ON STEWART'S SINGULAR VALUE DECOMPOSITION FOR PARTITIONED ORTHOGONAL MATRICES

by

Charles Van Loan

STAN-CS-79-767
September 1979

DEPARTMENT OF COMPUTER SCIENCE
School of Humanities and Sciences
STANFORD UNIVERSITY

ON STEWART'S SINGULAR VALUE DECOMPOSITION FOR PARTITIONED ORTHOGONAL MATRICES

Charles Van Loan

Department of Computer Science
Upson Hall, Cornell University
Ithaca, New York 14853

This research was supported in part by the National Science Foundation Grant No. MCS-78-11985.

ABSTRACT.

A variant of the singular value decomposition for orthogonal matrices due to G. W. Stewart is discussed. It is shown to be useful in the analysis of (a) the total least squares problem, (b) the Golub-Klema-Stewart subset selection algorithm, and (c) the algebraic Riccati equation.

1. Introduction.

In a recent survey article G. W. Stewart [8] presented the following variant of the singular value decomposition (SVD):

Theorem 1.

If $Q \in \mathbb{R}^{m \times m}$ is orthogonal and partitioned as follows,

$$Q = \begin{bmatrix} Q_{11} & Q_{12} \\ Q_{21} & Q_{22} \end{bmatrix} \begin{matrix} k \\ p \end{matrix} \qquad k+p=m, \quad k \geq p$$
$$\qquad\quad k \qquad p$$

then there exist orthogonal U_1 and V_1 in $\mathbb{R}^{k \times k}$ and orthogonal U_2 and V_2 in $\mathbb{R}^{p \times p}$ such that

$$\begin{bmatrix} U_1 & 0 \\ 0 & U_2 \end{bmatrix}^T \begin{bmatrix} Q_{11} & Q_{12} \\ Q_{21} & Q_{22} \end{bmatrix} \begin{bmatrix} V_1 & 0 \\ 0 & V_2 \end{bmatrix} = \begin{bmatrix} I_{k-p} & 0 & 0 \\ 0 & C & S \\ 0 & -S & C \end{bmatrix} \begin{matrix} k-p \\ p \\ p \end{matrix}$$
$$k-p \quad p \quad p$$

where

$$C = \text{diag}(c_1, \ldots, c_p) \qquad c_i = \cos(\theta_i)$$

$$S = \text{diag}(s_1, \ldots, s_p) \qquad s_i = \sin(\theta_i)$$

and $0 \leq \theta_1 \leq \cdots \leq \theta_p \leq \pi/2$.

For notational convenience we will sometimes express the above decomposition in the form

$$\bar{U}^T Q \bar{V} = \begin{bmatrix} C_0 & S_0^T \\ -S_0 & C \end{bmatrix} \begin{matrix} k \\ p \end{matrix}$$
$$\qquad\qquad k \quad p$$

where $\bar{U} = \text{diag}(U_1, U_2)$, $\bar{V} = \text{diag}(V_1, V_2)$, $C_0 = \text{diag}(I_{k-p}, C)$, and

$$S_0 = [\ 0 \quad S\ ]\ p \quad .$$
$$\qquad k-p \quad p$$

Note that $U_i^T Q_{ij} V_j$ displays the singular values of Q_{ij} . The quantities $c_1, \ldots, c_p$ will be referred to as the p-singular values of Q and the entire decomposition as the p-SVD . The p-singular values of Q are thus the singular values of Q's trailing pxp principle submatrix. 'The assumption $p \leq k$ is not restrictive.

The aim of this paper is to demonstratethat the p-SVD can play a useful role in the analysis of certain matrix computation problems. This is not a new endeavor; Davis and Kahan [2] made use of the p-SVD in their detailed paper about invariant subspace perturbation. Although this paper precedes Stewart [8], it was in the latter article that Theorem 1 was first made explicit.

We briefly indicate how Theorem 1 can be proved. For clarity, assume $p = k = 3$. Let $U_1^T Q_{11} V_1 = \text{diag}(c_1, c_2, c_3)$ be the SVD of Q_{11} . Since $\|Q_{11}\|_2 \leq \|Q\|_2 = 1$, it follows that $c1 \leq 1$.

