History and Generality of the CS Decomposition

C. C. Paige*

School of Computer Science
McGill University
Montreal, Quebec
Canada H3A 2A7

and

M. Wei

Department of Mathematics
East China Normal University
Shanghai 200062
China

Submitted by Peter Rosenthal

To Chandler Davis, for his exceptional scholarship, intellect, and dignity.

ABSTRACT

It is almost a quarter of a century since Chandler Davis and William Kahan brought together the key ideas of what Stewart later completed and defined to be the CS decomposition (CSD) of a partitioned unitary matrix. This paper outlines some germane points in the history of the CSD, pointing out the contributions of Jordan, of Davis and Kahan, and of Stewart, and the relationship of the CSD to the "direct rotation" of Davis and Kato. The paper provides an easy to memorize constructive proof of the CSD, reviews one of its important uses, and suggests a motivation for the CSD which emphasizes how generally useful it is. It shows the relation between the CSD and generalized singular value decompositions, and points out some useful nullity properties one form of the CSD trivially reveals. Finally it shows how via the QR factorization, the CSD can be used to obtain interesting results for partitioned nonsingular matrices. We suggest the CSD be taught in its most general form with no restrictions on the two by two partition, and initially with no mention of angles between subspaces.

*Supported by NSERC of Canada Grant OGP0009236

655 Avenue of the Americas, New York, NY 10010 0024-3795/92/$5.00

1. INTRODUCTION

In 1875 Jordan [14] published a remarkably advanced and thorough analysis of the angles between two subspaces in $\mathcal{R}^n$. Here we discuss this briefly, but concentrate mainly on the period when this very geometric understanding of angles between subspaces developed into the closely related but more general algebraic concept of the CS decomposition (CSD) of a two-block by two-block partitioned unitary matrix.

In 1969 Chandler Davis and William Kahan [5] published a concise overview of their paper [6]. In the overview they emphasized in a finite dimensional setting the essentials of what we now know as the CSD. Today it is slowly being recognized as one of the major tools of matrix analysis. Davis and Kahan's subsequently published paper [6] used the ideas in the context of infinite-dimensional Hilbert spaces, but here we consider only $\mathcal{C}^n$ with vector norm $\|u\| \equiv (u^H u)^{1/2}$ and subordinate matrix norm $\|A\|$.

The simplest useful example of a CSD is a 2×2 real orthogonal matrix partitioned into elements, since it already exhibits the CSD as one of the forms $\begin{bmatrix} c & -s \\ s & c \end{bmatrix}$ or $\begin{bmatrix} c & s \\ s & -c \end{bmatrix}$, where the abbreviations for $\cos\theta$ and $\sin\theta$ give the CSD its name. There is at present no universally accepted form for the CSD in general, but all forms effectively diagonalize the four subblocks of the unitary matrix, and are trivial variants of each other corresponding to permutations and sign changes, so in Section 2 we state a general form which encompasses all possibilities. In Section 3 we show to what extent in [5] and [6] Davis and Kahan derived the CSD, and summarize some of their other contributions to the topic, including emphasizing the very close relation of the direct rotation of Davis [2] (see also Kato [16]) with the CSD in the context of angles between subspaces. In Section 4 we briefly outline some of the history following Davis and Kahan's work, in particular indicating Stewart's key contribution [23], and the relevance of the approach of Paige and Saunders [17]. Section 5 summarizes the superb contribution of Jordan [14] in analyzing the angles between subspaces problem so thoroughly, while Section 6 indicates the historical development of the direct rotation. Section 7 discusses some significant work by statisticians on angles between subspaces. In Section 8 we give a simple proof of the general form of the CSD stated in Section 2, in Section 9 we discuss other variants of the CSD, and in Section 10 we consider how the direct rotation of Davis and Kato may be relevant to the CSD even if we are not initially dealing with angles between subspaces.

We then move from history to illustrating the power and generality of the CSD. In Section 11 we start by giving a simple but convincing algebraic motivation for learning and using the CSD, then discuss some of its uses to give an idea of its generality and importance. Section 12 shows how the CSD provides the generalized singular value decompositions (GSVD) of collections of subblocks of the partitioned unitary matrix. Section 13 shows an apparently new use of the CSD for analyzing relationships between submatrices of a nonsingular matrix and of its inverse. Section 14 shows how the form of the CSD given in Section 8 trivially reveals some useful relationships between the nullities of the subblocks of Q, and how via the technique in Section 13 these relationships are immediately seen to hold between the subblocks of a nonsingular matrix and its inverse. This is a new use of the CSD, showing how it easily provides proofs of some interesting recent results on nullities. We conclude by summarizing our historical findings and discussing what would be the most desirable form for the CSD in general. We suggest that for pedagogical reasons the CSD be presented as a decomposition of subblocks of unitary matrices with no initial mention of the angles between subspaces problem. This encourages use of the simple and fully general proof in Section 8 here. The angles between subspaces problem can then be treated as one very important use of the CSD. We suggest some simple answers to the question as to when we should try to use the CSD. Briefly, this powerful theoretical tool is worth trying whenever it might even remotely be applicable — it brings such clarity and simplicity.

2. THE GENERAL FORM OF THE CSD

The most general form of the CSD is as follows (we give a proof in Section 8). For *any* 2×2 partitioning

$$Q = \begin{array}{c} \begin{array}{cc} c_1 & c_2 \end{array} \\ \left[\begin{array}{cc} Q_{11} & Q_{12} \\ Q_{21} & Q_{22} \end{array} \right] \end{array} \begin{array}{c} \\ r_1 \\ r_2 \end{array}, \quad n = r_1 + r_2 = c_1 + c_2, \tag{1}$$

of a unitary matrix Q ($Q^H Q = QQ^H = I$), there exist unitary U_1, U_2, V_1, V_2 such that (here and elsewhere unnamed blocks are always zero)

$$U^H Q V \equiv \begin{bmatrix} U_1^H & \\ & U_2^H \end{bmatrix} \left[\begin{array}{c|c} Q_{11} & Q_{12} \\ \hline Q_{21} & Q_{22} \end{array} \right] \begin{bmatrix} V_1 & \\ & V_2 \end{bmatrix} = D \equiv \left[\begin{array}{c|c} D_{11} & D_{12} \\ \hline D_{21} & D_{22} \end{array} \right] \tag{2}$$

with each $r_i \times c_j$ $D_{ij} = U_i^H Q_{ij} V_j$ being real and essentially diagonal (that is, each row and each column having at most one nonzero element). Thus the four unitary U_1, U_2, V_1, V_2 essentially give the singular value decompositions (SVD) of all four subblocks

$$Q_{ij} = U_i D_{ij} V_j^H, \quad i,j = 1,2,$$

whereas eight different unitary matrices would be required if Q was a general matrix. This double use of each of U_1, U_2, V_1, V_2 in these "SVDs" is the key to the wide applicability of the CSD. We will discuss precise forms of possible D_{ij} in later sections.

3. THE CONTRIBUTIONS OF DAVIS AND KAHAN

Davis and Kahan [5, 6] were concerned with distances (in terms of angles) between subspaces. For our simplified description let $\mathcal{E}_1$ of dimension r_1 and $\mathcal{F}_1$ of dimension c_1 be any two subspaces of $\mathcal{C}^n$. For simplicity in [5, pp. 864–5] they assumed

$$r_1 \le c_1, \quad r_1 + c_1 \le n, \tag{3}$$

and in our terms effectively showed that for any unitary matrices $E = [E_1, E_2]$, $F = [F_1, F_2]$ such that $\mathcal{R}(E_1) = \mathcal{E}_1$ and $\mathcal{R}(F_1) = \mathcal{F}_1$, where $\mathcal{R}(\cdot)$ denotes range, there exist unitary U_1, U_2 and V_1 so that

$$U^H E^H F_1 V_1 \equiv \begin{array}{c} c_1 \\ \left[\begin{array}{c} U_1^H E_1^H F_1 V_1 \\ \hline U_2^H E_2^H F_1 V_1 \end{array}\right] \end{array} \begin{array}{c} \\ r_1 \\ r_2 \end{array} = \left[\begin{array}{cc} \hat{C} & 0 \\ \hline 0 & I_s \\ \hat{S} & 0 \\ 0 & 0 \end{array}\right] \equiv \hat{D}_1, \tag{4}$$

$$\hat{C} = \mathrm{diag}(\cos\theta_1, \ldots, \cos\theta_{r_1}), \quad \hat{S} = \mathrm{diag}(\sin\theta_1, \ldots, \sin\theta_{r_1}), \quad 0 \le \theta_i \le \pi/2,$$

pointing out of the θ_i that "These angles, 'the angles between the subspaces', are the invariants that characterize the separation between the given subspaces". If in (2) we define $Q \equiv E^H F$ we can see Davis and Kahan derived the structure of D_{11} and D_{21} in the CSD, subject only to the innocuous restriction (3).

