NUMERICAL ANALYSIS PROJECT
MANUSCRIPT NA-92-13

SEPTEMBER 1992

The Canonical Correlations of Matrix Pairs and Their Numerical Computation

by

Gene H. Golub
and
Hongyuan Zha

NUMERICAL ANALYSIS PROJECT
COMPUTER SCIENCE DEPARTMENT
STANFORD UNIVERSITY
STANFORD, CALIFORNIA 94305

THE CANONICAL CORRELATIONS OF MATRIX PAIRS AND THEIR NUMERICAL COMPUTATION

GENE H.GOLUB AND HONGYUAN ZHA

ABSTRACT. **This paper is concerned with the analysis of canonical correlations of matrix pairs and their numerical computation. We first develop a decomposition theorem for matrix pairs having the same number of rows which explicitly exhibits the canonical correlations. We then present a perturbation analysis of the canonical correlations, which compares favorably with the classical first order perturbation analysis. Then we propose several numerical algorithms for computing the canonical correlations of general matrix pairs; emphasis is placed on the case of large sparse or structured matrices.**

1. INTRODUCTION

Given two vectors $u \in \mathcal{R}^n$ and $v \in \mathcal{R}^n$, a natural way to measure the closeness of two ***one*** dimensional linear subspaces spanned by u and v respectively, is to consider the acute angle formed by the two vectors, the cosine of which is given by

$$\sigma(u, v) := \frac{|u^T v|}{\|u\|_2 \|v\|_2}.$$

We observe that $\sigma(u, v)$ = 0, when u and v are orthogonal to each other; and $\sigma(u,$ v) $=$ 1, when the two linear subspaces are identical. Given two linear subpaces that are spanned by the columns of matrices $A \in \mathcal{R}^{m\times n}$ and $B \in \mathcal{R}^{m\times l}$, we are concerned with the problem of how to measure the closeness of span(A) and span(B), the range spaces of A and B. One natural extension of the one dimensional case is to choose a vector from span(A), i.e., a linear combination of the columns of A, say Ax, and similarly By from span{ B}, and form $\sigma(By, Az)$. The closeness of span(A) and $\mathrm{span}\{B\}$ can be measured by the following

$$d(A, B) = \min_{x\in\mathcal{R}^n,\ y\in\mathcal{R}^l} \sigma(By, Ax).$$

However, the two linear subspaces or rather the matrix pair ***(A,*** B) have more structure to reveal than that defined by the minimum. In 1936, Hotelling proposed to recursively define a sequence of quantities which is now called ***canonical correlations*** of a matrix pair ***(A, B)*** [8].

1991 *Mathematics Subject Classification.* **primary 15A18, 15A21, 65F15; secondary 62H20.**

Key words and phrases. **canonical correlation, singular value decomposition, perturbation analysis, large sparse matrix, structured matrix.**

The work of the first author was supported in part by NSF grant DRC-8412314 and Army contract DAAL-03-90-G-0105.

The work of the second author was supported in part by Army contract DAAL-03-90-G-0105.

Definition 1.1. Let $A \in \mathcal{R}^{m\times n}$ and $B \in \mathcal{R}^{m\times l}$, and assume that

$$p = \text{rank(A)} \geq \text{rank(B)} = q.$$

The canonical correlations $\sigma_1(A, B), \cdots, \sigma_q(A, B)$ of the matrix pair (A, B) are defined recursively by the formulae

$$(1.1)\quad \sigma_k(A, B) = \max_{\substack{Ax\neq 0,\ By\neq 0,\\ Ax\perp\{Ax_1,\cdots,Ax_{k-1}\},\\ By\perp\{By_1,\cdots,By_{k-1}\}.}} \sigma(By, Ax) =: \sigma(By_k, Ax_k), \quad k = 1, \cdots, q.$$

It is readily seen that

$$\sigma_1(A, B) \geq \cdots \geq \sigma_q(A, B),$$

and

$$d(A, B) = \sigma_q(A, B).$$

The unit vectors

$$Ax_i/\|Ax_i\|_2,\ By_i/\|By_i\|_2,\ (i = 1, \cdots, q)$$

in (1.1) are called the canonical vectors of (A, B); and

$$x_i/\|Ax_i,\ y_i/\|By_i\|_2,\ i = 1, \cdots, q$$

are called the canonical weights. Sometimes the angles $\theta_k \in [0, r/2]$ satisfying $\cos\theta_k = \sigma_k(A, B)$ are called the principal angles between span(A) and span$\{B\}$ [7].[1] The basis of span(A) or span$\{B\}$ that consists of the canonical vectors are called the canonical basis.

There are various ways of formulating the canonical correlations, which are all equivalent. They shed insights on the problem from different perspectives, and as we will see later, some of the formulations are more suitable for numerical computation than others. The applications of the canonical correlations are enormous such as system identification, information retrieval, statistics, econometrics, psychology, educational research, anthropology and botany [1] [17] [9]. There are also many variants **and generalizations** of the canonical correlations: to the case of more than two matrices (surveyed by Kettenring [11], see also [17]); to sets of random functions [2]; to nonlinear transformations [17]; and to problems with (in)equality constraints. Several numerical algorithms have been proposed for the computation of the canonical correlations and the corresponding canonical vectors (see Björck and Golub's paper [4] and references therein); however, in the literature there is very little discussion of the case of large sparse and structured matrix pairs, which will receive a fairly detailed treatment in Section 4.

The organization of the paper is as follows: in Section 2, we present several different formulations of the canonical correlations; in Section 3, we develop a decomposition theorem for general matrix pairs having the same number of rows: this decomposition not only explicitly exhibits the canonical correlations of the matrix pair, it also reveals some of its other intrinsic structures. We also discuss the relation between the canonical correlations and the corresponding eigenvalue problem

[1] As is pointed by G.W. Stewart [15], the concept of canonical angles between two linear subspaces is much older than canonical correlations, and can be traced back to C. Jordan [10, p.129 Equation(60)].

and the RSVD [20]. In Section 4, we present perturbation analyses of the canonical correlations; the results compare favorably with the classical first order counterpart developed in [4]. We derive perturbation bounds for the normwise as well as componentwise perturbations. In Section 5, we propose several numerical algorithms for computing the canonical correlations. For the case of dense matrices, we also discuss the updating problem. The emphasis of the section is placed on the case of large sparse or structured matrix pairs. We will first present an algorithm using alternating linear least squares approach which has a nice geometric interpretation. We also relate this algorithm to a modified power method and derive its convergence rate. Then we adapt the Lanczos bidiagonalization process to compute a few of the largest canonical correlations. Our algorithms have the attractive feature that it is not necessary to compute the orthonormal basis of the column space of A and B as is used in Björck-Golub's algorithm, and thus one can fully take advantage of the sparsity or special structures (e.g, Hankel or Toeplitz structures) of the underlying matrices. Numerical examples will also be given to illustrate the algorithms.

2. Several Different Formulations

There are quite a few different ways of defining and formulating canonical correlations: Hotelling's original derivation is based on matrix algebra and analysis [8]; Rao and Yanai used the theory of orthogonal projectors [14]; Escoufier proposed a general frame work for handling data matrix by matrix operators, which also includes the canonical correlations as a special case [6]; Björck and Golub used matrix decomposition of the given data matrices [4]. In this section, we give some of the formulations and indicate their equivalence.