Let U_2 be an orthogonal matrix such that the first column of $U_2^T(Q_{21}V_1)$ is a non-positive multiple of e_1, the first column of the 3X3 identity. Similarly, let V_2 be orthogonal so that the first row of $(U_1^T Q_{12})V_2$ is a non-negative multiple of e_1^T. It then follows that

$$\text{diag}(U_1^T, U_2^T)\ Q\ \text{diag}(V_1, V_2) = \left[\begin{array}{ccc:ccc} c_1 & 0 & 0 & a & 0 & 0 \\ 0 & c_2 & 0 & b & x & x \\ 0 & 0 & c_3 & d & x & x \\ \hdashline r & u & v & f & g & h \\ 0 & x & x & k & x & x \\ 0 & x & x & j & x & x \end{array}\right]$$

where $a \geq 0$, $r \leq 0$ and "x" denotes an arbitrary scalar.

Since this transformed matrix is orthogonal, both row 1 and column 1 have unit 2-norm and thus, if $s_1 = \sqrt{1-c_1^2}$ then $a = s_1$ and $r = -s_1$. This implies that $f = c_1$ because columns 1 and 4 must have a zero inner product. It then follows from the unit length of row 4 and column 4 that u,v,g,h,b,d,k, and j are all zero. This leaves us with 2 X 2 blocks -- Q.E. D. by induction.

2. The p-SVD, Direct Rotations, and Angles between Subspaces.

In this section we relate the p-SVD to certain well known relationships that exist between subspaces. As we mentioned, Davis and Kahan [2] used p-SVD ideas in their study of the invariant subspace perturbations.

In their analysis of this problem, it is necessary to be able to rotate a given p-dimensional subspace A into another p-dimensional subspace B in the most "economical" fashion. More precisely, if

$$Z = [\underset{n-p}{Z_1} \mid \underset{p}{Z_2}] \quad W = [\underset{n-p}{W_1} \mid \underset{p}{W_2}]$$

are $n\times n$ orthogonal matrices with $A = \text{Range}(Z_2)$ and $B = \text{Range}(W_2)$, then we wish to determine an orthogonal $T_{min} \in \mathbb{R}^{n\times n}$ that minimizes $\|T - I_n\|_F$ subject to the constraint $T Z_2 = W_2$. (Here, $\|C\|_F^2 = \text{trace}(C^TC)$, a unitarily invaniant norm.)

It is clear that any orthogonal $T \in \mathbb{R}^{n\times n}$ satisfying $T Z_2 = W_2$ must have the form

$$T = [W_1V_1 \mid W_2][Z_1U_1 \mid Z_2]^T \equiv \hat{W}\hat{Z}^T$$

where U_1 and V_1 are orthogonal matrices in $\mathbb{R}^{(n-p)\times(n-p)}$. From the identity

$$\|Z_1^T W_2\|_F^2 + \|Z_2^T W_1\|_F^2 = \|Z_2Z_2^T - W_2W_2^T\|_F^2$$

it follows that

$$\|T - I_n\|_F^2 = \|\hat{Z}^T T \hat{Z} - I_n\|_F^2 = \|\hat{Z}^T\hat{W} - I_n\|_F^2$$

$$= \|U_1^T(Z_1^T W_1)V_1 - I_{n-p}\|_F^2 + \|U_1^T(Z_1^T W_2)\|_F^2$$

$$+ \|(Z_2^T W_1)V_1\|_F^2 + \|Z_2^T W_2 - I_p\|_F^2$$

$$= \|U_1^T(Z_1^T W_1)V_1 - I_{n-p}\|_F^2 + \|Z_2Z_2^T - W_2W_2^T\|_F^2 + \|Z_2^TW_2 - I_p\|_F^2$$

This expression is minimized by choosing U_1 and V_1 so $U_1^T(Z_1^T w_1)V_1$ is diagonal. (See [7]). Moreover, if

$$\begin{bmatrix} U_1 & 0 \\ 0 & U_2 \end{bmatrix}^T \underbrace{\begin{bmatrix} Z_1^T W_1 & Z_1^T W_2 \\ Z_2^T W_1 & Z_2^T W_2 \end{bmatrix}}_{Z^T W} \begin{bmatrix} V_1 & 0 \\ 0 & V_2 \end{bmatrix} = \begin{bmatrix} C_0 & S_0^T \\ -S_0 & C \end{bmatrix}$$

