Analysis of Some Matrix Problems Using the CS Decomposition

Charles Van Loan

TR 84-603
March 1984

Department of Computer Science
Cornell University
Ithaca, New York 14853

This research was partially supported by the Office of Naval Research contract N00014-83-K-0640.

Analysis of Some Matrix Problems Using the CS Decomposition

Charles Van Loan

TR 84-603
March 1984

Department of Computer Science
Cornell University
Ithaca, New York 14853

This research was partially supported by the Office of Naval Research contract N00014-83-K-0640.

ANALYSIS OF SOME MATRIX PROBLEMS USING THE CS DECOMPOSITION

Charles Van Loan
Department of Computer Science
Cornell University
Ithaca, New York 14853

ABSTRACT

The gist of the CS decomposition is that the blocks of a partitioned orthogonal matrix have related singular value decompositions. In this paper we develop a perturbation theory for the CS decomposition and use it to analyze (a) the total least squares problem, (b) the Golub-Klema-Stewart subset selection algorithm, (c) the algebraic Riccati equation, and (d) the generalized singular value decomposition.

This research was partially supported by the Office of Naval Research contract N00014-83-K-0640.

1. Introduction

In a survey article concerned with perturbation theory, Stewart [13] presented the following decomposition:

Theorem 1.1 (*CS Decomposition*)

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

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

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

$$\begin{bmatrix} U_1 & 0 \\ 0 & U_2 \end{bmatrix}^T \begin{bmatrix} Q_{11} & Q_{12} \\ Q_{21} & Q_{22} \end{bmatrix} \begin{bmatrix} V_1 & 0 \\ 0 & V_2 \end{bmatrix} = \begin{bmatrix} C_0 & S \\ -S & C \end{bmatrix}$$

where

$$\begin{aligned} S &= diag(s_1, \ldots, s_p) & s_i &= sin(\theta_i) \\ C &= diag(c_1, \ldots, c_p) & c_i &= cos(\theta_i) \\ C_0 &= diag(c_1, \ldots, c_p, \underbrace{1,\ldots,1}_{k-p}) \end{aligned}$$

and $\frac{\pi}{2} \geq \theta_1 \geq \cdots \geq \theta_p \geq 0$. Note that $U_i^T Q_{ij} V_j$ displays the singular values of Q_{ij}. We refer to the quantities $c_1,..,c_p$ as the p-singular values of Q. Thus, the p-singular values are the singular values of the $p \times p$ trailing principal submatrix of Q. The assumption $p \leq k$ is made for clarity and is in no way restrictive.

The aim of this paper is to highlight the useful role that the CS decomposition (CSD) plays in the analysis and solution of several important matrix problems. This is not a new endeavor. Davis and Kahan [4] and Stewart [13] make use of the CSD in detailed papers about invariant subspace perturbation.

We briefly suggest how the CSD can be proved. For clarity assume $p = k = 3$. Let $U_1^T Q_{11} V_1 = diag(c_1, c_2, c_3)$ be the singular value decomposition (SVD) of Q_{11} with $0 \leq c_1 \leq c_2 \leq c_3$. Note that $c_3 = ||Q_{11}||_2 \leq ||Q||_2 = 1$. Let U_2 be an orthogonal matrix such that the first column of $U_2^T(Q_{21}V_1)$ is a nonpositive multiple of e_1, the first column of I_3. Similarly, let V_2 be orthogonal so that the first row of $(U_1^T Q_{12})V_2$ is a nonnegative multiple of e_1^T. It follows that

$$diag(U_1^T, U_2^T)\ Q\ diag(V_1, V_2) \quad = \quad \begin{bmatrix} c_1 & 0 & 0 & a & 0 & 0 \\ 0 & c_2 & 0 & b & x & x \\ 0 & 0 & c_3 & d & x & x \\ r & u & v & f & g & h \\ 0 & x & x & k & x & x \\ 0 & x & x & j & x & x \end{bmatrix} \tag{1.1}$$

where $a \geq 0$, $r \geq 0$, and "x" denotes an arbitrary scalar whose exact value does not concern us.