The Singular Value Decomposition (SVD) Formulation. Let the QR decomposition of A and B be

$$A = Q_A R_A, \quad B = Q_B R_B,$$

where Q_A and Q_B are orthonormal matrices, and R_A and R_B are nonsingular upper triangular matrices, then

$$\sigma(By, Ax) = \frac{y^T B^T A x}{\|By\|_2 \|Ax\|_2} = \frac{y^T R_B^T Q_B^T Q_A R_A x}{\|R_B y\|_2 \|R_A x\|_2} =: v^T Q_B^T Q_A u,$$

where we have designated $u = R_A x / \|R_A x\|_2$ and v = $R_B y / \|R_B y\|_2$. Using a characterization of the SVD [7, p. 428], we see that the canonical correlations are the singular values of $Q_B^T Q_A$, and if

$$Q_B^T Q_A = P^T \operatorname{diag}(\sigma_1(A, B), \cdots, \sigma_q(A, B)) Q$$

denotes the SVD of $Q_B^T Q_A$, then

$$Q_A P(:, 1:q) = [u_1, \cdots, u_q], \quad \text{and } Q_B Q = [v_1, \cdots, v_q]$$

give the canonical vectors of (A, B). Note that since $Q_B^T Q_A$ is a section of an orthogonal matrix, $\sigma_k(A, B) \leq 1$, $k = 1, \cdots, q$. We also note that the canonical vectors are not unique if, say $\sigma_k (A, B) = \sigma_{k+1}(A, B)$. However, the above formulation is rather general in the sense that it can also handle the case when A and/or B are rank deficient.

A Trace Mazimiration Formulation. Let us consider the following maximization problem:

$$\max_{\substack{L^T B^T BL = I_p \\ M^T A^T AM = I_p}} \operatorname{trace}(L^T B^T AM), \tag{2.2}$$

where for simplicity we have further assumed that $p = q$; otherwise we can append zero columns to B to make the pair (A, B) satisfy this assumption. Again using the QR decomposition of A and B, we see that the two equality constraints in (2.2) imply that $R_A M$ and $R_B L$ are orthogonal matrices, and we arrive at the following equivalent maximization problem

$$\max_{U \text{ and } V \text{ are orthogonal}} \operatorname{trace}(U^T (Q_B^T Q_A) V). \tag{2.3}$$

To this end, we cite a well-known result of Von Neumann [16].

Lemma 2.1. *Let the singular values of A and B be*

$$\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_n \text{ and } \tau_1 \geq \tau_2 \geq \cdots \geq \tau_n.$$

Then

$$\max_{U \text{ and } V \text{ are orthogonal}} \operatorname{trace}(BU^T AV) = \Sigma_i \sigma_i \tau_i.$$

The above problem (2.3) is a special case of the lemma by choosing $B = I$ and $A = Q_B^T Q_A$.

Remark 2.1. Since $L^T B^T BL = I_p$, $M^T A^T AM = I_p$, the maximization problem (2.2) is equivalent to the following minimization problem:

$$\min_{\substack{L^T B^T BL = I_p \\ M^T A^T AM = I_p}} \|AM - BL\|_F, \tag{2.4}$$

which can be interpreted as finding an orthonormal basis of span(A) and span{ B} respectively, such that their difference measured in the Frobenius norm are minimized. It is equivalent to the following orthogonal ... ***ocrustes problem.*** Let Q_A and Q_B be any orthonormal basis of span{ A} and span{ B}, respectively. Then (2.2) is equivalent to

$$\min_{U \text{ is orthogonal}} \|Q_A - Q_B U\|_F.$$

We note that the above is a special Procrustes problem where Q_A and Q_B are orthonormal, while in the general case, Q_A and Q_B can be replaced by two general matrices [7, Section 12.4.1].

A Lagrange Multiplier Formulation [S]. For the constrained minimization problem (1.1), write the Lagrange multiplier function

$$f(x, y, \lambda, \mu) = y^T B^T Ax - \lambda(\|Ax\|_2^2 - 1) - \mu(\|By\|_2^2 - 1).$$

Differentiating with respect to x, y, λ, and μ leads to:

$$\begin{aligned} B^T Ax - \mu B^T By &= 0, \\ A^T By - \lambda A^T Ax &= 0, \\ y^T B^T By &= 1, \\ x^T A^T Ax &= 1. \end{aligned} \tag{2.5}$$

It follows that $\lambda = \mu$ and

$$\begin{pmatrix} O & B^T A \\ A^T B & O \end{pmatrix} \begin{pmatrix} y \\ x \end{pmatrix} = \lambda \begin{pmatrix} B^T B & O \\ O & A^T A \end{pmatrix} \begin{pmatrix} y \\ x \end{pmatrix}.$$

Therefore finding the canonical correlations, which are the stationary values, corresponds to solving for the eigenvalues of the above generalized eigenvalue problem. On the other hand, since

$$\| \frac{Ax}{\|Ax\|_2} - \frac{By}{\|By\|_2} \|_2^2 = 2(1 - \frac{y^T B^T Ax}{\|By\|_2 \|Ax\|_2}),$$

the first canonical correlation can also be computed by solving the minimization problem

$$\min_{Ax,\, By \neq 0} \| \frac{Ax}{\|Ax\|_2} - \frac{By}{\|By\|_2} \|_2.$$

One way of solving the minimization problem is to first fix y, and find the optimal x; then fix x at this optimal value, and then solve for y and so on. At each iteration step, we can reformulate the problem as

$$\min_{\text{subject to } \|Az\|_2 = 1} \|w - Az\|_2,$$

where w is of unit length. Using the Lagrange multiplier method, we seek to minimize

$$f(z, \lambda) = \|w - Az\|_2^2 + \lambda(\|Az\|_2^2 - 1).$$

Writing down the first order condition for the stationary values, we obtain

$$A^T Az = A^T w/(1 + \lambda), \; z^T A^T Az = 1;$$

and the solution is given by

$$\begin{aligned} \lambda &= (A^T w)^T (A^T A)^{-1} (A^T w) - 1 \\ &= w^T P_A w - 1, \end{aligned} \qquad \begin{aligned} z &= (A^T A)^{-1} A^T w/(1 + \lambda) \\ &= A^\dagger w/(1 + \lambda), \end{aligned}$$

where $P_A^2 = P_A$ is the orthogonal projection onto span(A). We note that z is in the direction of $A^\dagger w$, and is the least squares solution of

$$\min_x \|w - Az\|_2.$$

Actually, this approach will lead to the alternating least squares (ALS) method that we will discuss in Section 5.

3. A DECOMPOSITION THEOREM

It is readily checked from Definition 1.1 that the canonical correlations are invariant under the following group transformation

$$A \longrightarrow QAX_A, \quad B \longrightarrow QBX_B,$$

where Q is orthogonal and X_A and X_B are nonsingular. The following theorem gives the maximum invariants of a matrix pair (A, B) under the above group transformation. It also provides information on other structures of the matrix pair as well. It can be considered as a recast of Theorem 5.2 in [16, pp. 40-42] (cf. [4, Equation (15)] [18, Equation (2.2)]).