is the p-SVD of $Q = Z^T W$, then

$$\begin{aligned} T_{min} &= \hat{W} \ \hat{Z}^T = \hat{Z}(\hat{Z}^T\hat{W})\hat{Z}^T \\ &= Z(Z^T W) \operatorname{diag}(V_1 U_1^T, I_p) \ Z^T \\ &= [Z_1 U_1 | Z_2 U_2] \left[\begin{array}{cc|} C_0 & S_0^T \\ -S_0 & C \end{array} \right] \begin{bmatrix} U_1^T Z_1^T \\ V_2^T Z_2^T \end{bmatrix} \end{aligned}$$

and

$$\|T_{min} - I_n\|_F^2 = \|C - I_p\|_F^2 + 2\|S\|_F^2 + \|C(V_2^T U_2) - I_p\|_F^2 \quad .$$

The p-singular values $\{\cos(\theta_i)\}_{i=1}^p$ of $Z^T w$ provide a measure of how different the subspaces A and B are. The θ_i are referred to as the "principle angles" between A and B and a stable, efficient algorithm for their computation is given in a paper by Bjork and Golub [1]. Wedin [9] has developed a perturbation theory for the principle angles. T_{min} is referred to in [2] as a "direct rotation" from A to B .

3. A Wielandt-Hoffman Theorem for p-Singular Values.

If an orthogonal matrix Q is perturbed, how are its p-singular values effected? The following theorem answers this question.

Theorem 2.

If Q and $\hat{Q}$ are $m \times m$ orthogonal matrices having p-singular values $\{\cos(\theta_i)\}_{i=1}^{p}$ and $\{\cos(\hat{\theta}_i)\}_{i=1}^{p}$ respectively, then

$$4\sum_{i=1}^{p}[1 - \cos(\theta_i - \hat{\theta}_i)] = 8\sum_{i=1}^{p} \sin^2\left(\frac{\theta_i - \hat{\theta}_i}{2}\right) < \|Q - \hat{Q}\|_F^2 .$$

Proof.

If the p-SVD's of Q and $\hat{Q}$ are given by

$$\begin{bmatrix} U_1 & 0 \\ 0 & U_2 \end{bmatrix}^T \begin{bmatrix} Q_{11} & Q_{12} \\ Q_{21} & Q_{22} \end{bmatrix} \begin{bmatrix} V_1 & 0 \\ 0 & V_2 \end{bmatrix} = \begin{bmatrix} C_0 & S_0^T \\ -S_0 & C \end{bmatrix} \begin{matrix} k \\ p \end{matrix}$$
$$\hphantom{=} \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad k \quad p$$

and

$$\begin{bmatrix} \hat{U}_1 & 0 \\ 0 & \hat{U}_2 \end{bmatrix}^T \begin{bmatrix} \hat{Q}_{11} & \hat{Q}_{12} \\ \hat{Q}_{21} & \hat{Q}_{22} \end{bmatrix} \begin{bmatrix} \hat{V}_1 & 0 \\ 0 & \hat{V}_2 \end{bmatrix} = \begin{bmatrix} \hat{C}_0 & \hat{S}_0^T \\ -\hat{S}_0 & \hat{C} \end{bmatrix} \begin{matrix} k \\ p \end{matrix}$$
$$\hphantom{=} \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad k \quad p$$

respectively, then

$$\begin{aligned} \|Q - \hat{Q}\|_F^2 &= \|U_1 C_0 V_1^T - \hat{U}_1 \hat{C}_0 \hat{V}_1^T\|_F^2 \\ &+ \|U_1 S_0^T V_2^T - \hat{U}_1 \hat{S}_0^T \hat{V}_2^T\|_F^2 \\ &+ \|U_2 S_0 V_1^T - \hat{U}_2 \hat{S}_0 \hat{V}_1^T\|_F^2 \\ &+ \|U_2 C V_2^T - \hat{U}_2 \hat{C} \hat{V}_2^T\|_F^2 \quad . \end{aligned}$$