Davis and Kahan did not give the structure of D_{12} and D_{22} in (2) in either [5] or [6], but instead showed that if

$$W \equiv EU\hat{D}U^H E^H, \tag{5}$$

$$\hat{D} \equiv \begin{array}{cc} c_1 & c_2 \end{array} \atop [\hat{D}_1, \hat{D}_2] \equiv \left[\begin{array}{c|c} \hat{D}_{11} & \hat{D}_{12} \\ \hline \hat{D}_{21} & \hat{D}_{22} \end{array}\right] \equiv \left[\begin{array}{cc|cc} \hat{C} & & -\hat{S} & \\ \hline & I_s & & \\ \hat{S} & & \hat{C} & \\ & & & I_c \end{array}\right] \begin{array}{c} r_1 \\ \\ r_2 \\ \\ \end{array}, \tag{6}$$

then unitary W is the "direct rotation" of $\mathcal{R}(E_1)$ into $\mathcal{R}(F_1)$. They pointed out in [6, p.18] that the direct rotation was introduced by Davis [2] and Kato [16, First Edition, Ch. 1 §4.6 & §6.8], both influenced by Sz.-Nagy [19, §136], (this book was written in two parts, Sz.-Nagy writing the second part of §§64–155). They also mentioned that "Most of the novelty of the present treatment" (of distances between subspaces) "is in matters concerning the direct rotation". They emphasized the geometric significance of the direct rotation and in [6], for $r_1 = c_1$, proved several important properties, in particular that of all unitary matrices taking $\mathcal{R}(E_1)$ onto $\mathcal{R}(F_1)$, the direct rotation differs least from the identity [6, p.10 & §4]. We will refer to $\hat{D}$ as the "core" of the direct rotation W. To touch briefly on these properties, note that

$$\|W - I\| = \|\hat{D} - I\|, \tag{7}$$

and no unitary transformations restricted to individual subblocks of $\hat{D}$ will decrease the norm of this difference, and this is clearly true in (6) for $r_1 < c_1$ as well as $r_1 = c_1$.

By finding this detailed structure of the direct rotation, Davis and Kahan did everything *but* prove the rest of the CSD, because as is now well known, this $\hat{D}$ does in fact correspond to one form of D in the CSD (2) of $Q \equiv E^H F$. We will show this here for the special case of $r_1 = c_1$, and in Section 9 for the general case. But before this, note from (4) that

$$F_1 V_1 = E U \hat{D}_1,$$

so from the definition of W, and since $r_1 \leq c_1$,

$$\begin{aligned} W E_1 U_1 &= E U \hat{D} U^H E^H E_1 U_1 \\ &= E U \hat{D} \begin{bmatrix} I_{r_1} \\ 0 \end{bmatrix} = E U \hat{D}_1 \begin{bmatrix} I_{r_1} \\ 0 \end{bmatrix} = F_1 V_1 \begin{bmatrix} I_{r_1} \\ 0 \end{bmatrix}, \end{aligned} \tag{8}$$

showing $\mathcal{R}(W E_1) \subseteq \mathcal{R}(F_1)$, as required of the direct rotation. In [6] only the case $r_1 = c_1$ was considered, and then clearly $W E_1 U_1 = F_1 V_1$ and $\mathcal{R}(W E_1) = \mathcal{R}(F_1)$. In this case it is easy for us to show the correspondence between $\hat{D}$ and the CSD of $E^H F$. Since $EU = [E_1 U_1, E_2 U_2]$, $F = [F_1, F_2]$ and WEU are unitary matrices,

$$\mathcal{R}(W E_2 U_2) = \mathcal{R}(W E_1 U_1)^{\perp} = \mathcal{R}(F_1 V_1)^{\perp} = \mathcal{R}(F_2),$$

and we see

$$WEU = [W E_1 U_1, W E_2 U_2] = [F_1 V_1, F_2 V_2] = FV$$

for some unitary V_2, and so $WEU = EU\hat{D}U^H E^H EU = FV$ gives

$$U^H E^H F V = \hat{D}, \tag{9}$$

which from the structure in (6), see (2), is clearly the CSD of $Q \equiv E^H F$ with $r_1 = c_1$.

So although Davis and Kahan did not prove the full CSD, under the mild restriction (3) they did prove half of it (the form of D_{11} and D_{21}) and give the form of the rest of it. We see also that the core of the direct rotation between subspaces gives one important form for the CSD, and Davis and Kahan proved valuable properties of this, properties which apply immediately to this form of the CSD.

4. SUBSEQUENT CONTRIBUTIONS OF STEWART AND OTHERS

The overview [5] was received on February 4, 1969, two months after the main paper [6] (received December 9, 1968), and as it was written later, was not referenced by that main paper. The contributions of the main paper were quickly recognized, see for example [21, 30, 1, 22] (none of which cited [5]), and it became a key paper in matrix analysis. In particular Björck and Golub [1, equation (15)] derived the direct rotation and, referring to [6], emphasized several of its properties and its importance for angles between subspaces and canonical correlations. They also proved that the cosines of the principle angles and the principle vectors associated with these subspaces came from the SVD of (in our notation) $E_1^H F_1$ in (4), see [1, Theorem 1], but unlike [5] did not explicitly point out the SVD of $E_2^H F_1$ in (4), or the other subblocks of $E^H F$. It was left to Stewart [23] to recognize the greater generality of this angles between subspaces tool that Davis and Kahan in [6] and Björck and Golub in [1] had used so effectively. Unaware of [5], Stewart in an appendix to [23] gave a proof of the CSD for the case of $r_1 = c_1 \leq n/2$, pointing out that the result was implicit in [6] and in [1]. This contribution was extremely important, not so much because it appears to be the first complete proof given for a CSD, but because it simply stated the CSD as a general decomposition of unitary matrices, rather than just a tool for analyzing angles between subspaces — this was something [5] had not quite done. This elegant and unequivocal statement brought the CSD to the notice of a wider audience, as well as emphasizing its broader usefulness. Stewart widely advocated the use of the CSD, and came up with this appropriate name. He first used the name in the early 1982 presentation of [25], and it first appeared in print in [24].

Van Loan, see for example [28], was another champion of the CSD, and it was partly as a result of his enthusiasm for it that the CSD became the crucial tool in the formulation of the generalized (now *quotient*) singular value decomposition (QSVD) proposed in [17] (a reformulation and minor generalization of Van Loan's BSVD in [27]). For that paper the authors were aware of [6] and [23], but not of [5], and finding it necessary for their theory of the QSVD they proved the CSD (2) with no restrictions on the dimensions of the subblocks of Q. In retrospect one small contribution of that proof compared with [5] and [23] is that it has no distinct cases to deal with, and so its CSD statement is a little more general and the proof a little more simple. This CSD statement also makes more obvious some previously unnoticed rank relations, see Section 14. The paper gave another important application of the CSD, to the theory of the QSVD of general $r_1 \times m$ A_1 and $m \times r_2$ A_2. Section 12 follows those ideas to reveal how the CSD gives the *generalized* SVDs (GSVD, see [7]) of collections of subblocks of Q in (1).

The CSD was stated in the first (1983, §2.4) and second editions of the high level text by Golub and Van Loan, see [9, §2.6], in the context of distances between subspaces, and distances between orthogonal projectors. A proof of the CSD was given in the monograph by Stewart and Sun [26, §I.5] which used it in the context of distances between subspaces, and showed how it gives useful results on the singular values of products and differences of orthogonal projectors. On page 46 Stewart and Sun gave the direct rotation W and stated the optimality of $\|W - I\|$ (see (7) here), leaving the proofs as exercises. Both these books referenced [6, 23, 17], but not [5]. The ideas of the CSD were used in the undergraduate text [29, §7.5], in the context of angles and distances between subspaces, but the actual CSD was not stated — this book referenced none of the papers, but gave an introduction to the basic ideas.

