SU 376P30-10

NUMERICAL METHODS FOR COMPUTING ANGLES BETWEEN LINEAR SUBSPACES

BY

ÅKE BJÖRCK

GENE H. GOLUB

STAN-CS-225-71

JULY, 1971

COMPUTER SCIENCE DEPARTMENT
School of Humanities and Sciences
STANFORD UNIVERSITY

NUMERICAL METHODS FOR COMPUTING ANGLES BETWEEN LINEAR SUBSPACES

Åke Björck [1)]

Gene H. Golub [2)]

6 -71

[1)] Linköping University, s-581 83 Linköping, Sweden

[2)] The work of this author was in part supported by the Atomic Energy Commission and by the National Science Foundation.

Abstract

Assume that two subspaces F and G of a unitary space are defined as the ranges (or nullspaces) of given rectangular matrices A and B. Accurate numerical methods are developed for computing the principal angles $\theta_k(F,G)$ and orthogonal sets of principal vectors $u_k \in F$ and $v_k \in G$, $k = 1,2,\ldots, q = \dim(G) \leq \dim(F)$. An important application in statistics is computing the canonical correlations $\sigma_k = \cos \theta_k$ between two sets of variates. A perturbation analysis shows that the condition number for θ_k essentially is $\max(\kappa(A),\kappa(B))$, where κ denotes the condition number of a matrix. The algorithms are based on a preliminary &R-factorization of A and B (or A^H and B^H), for which either the method of Householder transformations (HT) or the modified Gram-Schmidt method (MGS) is used. Then $\cos \theta_k$ and $\sin \theta_k$ are computed as the singular values of certain related matrices. Experimental results are given, which indicates that MGS gives θ_k with equal precision and fewer arithmetic operations than HT. However, HT gives principal vectors, which are orthogonal to working accuracy, which is not in general true for MGS. Finally the case when A and/or B are rank deficient is discussed.

1. Introduction

Let F and G be given subspaces of a unitary space E^m, and assume that

(1) $$p = \dim(F) \geq \dim(G) = q \geq 1.$$

The smallest angle $\theta_1(F,G) = \theta_1 \epsilon [0,\pi/2]$ between F and G is defined by

$$\cos \theta_1 = \max_{u \epsilon F} \max_{v \epsilon G} u^H v , \quad ||u||_2 = 1 \quad ||v||_2 = 1$$

Assume that the maximum is attained for $u = u_1$ and $v = v_1$. Then $\theta_2(F,G)$ is defined as the smallest angle between the orthogonal complement of F with respect to u_1 and that of G with respect to v_1. Continuing in this way until one of the subspaces is empty, we are led to the following definition.

DEFINITION The principal angles $\theta_k \epsilon [0,\pi/2]$ between F and G are recursively defined for $k = 1,2,\dots,q$ by

(2) $$\cos \theta_k = \max_{u \epsilon F} \max_{v \epsilon G} u^H v = u_k^H v_k , \quad ||u||_2=1 \quad ||v||_2=1$$

subject to the constraints

$$u_j^H u = 0, \; v_j^H v = 0, \; j = 1,2,\dots,k-1.$$

The vectors $(u_1,\dots,u_q)$ and $(v_1,\dots,v_q)$ are called principal vectors of the pair of spaces.

We note that the principal vectors need not be uniquely defined, but the principal angles always are. The vectors $V = (v_1,\dots,v_q)$ form a unitary basis for G and the vectors $U = (u_1,\dots,u_q)$ can be complemented with (p-q) unitary vectors so that $(u_1,\dots,u_p)$ form a unitary basis for F.

It can also be shown that

$$u_j^H v_k = 0, \quad j \neq k, \quad j = 1,\ldots,p, \quad k = 1,\ldots,q.$$

For an introduction to these concepts we refer to [1]. An up to date list of references can be found in [7].

Principal angles and vectors have many important applications in statistics and numerical analysis. In [5] the statistical models of canonical correlations,factor analysis and stochastic equations are described in these terms. By taking the vectors u_k corresponding to $\cos \theta_k = 1$ we get a unitary basis for the intersection of the two spaces F and G. This has applications in the generalized eigenvalue problem [11]. Other applications are found in the theory of approximate least squares [6] and in the computation of invariant subspaces of a matrix [18].

The purpose of this paper is to develop new and more accurate methods for computing principal angles and vectors, when the subspaces are defined as the ranges (or nullspaces) of two given matrices A and B. In section 2 we describe the standard method of computing canonical correlations, and show why this method may give rise to a serious loss of accuracy. Assuming that unitary bases for F and G are known we derive in section 3 formulas for computing principal angles and vectors from the singular values and vectors for certain matrices. To find out how accurately the angles are defined in presence of uncertainties in A and B, first order perturbation results are given in section 4. In section 5 different numerical methods for computing the unitary bases, and the use of the formulas from section 3, are discussed with respect to efficiency and accuracy. The special problems which arise when A and/or B are exactly or nearly rank deficient are discussed in section 6. Finally some numerical results are given in section 7.

2. Canonical correlations

For a matrix A we denote the range of A by R(A) and the nullspace of A by N(A),

$$(3) \qquad R(A) = \{u \mid Ax = u\} , \qquad N(A) = \{x \mid Ax = 0\} .$$

In the problem of canonical correlations we have F = R(A), G = R(B) where A and B are given rectangular matrices. Then, the canonical correlations are equal to cos θ_k, and it can be shown that

$$(4) \qquad \cos \theta_k = \sigma_k, \quad u_k = Ay_k, \quad v_k = Bz_k, \quad k = 1,2,\ldots,q,$$

where $\sigma_k \geq 0$ are eigenvalues and y_k, z_k properly normalized eigenvectors to the generalized eigenvalue problem

$$(5) \qquad \begin{pmatrix} 0 & A^HB \\ B^HA & 0 \end{pmatrix} \begin{pmatrix} y \\ z \end{pmatrix} = \sigma \begin{pmatrix} A^HA & 0 \\ 0 & B^HB \end{pmatrix} \begin{pmatrix} y \\ z \end{pmatrix} .$$

Assume for convenience that A and B have full column rank. The standard method [4] of computing canonical correlations is to compute A^HA, B^HB, A^HB and perform the Choleski decompositions

$$A^HA = R_A^HR_A, \qquad B^HB = R_B^HR_B,$$

where R_A and R_B are upper triangular.