Theorem 3.1. *Let $A \in R^{m\times n}$ and $B \in R^{m\times l}$, and assume that*

$$p = \text{rank(A)} \geq \text{rank(B)} = q.$$

Then there exists orthogonal matrix Q and nonsingular matrices X_A and X_B such that

$$A = Q[\Sigma_A, O]X_A, \quad B = Q[\Sigma_B, O]X_B,$$

where $\Sigma_A \in R^{m\times p}$ and $\Sigma_B \in R^{m\times q}$ are of the following form

$$\Sigma_A = \begin{pmatrix} I_i & & \\ & C & O \\ \hline O & & \\ & S & \\ & & I_k \end{pmatrix}, \quad \Sigma_B = \begin{pmatrix} I_q \\ \hline O \end{pmatrix}, \tag{3.6}$$

with

$$\begin{aligned} &C = \operatorname{diag}(\alpha_{i+1} \cdots \alpha_{i+j}),\ 1 > \alpha_{i+1} \geq \cdots \geq \alpha_{i+j} > 0, \\ &S = \operatorname{diag}(\beta_{i+1}, \cdots, \beta_{i+j}),\ 0 < \beta_{i+1} \leq \cdots \leq \beta_{i+j} < 1, \\ &\alpha_{i+1}^2 + \beta_{i+1}^2 = 1, \cdots, \alpha_{i+j}^2 + \beta_{i+j}^2 = 1, \end{aligned} \tag{3.7}$$

and $p = i + j + k$. The canonical correlations of (A, B) are the diagonal elements of $\Sigma := \operatorname{diag}(I_i, C, 0)$. Moreover, we have

$$\begin{aligned} i &= \text{rank(A)} + \text{rank(B)} - \operatorname{rank}([A, B]), \\ j &= \operatorname{rank}([A, B]) + \operatorname{rank}(B^T A) - \text{rank(A)} - \text{rank(B)} \\ k &= \text{rank(A)} - \operatorname{rank}(B^T A). \end{aligned} \tag{3.8}$$

Proof. Using the QR decomposition, we can transform A and B to

$$A = [Q_A, O]R_A, \quad B = [Q_B, O]R_B$$

where $Q_A \in R^{m\times p}$ and $Q_B \in R^{m\times q}$ are orthonormal, and R_A and R_B are nonsingular. We then find U orthogonal such that

$$Q_B = U \begin{pmatrix} I_q \\ \hline O \end{pmatrix}.$$

We partition $U^T Q_A$ as $U^T Q_A = [A_1^T, A_2^T]^T$ with $A_1 \in R^{q \times q}$. Let the SVD of A_1 be

$$A_1 = U_1 \begin{pmatrix} I_i & & \\ & C & \\ & & 0 \end{pmatrix} V_2^T;$$

C defined as in (3.7), and $A_2 V_2 = [A_{11}, A_{12}, A_{13}]$ be partitioned compatibly. [2] Then we have $A_{11} = 0$ and A_{13} is orthonormal and can be written as $A_{13} = U_2[O, I_k]^T$ with U_2 an orthogonal matrix. Hence $U_2^T A_{12} = [\tilde{A}_{12}^T, O]^T$, with its last k rows equal zero. Since the columns of A_{12} are orthogonal to each other, we can find an orthogonal matrix U_3 such that $\tilde{A}_{12} = U_3[O, S]^T$. The relations in (3.7) follow from the fact that $U^T Q_A$ is orthonormal. Accumulating all the transformations establishes the decomposition (3.6). For the rank expressions of the integer indices we observe that

rank(A) = i + j + k, rank(B) = q, $\mathbf{rank}(B^T A)$ = i + j, $\mathbf{rank}([A, B]) = j + k + q$.

Some elementary calculation leads to the result (3.8). ❑

Remark 3.1. The dimension of $\mathcal{R}(A) \cap \mathcal{R}(B)$ is exactly the number of those canonical correlations of *(A, B)* which are equal to one.

Corollary 3.1. *Let* $Q = (Q_1, Q_2, Q_3, Q_4, Q_5, Q_6)$ *be compatibly partitioned with the block* row partitioning *of* Σ_A, i.e., $Q_1 \in R^{m \times i}$, $Q_2 \in R^{m \times j}$ *and so on. Then*

$\mathbf{span}\{Q_1, Q_2 C + Q_5 S, Q_6\} = \mathcal{R}(A)$;
$\mathbf{span}\{Q_1, Q_2, Q_3\} = \mathcal{R}(B)$;
$\mathbf{span}\{Q_3, -Q_2 S + Q_5 C, Q_4\} = \mathcal{R}(A)^\perp$;
$\mathbf{span}\{Q_4, Q_5, Q_6\} = \mathcal{R}(B)^\perp$;
$\mathbf{span}\{Q_1\} = \mathcal{R}(A) \cap \mathcal{R}(B)$;
$\mathbf{span}\{Q_3\} = \mathcal{R}(A)^\perp \cap \mathcal{R}(B)$;
$\mathbf{span}\{Q_4\} = \mathcal{R}(A)^\perp \cap \mathcal{R}(B)^\perp$;
$\mathbf{span}\{Q_6\} = \mathcal{R}(A) \cap \mathcal{R}(B)^\perp$.

Proof. We prove $\mathbf{span}\{Q_3\} = \mathcal{R}(A)^\perp \cap$ *'R(B);* the other formulae can be similarly established. It is easy to see that $\mathbf{span}\{Q_3\} \subseteq \mathcal{R}(A)^\perp \cap \mathcal{R}(B)$; *However*

$$G := (Q_1, Q_2, Q_3)^T (Q_3, -Q_2 S + Q_5 C, Q_4) = \begin{pmatrix} O & O & O \\ O & -S & O \\ I_k & O & O \end{pmatrix}.$$

It follows that the number of singular values of G that are equal to one is exactly the column dimension of Q_3; the result follows from the comment in Remark 3.1. ❑

Corollary 3.2. *We also have the following expressions for the dimensions of some of the linear subspaces in Corollary* 3.1:

$\dim(\mathcal{R}(A) \cap \mathcal{R}(B))$ = rank(A) + rank(B) $-$ $\mathbf{rank}([A, B])$;
$\dim(\mathcal{R}(A)^\perp \cap \mathcal{R}(B)) =$ rank(B) $-$ $\mathbf{rank}(B^T A)$;
$\dim(\mathcal{R}(A)^\perp \cap \mathcal{R}(B)^\perp) = m -$ rank($[A, B]$);
$\dim(\mathcal{R}(A) \cap \mathcal{R}(B)^\perp) =$ rank(A) $-$ $\mathbf{rank}([A, B])$.

[2] Since A_1 is a section of an orthogonal matrix, all its singular values are less than or equal to one.

Proof. All the expressions can be proved by using the above corollary and the rank formulae in Theorem 3.1. □

Corollary 3.3. ***The finite and zero eigenvalues of***

$$\begin{pmatrix} O & B^TA \\ A^TB & O \end{pmatrix} x = \lambda \begin{pmatrix} B^TB & O \\ O & A^TA \end{pmatrix} x$$

are $\pm\sigma_1(A, B), \cdots, \pm\sigma_q(A, B)$. ***And if*** $B^TA = I$, ***then the nonzero singular values of the matrix product*** AB^T ***are*** $1/\sigma_1(A, B), \ldots, 1/\sigma_q(A, B)$.[3]

Proof. The result can be proved by using the decomposition in Theorem 3.1 and direct computation. It follows that the canonical correlations can also be found by the RSVD of the matrix triplet (B^TA, B^T, A) [20, p. 193]; if $B^TA = I$, the RSVD of (B^TA, B^T, A) reduces to the PSVD of (B^T, A) [20, corollary 4.2]. □

4. PERTURBATION ANALYSES

In this section we establish some perturbation bounds for the canonical correlations; some of the techniques used here are first devised by Paige in his analysis of the generalized singular value decomposition [12]. We also mention that Björck and Golub developed a first order perturbation analysis in their paper[4]. Before we discuss the general case, let us first consider a simple example:

Example 4.1. ***We consider the matrix pair:***

$$A = (\ ;\ ),\quad B = \begin{pmatrix} 1 & 1 \\ 0 & \epsilon \\ 1 & 1 \end{pmatrix},$$

where ϵ ***is a small quantity. Since*** $A = (B(:, 2) - B(:, 1))/\epsilon$, ***hence*** $\sigma(A, B) = 1$. ***But if we perturb B to***

$$\tilde{B} = \begin{pmatrix} 1 & 1-\epsilon \\ 0 & 0 \\ 1 & 1 \end{pmatrix} = B - \begin{pmatrix} 0 & \epsilon \\ 0 & \epsilon \\ 0 & 0 \end{pmatrix},$$

since A is orthogonal to the columns of $\tilde{B}$, ***we have*** $\sigma(A, \tilde{B}) = 0$. ***Therefore a small change in the matrix pair (A, B) causes a large change of its canonical correlations. We note that B and*** $\tilde{B}$ ***are of the same rank.***

We observe that $\mathrm{cond}(B) \approx 1/\epsilon$. This example suggests that the canonical correlations are sensitive to perturbations if the condition number of A or B is large. Now we turn to the discussion of the general case. Using the QR decomposition with column pivoting, A and B can be factorized as

$$A = Q_A R_A, \quad B = Q_B R_B$$

where Q_A and Q_B are orthonormal, and R_A and R_B are of full row rank. The canonical correlations are simply the singular values of $Q_A^TQ_B$ (cf. [4]). Let the SVD of $Q_A^TQ_B$ be

$$Q_A^TQ_B = U\Sigma V^T.$$

[3] The first result is proved in [8], and the second is also implicit in [20].