Since this transformed matrix is orthogonal, both row 1 and column 1 have unit 2-norm. Thus, $a = \sqrt{1 - c_1^2} \equiv s_1$ and $r = -\sqrt{1 - c_1^2} = -s_1$. If $s_1 \neq 0$, then by looking at the inner product of column 1 with columns 2 through 6, we conclude that $u = v = g = h = 0$ and $f = c_1$. Likewise, the orthogonality of row 1 with rows 2, 3, 5, and 6 implies $b = d = k = j = 0$. With these observations, we are left with a 2×2 subproblem in (1.1). An obvious induction completes the proof for the case $s_1 \neq 0$. On the other hand, if $s_1 = 0$, then $1 = c_1 \leq c_2 \leq c_3 \leq 1$. Thus, Q_{11} is orthogonal and so $Q_{21} = 0$. It follows that Q_{12} is also zero and thus

$$Q = \begin{bmatrix} Q_{11} & 0 \\ 0 & Q_{22} \end{bmatrix}$$

In this case, the CSD amounts to a pair of decoupled SVD's, one of Q_{11} and the other of Q_{22}.

In following we collect some well-known properties of the CSD, present some perturbation theory, and show how various problems in matrix computations are better understood and better solved through the CSD. This work is an embellishment of a technical report written several years ago by the author [17] and was prompted by current research in the parallel matrix computation area [2,3].

2. The CSD, Direct Rotations, and Angles Between Subspaces

In this section we relate the CSD to certain well-known relationships that exist between subspaces. As we mentioned, David and Kahan [4] use CSD ideas in their study of 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

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

are n-by-n orthogonal matrices with $A = Range(Z_2)$ and $B = Range(W_2)$, then an orthogonal T_{min} is sought that minimizes $\|T - I_n\|_F$ subject to the constraint $TZ = W$. Here, $\|\cdot\|_F$ is the Frobenius norm, i.e., $\|C\|_F^2 = \mathrm{trace}(C^T C)$.

It is clear that any orthogonal $T \in \mathbf{R}^{n\times n}$ that satisfies $TZ = W$ must have the form

$$T = \hat{W}\hat{Z}^T$$

where

$$\hat{W} = [W_1 V_1, W_2] \qquad V_1^T V_1 = I_{n-p}$$
$$\hat{Z} = [Z_1 U_1, Z_2] \qquad U_1^T U_1 = I_{n-p}$$

Since

$$Z^T(Z_2 Z_2^T - W_2 W_2^T)W = \begin{bmatrix} 0 & -Z_1^T W_2 \\ Z_2^T W_1 & 0 \end{bmatrix}$$

we have

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

and so

$$\begin{aligned} \|T - I_n\|_F^2 &= \|\hat{Z}^T T \hat{Z} - I_n\|_F^2 = \|\hat{Z}^T \hat{W} - I_n\|_F^2 \\ &= \|U_1^T(Z_1^T W_1)V_1 - I_{n-p}\|_F^2 + \|U_1^T(Z_1^T W_2)\|_F^2 + \|(Z_2^T W_1)V_1\|_F^2 + \|Z_2^T W_2 - I_p\|_F^2 \\ &= \|U_1^T(Z_1^T W_1)V_1 - I_{n-p}\|_F^2 + \|Z_2 Z_2^T - W_2 W_2^T\|_F^2 + \|Z_2^T W_2 - I_p\|_F^2 . \end{aligned}$$

Recall that our object is to make this quantity as small as possible by judiciously choosing the orthogonal matrices U_1 and V_1. It is not hard to show that this is accomplished if $U_1^T(Z_1^T W_1)V_1$ is diagonal with non-negative entries. Moreover, if

$$\begin{bmatrix} U_1 & 0 \\ 0 & U_2 \end{bmatrix}^T \begin{bmatrix} Z_1^T W_1 & Z_1^T W_2 \\ Z_2^T W_1 & Z_2^T W_2 \end{bmatrix} \begin{bmatrix} V_1 & 0 \\ 0 & V_2 \end{bmatrix} = \begin{bmatrix} C_0 & S \\ -S & C \end{bmatrix}$$