The eigenvalue problem (5) is then equivalent to the eigenvalue problems for a pair of Hermitian matrices

$$MM^H\hat{y}_i = \sigma_i^2 \hat{y}_i, \qquad M^HM \hat{z}_i = \sigma_i^2 \hat{z}_i$$

where

$$M = (R_A^H)^{-1}(A^HB)R_B^{-1}, \quad \hat{y}_i = R_Ay_i, \quad \hat{z}_i = R_Bz_i$$

These can be solved by standard numerical methods.

When q = 1 and B = b, the principal angles and vectors are closely related to the least squares problem of minimizing $||b - Ax||_2$. In fact, with the notations above (but dropping subscripts), we have

$$y = x/||Ax||_2, \quad z = 1/||b||_2, \quad \sigma = ||Ax||_2/||b||_2,$$

and (5) is reduced to

$$A^H bz = \sigma A^H A y, \qquad b^H A y = \sigma b^H b z.$$

But the first equation here is the normal equations for $x = \sigma y/z$. Thus the classical algorithm reduces for q = 1 to solution of the normal equations by Choleski's method.

Lately it has been stressed by several authors that forming the normal'equations in single precision involves a loss of information which cannot be retrieved. For linear least squares problems other methods without this disadvantage have been developed [2],[13] and [14]. Our aim in this paper is to generalize these methods to the case when q > 1.

3. Solution using singular values

In most applications each subspace is defined as the range, or the complement of the range, of a given matrix. In this case a unitary basis for the subspace may be computed in a numerically stable way by well known methods for the &R-decomposition of a matrix. These methods will produce for an mxn matrix A, with $m \geq n$, a decomposition

$$A = (Q'|Q'') \begin{pmatrix} S \\ \hline 0 \end{pmatrix} \begin{matrix} \}p\times n \\ \}(m-p)xn \end{matrix} \quad ,$$

where rank $(S) = p$ and $Q = (Q'|Q'')$ is unitary. Then Q' gives a unitary basis for the range of A, R(A), and Q'' a unitary basis for the complement $\overline{R(A)}$. Notice that the case when a subspace is defined as the null-space $N(A^H)$ of a matrix A^H is included, since $N(A^H) = \overline{R(A)}$. The computation of unitary bases will be discussed in more detail in section 5 and 6, and we assume here that such bases have been obtained.

Recently an efficient and numerically stable algorithm for computing the singular value decomposition [9] (SVD) of a matrix has been developed [14]. This algorithm will be our basic tool for computing principal angles and vectors. The relation between singular values and our problem is clear from the following theorem.

THEOREM 1. Assume that the columns of Q_A and Q_B form unitary bases for two subspaces of a unitary space E^m. Put

(7) $$M = Q_A^H Q_B,$$

and let the SVD of this $p\times q$ matrix be

(8) $$M = Y\ C\ Z^H, \qquad C = \mathrm{diag}(\sigma_1, \ldots, \sigma_q),$$

where

$$Y^H Y = Z^H Z = Z Z^H = I_q$$

If we assume that

$$\sigma_1 \geq \sigma_2 \geq \ . \ . \ . \geq \sigma_q$$

then the principal angles and principal vectors associated with this part of subspaces are given by

(9) $$\cos \theta_k = \sigma_k(M), \qquad U = Q_A Y, \qquad V = {}_{QB} Z,$$

Proof: It is known [15]that the singular values and singular vectors of a matrix M can be characterized by

(10) $$\sigma_k = \max_{||y||_2 = ||z||_2 = 1} (y^H Mz) = y_k^H M z_k,$$

subject to

$$y^H y_j = z^H z_j = 0, \quad j = 1, \ldots, k-1.$$

If we put

$$U = Q_A y \in R(Q_A), \qquad v = Q_B z \in R(Q_B),$$

then it follows that $||u||_2 = ||y||_2$, $||v||_2 = ||z||_2$ and

$$y^H y_j \ . \ u^H u_j, \quad {}^H{}_j \ . \ v^H y_j. \ .$$

Since $\qquad Y^H Mz = y^H Q_A^H Q_B z = u^H v,$ (10) is equivalent to

$$\sigma_k = \max_{||u||_2 = ||v||_2 = 1} (u^H v) = u_k^H v_k$$

subject to

$$u^H u_j = v^H v_j = 0, \quad j = 1, \ldots, k-1$$

Now (9) follows directly from the definition of principal angles and vectors (2), which concludes the proof.