In all these books only the case $r_1 = c_1$ was considered, even though the original work in [5] did not have this restriction, and [17] showed the CSD with no restrictions at all on the dimensions of subblocks is required in general problems, and gave a brief proof of this general formulation.

5. THE ORIGINAL CONTRIBUTION OF JORDAN

G. W. Stewart informed the authors that Jordan dealt with angles between subspaces in 1875 [14], and [6] also referenced Jordan's paper. Jordan's work was so advanced and complete it is worth summarizing some of the details here. He dealt with $\mathcal{R}^n$ and used the terminology that a linear equation defined a *plane*, k simultaneous (independent and not incompatible) linear equations defined a *k-plane* (which we see is a linear variety or affine subspace of dimension $n-k$), $n-1$ a *line* and n a *point*. For these he used "le nom générique de *multiplans*". In parts IV and V he studied those relations between two "multiplanes" that were independent of the choice of (rectangular) axes, and stated on page 104:

> Our main results can be stated as the following propositions:
>
> - A system made of a k-plane P_k and an l-plane P_l having a common point has r distinct invariants, r being the minimum of k, l, $n-k$, $n-l$. One can consider these invariants as defining the angles of the two multiplanes.
> - The various planes perpendicular to P_k and to P_l form respectively by their intersections an $n-k$ plane P_{n-k} and an $n-l$ plane P_{n-l}, having between them the same angles as P_k and P_l.
> - If P_k and P_l do not have a common point, we will have one invariant more, that is their shortest distance. This invariant is expressed by a fraction whose numerator and denominator are sums of squares of determinants.

It is clear from the first proposition that Jordan obtained and understood angles between subspaces just as we know them today. In fact in [14, p.138, (78)&(79)] for the case $k = l$ he obtained the now well known equations (later essentially given by Hotelling [10]) for $\gamma = \cos\theta$

$$\begin{bmatrix} 0 & AB^T \\ BA^T & 0 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \gamma \begin{bmatrix} AA^T & 0 \\ 0 & BB^T \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix},$$

where $k \times n$ A gave P_k as the null space of A and $k \times n$ B gave P_l as the null space of B. Here we have shifted to spaces and written the equations in matrix form for clarity.

Jordan's second proposition is less clearly stated (or our translation of 19th century scientific French is inadequate), but from the rest of the paper it is clear it means (again using subspaces and modern terminology for clarity): If P_k and P_l are two subspaces of $\mathcal{R}^n$, then the angles between $P_k^\perp$ and $P_l^\perp$ are the same as those between P_k and P_l.

6. HISTORY OF THE DIRECT ROTATION

In his papers [2, 3, 4] Davis was unaware of [14], but did for example cite [20] which discussed only rudimentary ideas of angles between subspaces. It was not until the collaboration with Kahan [6] that they cited Jordan's work. So instead of working directly from Jordan's ideas, it is apparent that Davis developed the direct rotation as an original way of treating the separation of subspaces. Since the direct rotation led to the CSD, we will indicate its history here, as usual restricting ourselves to $\mathcal{C}^n$. In doing so, we will reveal the structure of the predecessors of the direct rotation in the notation of Section 3.

Thus $E = [E_1, E_2]$, $F = [F_1, F_2]$ and $Q = E^H F$ are unitary matrices with $r_i \times c_j$ $Q_{ij} = E_i^H F_j$, and again we assume $r_1 \le c_1$ and $r_1 + c_1 \le n$. For clarity we will also define the projectors $A = E_1 E_1^H$, $\tilde{A} = E_2 E_2^H$, $B = F_1 F_1^H$ and $\tilde{B} = F_2 F_2^H$. We can use the distance between projectors $\|A - B\|$ as a measure of distance between $\mathcal{R}(E_1)$ and $\mathcal{R}(F_1)$, and the results we mention here will require

$$\|A - B\| < 1. \tag{10}$$

Now from (4) we can show $A - B = E_1 E_1^H - F_1 F_1^H$ has $r_2 - c_1$ zero eigenvalues and $c_1 - r_1$ unit eigenvalues, with the rest being $\pm \sin\theta_i$, $i = 1, \ldots, r_1$. Thus to satisfy (10) we need $r_1 = c_1$, meaning $\mathcal{R}(E_1)$ and $\mathcal{R}(F_1)$ have the same dimension, and $|\sin\theta_i| < 1$, so that $\hat{D}_{11}$ in (6) is nonsingular $\hat{C}$, I_s is nonexistent and $\hat{D}_{22}$ is nonsingular.

For the case (10) Sz.-Nagy [19, §136] introduced

$$\tilde{W} \equiv B[I + A(B - A)A]^{-1/2} A \tag{11}$$

to map $\mathcal{R}(E_1)$ onto $\mathcal{R}(F_1)$, with $\tilde{W}^H$ mapping $\mathcal{R}(F_1)$ onto $\mathcal{R}(E_1)$, see below. The term in square brackets is from (4)

$$\tilde{A} + ABA = E_2 E_2^H + E_1 E_1^H F_1 F_1^H E_1 E_1^H = E \operatorname{diag}(U_1 \hat{C}^2 U_1^H, I_{r_2}) E^H,$$

and the nonsingularity of $\hat{C}$ gives

$$\begin{aligned} \tilde{W} &= F_1 F_1^H E \operatorname{diag}(U_1 \hat{C}^{-1} U_1^H, I_{r_2}) E^H E_1 E_1^H \\ &= F_1 F_1^H E_1 U_1 \hat{C}^{-1} U_1^H E_1^H = F_1 V_1 U_1^H E_1^H. \end{aligned}$$

We see $\tilde{W}$ is singular but neither Hermitian nor idempotent, while

$$\tilde{W} E_1 = F_1 V_1 U_1^H, \qquad \tilde{W}^H F_1 = E_1 U_1 V_1^H.$$

Thus $\tilde{W}$ preserves norms and inner products of elements in $\mathcal{R}(E_1)$, and maps any element of $\mathcal{R}(E_2)$ to zero. Clearly $\tilde{W}^H$ treats $\mathcal{R}(F_1)$ and $\mathcal{R}(F_2)$ analogously. Sz.-Nagy called $\tilde{W}$ and $\tilde{W}^H$ "partially isometric".

Davis wanted a *unitary* mapping, and influenced by this work of Sz.-Nagy defined for $\|A-B\|<1$ (see [2, (2.11)–(3.12)])

$$\begin{aligned} H &\equiv ABA+\tilde{A}\tilde{B}\tilde{A}=(AB+\tilde{A}\tilde{B})(BA+\tilde{B}\tilde{A}) \\ &= I-A-B+AB+BA=\tilde{B}\tilde{A}+AB \qquad (12) \\ &= BAB+\tilde{B}\tilde{A}\tilde{B}=(BA+\tilde{B}\tilde{A})(AB+\tilde{A}\tilde{B}), \\ \hat{W} &\equiv H^{-1/2}(BA+\tilde{B}\tilde{A}), \qquad (13) \end{aligned}$$

since in this case he showed H is nonsingular. It followed that $\hat{W}\hat{W}^H = H^{-1/2}HH^{-1/2}=I$, so $\hat{W}$ is unitary, and Davis showed $\hat{W}$ takes $\mathcal{R}(E_1)$ onto $\mathcal{R}(F_1)$ and $\mathcal{R}(E_2)$ onto $\mathcal{R}(F_2)$. Thus for $\mathcal{R}(E_1)$ and $\mathcal{R}(F_1)$, $\hat{W}$ is the unitary version of (11).

While one of the lasting contributions of Davis, and then of Davis and Kahan, was to develop and understand $\hat{W}$ and its importance for the angles between subspaces problem, leading to the direct rotation in (5) and the partial CSD in (4), here we use their results to travel swiftly in reverse and show $\hat{W}$ is a special case of the direct rotation W in (5). This will also reveal the interesting structure of H.

First we note from (12)

$$(E_1^H H E_2)^H = E_2^H H E_1 = 0,$$

and from (4) remembering $\hat{D}_{11}=\hat{C}$

$$E_1^H H E_1 = E_1^H B E_1 = E_1^H F_1 F_1^H E_1 = U_1 \hat{D}_{11}^2 U_1^H.$$

But since $r_1=c_1$ we have shown there exists unitary V_2 giving (9), so

$$E_2^H H E_2 = E_2^H \tilde{B} E_2 = E_2^H F_2 F_2^H E_2 = U_2 \hat{D}_{22}^2 U_2^H,$$

and these results combine to give

$$H = EU\ \mathrm{diag}(\hat{D}_{11},\hat{D}_{22})^2 U^H E^H \equiv EU\mathcal{D}^2U^HE^H, \ \text{ say,} \qquad (14)$$

where H is clearly positive definite giving $H^{-1/2}=EU\mathcal{D}^{-1}U^HE^H$.