is the CSD of $Q = Z^T W$, then

$$T_{\min} = \hat{W}\hat{Z}^T = [W_1 V_1 , W_2][Z_1 U_1 , Z_2]^T = W \, diag(V_1 U_1^T , I_p) Z^T$$

$$= Z(Z^T W) \begin{bmatrix} V_1 U_1^T Z_1^T \\ Z_2^T \end{bmatrix} = [Z_1 U_1 , Z_2 U_2] \begin{bmatrix} C_0 & S \\ -S & C \end{bmatrix} [Z_1 U_1 , Z_2 V_2]^T$$

and

$$||T_{\min} - I_n||_F^2 = \left|\left| \begin{bmatrix} C_0 & S \\ -S & C \end{bmatrix} - \begin{bmatrix} I_p & 0 \\ 0 & U_2^T V_2 \end{bmatrix} \right|\right|_2^2$$

$$= ||C - I_p||_F^2 + 2||S||_F^2 + ||C - U_2^T V_2||_F^2 \; .$$

The p-singular values $c_i = cos(\theta_i)$ of $Z^T W$ provide a measure of how different the subspaces A and B are. The θ_i are referred to as the *principal angles* between A and B and a stable efficient algorithm for their computation is given in Bjorck and Golub [1]. Wedin [20] has developed a perturbation theory for the principal angles. $T_{\min}$ is referred to as a *direct rotation* from A to B.

3. Some Perturbaton Theorems

If an orthogonal matrix Q is perturbed, how are its p-singular values effected? The following theorem, versions of which are discussed in Paige [9], Sun Ji-Guang [15], and Van Loan [17], answers this question.

Theorem 3.1

Suppose 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. 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

$$8 \sum_{i=1}^p \sin^2\left(\frac{\theta_i - \hat{\theta}_i}{2}\right) = 4 \sum_{i=1}^p [1 - cos(\theta_i - \hat{\theta}_i)] = 2 \sum_{i=1}^p \{(c_i - \hat{c}_i)^2 + (s_i - \hat{s}_i)^2\} \le ||Q - \hat{Q}||_F^2$$

Proof.

$$\hat{S} = diag(1, s_2, \ldots, s_p)$$
$$\hat{C}_0 = diag(0, c_2, \ldots, c_p, 1, \ldots, 1)$$

then

$$\|Q - \hat{Q}\|_F = \min_{Z \epsilon \Omega_p^m} \|Q - Z\|_F = 2\sqrt{1 - s_1} \le 2c_1 \ .$$

Proof.

If $Z \epsilon \Omega_p^m$, then Z has p-singular values $\{\cos(\frac{\pi}{2}), \cos(\psi_2), \ldots, \cos(\psi_p)\}$ and so from Theorem 3.1 we have

$$\|Q - Z\|_F^2 \ge 8 \sin^2((\theta_1 - \pi/2)/2) + 8 \sum_{i=2}^{p} \sin^2((\theta_i - \psi_i)/2) \ge 4(1 - sin(\theta_1))$$

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

If the CSD's of Q and $\hat{Q}$ are given by

$$\begin{bmatrix} U_1 & 0 \\ 0 & U_2 \end{bmatrix}^T \begin{bmatrix} Q_{11} & Q_{12} \\ Q_{21} & Q_{22} \end{bmatrix} \begin{bmatrix} V_1 & 0 \\ 0 & V_2 \end{bmatrix} = \begin{bmatrix} C_0 & S \\ -S & C \end{bmatrix}$$

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

respectively, then

$$\begin{aligned} \|Q - \hat{Q}\|_F^2 = \; & \|U_1 C_0 V_1^T - \hat{U}_1 \hat{C}_0 \hat{V}_1^T\|_F^2 + \|U_1 S V_2^T - \hat{U}_1 \hat{S} \hat{V}_2^T\|_F^2 \\ & + \|U_2 S V_1^T - \hat{U}_2 \hat{S} \hat{V}_1^T\|_F^2 + \|U_2 C V_2^T - \hat{U}_2 \hat{C} \hat{V}_2^T\|_F^2. \end{aligned} \tag{3.1}$$

Now the Wielandt-Hoffman theorem for singular values states that if matrices R and $\hat{R}$ have the same number of rows and columns and have singular values σ_i and $\hat{\sigma}_i$ respectively, then

$$\sum (\sigma_i - \hat{\sigma}_i)^2 \le \|R - \hat{R}\|_F^2$$

where it is assumed that both sets of singular values are ordered from largest to smallest. See Golub and Van Loan [7,p.287] or Wedin [21]. The theorem follows by applying this result to each of the four terms on the right hand side in (3.1) and using some elementary trigonometry □ .