For small angles θ_k is not well determined from $\cos\theta_k$ and we now develop formulas for computing $\sin\theta_k$. Let Q_A and Q_B be defined as in theorem 1. For convenience we change the notations slightly and write (8) and (9) as

(8') $$M = Y_A\ C\ Y_B^H, \qquad c = \mathrm{diag}\,(\cos\theta_k),$$

(9') $$U_A = Q_A Y_A, \qquad U_B = Q_B Y_B$$

We split Q_B according to

(11) $$Q_B = P_A Q_B + (I - P_A) Q_B\ ,$$

where $P_A = Q_A Q_A^H$ is the orthogonal projection onto $R(Q_A)$. Here

$$P_A Q_B = Q_A Q_A^H Q_B = Q_A M = Q_A Y_A C\ Y_B^H\ ,$$

and hence the SVD of the matrix $P_A Q_B$ is given by

(12) $$P_A Q_B = U_A\ C\ Y_B^H\ , \qquad C = \mathrm{diag}(\cos\theta_k).$$

Since $P_A(I - P_A) = 0$ we get from squaring (11)

$$Q_B^H(I - P_A)^2 Q_B = I - Q_B^H\ P_A^2\ Q_B = Y_B(I - C^2)Y_B^H$$

and it follows that the SVD of $(I-P_A)Q_B$ is given by

(13) $$(I-P_A)Q_B = W_A\ S\ Y_B^H, \qquad s = \mathrm{diag}(\sin\theta_k).$$

Comparing (13) with (12) it is evident that W_A gives the principal vectors in the complement $\overline{R(Q_A)}$ associated with the pair of subspaces $(\overline{R(Q_A)}, R(Q_B))$. When $p \ll m$ the SVD of $(I-P_A)Q_B$ can be computed more economically from that of M, using

(14) $$(I-P_A)Q_B Y_B = W_A \cdot S$$

We will for the rest of this section assume that in addition to (1) we have

$$p + q \leq m \ .$$

This is no real restriction, since otherwise we have $(m-p)+(m-q) \leq m$, and we can work with the complements of $R(Q_A)$ and $R(Q_B)$ instead. Then $\dim(\overline{R(Q_A)}) = m-p \geq q$, and we can choose the $m \times q$ matrix W_A in (13) so that $W_A^H U_A = 0$.

By analogy we have formulas similar to (12) and (13) related to the splitting $Q_A = P_B Q_A + (I-P_B)Q_A$,

$$(15) \qquad P_B Q_A = U_B \, C \, Y_A^H \quad , \qquad (I - P_B)Q_A = W_B \, S \, Y_A^H \ ,$$

where again since $m - q \geq p \geq q$ we can choose the $m \times q$ matrix W_B so that $W_B^H U_B = 0$. From (15) we get

$$U_A = Q_A Y_A = (U_B C + W_B S) Y_A^H Y_A = (U_B W_B) \begin{pmatrix} C \\ S \end{pmatrix}$$

If we put

$$P_{B,A} = U_A U_B^H = (U_B \ W_B) \begin{pmatrix} C \\ S \end{pmatrix} U_B^H \ ,$$

then, since $R(Q_B) = R(U_B)$, we have for any $x \in R(Q_B)$ that

$$P_{B,A} x \in R(Q_A), \ ||x||_2 = ||P_{B,A} x||_2$$

We can now alwaysfind an $m \times (m-2q)$ matrix Z_B such that $(U_B \ W_B \ Z_B)$ is a unitary basis in E^m. Then

$$(16) \qquad \mathcal{P}_{B,A} = (U_B \ W_B | \ Z_B) \left(\begin{array}{cc|c} C & -S & \\ S & C & 0 \\ \hline & 0 & I \end{array} \right) \left(\begin{array}{c} U_B^H \\ W_B^H \\ \hline Z_B^H \end{array} \right)$$

is the matrix of a unitary transformation, mapping $R(Q_B)$ into $R(Q_A)$. Its restriction to $R(Q_B)$ is $P_{B\ A}$, and it leaves all vectors in $R(Z_B)$ unchanged. This transformation'is called a <u>direct rotation</u> [7] from $R(Q_B)$ into $R(Q_A)$. It is distingui‘shed from other unitary transformations $\mathcal{P}$ taking $R(Q_B)$ into $R(Q_A)$ by the property that it minimizes each unitarily invariant norm of $(I - \mathcal{P})^H(I - \mathcal{P})$. If $R(Q_B) \cap \overline{R(Q_A)}$ is empty, then all $\theta_k < \Pi/2$ and the direct rotation is uniquely determined.

Similarly we can construct a direct rotation taking $R(U_A)$ into $(R(Q_B)$. It is obvious that the relations between the two subspaces are very completely characterized by the quantities C, S, U_A, W_A, U_B and W_B.

4. Perturbation of principal angles

We consider here how the principal angles between R(A) and R(B) changes when the elements in A and B are subject to perturbations. We assume in this analysis that the matrices A and B are m×p and m×q respectively, and have linearly independent columns. Consider first a perturbation of A only,

$$A_\varepsilon = A + \varepsilon E = (A + \varepsilon E_1) + \varepsilon E_2,$$

where we have split the perturbation in components in and orthogonal to R(A),

$$E_1 = P_A E, \; E_2 = (I-P_A)E.$$

Let the polar decomposition of $A+\varepsilon E_1$ be

$$A+\varepsilon E_1 = Q_A H_A \quad , \quad Q_A^H Q_A = I, \quad H_A \text{ positive definite.}$$

Then, since $R(A) = R(A+\varepsilon E_1)$, Q_A gives a unitary basis for R(A).

To get a unitary basis for $R(A_\varepsilon)$ we note that for small absolute values of ε , the matrix

(17) $$(A+\varepsilon E)H_A^{-1} = Q_A + \varepsilon F, \quad F = (I-P_A)EH_A^{-1}$$

is nearly orthogonal. Since $Q_A^H F = Q_A^H Q_A Q_A^H F = Q_A^H P_A F = 0$ we have

$$I - (Q_A+\varepsilon F)^H (Q_A+\varepsilon F) = -\varepsilon(F^H Q_A + Q_A^H F) + O(\varepsilon^2) = O(\varepsilon^2).$$

Then from a series expansion for the unitary factor Q_{A_ε} in the polar decomposition of $A_\varepsilon H_A^{-1}$ [3] it follows that

