given the Householder transformations $H_1,\ldots,H_{n-1}$ constructed above, observe that (2) follows from $\| H_{n-1}\cdots H_1u_n \| = 1$ and the relations $\langle H_{n-1}\cdots H_1u_n, e_j\rangle = 0$, $j = 1,\ldots,n-1$.

Corollary 9. If U is an orthogonal matrix, then there exist transformations $H_1,\ldots,H_{n-1} \in \mathcal{H}_n$ such that:

(1) $H_{n-1}\cdots H_1U^T = \left(\begin{array}{c|c} I_{n-1} & Z \\ \hline Z & \pm 1 \end{array}\right)$

(2) $\left(\begin{array}{c|c} I_{n-1} & Z \\ \hline Z & \pm 1 \end{array}\right) H_{n-1}\cdots H_1U^T = I$

(3) $U^T \left(\begin{array}{c|c} I_{n-1} & Z \\ \hline Z & \pm 1 \end{array}\right) H_{n-1}\cdots H_1 = I$

(4) for $p = 1,\ldots,n-2$, there exists a unit vector $x_{n-p} \in R^{n-p}$ such that

$$H_{p+1} = I - 2\begin{pmatrix} Z \\ x_{n-p} \end{pmatrix} (Z | x_{n-p}^T) = \left(\begin{array}{c|c} I_p & Z \\ \hline Z & I_{n-p} - 2x_{n-p}x_{n-p}^T \end{array}\right)$$

(5) exactly one of the following holds:

a) $H_{n-1}\cdots H_1U^T = I$

b) $H_nH_{n-1}\cdots H_1U^T = I$, where

$$H_n = \left(\begin{array}{c|c} I_{n-1} & Z \\ \hline Z & -1 \end{array}\right) = I - 2e_ne_n^T \in \mathcal{H}_n$$

(6) if (5) a) holds, then $U = (U^T)^{-1} = H_{n-1}\cdots H_1$ and if (5) b) holds, then $U = (U^T)^{-1} = H_nH_{n-1}\cdots H_1$

(7) $(I_k|Z)H_nH_{n-1}\cdots H_1 = (I_k|Z)H_k\cdots H_1$.

Proof. Since U is an orthogonal matrix, we also have $U^T \in \mathcal{O}_n$. Thus the columns of U^T form an orthonormal set of vectors $\{u_1,\ldots,u_n\}$. Choose transformations $H_1,\ldots,H_{n-1} \in \mathcal{H}_n$ satisfying (1) and (2) of Theorem 8. Because $U^T = (u_1|\cdots|u_n)$, parts (1) through (6) are immediate consequences of this theorem. Part (7) follows from the observation that

$$(I_k|Z)H_{k+j} = (I_k|Z)$$

for all $j = 1,\ldots,n-k$.

One should observe that Theorem 8 may be restated in the following form.

Theorem 10. Let $\{u_1,\ldots,u_n\}$ be an orthonormal set of vectors in R^n. Then there exist $H_1,\ldots,H_{n-1} \in \mathcal{H}_n$ such that

(1) if $i \le n-1$, then $H_i\cdots H_1u_j = e_j$ for all $n+1-i \le j \le n$

(2) $H_{n-1}\cdots H_1u_1 = \pm e_1$.

Corollary 11. If U is an orthogonal matrix, then there exist transformations $H_1,\ldots,H_{n-1} \in \mathcal{H}_n$ such that:

(1) $H_{n-1}\cdots H_1U^T = \left(\begin{array}{c|c} \pm 1 & Z \\ \hline Z & I_{n-1} \end{array}\right)$

(2) $\left(\begin{array}{c|c} \pm 1 & Z \\ \hline Z & I_{n-1} \end{array}\right)H_{n-1}\cdots H_1U^T = I$

(3) $U^T\left(\begin{array}{c|c} \pm 1 & Z \\ \hline Z & I_{n-1} \end{array}\right)H_{n-1}\cdots H_1 = I$

(4) for $p = 1,\ldots,n-2$, there exists a unit vector $x_{n-p} \in R^{n-p}$

such that

$$H_{p+1} = I - 2\begin{pmatrix} x_{n-p} \\ \hline Z \end{pmatrix}(x_{n-p}^T | Z) = \left(\begin{array}{c|c} I_{n-p} - 2x_{n-p}x_{n-p}^T & Z \\ \hline Z & I_p \end{array}\right)$$