We denote the perturbed quantities by adding-" ~" to the corresponding unperturbed ones. We ***assume*** that A and A, B and B are ***of the same rank.***

Let the orthogonal complement of Q_A and Q_B be denoted by $\hat{Q}_A$ and $\hat{Q}_B$ respectively. We define

$$\delta_2(A) = \min_{U \text{ is orthogonal}} \|Q_A - Q_{\tilde{A}}U\|_2,$$

$$\delta_F(A) = \min_{U \text{ is orthogonal}} \|Q_A - Q_{\tilde{A}}U\|_F.$$

We note that $\delta_F(A)$ is introduced in [12], and actually it is a special case of the following well known ***Procrustes problem*** [7]:

$$\min_{Q \text{ is orthogonal}} \|A - BQ\|_F,$$

where A and B are arbitrary matrices with same number of columns and rows. The solution of the Procrustes problem can be obtained using the SVD of B^TA: let $B^TA = U\Sigma V^T$, then the optimal Q is given by $Q = UV^T$ [7]. For $\delta_F(A)$, some interesting relations can be derived by invoking the CS-decomposition of $(Q_A, \hat{Q}_A)^T\hat{Q}_A$ [12] [16]:

$$\begin{pmatrix} Q_A^TQ_{\tilde{A}} \\ \hat{Q}_A^TQ_{\tilde{A}} \end{pmatrix} = \begin{pmatrix} U_AC_AW_A^T \\ V_AS_AW_A^T \end{pmatrix},$$

where U_A, V_A and W_A are orthogonal, and C_A and S_A are quasi-diagonal. Then we have [12]:

$$\begin{aligned} \delta_F(A)^2 &= 2\Sigma_i (1-\sigma_i) \le 2\Sigma_i (1-\sigma_i^2) \\ &= 2\|S\|_F^2 = 2\|\hat{Q}_A^TQ_{\tilde{A}}\|_F^2 \\ &= 2\Sigma_i (1-\sigma_1)(1+\sigma_1) \le 2\delta_F(A)^2. \end{aligned} \tag{4.9}$$

where we used $C_A = \mathrm{diag}(\sigma_1, \ldots, \sigma_q)$ with $\sigma_1 \le \cdots \le \sigma_q$.

For the $\delta_2(A)$, we proceed as follows:

$$\begin{aligned} \|Q_A - Q_{\tilde{A}}U\|_2^2 &= \sigma_{\max}(2I + Q_A^TQ_{\tilde{A}}(-U) + (Q_A^TQ_{\tilde{A}}(-U))^T) \\ &\ge 2 - \sigma_{\max}(Q_A^TQ_{\tilde{A}}(-U) + (Q_A^TQ_{\tilde{A}}(-U))^T) \\ &\ge 2 - 2\sigma_{\max}(Q_A^TQ_{\tilde{A}}(-U)) \\ &= 2 - 2\sigma_{\max}(Q_A^TQ_{\tilde{A}}) \\ &= 2(1-\sigma_q) \end{aligned} \tag{4.10}$$

hence $\delta_2(A) \ge \sqrt{2(1-\sigma_q)}$. However, by choosing the canonical basis in the CS-decomposition we have

$$\begin{aligned} \delta_2(A) &\le \left\| \begin{pmatrix} I \\ O \\ O \end{pmatrix} - \begin{pmatrix} C_A \\ S_A \\ O \end{pmatrix} \right\|_2 \\ &= \sqrt{\lambda_{\max}((I-C_A)^2 + S_A^2)} \\ &= \sqrt{2(1-\sigma_1)}. \end{aligned} \tag{4.11}$$

Therefore we ontain

$$\sqrt{2(1-\sigma_q)} \le \delta_2(A) \le \sqrt{2(1-\sigma_1)}.$$

It follows that

$$\begin{aligned} \delta_2(A)^2 &\le 2(1-\sigma_1) \le 2(1-\sigma_1^2) \\ &= 2\|S\|_2^2 = 2\|\hat{Q}_A^T Q_{\tilde{A}}\|_2^2. \end{aligned} \tag{4.12}$$

We should remark here that, there is generally no closed form solution to the following ***Procrustes*** problem:

$$\min_{Q \text{ is orthogonal}} \|A - BQ\|_2,$$

when A and B are orthonormal. With the above preparation, we are now ready to prove the following perturbation theorems.

Theorem 4.1. *Let A and $\tilde{A}$, and B and $\tilde{B}$ have the same rank, and let the condition numbers of A and B be*

$$\kappa(A) = \|A\|_2\|A^\dagger\|_2,\ \kappa(B) = \|B\|_2\|B^\dagger\|_2$$

then

$$\|\Sigma - \tilde{\Sigma}\|_2 \le \sqrt{2}\left\{\kappa(A)\frac{\|A-\tilde{A}\|_2}{\|A\|_2} + \kappa(B)\frac{\|B-\tilde{B}\|_2}{\|B\|_2}\right\}.$$

Proof. For any orthogonal matrices U and V, we have

$$\begin{aligned} \|\Sigma - \tilde{\Sigma}\|_2 &\le \|Q_A^T Q_B - U^T Q_{\tilde{A}}^T Q_{\tilde{B}} V\|_2 \\ &= \|(Q_A - Q_{\tilde{A}}U)^T Q_B + U^T Q_{\tilde{A}}^T (Q_B - Q_{\tilde{B}}V)\|_2. \\ &\le \|Q_A - Q_{\tilde{A}}U\|_2 + \|Q_B - Q_{\tilde{B}}V\|_2, \end{aligned} \tag{4.13}$$

Since U and V are arbitrary, we obtain

$$\begin{aligned} \|\Sigma - \tilde{\Sigma}\|_2 &\le \delta_2(A) + \delta_2(B) \\ &\le \sqrt{2}(\|\hat{Q}_A^T Q_{\tilde{A}}\|_2 + \|\hat{Q}_B^T Q_{\tilde{B}}\|_2). \end{aligned} \tag{4.14}$$

Let $\tilde{A} = A + \mathrm{AA}$, then

$$\hat{Q}_{\tilde{A}}^T \Delta A = -\hat{Q}_{\tilde{A}}^T Q_A R_A,\ \hat{Q}_A^T \Delta A = \hat{Q}_A^T Q_{\tilde{A}} R_{\tilde{A}}. \tag{4.15}$$

Since R_A and $R_{\tilde{A}}$ are of full row rank:

$$\hat{Q}_{\tilde{A}}^T Q_A = -\hat{Q}_{\tilde{A}}^T \Delta A R_A^\dagger,\ \hat{Q}_A^T \tilde{Q}_A = \hat{Q}_A^T \Delta A R_{\tilde{A}}^\dagger.$$