(18) $$Q_{A_\varepsilon} = Q_A + \varepsilon F + O(\varepsilon^2).$$

Premultiplying (18) by P_B we get

$$P_B Q_{A_\varepsilon} = P_B Q_A + \varepsilon P_B (I-P_A) E H_A^{-1} .$$

Using the well known inequalities for singular values, [12] p. 30,

$$\sigma_k(A+B) \le \sigma_k(A) + \sigma_1(B), \ \sigma_k(AB) = \sigma_k(A)\sigma_1(B),$$

$$k = 1,2,\ldots,q,$$

we obtain

$$|\sigma_k(P_B Q_{A_\varepsilon}) - \sigma_k(P_B Q_A)| \le \varepsilon\, \sigma_1((P_B(I-P_A))\sigma_1(EH_A^{-1}) + O(\varepsilon^2).$$

Now $P_B(I-P_A) = U_B \operatorname{diag}(\sin\theta_k) W_A^H$ and since

$$\sigma_1(H_A^{-1}) = 1/\sigma_p(A+\varepsilon E_1) = 1/\sigma_p(A) + O(\varepsilon),$$

we have to first order in ε

(19) $$|\Delta\cos\theta_k| \le \varepsilon \sin\theta_{max}\, \sigma_1(E)/\sigma_p(A).$$

If instead we premultiply (20) by $(I-P_B)$, and proceed in the same way we arrive at

(20) $$|\Delta\sin\theta_k| \le \varepsilon \cos\theta_{min}\, \sigma_1(E)/\sigma_p(A).$$

Now assume that both A and B are perturbed by δA and δB respectively, where

$$||\delta A||_2/||A||_2 \le \varepsilon A, \ ||\delta B||_2/||B||_2 \le \varepsilon_B.$$

Then to first order of approximation the perturbationsadd together and we get from (19) and (20)

(21) $$|\Delta\cos\theta_k| \le K\cdot\sin\theta_{max}, \quad |\Delta\sin\theta_k| \le K\cos\theta_{min},$$

$$K = \varepsilon_A \frac{\sigma_1(A)}{\sigma_p(A)} + \varepsilon_B \frac{\sigma_1(B)}{\sigma_q(B)} = \varepsilon_A \kappa(A) + \varepsilon_B \kappa(B)$$

Thus again neglecting terms of higher order, we have

$$|\Delta\theta_k| \le K\cdot\min\left(\frac{\sin\theta_{max}}{\sin\theta_k}, \frac{\cos\theta_{min}}{\cos\theta_k}\right) = K\cdot g(\theta_k).$$

The maximum of g(8) for $0 \le \theta \le \pi/2$ is attained for $\theta = \arctan r$

$$g_{max} = (1+r^2)^{\frac{1}{2}}\cos\theta_{min}, \quad r = \sin\theta_{max}/\cos\theta_{min}.$$

It follows that

$$g_{max} \le (1 + \cos^{-2}\theta_{min})^{\frac{1}{2}}\cos\theta_{min} \le \sqrt{2}$$

and finally

(22) $$|\Delta\theta_k| \le \sqrt{2}(\varepsilon_A\kappa(A) + \varepsilon_B\kappa(B)).$$

We conclude that when both $\kappa(A)$ and $\kappa(B)$ are small, then the angles θ_k are well determined.

We note that if the columns in A are scaled, then $\kappa(A)$ will change, but not R(A). Also the numerical algorithms for the &R-decomposition have the property that, unless column pivoting is used, they give the same numerical results independent of such a scaling. Therefore it is often more relevant to take in (21) as condition number for A the number

$$\kappa'(A) = \min_D \kappa(AD), \qquad D = \mathrm{diag}(d_1,\ldots,d_p).$$

It has been shown in [16] and [17] that $\kappa(AD)$ is not more than a factor of $p^{\frac{1}{2}}$ away from its minimum, if in AD all columns have equal L_2-norm. This suggests that A and B should'be assumed to be preconditioned so that

$$||a_i||_2 = ||b_j||_2 = 1, \quad i = 1,\ldots,p, \quad j = 1,\ldots,q.$$

We remark that $\kappa'(A)$ is essentially the spanning precision of the basis in R(A) provided by A as defined in [17].

5 Numerical methods

We assume in this section that the columns in A and B are linearly independent. The singular and near singular case will be briefly discussed in section 6. For convenience we also assume that A and B are real matrices, although all algorithms given here can easily be generalized to the complex case. Computed quantities will be marked by a bar.

In order to get the orthogonal bases for F and G we need the QR-decompositions of the matrices A and B. We now describe two efficient methods for computing these. In the method of Householder triangularizations (HT) [13] orthogonal transformations of the type $Q_k = I - 2w_k w_k^T$ are used, where

$$w_k = (0,\dots,0,w_{kk},\dots,w_{mk})^T, \quad ||w_k||_2 = 1.$$

The m×p matrix A is then reduced to triangular form using premultiplications

$$Q_p \cdots Q_2 Q_1 A = \begin{pmatrix} R_A \\ \hline 0 \end{pmatrix} \begin{matrix} \} \; p \\ \} \; m-p \end{matrix}$$

where w_k is chosen so that Q_k annihilates the appropriate elements in the k th column. Since $Q_k^{-1} = Q_k$, an orthogonal bases Q_A for R(A) can then be computed by premultiplying the first p columns in the unit matrix I_m by the same transformations in reversed order,

$$Q_A = Q_1 \; Q_2 \cdots Q_p \begin{pmatrix} I_p \\ \hline 0 \end{pmatrix} .$$

For this method a very satisfactory error analysis is given in [19].