(5) exactly one of the following holds:

a) $H_{n-1} \cdots H_1 U^T = I$

b) $H_n H_{n-1} \cdots H_1 U^T = I$, where

$$H_n = \left(\begin{array}{c|c} -1 & Z \\ \hline Z & I_{n-1} \end{array}\right) = I - 2e_1 e_1^T \in \mathcal{H}_n$$

(6) if (5) a) holds, then $U = H_{n-1} \cdots H_1$

if (5) b) holds, then $U = H_n H_{n-1} \cdots H_1$

(7) $(I_k | Z) H_n H_{n-1} \cdots H_1 = A(I_k | Z) H_{n-k} \cdots H_1$,

for some nonsingular $k \times k$ matrix A.

<u>Proof.</u> Parts (1) through (6) of this corollary follow directly from Theorem 10. Part (7) follows by observing that for $p = n-k+1$,

$$(I_k | Z) H_p = (I_k | Z) \left(\begin{array}{c|c} I_k - 2x_k x_k^T & Z \\ \hline Z & I_{n-k} \end{array}\right) = (I_k - 2x_k x_k^T)(I_k | Z)$$

and, for $p > n-k+1$ and $(n-p) + r = k$,

$$(I_k | Z) H_p = (I_k | Z) \left(\begin{array}{c|c|c} I_{n-p} - 2x_{n-p}x_{n-p}^T & Z & Z \\ \hline Z & I_r & Z \\ \hline Z & Z & I_s \end{array}\right) = \left(\begin{array}{c|c} I_{n-p} - 2x_{n-p}x_{n-p}^T & Z \\ \hline Z & I_r \end{array}\right)(I_k | Z).$$

<u>Theorem 12.</u> Let ψ be any real-valued separability criterion which is invariant under nonsingular transformation. If B is a rank k, $k \times n$ ψ-optimal solution of the feature selection problem, then there exist at most $m = \min\{k, n-k\}$ Householder transformations such that

$(I_k|Z)H_m \cdots H_1$ is ψ-optimal (i.e., $\psi_{(I_k|Z)H_m \cdots H_1} = \psi_B$).

<u>Proof.</u> Recall that $B = (I_k|Z)U$, for some orthogonal matrix U. The proof then consists of selecting m to be the smaller of k and $n-k$, and subsequently applying either Corollary 9 or Corollary 11.

II. Separability Criteria and Suggested Algorithms for Optimal (Suboptimal) Linear Combinations

Let ψ be any continuous real function of the matrix variable $(I_k|Z)H$. Since $\mathcal{H}_n$ is compact and $\psi_{(I_k|Z)H} : \mathcal{H}_n \to R$, there is an $H_1 \in \mathcal{H}_n$ such that

$$\psi_{(I_k|Z)H_1} = \underset{H \in \mathcal{H}_n}{\text{l.u.b.}}\ \psi_{(I_k|Z)H} .$$

Theorem 13. For each positive integer i, let the element H_i of $\mathcal{H}_n$ be chosen such that

$$\psi_{(I_k|Z)H_i H_{i-1} \cdots H_1} = \underset{H \in \mathcal{H}_n}{\text{l.u.b.}}\ \psi_{(I_k|Z)HH_{i-1} \cdots H_1} .$$

It follows that, for each i,

(1) $\psi_{(I_k|Z)H_i \cdots H_1} \le \psi_{(I_k|Z)H_{i+1}H_i \cdots H_1}$

(2) $\psi_{(I_k|Z)H_i \ldots H_1 H} \le \psi_{(I_k|Z)H_{i+1}H_i \ldots H_1}$ for every $H \in \mathcal{H}_n$

(3) $\psi_{(I_k|Z)HH_i \ldots H_1} \le \psi_{(I_k|Z)H_{i+1} \cdot H_i \cdots H_1}$ for every $H \in \mathcal{H}_n$

(4) $\psi_{(I_k|Z)H_i \ldots H_{i-p} HH_{i-(p+1)} \cdots H_1} \le \psi_{(I_k|Z)H_{i+1}H_i \cdots H_1}$

for every $H \in \mathcal{H}_n$

and $p = 0, \ldots, i-2$.