In the next section it is necessary to know how far a given $m \times m$ orthogonal matrix is to the set Ω_p^m defined by

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

i.e., the set of all $m \times m$ orthogonal matrices whose trailing $p \times p$ principal submatrix is singular.

Theorem 3.2

If Q is an $m \times m$ orthogonal matrix with CSD given by Theorem 1.1, and if $\hat{Q}$ is defined by

$$\hat{Q} = \begin{bmatrix} U_1 & 0 \\ 0 & U_2 \end{bmatrix} \begin{bmatrix} \hat{C}_0 & \hat{S} \\ -\hat{S} & \hat{C} \end{bmatrix} \begin{bmatrix} V_1 & 0 \\ 0 & V_2 \end{bmatrix}^T$$

with

$$\hat{C} = diag(0, c_2, \ldots, c_p)$$

4. Some Applications of the CSD

(a) Total Least Squares

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

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

subject to the constraint

$$Range(B + R) \qquad Range(A + E)\ .$$

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

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

is a TLS solution. Suppose

$$U^T F V = diag(\sigma_1, \ldots, \sigma_{n+p})$$

is the SVD of $F = D[A\ ,B]T$ with

$$\sigma_1 \geq \cdots \geq \sigma_n > \sigma_{n+1} \geq \cdots \geq \sigma_{n+p}\ .$$

If

$$V = \begin{bmatrix} V_{11} & V_{12} \\ V_{21} & V_{22} \end{bmatrix} \begin{matrix} n \\ p \end{matrix}$$
$$\qquad\quad n \quad\ p$$

and V_{22} is nonsingular, then it can be shown that the TLS problem has a unique solution given by

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

See Golub and Van Loan [6] for details.

Numerical difficulties arise in the TLS problem if V_{22} is close to singular. Consequently, we are interested in how close the TLS problem $\{A, B, S, T\}$ is to a TLS problem $\{\hat{A}, \hat{B}, S, T\}$ whose V_{22} matrix is singular.

Theorem 4.1

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

$$\frac{||D\ [\hat{A} - A\ ,\ \hat{B} - B]\ T\,||_F}{||\,D\ [A\ ,\ B]\ T\ ||_F} \leq 2\ cos(\theta_1)\ .$$

Proof.

Let $\hat{V}$ be the matrix in Ω_p^{n+p} closest to V. (See Theorem 3.2.) Now

$$U^T(FV\hat{V}^T)\hat{V} = U^T FV = diag(\sigma_1, \ldots, \sigma_{n+p})$$

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

$$D[\hat{A}, \hat{B}]T = D[A,B]TV\hat{V}^T = FV\hat{V}^T$$

we see that the TLS problem $\{\hat{A}, \hat{B}, D, T\}$ has no solution. The theorem follows since

$$||D[\,\hat{A} - A\ ,\ \hat{B} - B\,]\ T\ ||_F \leq ||\,D\ [\,A, B\,]T\ ||_F\ \ ||I - V\hat{V}^T\,||_F$$

and $||I - V\hat{V}^T\ ||_F \leq 2\cos(\theta_1)$ □.

Thus, the smallest p-singular value of V is a measure TLS sensitivity.