To obtain the structure of $\hat{W}$ we see $(BA+\tilde{B}\tilde{A})E_1 = BE_1$, while $(BA+\tilde{B}\tilde{A})E_2=\tilde{B}E_2$, so from (13), (4) and (9)

$$\begin{aligned} \hat{W}E_1 &= EU\mathcal{D}^{-1}U^HE^HF_1F_1^HE_1 = EU\mathcal{D}^{-1}\hat{D}_1\hat{D}_{11}U_1^H = EU\hat{D}_1U_1^H, \\ \hat{W}E_2 &= EU\mathcal{D}^{-1}U^HE^HF_2F_2^HE_2 = EU\mathcal{D}^{-1}\hat{D}_2\hat{D}_{22}U_2^H = EU\hat{D}_2U_2^H, \end{aligned}$$

where we have used the structure of $\mathcal{D}$ and $\hat{D}$ in (14) and (6) with I_s nonexistent. Combining these gives

$$\hat{W}E = EU\hat{D}U^H = WE$$

for W in (5), so $\hat{W}$ is the special case of the direct rotation W when $\|A - B\| < 1$.

Again following Sz.-Nagy [19, §136], Kato [15], see [16, Ch.1 §4.6 & §6.8], introduced the direct rotation (13), but did not develop the ideas significantly.

7. SOME CONTRIBUTIONS BY STATISTICIANS

The angles between subspaces problem is also of great interest to statisticians, and in 1936 Hotelling [10] rederived some of Jordan's results and showed their importance in that area. Around the time Davis and Kahan were writing [5, 6], two statisticians were exploring some similar ideas. In 1970 James and Wilkinson [11] stated:

> The characterizing geometrical properties are summarized by a canonical decomposition theorem for vector spaces, due essentially to Jordan and to Hotelling (1936). In §2 we give a formulation and proof of the theorem in terms of projection and shrinkage operators.

If $\mathcal{E}_1$ and $\mathcal{F}_1$ are *any* two subspaces of $\mathcal{R}^n$, A is the orthogonal projector onto $\mathcal{E}_1$ and B is the orthogonal projector onto $\mathcal{F}_1$, they showed that the nonzero eigenvalues λ_i of ABA and BAB are identical, and $\lambda_i = \cos^2\theta_i$, the θ_i being the angles between the subspaces. If the corresponding (orthonormal sets of) eigenvectors are given by $ABA\tilde{u}_i = \lambda_i\tilde{u}_i$ and $BAB\tilde{v}_i = \lambda_i\tilde{v}_i$ they showed these are biorthogonal sets, that is $\tilde{u}_i^T\tilde{v}_j = 0$, $i \neq j$, and $\tilde{u}_i^T\tilde{v}_i = \cos\theta_i$.

With a little imagination (and abbreviation for simplicity) we can see such geometric results imply part of the CSD, although James and Wilkinson did not formulate any part of the algebraic CSD. Thus if we use the notation of Section 3, we have $A = E_1E_1^H$, $B = F_1F_1^H$, and

$$ABA = E_1Q_{11}Q_{11}^HE_1^H, \qquad BAB = F_1Q_{11}^HQ_{11}F_1^H, \qquad Q_{11} \equiv E_1^HF_1.$$

The nonzero eigenvalues are clearly identical, and equal to the squares of the nonzero singular values of Q_{11}, the $\cos\theta_i$ as stated. If $\tilde{D}$ is the diagonal matrix of these nonzero singular values, and $\tilde{U}_1$ and $\tilde{V}_1$ the matrices of corresponding orthonormal eigenvectors of ABA and BAB, we have

$$E_1 Q_{11} Q_{11}^H E_1^H \tilde{U}_1 = \tilde{U}_1 \tilde{D}^2, \qquad F_1 Q_{11}^H Q_{11} F_1^H \tilde{V}_1 = \tilde{V}_1 \tilde{D}^2.$$

But clearly $\tilde{U}_1$ is orthogonal to E_2 and $\tilde{V}_1$ is orthogonal to F_2, so we may write $\tilde{U}_1 = E_1 U_{11}$ and $\tilde{V}_1 = F_1 V_{11}$ where U_{11} and V_{11} may not be square but $U_{11}^H U_{11} = I$, $V_{11}^H V_{11} = I$. We can now see that their biorthogonality results imply

$$\tilde{U}_1^H \tilde{V}_1 = U_{11}^H E_1^H F_1 V_{11} = U_{11}^H Q_{11} V_{11} = \tilde{D},$$

which is part of the D_{11} block of the CSD. James and Wilkinson also showed if $\tilde{A} \equiv I - A$ then $\tilde{A}B\tilde{A}$ has eigenvalues $\sin^2 \theta_i$. Note that $\tilde{A}B\tilde{A} = E_2 Q_{21} Q_{21}^H E_2^H$, so they also obtained knowledge pertinent to the D_{21} block of the CSD.

8. A PROOF OF THE CSD

Since the D_{ij} are essentially diagonal, and D is real and unitary in (2), its elements are severely restricted. Different statements of the CSD correspond to different allowable matrices D within these restrictions. Perhaps the simplest formulation and proof when there are no restrictions on the partitioning (1) is the following [17].

THEOREM 1. **The CSD.** *For any partitioning of unitary Q as in (1) there exist unitary U_1, U_2, V_1, V_2 such that*

$$U^H Q V = \left[\begin{array}{c|c} U_1^H Q_{11} V_1 & U_1^H Q_{12} V_2 \\ \hline U_2^H Q_{21} V_1 & U_2^H Q_{22} V_2 \end{array}\right] = \begin{array}{cc} \begin{array}{cc} c_1 & c_2 \end{array} & \\ \left[\begin{array}{c|c} D_{11} & D_{12} \\ \hline D_{21} & D_{22} \end{array}\right] & \begin{array}{c} r_1 \\ r_2 \end{array} \end{array}$$

$$= \left[\begin{array}{c|c} D_1 & D_2 \end{array}\right] = D = \left[\begin{array}{ccc|ccc} I & & & O_s^H & & \\ & C & & & S & \\ & & O_c & & & I \\ \hline O_s & & & I & & \\ & S & & & -C & \\ & & I & & & O_c^H \end{array}\right], \qquad (15)$$

$C \equiv diag(\gamma_1, \ldots, \gamma_s), \quad 1 > \gamma_1 \geq \ldots \geq \gamma_s > 0,$

$S \equiv diag(\sigma_1, \ldots, \sigma_s), \quad 0 < \sigma_1 \leq \ldots \leq \sigma_s < 1, \quad C^2 + S^2 = I.$

The O_s and O_c are matrices of zeros, and, depending on Q and the partition, possibly having no rows and or no columns. Some of the unit matrices could be nonexistent, and no two of them need be equal. The four C and S matrices are square with the same dimensions, and could be nonexistent.

Proof. Choose U_1 and V_1 to give the usual SVD of Q_{11}, resulting in D_{11}. Note $\|d_j\| = 1$ for any column (or row) of D, so no singular value of D_{11} can exceed unity. Choose unitary U_2 and V_2 to make $U_2^H Q_{21} V_1$ lower, and $U_1^H Q_{12} V_2$ upper, triangular with real nonnegative elements on their diagonals ending in the bottom right hand corners. The orthonormality of the columns of D shows D_{21} must have the stated form. The orthonormality of rows gives D_{12} except for the dimensions of the zero block denoted O_s^H. Unit length of columns and rows also shows the form of D_{22} in

$$D = \left[\begin{array}{ccc|ccc} I & & & O_{12} & & \\ & C & & & S & \\ & & O_c & & & I \\ \hline O_s & & & K & L & \\ & S & & M & N & \\ & & I & & & O_{22} \end{array}\right].$$

Orthogonality of the second and fourth blocks of columns shows $SM = 0$, and so $M = 0$ since S is nonsingular. Similarly from the second and fourth blocks of rows $L = 0$. Next from the fifth and second blocks of rows $SC + NS = 0$, so $N = -C$. Then we see $KK^H = I$ and $K^H K = I$, so K is unitary and can be transformed to I without altering the rest of D by for example replacing U_2^H by $\mathrm{diag}(K^H, I, I)U_2^H$. Finally the unit matrices in the (1,1) and (4,4) blocks show $O_{12} = O_s^H$, and similarly $O_{22} = O_c^H$. ∎

The form in (15) is reasonably easy to remember on noting the nonzero elements of D_{11} and D_{22} are on their main diagonals, and are nonincreasing in absolute value going down these diagonals. The nonzero elements of D_{21} and D_{12} are nonincreasing going up the diagonals *starting* in the bottom right hand corners. These "diagonal forms" of what are essentially SVDs are about as close as we can get to the usual form of SVDs.