Now the Wielandt-Hoffman for singular values states that if the n×n matrices R and $\hat{R}$ have SVD's $U\ \mathrm{diag}(\sigma_i)V^T$ and $\hat{U}\ \mathrm{diag}(\hat{\sigma}_i)\hat{V}^T$ respectively, then

$$\sum_{i=1}^{p} (\sigma_i - \hat{\sigma}_i)^2 \leq \|R - \hat{R}\|_F^2$$

This result follows by applying the "original" Wielandt-Hoffman Theorem [5] for eigenvalues to the symmetric matrices $\begin{bmatrix} 0 & R^T \\ R & 0 \end{bmatrix}$ and $\begin{bmatrix} 0 & \hat{R}^T \\ \hat{R} & 0 \end{bmatrix}_I$. (These matrices have eigenvalues $\pm\sigma_i$ and $\pm\hat{\sigma}_i$ respectively.) Thus, if $c_i = \cos(\theta_i)$, $s_i = \sin(\theta_i)$, $\hat{c}_i = \cos(\hat{\theta}_i)$ and $\hat{s}_i = \sin(\hat{\theta}_i)$, then

$$\|Q - \hat{Q}\|_F^2 \geq 2\sum_{i=1}^{p} (c_i - \hat{c}_i)^2 + (s_i - \hat{s}_i)^2$$

$$= 4\sum_{i=1}^{p} [1 - \cos(\theta_i - \hat{\theta}_i)]$$

$$= 8\sum_{i=1}^{p} \sin^2\left(\frac{\theta_i - \hat{\theta}_i}{2}\right) \quad . \quad \square$$

In the next section it will be necessary to know how far a given m x n orthogonal matrix Q is to the set Ω_p^m defined by

$$\Omega_p^m = \left\{ Z = \begin{bmatrix} Z_{11} & Z_{12} \\ Z_{21} & Z_{22} \end{bmatrix} \begin{matrix} m-p \\ p \end{matrix} \;\middle|\; Z^T Z = I_m \ , \ \det(Z_{22}) = 0 \right\},$$

$$\begin{matrix} m-p & p \end{matrix}$$

i.e., the set of all mxm orthogonal matrices whose trailing pxp principle submatrix is singular.

Theorem 3.

If Q is an mxm orthogonal matrix with p-SVD given by Theorem 1, and if $\hat{Q}$ is defined by

$$\hat{Q} = \left[\begin{array}{c|c} U_1 & 0 \\ \hline 0 & U_2 \end{array}\right] \left[\begin{array}{cc|c} I_{k-p} & 0 & 0 \\ 0 & \hat{C} & \hat{S} \\ \hline 0 & -\hat{S} & \hat{C} \end{array}\right] \left[\begin{array}{c|c} V_1 & 0 \\ \hline 0 & V_2 \end{array}\right]^T$$

with

$$\hat{C} = \text{diag}(c_1,\ldots,c_{p-1},0)$$

$$\hat{S} = \text{diag}(s_1,\ldots,s_{p-1},1)$$

then

$$\|Q - \hat{Q}\|_F = \min_{Z \in \Omega_p^m} \|Q - Z\|_F = 2\sqrt{1 - \sin(\theta_p)} \leq 2 \cos(\theta_p)$$

Proof.

Any $Z \in \Omega_p^m$ has p-singular values of the form $\{\cos(\hat{\theta}_1),\ldots,\cos(\hat{\theta}_{p-1}), 0\}$ and so from Theorem 2,

$$\|Z - Q\| \geq 8 \sum_{i=1}^{p-1} \sin^2\left(\frac{\hat{\theta}_i - \theta_i}{2}\right) + 8 \sin^2\left(\frac{\theta_p - \pi/2}{2}\right) \geq 4(1 - \sin(\theta_p)) \quad .$$

By setting $Z = \hat{Q}$, the lower bound is attained. The rest of the Theorem follows from elementary trigonometry. □

4. Some Applications.

We now apply the p-SVD to several computational problems.

(a) Total Least Squares [4].