Proof. As in the proof of Corollary 9, we may choose $H \in \mathcal{H}_n$ such that $(I_k|Z)H = (I_k|Z)$ and use the definition of $\psi_{(I_k|Z)H_{i+1}H_i \ldots H_1}$ to conclude that

$$\psi_{(I_k|Z)H_i \ldots H_1} = \psi_{(I_k|Z)HH_i \ldots H_1} \le \psi_{(I_k|Z)H_{i+1}H_i \cdots H_1} .$$

Now let $H \in \mathcal{H}_n$ and define $\bar{H} = H_i(\ldots(\ldots(H_1 H H_1)\ldots)\ldots)H_i$. By Theorem 10, inductively conclude that $\bar{H} \in \mathcal{H}_n$, and therefore that

$$\psi(I_k|Z)\bar{H}H_i\ldots H_1 \leq \psi(I_k|Z)H_{i+1}H_i\ldots H_1.$$

However, $\bar{H}H_i\ldots H_1 = H_i\ldots H_1 H H_1\ldots H_i H_i \ldots H_1 = H_i H_{i-1}\ldots H_1 H$, so (2) follows. Clearly, (3) holds by the definition of H_{i+1}. In order to conclude (4), suppose $0 \leq p \leq i-2$ and $H \in \mathcal{H}_n$. For this case, define

$$\bar{H} = H_1 \ldots H_{i-p} H H_{i-(p+1)} \ldots H_1.$$

Since $\bar{H} \in \mathcal{H}_n$, statement (4) follows from (2).

Theorem 14. If the hypothesis of Theorem 13 is satisfied and the sequence

$$\{\psi_{(I_k|Z)H_i\ldots H_1}\}_{i=1}^{\infty} \text{ is bounded above,}$$

then

$$\lim_{i\to\infty} \psi_{(I_k|Z)H_i\ldots H_1} = \text{l.u.b.}\{\psi_{(I_k|Z)H_i\ldots H_1} \mid i \text{ is a positive integer}\}.$$

Remark. It is clear that Theorems 13 and 14 remain true if "l.u.b." is replaced by "g.l.b.", "bounded above" is replaced by "bounded below", and the inequalities are reversed.

Note that by definition, Divergence, Bhattacharyya distance, probability of misclassification, etc., all satisfy the hypothesis of Theorems 12 and 13. These theorems give rise to a sequential monotone procedure for possibly obtaining a ψ-extremal rank k linear combination matrix. At each stage in the procedure, the extremal problem is a function of only n variables. At this point, we conjecture that the process should terminate in at most $\min\{k,n-k\}$ steps. This conjecture is clearly in line with the $\min\{k,n-k\}$ representation of the actual ψ-extremal solution. In addition, all tests of the algorithm on real data further lead one to believe that the conjecture is fact. It is evident that this procedure is at worst sub-optimal.

S. Marani, in [6], gives details of computational results obtained using the sequential procedure and a very crude differential correction scheme to solve the n-dimensional variational problem for the generators of the H_i's at each stage of the procedure. The initial guess used at each stage of the process was arbitrarily set at $X = (\frac{1}{\sqrt{n}}, \frac{1}{\sqrt{n}}, \ldots, \frac{1}{\sqrt{n}})^T$. Even using this rough initial guess and the differential correction scheme, convergence seems to be fairly rapid. Moreover, one would reasonably expect to reduce iteration time using an improved scheme for successive initial guesses at each stage of the procedure, together with a more sophisticated iteration procedure.

The results obtained by Marani, despite use of the arbitrary initial guess at each of the stages, match the results of known Divergence-optimal solutions calculated by J. Quirein in [2]. Note that the total number of scalar variables involved in this process cannot exceed

$$\begin{aligned} V &= n + (n-1) + (n-2) + \ldots + (n-(m-1)) \\ &= nm - \tfrac{1}{2}m(m-1) , \end{aligned}$$

where $m = \min\{k,n-k\}$.

In [3], Decell and Mayekar have developed an analytical expression, in the case ψ = Divergence, for the variational equations as a function of the Householder generators for every H_i. This expression is utilized in Marini's calculations. We should further point out that these results only depend on

1. The continuity of ψ
2. The compactness of $\mathcal{H}_n$
3. The invariance of ψ under nonsingular transformation.

The following are several related open questions:

1. Does the process terminate in at most min{k,n-k} steps?
2. Does it terminate at an absolute ψ-extremum (rank-k maximal statistic)?
3. Given ψ_1 and ψ_2, under what conditions is a ψ_1-extremal solution also a ψ_2-extremal solution?
4. If the process does not terminate in a finite number of steps, what can be said of the cluster points of the sequence $\{H_i \ldots H_1\}_{i=2}^{\infty}$? (Recall that $\mathcal{O}_n$ is compact).