Assume that floating point arithmetic with a mantissa of t binary digits is used, and that inner-products are accumulated in double precision wherever possible. Then there exists an _exactly_ orthogonal matrix Q such that the computed matrices satisfy

$$(23)\qquad Q^T(A + E_A) = \begin{pmatrix} \bar{R}_A \\ 0 \end{pmatrix}, \qquad \bar{Q}_A = Q\begin{pmatrix} I_p \\ 0 \end{pmatrix} + F_A = Q_{\tilde{A}} + F_A,$$

$$\|E_A\|_F = 12.5\ p\ 2^{-t}\|A\|_F,\ \|F_A\|_F = 12.5\ p^{3/2}\ 2^{-t},$$

where $Q_{\tilde{A}}$ is an exactly orthogonal basis f'or $R(A+E_A)$. From this and a similar estimate for $\bar{Q}_B$ we get

$$(24)\qquad |\sigma_k(\bar{M}) - \sigma_k(\tilde{M})| \le \sigma_1(\bar{M} - \tilde{M}) \le 13.0(p^{3/2} + q^{3/2})2^{-t},$$

where $\tilde{M} = Q_{\tilde{A}}^T Q_{\tilde{B}}$ and the constant 13 .0 accounts f'or the rounding errors in computing the product $\bar{Q}_A^T \bar{Q}_B$. We have $\sigma_k(\tilde{M}) = \cos \tilde{\theta}_k$, where $\tilde{\theta}_k$ arc the exact angles between $(A+E_A)$ and $(B+E_B)$. Thus, the difference between $\tilde{\theta}_k$ and θ_k can be estimated from (22),

$$(25)\qquad |\tilde{\theta}_k - \theta_k| \le 12.5\ \sqrt{2}\ (p\kappa(A)+q\kappa(B))2^{-t}\ .$$

Finally, the errors $\bar{\sigma}_k(\bar{M})-\sigma_k(\bar{M})$ in computing the singular values of $\bar{M}$, using the procedure in [14] , will be of the same order of magnitude as those in (24).

The error estimate given above is satisfactory, except when $\theta_k << 1$. In this case, the errors in cos θ_k from (24) will give rise to errors in θ_k which may be much larger than those in (25). We return later to the problem of accurately computing small angles.

An orthogonal basis Q'_A for $\overline{R(A)} = N(A^T)$ can be obtained by applying the transformations Q_k, $k = p, \ldots, 1$ to the last (m-p) columns in I_m,

$$Q'_A = Q_1 Q_2 \ldots Q_p \left(\frac{0}{I_{m-p}}\right) .$$

Also in this case the estimate (23) for $\bar{Q}'_A$, (24) and (25) still hold if the factor $p^{3/2}$ is everywhere replaced by $p(m-p)^{1/2}$.

The QR-decomposition of a matrix A can also be computed using the modified Gram-Schmidt method (MGS) [2]. The matrix A is then transformed in p steps, $A = A1, A_2, \ldots, A_{p+1} = Q_A$ where

$$A_k = (q_1, \ldots, q_{k-1}\ a_k^{(k)}, \ldots, a_p^{(k)}).$$

The matrix A_{k+1}, $k = 1, 2, \ldots, p$ is computed by

$$qk = a_k^{(k)} / || a_k^{(k)} ||_2 , \quad a_j^{(k+1)} = (I - q_k q_k^T) a_j^{(k)}, \ j > k ,$$

and the elements in the k th row of R_A are

$$r_{kk} = || a_k^{(k)} ||_2 , \quad r_{kj} = q_k^T a_j^{(k)} , \quad j > k.$$

It has been shown in [2] p. 10, 15 that the computed matrices $\bar{R}_A$ and $\bar{Q}_A$ satisfy

$$A + E_A = \bar{Q}_A \bar{R}_A , \quad ||E_A||_F \leq 1.5(p-1)2^{-t} ||A||_F ,$$

(26)

$$|| Q_{\tilde{A}} - \bar{Q}_A ||_2 \leq 2p(p+1)\kappa(A) \cdot 2^{-t} ,$$

where $Q_{\tilde{A}}$ is an exactly orthogonal basis for $R(A+E_A)$ and quantities of order 2^{-2t} have been neglected. With MGS $\bar{Q}_A$ will in general not be orthogonal to working accuracy, and we cannot therefore hope to get principal vectors which are nearly orthogonal. Also the condition numbers $\kappa(A)$ and $\kappa(B)$ will enter in the estimate corresponding to (24). However, since $\kappa(A)$ and $\kappa(B)$ already appear in (25), we can hope to get the principal angles as accurately as with HT. Experimental results reported in section 7 will confirm that this actually seems to be the case.

An advantage with MGS is that the total number of multiplications required to compute $\bar{R}_A$ and $\bar{Q}_A$ is less than for HT, i.e.

$$\text{MGS: } p^2m\ , \qquad \text{HT: } 2p^2(m-\frac{p}{3}).$$

If only the principal angles are wanted, then the number of multiplications in the SVD-algorithm is approximately

$$2q^2(p - \frac{q}{3}).$$

Thus, when $m >> p$, the dominating work is in computing Q_A and Q_B and in this case MGS requires only half as much work as HT.
If also the principal vectors are wanted, we must compute the full SVD of M. Assuming two iterations per singular value, this requires approximately

$$7q^2(p + \frac{10}{21}q)$$

multiplications. To compute U_A and U_B a further $mq(p+q)$ multiplications are needed.

To get a basis for $\overline{R(A)}$ using MGS we have to apply the method to the bordered matrix $(A|I_m)$, and after m steps pick out (m-p) appropriate columns. Especially when $(m-p) << m$, the number of multiplications compares unfavourably with HT,

$$\text{MGS: } m^2(m+2p), \qquad \text{HT: } 2mp(m-p) + \frac{2}{3}p^3.$$

In some applications, e.g. canonical correlations, we want to express the principal vectors as linear combinations of the columns in A and B, respectively. We have $U_A = Q_AY_A = A(R_A^{-1}Y_A)$, and hence

$$U_A = A\,X_A, \quad U_B = B\,X_B,$$

where

(27) $$X_A = R_A^{-1} Y_A, \; X_B = R_B^{-1} Y_B \;.$$

We remark that if we let $\bar{X}_A$ and $\bar{X}_B$ denote the computed matrices, then $A\,\bar{X}_A$ and $B\,\bar{X}_B$ will not in general be orthogonal to working accuracy even when HT is used.