Given $A \in \mathbb{R}^{m\times n}$, $B \in \mathbb{R}^{m\times p}$ $(m \geq n+p)$ and nonsingular "weighting matrices" $D = \text{diag}(d_1,\ldots,d_m)$ and $T = \text{diag}(t_1,\ldots,t_{n+p})$, the total least squares problem (TLS) involves minimizing

$$\|D\,[E\,|R]T\,\|_F \qquad E \in \mathbb{R}^{m\times n},\quad R \in \mathbb{R}^{m\times p}$$

subject to the constraint

$$\text{Range}\ (B + R) \subset \text{Range}\ (A + E)\ .$$

If a minimizing $\hat{E}$ and $\hat{R}$ can be found, then any $X \in \mathbb{R}^{n\times p}$ satisfying

$$(A + \hat{E})X = B + \hat{R}$$

is a TLS solution. Note that this last equation implies

$$\{D[A|B]T + D[\hat{E}|\hat{R}]T\}T^{-1}\begin{bmatrix} X \\ -I_p \end{bmatrix} = 0$$

and thus, the TLS problem involves finding the nearest matrix to $D[A|B]T$ that has a null space of dimension p . If

$$[\underset{m-p}{U_1}|\underset{p}{U_2}]^T\, D[A|B]T\, [\underset{n}{Q_1}|\underset{p}{Q_2}] = \text{diag}(\sigma_1,\ldots,\sigma_{n+p})$$

is the SVD of $D[A|B]T$, then

$$D[\hat{E}|\hat{R}]T = - U_2 \operatorname{diag}(\sigma_{n+1},\ldots,\sigma_{n+p})Q_2^T .$$

The minimizing $[\hat{E}|\hat{R}]$ is unique if a, > σ_{n+1}. Moreover, if

$$Q = [Q_1|Q_2] = \begin{bmatrix} Q_{11} & Q_{12} \\ Q_{21} & Q_{22} \end{bmatrix} \begin{matrix} n \\ p \end{matrix}$$

$$\qquad\qquad\qquad\qquad n \quad p$$

then

$$T^{-1}\begin{bmatrix} X \\ -I_p \end{bmatrix} = T^{-1}\begin{bmatrix} Q_{12} \\ Q_{22} \end{bmatrix} Q_{22}^{-1} \operatorname{diag}(t_{n+1}^{-1},\ldots,t_{n+p}^{-1})$$

provided Q_{22} is nonsingular. In this case

$$X_{TLS} = \operatorname{diag}(t_1,\ldots,t_n)\ Q_{12}Q_{22}^{-1}\ \operatorname{diag}(t_{n+1}^{-1},\ldots,t_{n+p}^{-1}) .$$

Numerical difficulties arise in the TLS problem if Q_{22} is close to singularity. Consequently we are interested in how close the TLS problem $\{A,B,D,T\}$ is to a corresponding problem $\{\hat{A},\hat{B},D,T\}$ with no solution.

<u>Theorem 4</u>.

Let A,B,D, and T be as above and suppose $F = D[A|B]T$ has SVD $U^TFQ = \operatorname{diag}(\sigma_1,\ldots,\sigma_{n+p})$ with $\sigma_n > \sigma_{n+1}$. If $\{\cos(\theta_i)\}_{i=1}^p$ are the p-singular values of Q , then there exists a TLS problem $\{\hat{A},\hat{B},D,T\}$ with no solution satisfying

$$\frac{\|D[\hat{A}|\hat{B}]T - D[A|B]T\|_F}{\|D[A|B]T\|_F} \leq 2\cos(\theta_p) .$$

Proof.

Let $\hat{Q}$ be the matrix in Ω_p^{n+p} closest to Q as in Theorem 3 . Now

$$U^T FQ = U^T(F Q\hat{Q}^T)\hat{Q} = \mathrm{diag}(\sigma_1,\ldots,\sigma_{n+p})$$

and so by defining $[\hat{A}|\hat{B}]$ from

$$D[\hat{A}|\hat{B}]T = FQ\hat{Q}^T = D[A|B]T\, Q\hat{Q}^T$$

we see that the TLS problem $(\hat{A},\hat{B},D,T)$ has no solution and

$$\|D[\hat{A}|\hat{B}]T - D[A|B]T\|_F \le \|D[A|B]T\|_F \, \|Q-\hat{Q}\|_F \quad .$$

The Theorem follows since $\|Q-\hat{Q}\|_F = \le 2\cos(\theta_p)$ ❑

(b) Golub-Klema-Stewart Subset Selection [3].