(b) Golub-Klema-Stewart Subset Solution

Consider the problem min $||Ax - b\,||_2$ where $b \in \mathbf{R}^m$ and $A \in \mathbf{R}^{m \times n}$ has SVD

$$U^T AV = diag(\sigma_1, \ldots, \sigma_n) \tag{4.1}$$

with $\sigma_1 \geq \ldots \geq \sigma_n$, $U = [u_1, \ldots, u_m]$ and $V = [v_1, \ldots, v_n]$. If $\sigma_r >> \sigma_{r+1} \approx 0$ then A is close to a rank r matrix. One way of "coping" with the ill-conditioning is to solve the nearest rank r LS problem, i.e., min $||A_r x - b\,||_2$ where $A_r = \sum_{i=1}^{r} \sigma_i u_i v_i^T$. This least squares problem has minimum norm solution $x_r = \sum_{i=1}^{r} (u_i^T b / \sigma_i) v_i$. A shortcoming of this approach, however, is that the predictor Ax_r generally 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 independent columns of A. This is the problem of subset selection.

Suppose $P \in \mathbf{R}^{n \times n}$ is a permutation matrix and that $y \in \mathbf{R}^r$ minimizes $||B_1 y - b||_2$ where

$$AP = \underset{\;\;r \quad n-r}{[\, B_1 \, , \, B_2 \,]} \, .$$

If

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

where V is from the SVD (4.1) and

$$\hat{x} = P \begin{bmatrix} y \\ 0 \end{bmatrix} ,$$

then it can be shown that

$$||r_{\hat{x}} - r_y||_2 \le \frac{\sigma_{r+1}}{\sigma_r} ||\tilde{V}_{11}^{-1}||_2 \, ||b||_2 \, . \tag{4.2}$$

See Golub and Van Loan [7,p.414ff]. Here, r_z denotes the residual of z, i.e., $r = b - Az$. Since

$$||r_{\hat{x}} - r_{x_r}||_2 = ||A\hat{x} - U_1 U_1^T b||_2$$

it can be argued that $\hat{x}$, vis-a-vis P, should be chosen to make the upper bound in (4.2) as small as possible since $U_1 U_1^T b$ is that component of b which can be stably approximated by the columns of A.

A heuristic method for doing this is proposed by Golub, Klema, and Stewart [5]. Suppose Q-R with column pivoting is applied to $[V_{11}^T \, , \, V_{21}^T]$:

$$[V_{11}^T \, , \, V_{21}^T] P = Z \, [R_1 , R_2] .$$

Here, $Z \in \mathbf{R}^{r \times r}$ is orthogonal, $R = [R_1 , R_2] \in \mathbf{R}^{p \times r}$ is upper triangular, and P is a permutation. The effect of the permutation is to make $R_1 \in \mathbf{R}^{r \times r}$ well-conditioned. This is desirable since $\tilde{V}_{11} = R_1^T Z$ and so $||\tilde{V}_{11}^{-1}||_2 = ||R_1^{-1}||_2$.

It is typical that $n-r \ll r$. However, from the CSD we know that $||\tilde{V}_{11}^{-1}||_2 = ||\tilde{V}_{22}^{-1}||_2$. Thus, we can determine P much more efficiently from the "skinny"

QR-with-column-pivoting problem

$$[V_{22}^T \, , \, V_{12}^T]\tilde{P} \;=\; \tilde{Z}\,[\underset{n-r}{\tilde{R}_1} \, , \, \underset{r}{\tilde{R}_2}] \quad r$$

Upon completion, we set $P = [\tilde{P}_2 \, , \, \tilde{P}_1]$ where $\tilde{P} = [\tilde{P}_1 \, , \, \tilde{P}_2]$.

Thus, this method of subset selection is made much more efficient through exploitation of the CSD.

(c) The Algebraic Riccati Equation

Suppose $A, B, C \in \mathbf{R}^{n \times n}$ with A and C each symmetric and non-negative definite. Well-known conditions of stabilizability and detectability [23] guarantee that if