We now turn to the problem of accurately determining small angles. One method is to compute $\sin\theta_k$ from the SVD (13) of the matrix

$$G = (I - P_A)Q_B = Q_B - Q_A M \;.$$

If we let $\tilde{G}$ denote the corresponding matrix computed from $Q_{\tilde{A}}$ and $Q_{\tilde{B}}$, then

$$\bar{Q}_B + \bar{Q}_A(\bar{Q}_A^T\bar{Q}_B) = \tilde{G} + (I-\bar{Q}_A\bar{Q}_A^T)F_B + (Q_{\tilde{A}}F_A^T + F_A Q_{\tilde{A}}^T)Q_{\tilde{B}} \;.$$

Neglecting second order quantities,

$$||\bar{G} - \tilde{G}||_2 \le ||F_B||_2 + 2||F_A||_2 + 2q^{1/2}2^{-t},$$

where the last term accounts for the final rounding of the elements in $\bar{M}$ and $\bar{G}$. Thus, if $\bar{Q}_A$ and $\bar{Q}_B$ are computed by HT, we have from (23)

(28) $$|\,\sigma_k(\bar{G}) - \sigma_k(G)\,| \le 13.2(q^{3/2} + 2p^{3/2})2^{-t} \;.$$

It follows that the singular values of the computed matrix $\bar{G}$ will differ little from $\sin\tilde{\theta}_k$, and thus small angles will be as accurately determined as is allowed by (25).

Since the matrix G is $m \times q$, computing the singular values of G will require about $2mq^2$ multiplications. If however, U_A and U_B are available we can obtain $\sin\theta_k$ accurately with fewer operations. We have

(29) $$(U_B - U_A C)^T (U_B - U_A C) = I + C^2 - 2C^2 = \mathrm{diag}(\sin^2\theta_k)$$

and

(30) $$(U_B - U_A)^T (U_B - U_A) = 2(I - C) \ .$$

From the last equation we can compute $2 \sin \frac{1}{2}\theta_k = (2(1 - \cos\theta_k))^{1/2}$, which since $0 \leq \frac{1}{2}\theta_k \leq \pi/4$ accurately determines both $\sin\theta_k$ and $\cos\theta_k$.

We finally remark about an appeerent imperfection of MGS. When A = B (exactly) we will get $\bar{Q}_A = \bar{Q}_B$. The exact angles equals zero, but since we only have the estimate

$$||I - \bar{Q}_A^T \bar{Q}_A||_2 \leq 2p(p+1)\,\kappa(A)2^{-t},$$

the singular values of $\bar{M} = \bar{Q}_A^T \bar{Q}_A$ may not be near one, which is the case if HT is used. However, since $\bar{M}$ is symmetric, SVD will give $Y_A \approx Y_B$ and therefore also $U_A \approx U_B$. It follows that if (30) is used, also MGS will yield angles which are near zero in this case. If however only $A \approx B$, then the rounding errors in computing Q_A and Q_B will not be correlated, and in an ill-conditioned case, we will probably not get angles near zero either with HT or MGS.

6. The singular case

We now consider the case when A and/or B does not have full column rank. In this case, the problem of computing principal angles and vectors is not well posed, since arbitrarily small perturbations in A and B will change the rank of A and/or B. The main computational difficulty then lies in assigning the correct rank to A and B. The most satisfactory way of doing this generally is the following [8]. Let the m x p matrix A have the SVD

$$A = Q_A D_A V_A^T , \quad D_A = \mathrm{diag}(\sigma_k(A)).$$

Let ε be a suitable tolerance and determine $p' \le p$ from

$$(31) \qquad \sum_{i=p'+1}^{n} \sigma_i^2(A) \le \varepsilon^2 < \sum_{i=p'}^{n} \sigma_i^2(A) .$$

We then approximate A with an m x p matrix A' such that rank $(A') = p'$,

$$A' = (Q'_A \; Q''_A) \begin{pmatrix} D'_A & 0 \\ 0 & 0 \end{pmatrix} (V'_A \; V''_A)^T, \quad D'_A = \mathrm{diag}(\sigma_1 , \; . \; . \; . \; . \; \sigma_{p'}),$$

where

$$Q_A = (Q'_A \; Q''_A), \qquad V_A = (V'_A \; V''_A)$$

have been partitioned consistently with the diagonal matrix. The matrix B is approximated in the same way.

If instead of (1) we assume that

$$p' = \mathrm{rank}(A') \ge \mathrm{rank}(B') = q' \ge 1,$$

then we can compute the principal angles and vectors associated with $R(A')$ and $R(B')$ by the previously derived algorithms, where now Q'_A and Q'_B should replace QA and Q_B.

In order to express the principal vectors of $R(A')$ as linear combinations of columns in A', we must solve the compatible system

$$A' X_A = U_A = Q'_A Y_A .$$

Since V''_A is an orthogonal basis for N(A), the general solution can be written

$$X_A = V'_A D'^{-1}_A Y_A + V''_A C_A ,$$

where C_A is an arbitrary matrix. It follows that by taking CA = 0 we get the unique solution which minimizes $||X_A||_F$, c.f. [14]. Thus we should take

$$(32) \qquad X_A = V'_A D'^{-1}_A Y_A, \; X_B = V'_B D'^{-1}_B Y_B ,$$

where X_A is $p \times p'$ and XB is $q \times q'$.