Consider the problem

$$\min\|Ax - b\|_2 \quad A \in \mathbb{R}^{m\times n} \quad , \quad b \in \mathbb{R}^m$$

where A has SVD

$$\underset{r \quad n-r}{[U_1|U_2]^T} \, A \, \underset{r \quad m-r}{[Q_1,Q_2]} = \mathrm{diag}(\sigma_1,\ldots,\sigma_n)$$

and $\sigma_r >> \sigma_{r+1} \approx 0$. This implies that A is close to a rank r matrix. One way of "coping" with the ill-conditioning is to minimize $\|A_r x - b\|_2$

where $A_r = U_1 \operatorname{diag}(\sigma_1, \ldots, \sigma_r) Q_1^T$.

This least squares problem has the solution

$$X_r = \sum_{i=1}^{r} \frac{u_i^T b}{\sigma_i} q_i$$

where q_i and u_i denote the i-th columns of Q and U respectively. A shortcoming of this approach is that the predictor Ax_r of b involves all n columns of A . Since rank degeneracy implies redundancy in the underlying linear model, it may be desirable to approximate b with r suitably chosen columns of A .

A method for doing this is suggested in [3]. Suppose $P \in \mathbb{R}^{n \times n}$ is a permutation matrix and that $\hat{y} \in \mathbb{R}^r$ minimizes $\|B_1 y - b\|_2$ where

$$AP = [\underset{r}{B_1} | \underset{n-r}{B_2}] .$$

If

$$P^T Q = \begin{bmatrix} \tilde{Q}_{11} & \tilde{Q}_{12} \\ \tilde{Q}_{21} & \tilde{Q}_{22} \end{bmatrix} \begin{matrix} r \\ n-r \end{matrix}$$

and

$$\hat{x} = P \begin{bmatrix} y \\ 0 \end{bmatrix} \begin{matrix} r \\ n-r \end{matrix}$$

then it can be shown that

$$\|r_{\hat{X}} - r_{x_r}\|_2 \leq \frac{\sigma_{r+1}}{\sigma_r} \|\tilde{Q}_{11}^{-1}\|_2 \|b\|_2 .$$

Here, the notation r_z means $r_z = b - Az$, the residual of z .
Since

$$\|r_{\hat{x}} - r_{x_r}\|_2 = \|A\hat{x} - U_1U_1^T b\|_2 ,$$

it can be argued that $\hat{x}$, vis-a-vis P , should be chosen to make this quantity as small as possible because $U_1U_1^Tb$ represents the "stable" component of b . In [3] this task is approximately accomplished by chosing P so that the resulting $\tilde{Q}_{11}$ is well-conditioned. This is done by applying the Q-R with column pivoting algorithm to Q_1 :

$$Q_1^TP^T = Z[R_1|R_2] \quad Z^TZ = I_r , \quad R_1 = \nabla$$
$$\qquad\quad r \quad n-r$$

In many applications, however, $n - r << r$. Since the p-SVD implies $\|\tilde{Q}_{11}^{-1}\|_2 = \|\tilde{Q}_{22}^{-1}\|_2$, we can essentially determine P by triangularizing the "skinny" matrix Q_2 thus saving work . Note, that if $r = n-1$, then P should merely interchange rows k and n of Q where $|q_{kn}| = \max_{1 \le i \le n} |q_{in}|$.

(c) The Algebraic Riccati Equation.

Suppose $A,B,C \in \mathbb{R}^{n\times n}$ are such that $A^T = A \geq 0$, $C^T = C \geq 0$. Well-known conditions of stabilizability and detectability [10] guarantee that if

$$M = \begin{bmatrix} B & A \\ C & -B^T \end{bmatrix}$$

then there exists $T,Y,Z \in \mathbb{R}^{n\times n}$ such that

$$M \begin{bmatrix} Y \\ Z \end{bmatrix} = \begin{bmatrix} Y \\ Z \end{bmatrix} T$$

where Z is nonsingular and T's eigenvalues are in the right-half-plane. Furthermore, it can be shown that $X = YZ^{-1}$ is the unique, non-negative definite, symmetric solution to the algebraic Riccati equation

$$A + BX + XB^T - xcx = 0 \ .$$

The matrix M is said to have Hamiltonian structure and in [6] the following decomposition is proved:

$$(1) \qquad \begin{bmatrix} Q_{11} & -Q_{21} \\ Q_{21} & Q_{11} \end{bmatrix}^T \begin{bmatrix} B & A \\ C & -B^T \end{bmatrix} \underbrace{\begin{bmatrix} Q_{11} & -Q_{21} \\ Q_{21} & Q_{11} \end{bmatrix}}_{Q} = \begin{bmatrix} T & R \\ 0 & -T^T \end{bmatrix}$$

where T is upper quasi-triangular, R is symmetric and Q orthogonal. If T has its eigenvalues in the right-half-plane, then $X = Q_{11}Q_{21}^{-1}$ solves the Riccati equation.