9. VARIANTS OF THE CSD

To obtain other variants of the CSD we note it is possible to permute the first r_1 or last r_2 rows of D in (15), or the first c_1 or last c_2 columns, in any fashion, and to change the sign of any row or column. Beyond that any allowable unitary transformation that alters D would destroy the real and essentially diagonal form of some block of D, and so D is unique except for these variations.

The main variant we will consider apart from (15) is that in which D in (2) is $\hat{D}$, the core of the direct rotation (5). Because of its analogy with a rotation matrix, this is easier to remember than (15), but to obtain this precise form we see we have to assume in (1) and (6) that

$$r_1 \leq c_1 \leq r_2, \tag{16}$$

which can be seen to be equivalent to (3). This can always be satisfied if we are free to reorder and transpose blocks, for suppose r_{max} is the maximum of r_1 and r_2 *etc.*, then we can either obtain (16) by permuting blocks, or

$$c_{min} \leq r_{min} \leq r_{max} \leq c_{max},$$

in which case we transpose Q (exchanging r and c) and then permute blocks if necessary. Then from (15) we can then write $O_c = [\hat{O}_c, \tilde{O}_c]$ and $O_s^H = [\tilde{O}_s^H, \hat{O}_s]$ with $\hat{O}_c$ and $\hat{O}_s$ square, and we can multiply the last two blocks of columns of the rightmost matrix in (15) by $-I$ and rewrite it as

$$\left[\begin{array}{ccc|c||c|ccc}
I & & & & \tilde{O}_s^H & \hat{O}_s & & \\
 & C & & & & & -S & \\
 & & \hat{O}_c & \tilde{O}_c & & & & -I \\
\hline\hline
\tilde{O}_s & & & & I_c & & & \\
\hline
\hat{O}_s & & & & & I & & \\
 & S & & & & & C & \\
 & & I & & & & & \hat{O}_c \\
\hline
 & & & I_s & & & & \tilde{O}_c^H
\end{array}\right] \equiv \left[\begin{array}{c|c||c|c}
\hat{C} & & & -\hat{S} \\
\hline\hline
 & & I_c & \\
\hline
\hat{S} & & & \hat{C} \\
\hline
 & I_s & &
\end{array}\right]$$

where $\hat{C} \equiv \mathrm{diag}(I, C, \hat{O}_c)$ and $\hat{S} \equiv \mathrm{diag}(\hat{O}_s, S, I)$ are square. Permuting rows and columns within the main blocks gives the "direct rotation version"

$$D \equiv \left[\begin{array}{c|c} D_{11} & D_{12} \\ \hline D_{21} & D_{22} \end{array}\right] = \left[\begin{array}{cc|cc} \hat{C} & & -\hat{S} & \\ \hline & I_s & & \\ \hat{S} & & \hat{C} & \\ & & & I_c \end{array}\right] \tag{17}$$

of (2), which is the form of the CSD partly proven by Davis and Kahan [5], and fully proven by Stewart [23] for the case $r_1 = c_1$.

10. THE DIRECT ROTATION AND THE GENERAL CSD

Section 9 showed that one form of the CSD of $Q \equiv E^H F$ does give the core $\hat{D}$ of the direct rotation (5), even when $r_1 < c_1$. The point of $\hat{D}$ here is that, unlike D in (15), it is closest to I among all CSDs of Q. Thus the equivalent to (17) for any partitioning can easily be found from (15) by making allowable changes to (15) till it is as close to I as possible. This means leaving D_{11} alone, moving the elements on the diagonal of D_{22} to be on the diagonal of D, and ensuring the part of the diagonal of D which passes through D_{21} (or D_{12}) becomes I (e.g. the I_s in (17)). It follows that (15) is a satisfactory starting point for finding the corresponding direct rotation version for any partitioning of Q.

It is useful to consider the geometric meaning and relevance of direct rotations W when we are given Q rather than two subspaces. We can arbitrarily factor $Q = E^H F = E^H(EQ)$ using any unitary E, and if the direct rotation version (17) of the CSD of Q with $r_1 \leq c_1$ is

$$U^H Q V = \hat{D},$$

then the corresponding direct rotation is from (5)

$$W_E = EU\hat{D}U^H E^H, \tag{18}$$

where if $E = [E_1, E_2]$, $Q = [Q_1, Q_2]$, $F = EQ$, with $n \times r_1$ E_1 and $n \times c_1$ Q_1, (8) shows

$$\mathcal{R}(W_E E_1) \subseteq \mathcal{R}(EQ_1). \tag{19}$$

Clearly every direct rotation corresponding to Q with this partitioning has the same core $\hat{D}$, which is the right hand side of the direct rotation version of the CSD of Q. Also paraphrasing [5, p.866], *of all unitary W satisfying (19), the one which minimizes each unitary-invariant norm of $(I-W)^H(I-W)$, including in particular $\|I-W\|$ and the Frobenius norm $\|I-W\|_F$, is the direct rotation (18).*

The most relevant choices of E are I and $Q^H \equiv [Q^{(1)}, Q^{(2)}]$, $n \times r_1$ $Q^{(1)}$. We see the direct rotation W_I corresponding to $E = I$ maps the subspace $\mathcal{R}(\begin{bmatrix} I_{r_1} \\ 0 \end{bmatrix})$ into $\mathcal{R}(Q_1)$, while W_{Q^H} maps $\mathcal{R}(Q^{(1)})$ into $\mathcal{R}(\begin{bmatrix} I_{c_1} \\ 0 \end{bmatrix})$, and these are the unitary transformations that differ least from the identity in doing this. Thus the idea of the direct rotation of Davis and Kato is relevant for general unitary Q, in that certain particular direct rotations among those possible for a given partitioning of Q have geometric meaning and useful optimality properties.

11. MOTIVATION FOR AND USE OF THE CSD

A brief discussion of why the CSD is so powerful and a particular example of its use will start to show how general and useful it is, and give some feeling for how to use it in novel situations.

Perhaps one of the most convincing motivations for using the CSD is the realization it provides an extremely simple and practical algebraic characterization of the subblocks of a unitary matrix Q. A matrix Q is unitary if and only if Q is square and $Q^HQ = I$, but it is awkward to include the squareness when dealing with subblocks. Thus a popular way of characterizing the four subblocks $r_i \times c_i$ Q_{ij}, $i, j = 1, 2$, of unitary Q used to be via the eight equations arising from the subblocks of

$$Q^HQ = \begin{bmatrix} I_{c_1} & 0 \\ 0 & I_{c_2} \end{bmatrix}, \quad QQ^H = \begin{bmatrix} I_{r_1} & 0 \\ 0 & I_{r_2} \end{bmatrix}, \tag{20}$$

but these are remarkably awkward to manipulate. However if we write

$$Q_{ij} = U_i D_{ij} V_j^H, \quad U_i \text{ and } V_j \text{ unitary}, \qquad i, j = 1, 2, \tag{21}$$

with the D_{ij} having the form in (15), then we not only immediately satisfy the eight equations (20), but we have some remarkable additional structure to use: the essential SVDs of the Q_{ij} and the repeated appearance of the U_i and V_j. The four decompositions (21) are far easier to work with than the eight equations (20), and the structure leads to short and simple proofs.