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

then there exist $T, Y, Z \in \mathbf{R}^{n \times n}$ 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 precisely those eigenvalues of M that are situated in the open right half-plane. Furthermore, it can be shown that $X = YZ^{-1}$ is the typically sought, unique, non-negative definite, symmetric solution to the algebraic Riccati equation

$$A + BX + XB^T - XCX \;=\; 0 \; . \tag{4.3}$$

The matrix M is said to have Hamiltonian structure and in [11] conditions are given that ensure the existence of an orthogonal

$$Q \;=\; \begin{bmatrix} Q_1 & -Q_1 \\ Q_2 & Q_2 \end{bmatrix} \tag{4.4}$$

such that

$$\begin{bmatrix} Q_1 & -Q_2 \\ Q_2 & Q_1 \end{bmatrix}^T \begin{bmatrix} B & A \\ C & -B^T \end{bmatrix} \begin{bmatrix} Q_1 & -Q_2 \\ Q_2 & Q_1 \end{bmatrix} \;=\; \begin{bmatrix} T & R \\ 0 & -T^T \end{bmatrix} .$$

where T is upper quasi-triangular. Q can be chosen so that T's eigenvalues are in the right half plane. It follows that $X = Q_1 Q_2^{-1}$.

The orthogonal matrix Q in (4.4) is said to have *symplectic* form. Orthogonal symplectic similarity preserves Hamiltonian structure. Moreover, the CSD of a partitioned orthogonal symplectic matrix has a very special form.

Theorem 4.2

If

$$Q = \begin{bmatrix} Q_1 & -Q_2 \\ Q_2 & Q_1 \end{bmatrix} \qquad Q_1, Q_2 \in \mathbf{R}^{n \times n}$$

is orthogonal, then there exist $n \times n$ orthogonal matrices U and V such that

$$\begin{bmatrix} U & 0 \\ O & U \end{bmatrix}^T \begin{bmatrix} Q_1 & -Q_2 \\ Q_2 & Q_1 \end{bmatrix} \begin{bmatrix} V & 0 \\ 0 & V \end{bmatrix} = \begin{bmatrix} \Sigma & -\Delta \\ \Delta & \Sigma \end{bmatrix}$$

where

$$\Sigma = diag(\sigma_1, \ldots, \sigma_n) \qquad \sigma_1 \geq \ldots \geq \sigma_n$$
$$\Delta = diag(\delta_1, \ldots, \delta_n)$$

Proof.

See [8] □.

Note that in this specialized CSD we allow Δ to have negative diagonal entries. If Σ is nonsingular, then the sought after solution to (4.3) is given by

$$X = Q_1 Q_2^{-1} = U diag(\sigma_i / \delta_i) U^T$$

It turns out, that in typical control theory applications, the δ_i are positive. Moreover,it is shown in [11] that perturbations of A,B, and C of order $|\delta_n|$ 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. The reciprocal of δ_n therefore serves to measure the "condition" of the solution X.

(d) The Generalized Singular Value Decomposition

Suppose $A \in \mathbf{R}^{m \times n}$ ($m \geq n$) and $B \in \mathbf{R}^{p \times n}$ are given and that

$$\begin{bmatrix} A \\ B \end{bmatrix} = \begin{bmatrix} Q_{11} \\ Q_{21} \end{bmatrix} R \ . \tag{4.5}$$

is its Q-R decomposition with $Q_{11} \in \mathbf{R}^{m \times n}$, $Q_{21} \in \mathbf{R}^{p \times n}$, and $R \in \mathbf{R}^{n \times n}$. Assume that R is nonsingular, i.e., that the null spaces of A and B intersect trivially. Paige and Saunders [10] have shown that there exists orthogonal matrices $U_1 \in \mathbf{R}^{m \times m}$,$U_2 \in \mathbf{R}^{p \times p}$ and $V \in \mathbf{R}^{n \times n}$ and

$$\begin{aligned} U_1^T Q_1 V &= diag(c_1, \ldots, c_n) \\ U_2^T Q_2 V &= diag(s_1, \ldots, s_q) \quad q = \min\{p,n\} \end{aligned} \tag{4.6}$$

where

$$0 \le c_1 \le \cdots \le c_q \le c_{q+1} = \cdots = c_n$$

and

$$1 \ge s_1 \ge \cdots \ge s_q \ge 0 \ .$$

Combining (4.5) and (4.6) we find that

$$A = U_1 C X^T \qquad B = U_2 S X^T \tag{4.7}$$

where

$$X^T = V^T R \ .$$

We refer to (4.7) as the generalized singular value decomposition of A and B. This factorization has numerous applications, see [7,16,18].