Therefore

$$\|\hat{Q}_{\tilde{A}}^T Q_A\|_2 = \|\hat{Q}_A^T \tilde{Q}_A\|_2 \le \|\Delta A\|_2 \min\{\|A^\dagger\|_2, \|\tilde{A}^\dagger\|_2\}.$$

We can also establish similar results for B; therefore

$$\begin{aligned} &\|\Sigma - \tilde{\Sigma}\|_2 \\ &\le \sqrt{2}(\|A - \tilde{A}\|_2 \min\{\|A^\dagger\|_2, \|\tilde{A}^\dagger\|_2\} + \|B - \tilde{B}\|_2 \min\{\|B^\dagger\|_2, \|\tilde{B}^\dagger\|_2\}) \\ &\le \sqrt{2}\{\kappa(A)\|A - \tilde{A}\|_2/\|A\|_2 + \kappa(B)\|B - \tilde{B}\|_2/\|B\|_2\}. \end{aligned} \tag{4.16}$$

which establishes the result. □

Using the same technique, we can also prove the following result for the case of Frobenius norm:

Theorem *4.2. Let A and $\tilde{A}$, and B and $\tilde{B}$ have the same rank, then*

$$\|\Sigma - \tilde{\Sigma}\|_F \leq \sqrt{2}(\|A - \tilde{A}\|_F \min\{\|A^\dagger\|_2, \|\tilde{A}^\dagger\|_2\} + \|B - \tilde{B}\|_F \min\{\|B^\dagger\|_2, \|\tilde{B}^\dagger\|_2\})$$

Remark 4.1. In [4], Björck and Golub derived the following first order perturbation bound:

$$\|\Sigma - \tilde{\Sigma}\|_2 \leq \epsilon_A \sin\theta_{\max}(A, \tilde{B}) + \epsilon_B \sin\theta_{\max}(A, B) + O(\delta^2),$$

where

$$\|A - A\|_2 \leq \epsilon_A \|A\|_2, \quad \|B - \tilde{B}\|_2 \leq \epsilon_B \|B\|_2, \quad \delta = \epsilon_A \kappa(A) + \epsilon_B \kappa(B).$$

Our result (Theorem 4.1) compares favorably to the above.

The above theorems well the perturbation result in Example 4.1, but they do not tell the whole story as is demonstrated by the following example. First, let us consider the following matrix pair

Example *4.2. Given the matrix pair,*

$$A = \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix}, \quad B = \begin{pmatrix} 1 & 1 \\ 0 & 10^{-10} \\ 1 & 1 \end{pmatrix}.$$

The computed Q in the QR decomposition of B is

```
Q =
   -0.70710678118655    0.00000125385069
                    0  -0.999999999998~3
   -0.70710678118655   -0.00000125385069
```

and the computed canonical correlation is

$$1.771321265309725 4e\text{-}06.$$

All the computation in this section was carried out on a Sun 3/50 workstation using MATLAB version 3.5e with machine precision eps≈ 2.22e-16. Since $\text{cond}(B) \approx 10^{10}$, this result coincides with the prediction given by the bounds in Theorem 4.1. Now let us consider another matrix pair,

Example 4.3. Given *the matrix pair*

$$A_1 = \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix}, \quad B_1 = \begin{pmatrix} 1 & 10^{10} \\ .4 & .9 \\ 1 & 10^{10} \end{pmatrix}.$$

The matrix Q in ***the QR decomposition of*** B_1 is

```
Q =
   -0.680~13817~3977    0.19~5008972988
   -0.27216552697591   -0.96225044864938
   -0.68041381743977    0.19245008972987.
```

We also compute $\sigma(A, B)$ as

$$7.654331812476101e\text{-}16.$$

But since $\mathrm{cond}(B_1) \approx 10^{10}$, and the machine precision is approximately 10^{-16}, the bounds in the above theorems will predict a perturbation of size

$$\mathrm{cond}(B) \text{ x eps} \approx 10^{-6},$$

which is much larger than the computed result. However, theoretically if we scale the last column of B, we get a well conditioned matrix, and column scaling does not change the canonical correlations. The perturbation bounds in the above theorems are not invariant under the column scaling of A and B, therefore we need a refined version of the perturbation bounds.

Before we proceed, we introduce some more notation. If $A = (a_{ij})$, then we write $|A| := (|a_{ij}|)$; We denote $|A| \leq |B|$, if $|a_{ij}| \leq |b_{ij}|$; it is easy to verify that if $A = BC$, then $|A| \leq |B||C|$.

We define the column-scaling independent condition number of A as

$$\kappa_S(A) = \| |R||R^{-1}| \|_2,$$

if the QR decomposition of A is A = QR. Obviously, $\kappa_S(A)$ is independent of the column scaling of A; i.e.,

$$\kappa_S(AD) = \kappa_S(A)$$

for all positive definite diagonal matrix D.

Theorem 4.3. *Let A and B be of full column rank,* $\tilde{A} = A + \mathrm{AA}$ *and* $\tilde{B} = A + \mathrm{AB}$ *with* $|\Delta A| \leq \epsilon |A|$ *and* $|\Delta B| \leq \epsilon |B|$, *and* $\tilde{A}$ *and* $\tilde{B}$ are *also of full column rank. Then*

$$\|\Sigma - \tilde{\Sigma}\|_2 \leq \sqrt{2}\epsilon(\sqrt{p(m-p)}\kappa_S(A) + \sqrt{q(m-q)}\kappa_S(B)).$$

Proof. From (4.15), we have

$$\hat{Q}_{\tilde{A}}^T Q_A = -\hat{Q}_{\tilde{A}}^T \Delta A R_A^{-1}.$$

It follows that

$$|\hat{Q}_{\tilde{A}}^T Q_A| \leq |\hat{Q}_{\tilde{A}}^T||\Delta A||R_A^{-1}| \leq \epsilon |\hat{Q}_{\tilde{A}}^T||Q_A||R_A| \cdot |R_A^{-1}|.$$

Hence

$$\begin{aligned} \|\Sigma - \tilde{\Sigma}\| &\leq \sqrt{2}\,(\|\hat{Q}_A^T Q_{\tilde{A}}\|_F + \|\hat{Q}_B^T Q_{\tilde{B}}\|_F) \\ &\leq \sqrt{2}\,\epsilon\left(\| |\hat{Q}_{\tilde{A}}^T||Q_A| \|_F \kappa_S(A) + \| |\hat{Q}_{\tilde{B}}^T||Q_B| \|_F \kappa_S(B)\right), \end{aligned} \tag{4.17}$$

and the result is established. □

5. Numerical Algorithms

In this section, we discuss numerical computation of the canonical correlations and the corresponding canonical vectors. For simplicity, throughout the section we assume that both A and B are of full column rank, i.e., $A \in R^{m\times p}$ and $B \in R^{m\times q}$, and

$$p = \mathrm{rank(A)} \geq \mathrm{rank(B)} = q.$$

5.1. Dense Matrices. The following algorithm based on SVD was proposed by Björck and Golub [4], [7, Chapter 12].

Algorithm 5.1. *Given A and B, the following procedure computes the orthonormal matrices* $U = [u_1, \cdots, u_q]$ *and* $V = [v_1, \cdots, v_q]$ *and* $\sigma_1(A, B), \cdots, \sigma_q(A, B)$ *where* $\{\sigma_k(A, B)\}$ *are the canonical correlations of (A, B) and the* u_k *and* v_k *are the associated canonical vectors.*

i) *Compute the QR decomposition of A and B:*

$$A = Q_A R_A, \quad \text{where } Q_A^T Q_A = I_p,$$

$$B = Q_B R_B, \quad \text{where } Q_B^T Q_B = I_q.$$

ii) Form $C = Q_B^T Q_A$, *and compute the SVD of* C: $C = P^T \operatorname{diag}(\sigma_i(A, B))Q$.

$$Q_A P(:, 1:q) = [u_1, \cdots, u_q], \; Q_B Q = [v_1, \cdots, v_q].$$

As discussed in Section 4, there is no need to scale the columns of A and B before we compute their QR decompositions. Some numerical experiments were reported in [4], where QR decomposition with column pivoting was used to handle the rank deficient case.