The transformation Q is said to have symplectic form. Orthogonal symplectic matrices 'preserve Hamiltonian structure and moreover, their p-SVD is of very special form:

<u>Theorem 5.</u>

If $$Q = \begin{bmatrix} Q_{11} & -Q_{21} \\ Q_{21} & Q_{11} \end{bmatrix} \begin{matrix} n \\ n \end{matrix}$$
$$\begin{matrix} n & n \end{matrix}$$

is orthogonal, then there exist n×n orthogonal matrices U and V such that

$$\begin{bmatrix} U & 0 \\ 0 & U \end{bmatrix}^T \begin{bmatrix} Q_{11} & -Q_{21} \\ Q_{21} & Q_{11} \end{bmatrix} \begin{bmatrix} V & 0 \\ 0 & V \end{bmatrix} = \begin{bmatrix} \Sigma & \Delta \\ \Delta & \Sigma \end{bmatrix}$$

where

$$\Sigma = \mathrm{diag}(\sigma_1,\ldots,\sigma_n) \qquad 1 \geq \sigma_1 \geq \cdots \geq \sigma_n \geq 0$$

$$\Delta = \mathrm{diag}(\delta_1,\ldots,\delta_n) \qquad \Delta^2 + \Sigma^2 = I_n .$$

The proof is given in [5]. Note that Δ may have negative diagonal entries and that if $\delta_1 \neq 0$ then

$$X = Q_{11}Q_{21}^{-1} = U \ \mathrm{diag}(\sigma_i/\delta_i) \ U^T$$

(In the Riccati application, the δ_i are positive.)

In practice it is important to understand the significance of small δ_i since the accuracy of a computed X depends on the size of δ_1. In [6] this topic is pursued. Roughly speaking, it can be shown that perturbations in A,B, and C of order δ_1 can result in a Riccati equation $\hat{A} + \hat{B}X + X\hat{B}^T - X\hat{C}X = 0$ that has no symmetric positive definite solution.

Acknowledgement.

The author wishes to thank Professor Gene H. Golub for encouraging him to prepare this report.

REFERENCES.

[1] Bjork, A. and G. H. Golub, 'Numerical methods for computing Angles between Linear Subspaces", Math. Comp. 27 (1973), pp. 579-594.

[2] Davis, C. and W. M. Kahan, "The Rotation of Eigenvectors by a Perturbation. III", SIAM J. Numer. Anal. 7 (1970), pp. 1-46.

[3] Golub, G. H., V. Klema and G. W. Stewart, "Rank Degeneracy and Least Squares Problems", Dept. of Computer Science Technical Report 559, Stanford University, Stanford, CA. (1976).

[4] Golub, G. H. and C. Van Loan, "The Total Least Squares Problem", unpublished manuscript.

[5] Hoffman, A. J. and H. W. Wielandt, "The Variation of the Spectrum of a Normal Matrix", Duke Math. J. 20, (1953),pp. 37-39.

[6] Paige, C. and C. Van Loan, "A Hamiltonian-Schur Decomposition", Dept. of Computer Science Technical Report TR 79-377, Cornell University, Ithaca, New York, 1979.

[7] Schönemann, P., "A Generalized Solution of the Orthogonal Procrustes Problem", Psychometrika 31 (1966), 1-10.

[8] Stewart, G. W., "On the Perturbation of Pseudo-Inverses, Projections and Linear Least Squares Problems", SIAM Review 19 (1977), pp. 634-662.

[9] Wedin, Per Ake, 'A Geometrical Approach to Angles in Finite Dimensional Inner Product Spaces; the Triangular Inequality for Angles", Institute of Information Processing Rept. UMINF:66.78, University of Umear, Sweden (1979).

[10] Wonham, W. M., 'On a Matrix Riccati Equation of Stochastic Control", SIAM J. Control, 6 (1968), 681-697.