III. Several Useful Theorems

Theorem 15. Suppose that $H_x \in \mathcal{H}_n$. Then $UH_xU^T \in \mathcal{H}_n$, with Ux a generator of UH_xU^T, for all $U \in \mathcal{O}_n$. In particular, for every $H_y \in \mathcal{H}_n$, $H_yH_xH_y \in \mathcal{H}_n$. In this case, the Householder transformation $H_yH_xH_y$ is generated by the vector $x - 2\langle x,y\rangle y$.

Proof. Suppose $U \in \mathcal{O}_n$. Since $\| Ux \| = \| x \| = 1$, it follows that

$UH_xU^T = U(I - 2xx^T)U^T = I - 2(Ux)(Ux)^T \in \mathcal{H}_n$. If U is chosen to be some $H_y \in \mathcal{H}_n$, then the generator of $H_yH_xH_y$ is

$H_yx = (I - 2yy^T)x = x - 2y(y^Tx) = x - 2\langle x,y\rangle y$.

Theorem 16. For each H_x and H_y in $\mathcal{H}_n$, there exists some $H \in \mathcal{H}_n$ such that $HH_yH = H_x$.

Proof. By Corollary 7, there exists a transformation $H \in \mathcal{H}_n$ satisfying $Hy = x$. Therefore

$$HH_yH = H(I - 2yy^T)H = H^2 - 2Hyy^TH = H^2 - 2(Hy)(Hy)^T = I - 2xx^T = H_x.$$

Theorem 17. Suppose that $H_x, H_y \in \mathcal{H}_n$. Then the following statements are equivalent:

(1) $\langle x,y\rangle = 0$

(2) $x \neq \pm y$ and $H_xH_y = H_yH_x = H_x + H_y - I$.

Proof. Let $\langle x,y\rangle = 0$. Then clearly $x = \pm y$. Moreover,

$$H_xH_y = (I - 2xx^T)(I - 2yy^T) = I - 2xx^T - 2yy^T = I - 2xx^T + I - 2yy^T - I$$
$$= H_x + H_y - I = H_y + H_x - I = H_yH_x.$$

Conversely, suppose $H_xH_y = H_yH_x$. Then

$Z = H_xH_y - H_yH_x = 4(xx^Tyy^T - yy^Txx^T)$. Using the fact that $\langle x,y\rangle = x^Ty = y^Tx$, we obtain $4\langle x,y\rangle(xy^T - yx^T) = Z$. Since $x \neq \pm y$ implies $xy^T - yx^T \neq Z$, it follows that $\langle x,y\rangle = 0$.

<u>Theorem 18.</u> Suppose $H_x, H_y \in \mathcal{H}_n$. Then the following are equivalent:

(1) $\langle x,y\rangle = \pm 1$

(2) $x = \pm y$

(3) $H_x = H_y$.

<u>Proof.</u> Let $H_x = H_y$. Then $xx^T = yy^T$, and hence $x(x^Tx) = y(y^Tx)$. Since $\| x \|^2 = 1$, it follows that $x = (y^Tx)y$. Finally, $\| x \| = |y^Tx| \cdot \| y \|$ implies that $|\langle x,y\rangle| = |y^Tx| = 1$. The remaining parts of the proof are immediate.

REFERENCES

1. Chevalley, Claude, Theory of Lie Groups, Princeton University Press, Princeton, 1946.

2. Decell, H. P. Jr. and Quirein, J. A., "An Iterative Approach to the Feature Selection Problem", Report #26 NAS-9-12777, Department of Mathematics, University of Houston, March 1973.

3. Decell, Henry P. Jr. and Mayekar, Mike, "On the Variational Equations for Householder Transformations in Feature Selection", Report #39 NAS-9-12777, Department of Mathematics, University of Houston.

4. Householder, Alston S., "Unitary Triangularization of a Non-symmetric Matrix", J. Assoc. Comput. Mach., 5(1958), 339 - 342.

5. Householder, Alston S., The Theory of Matrices in Numerical Analysis, Blaisdell, New York, 1964.

6. Marani, Salma, "Routine to Find the Maximum Average Divergence", Report #36 NAS-9-12777, Department of Mathematics, University of Houston, August 1974.

7. Weyl, Hermann, The Classical Groups, Second Edition, Princeton University Press, Princeton, 1946.