As an example, let $A^\dagger$ denote the pseudo-inverse of A, and suppose we suspect

$$Q_{12}Q_{22}^\dagger + (Q_{11}^H)^\dagger Q_{21}^H = 0.$$

This is clearly true from $Q_{11}^H Q_{12} + Q_{21}^H Q_{22} = 0$ if Q_{11} and Q_{22} are square and nonsingular, but prior to the CSD it would not have been obvious how to prove it in general. Using the CSD we can just replace each Q_{ij} by its essential SVD (21), use $(U_i D_{ij} V_j^H)^\dagger = V_j D_{ij}^\dagger U_i^H$ for unitary U_i and V_j, and appeal to the structure of the D_{ij} in (15):

$$\begin{aligned}
&U_1^H[Q_{12}Q_{22}^\dagger + (Q_{11}^H)^\dagger Q_{21}^H]U_2 \\
&\quad = U_1^H[(U_1D_{12}V_2^H)(V_2D_{22}^\dagger U_2^H) + (U_1(D_{11}^H)^\dagger V_1^H)(V_1D_{21}^H U_2^H)]U_2 \\
&\quad = [D_{12}D_{22}^\dagger + (D_{11}^H)^\dagger D_{21}^H] \\
&\quad = \begin{bmatrix} O_s^H & & \\ & S & \\ & & I \end{bmatrix} \begin{bmatrix} I & & \\ & -C & \\ & & O_c^H \end{bmatrix}^\dagger + \begin{bmatrix} I & & \\ & C & \\ & & O_c^H \end{bmatrix}^\dagger \begin{bmatrix} O_s^H & & \\ & S & \\ & & I \end{bmatrix} \\
&\quad = \operatorname{diag}(O_s^H, -SC^{-1}, O_c) + \operatorname{diag}(O_s^H, C^{-1}S, O_c) = 0.
\end{aligned}$$

This illustrates how the repeated appearances of the U_i and V_j in (21) lead to simple proofs. But beyond showing the effectiveness of the CSD, both the particular example and the general principle (that is, that the CSD elegantly characterizes the subblocks of a unitary matrix) present convincing evidence for giving the fully general CSD in textbooks, and not just the restricted case of $r_1 = c_1$.

We have already seen [5, 6, 1] that the direct rotation and the closely related CSD are the correct tools for dealing with angles between subspaces. In that setting both are important — in the notation of Section 3 the CSD gives the structure of the partitioned unitary matrix $E^H F$, while the direct rotation is the unitary matrix closest to the identity which takes $\mathcal{R}(E_1)$ into $\mathcal{R}(F_1)$ when $r_1 \leq c_1$. The core of the direct rotation is the $\hat{D}$ in the direct rotation version of the CSD, and this gives cosines and sines of the principal angles between the subspaces. When $r_1 = c_1$ the CSD is used very effectively in the excellent advanced texts [9, 26] to examine such problems, as well as the related problem of distance between orthogonal projectors and many related perturbation problems, and we need go no further into this here.

12. THE CSD AND GENERALIZED SVDS

We have pointed out that the CSD gives the SVDs of the subblocks of unitary Q, but here we show that it also gives the generalized SVDs (GSVD) of combinations of subblocks. To understand this, note that if $A_1 \in \mathcal{C}^{r_1 \times r_2}$, and $A_2 \in \mathcal{C}^{r_2 \times r_2}$ is nonsingular, and we have the standard SVD $A_1 A_2^{-H} = U_1 \Sigma U_2^H$ (we use terms like A_2^{-H} rather than A_2^{-1} since we will later be looking at $A_1 A_2$, and if for example $A_2 = USV^H$ is an SVD, $A_2^{-H} = US^{-1}V^H$ is too, with U on the same side in each SVD), then for $j = 1, \ldots, \min\{r_1, r_2\}$, the jth row of $U_1^H A_1$ is a multiple of that of $U_2^H A_2^H$, since

$$U_1^H A_1 = \Sigma U_2^H A_2^H. \tag{22}$$

The quotient SVD (QSVD) of general $r_1 \times m$ A_1 and $m \times r_2$ A_2 is designed to correspond to the SVD of $A_1 A_2^{-H}$ when $m = r_2$ and A_2 is nonsingular. In line with (22) we seek unitary U_1 and U_2 so that $U_1^H A_1$ and $U_2^H A_2^H$ have, as far as is possible, corresponding rows parallel. Fortunately the CSD can be used in theory to give this U_1 and U_2, see [17]. Consider the reduction via unitary Q and W

$$Q^H \begin{bmatrix} A_1 \\ A_2^H \end{bmatrix} W = \begin{bmatrix} R & 0 \\ 0 & 0 \end{bmatrix}, \quad c_1 \times c_1 \;\; R \text{ nonsingular.} \tag{23}$$

Note we cannot assume $r_1 \le c_1$, much less $r_1 = c_1$ here. Let $U^H Q V = D$ in (15) be the CSD of Q, then

$$U^H \begin{bmatrix} A_1 \\ A_2^H \end{bmatrix} W = D V^H \begin{bmatrix} R & 0 \\ 0 & 0 \end{bmatrix} = D_1 \begin{bmatrix} V_1^H R & 0 \end{bmatrix}. \tag{24}$$

Applying $D_2^H D_1 = 0$ we obtain $D_{12}^H U_1^H A_1 + D_{22}^H U_2^H A_2^H = 0$, or equivalently

$$\begin{bmatrix} O_s & & \\ & S & \\ & & I \end{bmatrix} U_1^H A_1 = \begin{bmatrix} I & & \\ & C & \\ & & O_c \end{bmatrix} U_2^H A_2^H, \tag{25}$$

our desired analogy with (22), except here A_1 and A_2^H have been treated equally. D_{12}^H and D_{22}^H have identical row partitions, so the rows of $U_1^H A_1$ and $U_2^H A_2^H$ corresponding to the S and C blocks are the desired parallel rows. If A_2^H is nonsingular, then D_{21} is square and nonsingular in (24), and it follows from (15) that D_{12} is also. But then (25) shows D_{22} has no unit matrix, and $-D_{22}D_{12}^{-1}$ gives the singular values of $A_1 A_2^{-H}$. Thus in general U_1 and U_2 from the CSD of Q in (23) give the QSVD vectors of A_1 and A_2, while the elements of S and C give the nontrivial quotient singular values. This also points out that the form of CSD in (15) reveals some properties of the Q_{ij}, such as the common row spaces of Q_{1j} and Q_{2j}, that the direct rotation form (17) tends to hide.

Instead of viewing the QSVD of A_1 and A_2 as this relationship between the row spaces of A_1 and A_2^H, we can view it as a joint decomposition of A_1 and A_2^H. From (24) with $W \equiv [W_1, W_2]$ having $m \times c_1$ W_1, and nonsingular $X \equiv [W_1 R^H V_1, W_2]$ we see

$$A_1 = U_1 D_{11} \begin{bmatrix} V_1^H R & 0 \end{bmatrix} W^H = U_1 \Sigma_1 X^H, \quad \Sigma_1 \equiv \begin{bmatrix} D_{11} & 0 \end{bmatrix}, \tag{26}$$

$$A_2^H = U_2 D_{21} \begin{bmatrix} V_1^H R & 0 \end{bmatrix} W^H = U_2 \Sigma_2 X^H, \quad \Sigma_2 \equiv \begin{bmatrix} D_{21} & 0 \end{bmatrix}. \tag{27}$$

From this the QSVD of general $r_1 \times m$ A_1 and $m \times r_2$ A_2 can be viewed as this joint decomposition of A_1 and A_2^H into unitary $r_i \times r_i$ U_i, essentially diagonal $r_i \times m$ Σ_i and nonsingular $m \times m$ X^H. But the CSD gives $Q_{11} = U_1 D_{11} V_1^H$, $Q_{21} = U_2 D_{21} V_1^H$, which is just this joint decomposition of Q_{11} and Q_{21}, with the additional property that $X = V_1$ is unitary. Thus the CSD already gives the QSVDs of Q_{11} and Q_{21}^H, similarly of Q_{12} and Q_{22}^H, of Q_{11}^H and Q_{12}, and of Q_{21}^H and Q_{22}.

As well as considering the QSVD of $r_1 \times m$ A_1 and $m \times r_2$ A_2 as a generalization of the SVD of $A_1 A_2^{-H}$, we can also consider the product

SVD (PSVD) of A_1 and A_2 as the SVD of the product A_1A_2. But we have for example from (21) and (15)

$$Q_{11}Q_{21}^H = U_1D_{11}V_1^HV_1D_{21}^HU_2^H = U_1 \ \mathrm{diag}(O_s^H, CS, O_c)U_2^H,$$

which is essentially the SVD of $Q_{11}Q_{21}^H$, that is the PSVD of Q_{11} and Q_{21}^H, so the CSD also gives us the PSVDs of the above four pairs of subblocks of unitary Q.