The approach taken above also has the advantage that only one decomposition, the SVD, is used throughout. It can, of course, also be used in the non-singular case. However, computing the SVD of A and B, requires much more work than computing the corresponding QR-decompositions. In order to make the QR-methods work also in the singular case, column pivoting must be used. This is usually done in such a way [2], [10] and [13] that the triangular matrix $R = (r_{ij})$ satisfies

$$|r_{kk}|^2 = \sum_{i=k}^{j} |r_{ij}|^2 \quad , \quad k < j \le n.$$

Such a triangular matrix is called normalized, and in particular the sequence $|r_{11}|, |r_{22}|, \ldots . |r_{pp}|$ is non-increasing. In practice it is often satisfactory to take the numerical rank of A to be p' if for a suitable tolerance ε we have

$$(33) \qquad |r_{p'p'}| > \varepsilon \ge |r_{p'+1,p'+1}| .$$

We then approximate $A = Q_A R_A$ by a matrix $A' = Q_A R'_A$ of rank p' by putting

$$r'_{ij} = r_{ij}, \quad i \leq p', \quad r'_{ij} = 0, \quad i > p'.$$

It has been shown in [20] how to obtain the solution (32) of minimum length from this decomposition.

If we use the criterion (33), there is a risk of choosing p' too large. Indeed, from the inequalities [10]

$$3(4^k+6k-1)^{-1/2}|r_{kk}| \leq \sigma_k(A) \leq (n+k+1)^{1/2}|r_{kk}|$$

it is seen that $\sigma_k(A)$ may be much smaller than $|r_{kk}|$.

7. Test results

Some of the algorithms in section. 5 have been tested on the UNIVAC 1108 of Lund University. Single precision floating point numbers are represented by a normalized 27 bit mantissa, whence the machine precision is equal to $2^{-26} \approx 1.5 \cdot 10^{-8}$.

We have taken F = R(A), where A is the m x p matrix

$$A = \frac{1}{\sqrt{k}} \begin{pmatrix} e & 0 \ldots 0 \\ 0 & e \ldots 0 \\ \ldots & \\ 0 & 0 \ldots e \end{pmatrix}, \quad e = \left.\begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix}\right\} m/p = k ,$$

and k is an integer. Thus, A is already orthogonal, and $Q_A = A$. Further, G = R(B) where B is the m x p Vandermonde matrix

$$B = \begin{pmatrix} 1 & x_0 \ldots x_0^{p-1} \\ 1 & x_1 \ldots x_1^{p-1} \\ \ldots & \\ 1 & x_{m-1} \ldots x_{m-1}^{p-1} \end{pmatrix}, \quad x_i = -1 + \frac{2i}{m+1} .$$

The condition number $\kappa(B)$ is known to grow exponentially with p, when the ratio m/p is kept constant. These matrices A and B are the ones appearing in [6]. There is exactly one vector, $u = (1,1,\ldots,1)^T$, which belongs to both F and G, so there will be one minimum angle $\theta = 0$.

For the tests, the matrix B was generated in single precision. The procedures for the QR-decompositions use column pivoting and are apart from minor details identical with procedures published in [21] and [22] . Inner products were not accumulated in double precision. For checking purposes, a three term recurrence relation [6] was used in double precision, to compute an exact single precision orthogonal basis for R(B).

For m/p = 2 and p = 5(2)17, Q_A was computed both by the method of

Householder and the modified Gram-Schmidt, method . Then cos Ok, YA and Y_B were computed by the procedure in [14] , and finally U_A and U_B from (9'). The results-are shown in table 1, where

$$m(\sigma_k) = \max_k |\sigma_k - \bar{\sigma}_k|, \quad F(U) = ||I - U^T U||_F .$$

Notice,that because of rounding Q_B to single precision and rounding errors in the computation of the SVD, σ_k are not exact to single precision.

For the Gram-Schmidt method, the predicted lack of orthogonality in U_B when $\kappa(B)$ is large, is evident. However, there is no significant difference in the accuracy of cos $\bar{\theta}_k$ between the two methods. In table 2 we show for m = **26** and p = **13** the errors in cos $\bar{\theta}_k$ for each k.

For the same values of m and p, sin θ_k were computed from the singular values of both the matrix $(I-P_A)Q_B$ and the matrix $(I-P_B)Q_A$. The results in table 3 again show no significant difference between the two methods. For the Gram-Schmidt method, the values of sin θ_k differ somewhat between the two matrices, whereas the corresponding values for the Householder method are almost identical. This is confirmed by table 4, where, again for m = **26**, p = 13, results for each k are shown.

The authors are very pleased to acknowledge the help of Mr. Jan Svensson, who carried out the tests described in this section.

Table 1

		Househdlder			Gram-Schmidt		
m	P	$F(\bar{U}_A)\cdot 10^8$	$F(\bar{U}_B)\cdot 10^8$	$m(\cos\ \bar{\theta}_k)\cdot 10^8$	$F(\bar{U}_A)\cdot 10^8$	$F(\bar{U}_B)\cdot 10^8$	$m(\cos\ \bar{\theta}_k)\cdot 10^8$
10	5	11	15	4	15	12	10
14	7	27	35	10	24	76	12
18	9	37	28	26	33	202	21
22	11	30	46	40	47	2412	91
26	13	43	51	612	38	12129	913
30	15	57	63	1874	51	28602	1484
34	17	51	65	13051	56	344685	5417

Table 2

m = 26 p = 13

	Householder		Gram-Schmidt	
k	cos $\bar{\theta}_k$	$\Delta\cos\theta_k \cdot 10^8$	cos $\bar{\theta}_k$	$\Delta\cos\bar{\theta}_k \cdot 10^8$
1	0.99999979	2	0.99999989	12
2	0.99823279	8	0.99823304	25
3	0.99814388	- 33	0.99815032	613
4	0.99032719	15	0.99031791	- 913
5	o. 98988868	12	0.98989530	674
6	0.97646035	- 47	0.97646120	38
7	0.96284652	51	0.96284428	- 173
8	0.94148868	- 33	0.94148907	6
9	0.91758598	8	0.91758703	97
10	0.87013517	- 186	0.87013374	- 329
11	0.76366349	612	0.76365566	- 171
12	0.06078814	1	0.06078782	- 33
13	0.01558465	- 60	0.01558528	3