5.2. Updating Problems. Let B be augmented by a column vector b. We want to investigate the relation between the canonical correlations of (A, B) and those of $(A, [B, b])$. We will develop an algorithm for updating the canonical correlations. We summarize the result in the following theorem.

Theorem 5.1. *Let g be a unit vector that spans* $\mathcal{R}([B, b]) \cap \mathcal{R}(B)^{\perp}$.[4] *Let* Q_s *be the orthonormal basis of the subspace* $\mathcal{R}(A)^{\perp} \cap \mathcal{R}(B)^{\perp}$. *Define*

$$\eta = \|(I - Q_s Q_s^T)g\|_2.$$

Then[5]

(a) $\sigma_l(A, [B, b]) = 1, \quad l = 1, \cdots, i;$

(b) $\sigma_l(A, B) \le \sigma_l(A, [B, b]) \le \sigma_l(A, B)\sqrt{1 + (1 - \eta^2)\tan^2\theta_l},$
$l = i+1, \cdots, i+j;$
where θ_l *is the l-th canonical (principal) angle (see Definition* 1.1); *and*

(c) $0 \le \sigma_{i+j+1}(A, [B, b]) \le 1 - \eta^2, \sigma_l(A, [B, b]) = 0,$
$l = i + j + 2, \cdots, i + j + k.$

Proof. We consider the case when $g \neq 0$; the other case when $g = 0$ is trivial. Let the QR decomposition of A and B be

$$A = Q_A R_A, \; B = Q_B R_B.$$

Using Theorem 3.1, we write

$$Q_A = Q\Sigma_A U^T, \quad Q_B = Q\Sigma_B V^T$$

where Q, U and V are orthogonal, and Σ_A and Σ_B are given by (3.6). The QR decomposition of $[B, b]$ can be written as

$$[B, b] = [Q_B, g]R_{[B,b]}$$

[4] If $\mathcal{R}([B, b]) \perp \mathcal{R}(B)^{\perp}$, i.e., $b \in \mathcal{R}(B)$ then we take $g = 0$.

[5] The integer indices refer to those in Theorem 3.1.

with $R_{[B,b]}$ a nonsingular upper triangular matrix. Let Q $= (Q_1, Q_2)$ with $Q_1 \in R^{m\times q}$, then there exists a unit vector $\tilde{g} \in R^{m-q}$ such that g $= Q_2\tilde{g}$. Hence $\tilde{g} = Q_2^T g$. Let $\tilde{g}^T = (g_1^T, g_2^T, g_3^T)^T$ with $g_2 \in R^j$ and $g_3 \in R^k$. Using Corollary 3.1, we have $Q_s^T g = g_1$. On the other hand

$$H := \begin{pmatrix} Q_B^T \\ g^T \end{pmatrix} Q_A = \operatorname{diag}\{\tilde{V}, 1\} \begin{pmatrix} I_i & & \\ & C & \\ & & O \\ 0 & g_2^T S & g_3 \end{pmatrix} U^T, \tag{5.18}$$

where $\tilde{V}$ is an orthogonal matrix. Hence the singular values of H are the square roots of the eigenvalues of

$$\begin{pmatrix} I_i & O \\ O & X \end{pmatrix}, \text{ with } X = \begin{pmatrix} C^2 & O \\ O & O \end{pmatrix} + \begin{pmatrix} S & O \\ O & I \end{pmatrix} G_s \begin{pmatrix} S & O \\ O & I \end{pmatrix}$$

where $G_s = (g_2^T, g_3^T)^T(g_2^T, g_3^T)$. We have

$$\lambda_l(X) = 1 - \lambda_{j+k-l+1}\left(\begin{pmatrix} S & O \\ O & I \end{pmatrix}(I - G_s)\begin{pmatrix} S & O \\ O & I \end{pmatrix}\right)$$

and

(5.19)
$$\lambda_l\left(\begin{pmatrix} S & O \\ O & I \end{pmatrix}(I - G_s)\begin{pmatrix} S & O \\ O & I \end{pmatrix}\right) = \sigma_l^2\left((I - G_s)^{\frac{1}{2}}\begin{pmatrix} S & O \\ O & I \end{pmatrix}\right)$$
$$\geq \sigma_{\min}^2((I - G_s)^{\frac{1}{2}})\,\sigma_l^2\left(\begin{pmatrix} S & O \\ O & I \end{pmatrix}\right).$$

Since the diagonal elements of S are given in non-decreasing order and those of C are in non-increasing order, and

$$\sigma_{\min}^2((I - G_s)^{\frac{1}{2}}) = 1 - (\|g_2\|^2 + \|g_2\|^2) = \eta^2,$$

we have

$$\lambda_{j+k-l+1}\left(\begin{pmatrix} S & O \\ O & I \end{pmatrix}(I - G_s)\begin{pmatrix} S & O \\ O & I \end{pmatrix}\right) \geq \eta^2\sigma_l^2\left(\begin{pmatrix} S & O \\ O & I \end{pmatrix}\right).$$

Hence from $\sigma_l(A, B) = \alpha_l = \cos(\theta_l)$, it follows that

$$\begin{aligned} \sigma_l(A, [B, b]) &\leq \sqrt{1 - \beta_l^2(1 - (1 - \eta^2))} \\ &= \sqrt{\alpha_l^2 + (1 - \alpha_l^2)(1 - \eta^2)} \\ &= \sigma_l(A, B)\sqrt{1 + (1 - \eta^2)\tan^2\theta_l}. \end{aligned} \tag{5.20}$$

Now we have that X is a rank one update of $\operatorname{diag}(C^2, 0)$. Hence X has at most one additional nonzero eigenvalue. □

Here is how the numerical computation of the updating proceed. Suppose Householder transformations or Jacobi rotations are utilized to compute the orthonormal basis of A and B so that we have [7, Section 5.2.]

$$B = Q \begin{pmatrix} R_B \\ O \end{pmatrix} = (Q_B, Q_B^{\perp}) \begin{pmatrix} R_B \\ O \end{pmatrix}.$$

Let $Q^T b = (x^T, \hat{b}^T)^T$. Apply a Householder transformation or a sequence of Jacobi rotations Q_b, we get

$$Q_b^T \hat{b} = \begin{pmatrix} \|\hat{b}\|_2 \\ 0 \end{pmatrix}, \quad \hat{b} = Q_b \begin{pmatrix} \|\hat{b}\|_2 \\ 0 \end{pmatrix}.$$

The QR decomposition of $[B, b]$ can be written as

$$[B, b] = [Q_B, Q_B^{\perp} Q_b] \begin{pmatrix} R_B & x \\ O & \|\hat{b}\|_2 \\ O & O \end{pmatrix}.$$

Then $[Q_B, Q_B^{\perp}(\hat{b}/\|\hat{b}\|_2)]$ is the orthonormal basis of $[B, b]$.