The idea of the QSVD or PSVD of two matrices has been extended to the generalized SVD (GSVD) of any number of matrices with compatible dimensions by De Moor and Zha [7]. For $i = 1, \ldots, m$ let $n_{i-1} \times n_i$ A_i be m such matrices, then they proved there exist unitary U_1 and V_m, and nonsingular X_i and Z_i, $i = 1, \ldots, m-1$ such that

$$\begin{aligned} A_1 &= U_1D_1X_1^{-1}, \\ A_2 &= Z_1D_2X_2^{-1}, \\ A_3 &= Z_2D_3X_3^{-1}, \\ &\cdots \\ A_{m-1} &= Z_{m-2}D_{m-1}X_{m-1}^{-1}, \\ A_m &= Z_{m-1}S_mV_m^H, \end{aligned} \tag{28}$$

where the D_i and S_m are essentially diagonal, and for each i we can choose to have $Z_i = X_i^{-H}$ corresponding to a Q or quotient transition between A_i and A_{i+1}, or $Z_i = X_i$ corresponding to a P or product transition, see [7] for more details. Thus for three such matrices A_1, A_2 and A_3 we can have the QQ, QP, PQ or PP-GSVD, in analogy with the SVDs of $A_1A_2^{-H}A_3$, $A_1A_2^{-H}A_3^{-H}$, $A_1A_2A_3^{-H}$ or $A_1A_2A_3$ when the inverses exist.

Now if X_i is unitary then $X_i^{-H} = X_i$, and so the P and the Q transitions have the same Z_i. But the CSD already gives decompositions of the form (28) with the X_i and so $Z_i = X_i$ unitary, and all the GSVDs of *any* sequence of compatible subblocks $Q_{ij} = U_iD_{ij}V_j^H$, $Q_{kj}^H = V_jD_{kj}^HU_k^H$, *etc.* follow immediately from the CSD. For example

$$\begin{aligned} Q_{11}^H &= V_1D_{11}^HU_1^H \\ Q_{12} &= U_1D_{12}V_2^H \\ Q_{22}^H &= V_2D_{22}^HU_2^H \end{aligned}$$

give *all* the GSVDs of $c_1 \times r_1$ Q_{11}^H, $r_1 \times c_2$ Q_{12} and $c_2 \times r_2$ Q_{22}^H.

Thus while the CSD gives the separate SVDs of the subblocks of Q, it is the repetition of the U_i and V_j in these which makes the *combinations* GSVDs. These subtle properties contribute greatly to the power and usefulness of the CSD.

13. SUBMATRICES OF MATRICES AND THEIR INVERSES

Let $Z \in \mathcal{C}^{n\times n}$ be a nonsingular matrix, and G its inverse, partitioned as

$$Z = \begin{array}{cc} \begin{array}{cc} c_1 & c_2 \end{array} & \\ \begin{bmatrix} Z_{11} & Z_{12} \\ Z_{21} & Z_{22} \end{bmatrix} & \begin{array}{c} r_1 \\ r_2 \end{array} \end{array}, \quad G = \begin{array}{cc} \begin{array}{cc} r_1 & r_2 \end{array} & \\ \begin{bmatrix} G_{11} & G_{12} \\ G_{21} & G_{22} \end{bmatrix} & \begin{array}{c} c_1 \\ c_2 \end{array} \end{array}. \tag{29}$$

We can relate these subblocks via the QR factorization

$$\begin{bmatrix} Z_{11} & Z_{12} \\ Z_{21} & Z_{22} \end{bmatrix} = QR = \begin{bmatrix} Q_{11} & Q_{12} \\ Q_{21} & Q_{22} \end{bmatrix} \begin{bmatrix} R_{11} & R_{12} \\ & R_{22} \end{bmatrix}, \tag{30}$$

with unitary Q as in (1) and nonsingular $R_{11} \in \mathcal{C}^{c_1\times c_1}$, $R_{22} \in \mathcal{C}^{c_2\times c_2}$. Now

$$\begin{bmatrix} G_{11} & G_{12} \\ G_{21} & G_{22} \end{bmatrix} = R^{-1}Q^H = \begin{bmatrix} R_{11}^{-1} & -R_{11}^{-1}R_{12}R_{22}^{-1} \\ & R_{22}^{-1} \end{bmatrix} \begin{bmatrix} Q_{11}^H & Q_{21}^H \\ Q_{12}^H & Q_{22}^H \end{bmatrix},$$

and among these we have the useful relationships from the CSD (15)

$$\begin{aligned} G_{21} = R_{22}^{-1}Q_{12}^H = R_{22}^{-1}V_2D_{12}^HU_1^H, &\quad G_{22} = R_{22}^{-1}Q_{22}^H = R_{22}^{-1}V_2D_{22}^HU_2^H, \\ Z_{11} = Q_{11}R_{11} = U_1D_{11}V_1^HR_{11}, &\quad Z_{21} = Q_{21}R_{11} = U_2D_{21}V_1^HR_{11}. \end{aligned} \tag{31}$$

We will use these in the next section, but it is interesting to note that the CSD of Q not only gives the QSVD of Z_{11} and Z_{21}^H, but also the QSVD of G_{21}^H and G_{22}. For

$$Z_{11} = U_1D_{11}V_1^HR_{11}, \qquad Z_{21} = U_2D_{21}V_1^HR_{11},$$

is the equivalent of (26) and (27), as is

$$G_{21}^H = U_1D_{12}V_2^HR_{22}^{-H}, \qquad G_{22}^H = U_2D_{22}V_2^HR_{22}^{-H}.$$

This reveals the generalized singular vectors (columns of U_1 and U_2) are identical for the two QSVDs, and the singular value pairs are also closely related.

To find such relationships between other blocks we could consider the QL, RQ and LQ factorizations of Z, or just transpose and interchange Z and G.

This section has indicated that the CSD of Q in the QR factorization of a nonsingular matrix Z may be a useful theoretical tool in proving results on partitioned Z and Z^{-1}. We give an example in the next section.

14. NULLITIES OF SUBMATRICES

We define the column nullity cn and row nullity rn of $m \times n$ A to be the dimensions of the null spaces of A and A^H respectively. Thus if A has rank r then

$$cn(A) = n - r, \qquad rn(A) = m - r.$$

It is obvious from O_c and O_c^H in the CSD (15) that

$$Q_{11} \text{ and } Q_{22}^H \text{ have the same nullities.} \tag{32}$$

Similarly

$$Q_{21} \text{ and } Q_{12}^H \text{ have the same nullities.} \tag{33}$$

These simple properties are not quite so obvious from the direct rotation form (17) of the CSD.

Although (15) was derived in [17], those authors did not notice these nullity relations. These properties were noticed in [32, 31] where (32) was used in deriving properties of total least squares (TLS) solutions, but the result is a particular example of the following known result:

If G is the inverse of nonsingular Z partitioned as in (29), then

$$Z_{11} \textit{ and } G_{22} \textit{ have the same nullities.} \tag{34}$$

Since from (31) we see Z_{11} has the same nullities as Q_{11}, and G_{22} has the same nullities as Q_{22}^H, the result follows trivially from (32). We also see by transposing or interchanging Z and G where necessary that

$$\begin{aligned} &Z_{21} \text{ and } G_{21} \text{ have the same nullities,} \\ &Z_{12} \text{ and } G_{12} \text{ have the same nullities,} \\ &Z_{22} \text{ and } G_{11} \text{ have the same nullities.} \end{aligned} \tag{35}$$

The result (34) was needed in deriving the solution set of the generalized total least squares problem (GTLS) in [18, §2]. Charles Johnson [12] pointed out (35) has been proven (see for example [8, Theorem 2]) and used in several places recently. A quick look through the literature suggests the simplest proof from first principles is that given in [13, Lemma 4]. We have shown the result is "obvious", if we are already familiar with the CSD, the QR factorization, and know what to look for. Certainly it is immediately obvious for unitary matrices from the CSD (15).

15. CONCLUSION

In this paper we have made a clear distinction between the very *geometric* angles between subspaces problem and the *algebraic* CS decomposition — a unitary decomposition of subblocks of a unitary matrix. We have discussed the history from Jordan's early understanding of angles between subspaces through to today's clear understanding of the CSD. The geometric properties of angles between subspaces contributed greatly to the understanding of that problem, but limited us from seeing some of the more general algebraic power the CSD reveals. For this reason alone it seems pedagogically important to present the CSD initially without reference to angles between subspaces.