Table 3

		Householder[1]		Gram-Schmidt[1]	
m	p	$m(\sin \bar{\theta}_k)\cdot 10^8$	$m(\sin \tilde{\theta}_k)\cdot 10^8$	$m(\sin \bar{\theta}_k)\cdot 10^8$	$m(\sin \tilde{\theta}_k)\cdot 10^8$
10	5	3	2	4	3
14	7	16	7	27	4
18	9	51	49	48	6
22	11	68	68	135	97
26	13	704	709	390	288
30	15	2367	2358	1173	1140
34	17	--. 16285	16281	5828	4501

[1] sin $\bar{\theta}_k$ computed as $\sigma_k((I-P_A)Q_B)$, sin $\tilde{\theta}_k$ as $\sigma_k((I-P_B)Q_A)$

Table 4

m = 26 p = 13

	Householder[1]			Gram-Schmidt[1]		
k	sin $\bar{\theta}_k$	Δsin $\bar{\theta}_k$	Δsin $\tilde{\theta}_k$	sin $\bar{\theta}_k$	Δsin $\bar{\theta}_k$	Δsin $\tilde{\theta}_k$
1	0.00000002	0	3	0.00000002	0	1
2	0.05942237	- 24	- 24	0.05942257	4	5
3	0.06089812	129	129	o. 06089789	106	67
4	0.13875079	- 97	- 97	0.13875077	- 99	30
5	0.14184525	- 183	- 181	o. 14184804	96	
6	0.21569622	190	190	0.21569423	9	- 28
7	0.27004868	- 171	- 173	0.27004985	- 54	5
8	0.33704409	108	109	0.33704250	- 51	- 41
9	0.39753688	17	21	0.39753668	3	- 37
10	0.49281275	344	343	0.49280659	- 272	- 70
11	0.64561398	- 704	- 709	0.64562460	358	288
12	0.99815045	78	3	0.99814761	- 206	3
13	0.99987832	90	6	0.99988132	390	0

[1] sin $\bar{\theta}_k$ computed as $\sigma_k((I-P_A)Q_B)$, sin $\tilde{\theta}_k$ as $\sigma_k((I-P_B)Q_A)$

REFERENCES

1. S.N. Afriat, Orthogonal and oblique projectors and the characteristics of pairs of vector spaces, Proc. Camb. Phil. Soc. 53, 800-816 (1957)

2. Å. Björck, Solving linear least squares problems by Gram-Schmidt orthogonalization, BIT 7, 1-21 (1967)

3. Å Björck and C. Bowie, An iterative algorithm for computing the best estimate of an orthogonal matrix, SIAM J. Numer. Anal. Vol. 8, No 2

4. C. Cohen and A. Ben-Israel, On the computation of canonical correlations, Cahiers Centre Etudes Recherche Opér. 11, 121-132 (1969)

5. C. Cohen, An investigation of the geometry of subspaces for some multivariate statistical models, Thesis Dept. of Indust. Eng., Univ. of Illinois (1969)

6. G. Dahlquist, B. Sjöberg and P. Svensson, Comparison of the method of averages with the method of least squares, Math. Comp. 22, 833-846 (1968)

7. C. Davies and W. Kahan, The rotation of eigenvectors by a perturbation III, SIAM J. Numer. Anal. 7, 1-46 (1970)

8. C. Eckart and G. Young, The approximation of one matrix by another of lower rank, Psychometrika, 1, 211-218 (1936)

9. C. Eckart and G. Young, A principal axis transformation for non-hermitian matrices, Bull. Amer. Math. Soc. 45, 118-121 (1939)

10. D.K. Faddev, V.N. Kublanovskaya and V.N. Faddeeva, Sur les systèmes lineaires algebriques de matrices rectangulaires et mal-conditionnées, Colloq. Int. du C.N.R.S. Besancon 1966, no. 165, 161-170, Paris (1968)

11. G. Fix and R. Heiberger, An algorithm for the illconditioned generalized eigenvalue problem, Numer. Math. to appear

12. I.C. Gohberg and M.G. Krein, Introduction to the theory of linear nonselfadjoint operators, Moscow (1965), Transl. AMS Vol. 18, Providence, Rhode Island (1969)

13. G.H. Golub, Numerical methods for solving linear least squares problems, Numer. Math. 7, 206-216 (1965)

14. G.H. Golub and C. Reinsch, Singular value decompositions and least squares solutions, Numer. Math. **14**, 403-420 (1970)

15. I.J. Good, An essay on modern Bayesian methods, pp. 87-89, M.I.T. Press, (1965)

16. H. Hotelling, Relations between two sets of variates, Biometrika **28**, **321-377** (1936)

17. A. van der Sluis, Condition numbers and equilibration of matrices, Numer. Math. 14, 14-23(1969)

18. J.M. Varah, Computing invariant subspaces of a general matrix when the eigensystem is poorly conditioned, Math. of Comp. 24, 137-149 (1970)

19. J. Wilkinson, Error analysis of; transformations based on the use of matrices of the form I-2ww , Error in digital computation, Vol. II, L.B. Rall ed., 77-101, New York, John Wiley & Sons, (1965).

20. R. J. Hanson and C. L. Lawson, Extensions and applications of the Householder algorithm for solving linear least squares problems, Math. Comp. **23**, **787-812** (1969)

21. Å. Björck, Iterative refinement of linear least squares solutions iI, BIT 8, 8-30(1968)

22. P. Businger and G.H. Golub, Handbook series linear algebra. Linear least squares solutions by Householder transformations, Numer. Math. **7**, 269-276(1965)