Remark 5.1. We note that the modified Gram-Schmidt algorithm could also be used to compute the orthonormal basis of A and B. Although it is twice as fast as the Householder transformation based algorithms, it has the drawback that the orthonormality of the computed basis depends on the condition numbers $\kappa_2(A)$ and $\kappa_2(B)$.[6]

For updating the SVD in (5.18), both the secular equation method (or the bisection method) [7, Section 8.6.3.], or the two-way chasing method [21] can be used. For a more detailed account of the bisection method the reader is referred to [5].

5.3. Large Sparse or Structured Matrices. If the matrix pair (A, B) is large sparse or structured, then explicit computation of the orthonormal basis of A and B will usually gives rise to a dense matrix or destroys the underlying structure. The purpose of this subsection is to propose a class of algorithms that will avoid explicit formation of the orthonormal basis of span(A) and span{ B}. Let us first consider a simple case: let A consist of one column, say a. Also, let the orthogonal projection onto span(B) be P_B. Then the canonical correlation of the matrix pair (a, B) is given by

$$\sigma(a, B) = |(P_B a)^T a| / \|P_B a\|_2 \|a\|_2,$$

and the canonical vectors are $a/\|a\|_2$ and $P_B a/\|P_B a\|_2$. Since $P_B a$ can be obtained by solving the following least squares problem:

$$\min_{x \in R^q} \|a - Bx\|_2 := \|a - Bx_0\|_2, \; P_B a = Bx_0,$$

the sparsity or structure of the matrix B can be fully exploited. For example, if the LSQR algorithm (cf. [13])is used to solve the above least squares problem, the

[6] It is our belief that $\kappa_2(A)$ and $\kappa_2(B)$ in the error analysis by Björck [3] can be replaced by the condition numbers defined in Sectionsec: cc4.

matrix B is only used to form the matrix products Bx and $B^T y$ for given vectors x and y.[7]

In the general case, we propose the following alternating least squares (ALS) method to compute the largest canonical correlation of the matrix pair ***(A, B).***

Algorithm 5.2. Choose $b_0 \in$ span(B) *with* $\|b_0\|_2 = 1$.

For $k = 0, 1, 2, \cdots$ ***until convergence do***

(a) Solve linear ***least squares problem:***

$$\min_{x \in R^q} \|b_k - Ax\|_2 = \|b_k - Ax_k\|_2, \quad \textit{and form } a_k = Ax_k/\|Ax_k\|_2;$$

(b) Solve ***linear least squares problem:***

$$\min_{y \in R^q} \|a_k - By\|_2 = \|a_k - By_k\|_2, \quad \textit{and form } b_{k+1} = By_k/\|By_k\|_2;$$

Iterate.

Assume convergence in K steps. Now we compute

$$\sigma_1(A, B) = |b_K^T a_K|, \quad u_1 = a_K/\|a_K\|_2, \quad v_1 = b_K/\|b_K\|_2.$$

For the convergence criterion, we choose either

$$\|b_{k+1}^T a_{k+1}| - |b_k^T a_k\|, \quad \text{or } \min\{\|a_{k+1} - a_k\|_2, \|b_{k+1} - b_k\|_2\}$$

be below a certain given tolerance.

Remark 5.2. The alternating least squares method is an old and natural idea, which goes back to J. Von Neumann. It has been used extensively in the psychometrics literature, and a recent application can be found in [17].

Convergence analysis of the ALS method. We relate the ALS algorithm to a variant of the power method, and thus derive its convergence rate. First let us consider the power method. Since finding the canonical correlations is equivalent to computing the SVD of $Q_B^T Q_A$. Let

$$T = \begin{pmatrix} O & (Q_B^T Q_A)^T \\ Q_B^T Q_A & O \end{pmatrix}, \tag{5.21}$$

then the eigenvalues of T are $\{\pm\sigma_i(Q_B^T Q_A)\}$. Applying the power method to T, we have

$$z_{k+1} = Tz_k, \text{ with } z_0 \text{ an initital vector.} \tag{5.22}$$

Let $z_k = (x_k^T, y_k^T)^T$; equation (5.22) can be written as

$$x_{k+1} = Q_A^T Q_B y_k, \quad y_{k+1} = Q_B^T Q_A x_k.$$

We can use the most recent x_{k+1} to compute y_{k+1} so that

$$x_{k+1} = Q_A^T Q_B y_k, \quad y_{k+1} = Q_B^T Q_A x_{k+1}.$$

[7] For a detailed presentation of fast algorithms for computing a matrix-vector product with Hankel or Toeplitz matrices, the reader is referred to [19].

It follows that

$$Q_A x_{k+1} = Q_A Q_A^T Q_B y_k, \quad Q_B y_{k+1} = Q_B Q_B^T Q_A x_{k+1}.$$

The above two equations are equivalent, since we assume that A and B are of full column rank. Define

$$\tilde{x}_k = Q_A x_k, \quad \tilde{y}_k = Q_B y_k,$$

we have the following modified power method.

Algorithm 5.3. Choose $y_0 \in$ span(B) *with* $\|y_0\|_2 = 1$.
For $k = 0, 1, 2, \cdots$ until *convergence do*

$$x_{k+1} = Q_A Q_A^T \tilde{y}_k, \quad \tilde{x}_{k+1} = x_{k+1}/\|x_{k+1}\|_2;$$

$$y_{k+1} = Q_B Q_B^T \tilde{x}_{k+1} \; \tilde{y}_{k+1} = y_{k+1}/\|y_{k+1}\|_2;$$

Iterate.

To see-that Algorithm 5.3 is equivalent to the ALS algorithm, we observe that $b_0 \in R(B)$, and can be written as $b_0 = Q_B s$ for some vector s. The solution of the least squares problem

$$\min_{x \in R^q} \|b_0 - Ax\|_2 = \|b_0 - Ax_0\|_2$$

is given by $x_0 = A^{\dagger} b_0$. Hence $a_0 = \gamma_0 Q_A Q_A^T (Q_B s)$, where γ_0 is the normalization factor. By induction we can prove

$$a_k = \gamma_k Q_A [(Q_B^T Q_A)^T (Q_B^T Q_A)]^k Q_A^T Q_B s; \; b_k = \delta_k Q_B [(Q_B^T Q_A)(Q_B^T Q_A)^T]^k s.$$

where γ_k and δ_k are the normalization factors. Therefore the convergence rate of the ALS algorithm is dependent on

$$\kappa = (\sigma_2(Q_B^T Q_A)/\sigma_1(Q_B^T Q_A))^2 = (\sigma_2(A, B)/\sigma_1(A, B))^2.$$

Example 5.1. *We consider the matrix pair*

$$A = U \begin{pmatrix} 1 & 0 \\ 0 & .8 \\ 0 & .6 \end{pmatrix} P_1, \quad B = U \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{pmatrix} P_2,$$

where U is *an orthogonal matriz and* P_1 *and* P_2 *are nonsingular matrices. The* canonical *correlations* are $\sigma_2(A, B) = 1$, $\sigma_2(A, B) = 0.8$. *Therefore the convergence rate of the ALS algorithm* is *0.64. We compute* $\log_e(0.64)$

```
log_e(0.64)=-.44628710262842.
```

We have truncated the data *at both ends. The best computed linear polynomial fit to the computed data* gives *the slope*

```
-. 44614014758065,
```

which matches the convergence rate quite well.

In the literature, various ways to accelerate the power method are given [7, Chapter 10]. We can adapt these acceleration schemes to the ALS algorithm, but we will not go into the details here.

The drawback of the ALS algorithm together with its various acceleration schemes is that only the largest canonical correlation is computed. To compute several canonical correlations at the same time, we can use certain versions of subspace iteration and we use various acceleration schemes. We will not discuss these extensions here, but instead we will show how to adapt the Lanczos method to our problem by using a similar idea as the ALS method. We apply the Lanczos algorithm [7, Chapter 9] and start with the matrix T defined in (5.21).