Briefly we have traced the major development of the CSD through the following stages. First there was the brilliant work of Jordan [14] in analyzing the angles between subspaces problem so successfully. Next there was the influence of Sz.-Nagy [19, §136] leading to the direct rotation introduced by Davis [2] and Kato [16, §1.4.6, §1.6.8]. Then there was the development and use of the direct rotation by Davis [2, 3, 4] and by Davis and Kahan in [6] in dealing with angles between subspaces. The key step from understanding the geometric angles between subspaces problem to the formulation of the algebraic CSD started with the detailed understanding of the structure of this direct rotation by Davis and Kahan [6] leading to the partial formulation and proof of the CSD in [5]. Finally Stewart concluded the major work by unequivocally stating the CSD as a unitary decomposition of the subblocks of a partitioned unitary matrix.

While Stewart only proved the result for $r_1 = c_1 \leq n/2$, which is perfectly adequate for dealing with angles between subspaces of the same dimensions, Paige and Saunders [17] found this was insufficient for their theory of the generalized (now quotient) singular value decomposition, and gave a briefer proof and a slightly different form of the CSD for *any* two-block by two-block partitioning of a unitary matrix. The present paper adds a little to the work on the CSD by providing an easy proof and motivation, by pointing out some nullity and rank properties it immediately reveals, by emphasizing that the CSD gives GSVDs of collections of subblocks as well as SVDs of subblocks of the unitary matrix, and by showing how, via the QR factorization, the CSD contributes to the analysis of subblocks of a partitioned general nonsingular matrix and its inverse.

The proof in Section 8 holds for the trivial cases of any 2×1, 1×2 or 1×1 partition, and the form reveals some properties not obvious from the more restricted $(r_1 \leq c_1 \leq r_2)$ direct rotation form of Davis, Kahan and Stewart, though this latter form is easier to remember and has simple optimality properties for the angles between subspaces problem which are less easily stated for the form (15). But the angles between subspaces problem is only one, although very important, use of the CSD, and for other uses we find these restrictions and even the direct rotation form of the CSD awkward. Since the Paige and Saunders form (15) has no restrictions on dimensions, is easily proven requiring no special cases, and the Davis, Kahan and Stewart direct rotation form can be easily derived from it, it is probably advisable in the general case to present the Paige and Saunders formulation.

We hope we have given ample evidence here to justify the CSD being presented pedagogically in its general form, with no restrictions on the dimensions of the subblocks of the partition. It would also help such a presentation if the angles between subspaces problem was presented as a distinct problem, which of course can be treated elegantly via the CSD. That is, we feel it is clearer to present the CSD as the decomposition of subblocks of a partitioned unitary matrix, without reference to where these subblocks may have come. The angles between subspaces problem and the illuminating direct rotation can then be presented as one of the many applications of the CSD.

Finally we discuss the question of when should the CSD be tried in a theoretical investigation. An obvious answer is whenever the problem involves any of:

- angles between subspaces,
- 2×2 partitions of unitary matrices,
- orthogonal projectors, or
- 2×2 partitions of nonsingular matrices and their inverses.

Also the CSD is a powerful tool in perturbation analysis in general, see especially [23, 26]. Briefly, whenever some aspect of a problem can usefully be formulated in terms of two-block by two-block partitions of unitary matrices, the CSD will probably add insights and simplify the analysis.

We wish to thank Miguel Anjos, Bill Farebrother, Roger Horn, Charles Johnson, Pete Stewart and George Styan for valuable comments on this work and the relevant history. Xiaowen Chang helped us to put things together to meet the Special Issue deadline.

REFERENCES

1 Å. Björck and G.H. Golub. Numerical methods for computing angles between linear subspaces. *Math. Comp.*, 27:579–594 (1973).

2 C. Davis. Separation of two linear subspaces. *Acta Sci. Math. Szeged*, 19:172–187 (1958).

3 C. Davis. The rotation of eigenvectors by a perturbation. *J. Math. Anal. Appl.*, 6:159–173 (1963).

4 C. Davis. The rotation of eigenvectors by a perturbation II. *J. Math. Anal. Appl.*, 11:20–27 (1965).

5 C. Davis and W.M. Kahan. Some new bounds on perturbations of subspaces. *Bull. Amer. Math. Soc.*, 75:863–868 (1969).

6 C. Davis and W.M. Kahan. The rotation of eigenvectors by a perturbation III. *SIAM J. Numer. Anal.*, 7:1–46 (1970).

7 B. De Moor and H. Zha. A tree of generalizations of the ordinary singular value decomposition. *Linear Algebra Appl.*, 147:469–500 (1991).

8 M. Fiedler and T. Markham. Completing a matrix when certain of its entries are specified. *Linear Algebra Appl.*, 74:225–237 (1986).

9 G.H. Golub and C.F. Van Loan. *Matrix Computations.* The Johns Hopkins University Press, Baltimore, MD, second edition, (1989).

10 H. Hotelling. Relation between two sets of variates. *Biometrika*, 28:322–377 (1936).

11 A. T. James and G. N. Wilkinson. Factorization of the residual operator and canonical decomposition of nonorthogonal factors in the analysis of variance. *Biometrika*, 58:279–294 (1971).

12 C. R. Johnson (1992). Personal Communication.

13 C. R. Johnson and M. Lundquist. Operator matrices with chordal inverse patterns. IMA Preprint Series 885, University of Minnesota, Minneapolis, MN, (1991).

14 C. Jordan. Essai sur la géométrie à n dimensions. *Bulletin de la Société Mathématique de France*, 3:103–174 (1875).

15 T. Kato. Notes on projections and perturbation theory. Technical report 9, Department of Mathematics, University of California, Berkeley, CA (1955).

16 T. Kato. *Perturbation Theory for Linear Operators.* Springer-Verlag, New York, NY, second edition (1976). First edition 1966.

17 C.C. Paige and M. Saunders. Towards a generalized singular value decomposition. *SIAM J. Numer. Anal.*, 18:398–405 (1981).

18 C.C. Paige and M. Wei. Analysis of the generalized total least squares problem $AX \approx B$ when some columns of A are free of error. *Numer. Math.* To appear.

19 F. Riesz and B. Sz.-Nagy. *Functional Analysis.* Frederick Unger, New York, NY (1955). English translation of *Leçons d'analyse fonctionnelle*, 1952, 1953.

20 P. H. Schoute. *Mehrdimensionale Geometrie, Vol. 1: Die linearen Räume.* G. J. Göschensche Verlagshandlung, Leipzig (1902).

21 G.W. Stewart. Error bounds for approximate invariant subspaces of closed linear operators. *SIAM J. Numer. Anal.*, 8:796–808 (1971).

22 G.W. Stewart. Error and perturbation bounds for subspaces asso-

ciated with certain eigenvalue problems. *SIAM Rev.*, 15:727–764 (1973).

23 G.W. Stewart. On the perturbation of pseudo-inverses, projections, and linear least squares problems. *SIAM Rev.*, 19:634–662 (1977).

24 G.W. Stewart. Computing the CS decomposition of a partitioned orthogonal matrix. *Numer. Math.*, 40:297–306 (1982).

25 G.W. Stewart. A method for computing the generalized singular value decomposition. In B. Kågström and A. Ruhe, editors, *Matrix Pencils*, pages 207–220. Springer-Verlag, New York (1983).

26 G.W. Stewart and J. Guang Sun. *Matrix Perturbation Theory.* Academic Press, New York, NY (1990).

27 C.F. Van Loan. Generalizing the singular value decomposition. *SIAM J. Numer. Anal.*, 13:76–83 (1976).

28 C.F. Van Loan. On Stewart's singular value decomposition for partitioned orthogonal matrices. Technical Report CS79-767, Computer Science, Stanford University, Stanford, CA (1979).

29 D. S. Watkins. *Fundamentals of Matrix Computations.* John Wiley and Sons, New York (1991).

30 P.Å. Wedin. Perturbation bounds in connection with the singular value decomposition. *BIT*, 12:99–111 (1972).

31 M. Wei. Algebraic relations between the total least squares and the least squares problems with more than one solution. *Numer. Math.*, 62:123–148 (1992).

32 M. Wei. The analysis for the total least squares problem with more than one solution. *SIAM J. Matrix Anal. Appl.*, 13:746–763 (1992).