To prove (4.6), Paige and Saunders [10] establish a CSD for nonsquare partitionings of an orthogonal matrix, i.e.,

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

We refer the reader to their paper for more details and mention that our perturbation results in §3 have easily established "nonsquare" versions.

BIBLIOGRAPHY

[1] A. Bjorck and G.H. Golub , *Numerical methods for computing angles between linear subspaces*, Math. Comp. 27 (1973), 579-594.

[2] R.Brent, F. Luk, and C. Van Loan , *Computation of the singular value decomposition using mesh-connected processors*, Cornell Computer Science Technical Report TR 82-528, Ithaca, NY 14853.

[3] R.Brent, F. Luk, and C. Van Loan , *Computation of the generalized singular value decomposition using mesh-connected processors*, Cornell Computer Science Technical Report TR 83-563, Ithaca, NY 14853.

[4] C.Davis and W.M. Kahan, *The rotation of eigenvectors by perturbation III*, SIAM J. Numer.Anal. 7 (1970), 1-46.

[5] G.H.Golub, V. Klema, and G.W. Stewart , *Rank degeneracy and least squares problems*, Dept. of Compute Science Technical Report 559, Stanford University, Stanford , California 94305.

[6] G.H.Golub and C. Van Loan, *An analysis of the total least squares problem* , SIAM J. Numer. Analysis, 17 (1980), 883-893.

[7] G.H.Golub and C. Van Loan, *Matrix Computations* , The Johns Hopkins University Press, Baltimore, Md., 1983.

[8] A.J.Hoffman and H.W. Wielandt , *The variation of the spectrum of a normal matrix* , Duke Math. J. , 20 (1953) , 37-39.

[9] C.C.Paige *A note on a result of Sun-Ji-Guang: sensitivity of the CS and GSV decompositions*, SIAM J. Numer. Anal. 21 (1984), 186-191.

[10] C.C.Paige and M. Saunders, *Towards a generalized singular value decomposition*, SIAM J. Num. Anal., 18 (1981), 398-405.

[11] C.C.Paige and C. Van Loan , *A Schur decomposition for Hamiltonian matrices*, J. Lin. Alg. & Its Applic. , 41 (1981), 11-32.

[12] P.Schonemann *A generalized solution of the orthogonal procrustes problem*, Psychometrica 31 (1966) , 1-10.

[13] G.W.Stewart *On the perturbation of pseudo-inverses, projections, and linear least squares*, SIAM Review 19 (1977), 634-662.

[14] G.W.Stewart *Computing the CS decomposition of a partitioned orthogonal matrix*, Numer. Math. 40 (1982), 297-306.

[15] Sun Ji-Guang, *Perturbation analysis for the generalized singular value decomposition*, SIAM J. Numer. Anal. 20 (1983), 611-625.

[16] C.Van Loan *Generalizing the singular value decomposition* , SIAM J. Numer. Anal. , 13 (1976), 76-83.

[17] C. Van Loan *On Stewart's singular value decomposition for partitioned orthogonal matrices* , Technical Report CS-79-767, Department of Computer Science, Stanford University, 1979.

[18] C. Van Loan A generalized SVD analysis of some weighting methods for equality constrained least squares, in *Matrix Pencils Proceedings , Pite Havsbad, 1982* , B. Kagstrom and A. Ruhe (eds), Lecture Notes in Mathematics #973, Springer-Verlag, New York, 1983.

[19] C.Van Loan *Computing the CS and the generalized singular value decomposition* , Technical Report 84-604 , Department of Computer Science, Cornell University, 1984 .

[20] P.A.Wedin *A geometrical approach to angles in finite dimensional inner product spaces; the triangular inequality for angles,* Institute of Information processing report UMINF: 66.78 , University of Umea, Umea, Sweden.

[21] P.A.Wedin *Perturbation bounds in connection with the singular value decomposition,* BIT, 12 (1972), 99-111.

[22] P.A.Wedin *Perturbation bounds for pseudo-inverses,* BIT, 13 (1973).

[23] W.M.Wonham *On a matrix Riccati equation of stochastic control,* SIAM J. Control, 6 (1968), 681-697 .