Algorithm *5.4. (Lanczos Algorithm)*

v_1 is *given with* $\|v_1\|_2 = 1$
$p0 = v_1, \beta_0 = 1, \quad j = 0, u_0 = 0$
while $\beta_j \neq 0$
$v_{j+1} = p_j/\beta_j; j = j + 1$
$r_j = Q_B^T Q_A v_j - \beta_{j-1} u_{j-1}$
$\alpha_j = \|r_j\|_2; \quad u_j = r_j/\alpha_j$
$p_j = Q_A^T Q_B u_j - \alpha_j v_j$
$\beta_j = \|p_j\|_2$
end

We observe that the operator $Q_B^T Q_A$ is not available, since we do not explicitly form the orthonormal bases for span(A) and span{ ***B}.*** The device we use is to make a bases transformation. Let us transform the vectors generated in Algorithm 5.4 to the column spaces of ***A*** and ***B,*** i.e., span(A) and span(B), and denote

$$\tilde{u}_j = Q_B u_j, \ \tilde{v}_j = Q_A v_j, \tilde{r}_j = Q_B r_j, \ \tilde{p}_j = Q_A p_j.$$

Rewrite Algorithm 5.4 in the new basis, we obtain the following algorithm

Algorithm 5.5. (Modified ***Lanctos Algorithm)***

Choose $v = As$; ***set*** $\tilde{v}_1 = v/\|v\|_2$
$\tilde{p}_0 = v_1, \beta_0 = 1, j = 0, \tilde{u}_0 = 0$
while $\beta_j \neq 0$
$\tilde{v}_{j+1} = \tilde{p}_j/\beta_j; \quad j = j + 1$
$\tilde{r}_j = Q_B Q_B^T \tilde{v}_j - \beta_{j-1} \tilde{u}_{j-1}$
$\alpha_j = \|\tilde{r}_j\|_2; \quad \tilde{u}_j = \tilde{r}_j/\alpha_j$
$\tilde{p}_j = Q_A Q_A^T \tilde{u}_j - \alpha_j \tilde{v}_j$
$\beta_j = \|\tilde{p}_j\|_2$
end

Note that β_j and α_j in Algorithm 5.5 is the same as those in Algorithm 5.4. The computation of $Q_B Q_B^T \tilde{v}_j$ and $Q_A Q_A^T \tilde{u}_j$ are again carried out by solving the least squares problems:

$$\min_{y \in R^q} \|\tilde{v}_j - By\|_2 = \|\tilde{v}_j - By_j\|_2. \tag{5.23}$$

Then $Q_B Q_B^T \tilde{v}_j = By_j$. Similarly,

$$\min_{x \in R^p} \|\tilde{u}_j - Ax\|_2 = \|\tilde{u}_j - Ax_j\|_2, \tag{5.24}$$

and $Q_A Q_A^T \tilde{u}_j = A x_j$.

Remark 5.3. The above algorithm can be easily adapted to computing the canonical correlations between two linear subspaces defined either by the range space or null space of matrices. If for example, one of the subspace is defined by the null space of A, then instead of using $Q_A\, Q_A^T$ in the above, we use $I - Q_A Q_A^T$.

We have also tested the Modified Lanczos Algorithm. The matrix pair is given as follows

$$A = U \begin{pmatrix} C \\ S \end{pmatrix} P_1, \quad B = U \begin{pmatrix} I_n \\ O \end{pmatrix} P_2,$$

where U is orthogonal and P_1 and P_2 are nonsingular, with

$$C = \mathrm{diag}(0, 1/n, 2/n, \ldots, (n-1)/n), \quad \text{and S} = \sqrt{(I_n - C^2)}.$$

Therefore the canonical correlations of (A, B) are $0, 1/n, 2/n, \ldots, (n-1)/n$. For the particular example in Figure 1, we chose n = 100. We do not solve the least squares problems in (5.23) and (5.24) exactly, instead we simulate the LSQR algorithm [13] by first using a direct method to solve the least squares problem and add noise to the solution. More numerical experiments using the LSQR will be carried out in the future. In Figure 1, the relative errors of the first three computed canonical correlations are plotted against the iteration numbers.

There remains a number of problems associated with this technique such as determining a preconditioner for solving the least squares problem. Nevertheless, we feel that the approach is of great potential use in computing canonical correlations of large or sparse matrix pairs and it certainly deserves further investigation.

REFERENCES

1. T. W. Anderson, *An introduction to multivariate* statistical analysis, John Wiley and Sons, New York, 1958.
2. P. Besse, *Edudie* descriptive *d'un process*, Ph.D. thesis, Paul-Sabatier University, 1979.
3. A. Björck, Solving linear *least squares problems by Gram-Schmidt orthogonalization*, BIT 7 (1967), 1-21.
4. A. Björck and G. H. Golub, *Numerical methods* for computing *angles between* linear *rubspaces*, Mathematics of Computation 27 (1973), 579-594.
5. J. Demmel and W. Gragg, *On computing accurate* singular *values and* eigenvalues of acyclic *matrices*, IMA Preprint Series 962, IMA, University of Minnesota, 1992.
6. Y. Escoufier, *Operators related to a data matrix*, In Recent Developments in Statistics (Amsterdam), North Holland, 1976.
7. G. H. Golub and C. F. Van *Loan, Matrix computations*, 2nd ed., Johns Hopkins University Press, Baltimore, Maryland, 1989.
8. H. Hotelling, Relation *between two sets* of variates, Biometrika 28 (1936), 322-377.
9. A. Israëls, Eigenvalve *techniques for qualitative data*, DSWO Press, Leiden, 1987.
10. C. Jordan, *Essai sur la géométrie à n dimensions*, Bulletin de la Société Mathématique 3 (1875), 103-174.
11. J. R. Kettenring, Canonical *analysis of several sets of variates*, Biometrika 58 (1971), 433-451.
12. C. C. Paige, *A note on a result* of Sun Ji-guang: Sensitivity *of the cs and gsv decomposition*, SIAM Journal on Numerical Analysis 21 (1984), 186-191.
13. C. C. Paige and M. A. Saunders, *LSQR: an algorithm for sparse* linear *equations and sparse least squares*, ACM Transaction on Mathematical Software 8 (1982), 43-71.
14. C. R. Rao and H. *Yanai, General* definition *and* decomposition *of projectors and some* application to *statistical problems*, J. Statistical Planning and Inference 3 (1979), 1-17.
15. G. W. Stewart, Remarks made at an IMA workshop, 1992.

16. G. W. Stewart and G.-J. Sun, *Matrix perturbation theory*, Academic Press, Boston, 1990.
17. E. van der Burg, Nonlinear canonical correlation *and some related technique*, DSWO Press, Leiden, 1988.
18. P.-A. Wedin, *On angles between subspacer*, Matrix Pencils (New York) (B. Kågström and A. Ruhe, eds.), Springer, 1983, pp. 263-285.
19. G. Xu and T. Kailath, *Fast signal-subspace decompsotion — Part I: Ideal covariance* matrices, Manuscript submitted to ASSP. Information Systems Laboratory, Stanford University, 1990.
20. H. Zha, *The restricted* singular value decomposition of mattit *triplets*, SIAM Journal on Matrix Analysis and Applciations 12 (1991), 172-194.
21. ———, *A two-way chasing scheme for* redwing a rymmettic *arrowhead matrix to tridiagonal* form, Numerical Linear Algebra with Applications 1 (1992), 49-57.

COMPUTER SCIENCE DEPARTMENT, STANFORD UNIVERSITY, STANFORD, CA 94305-2140
E-mail *address:* golub@cholesky.stanford.edu

SCIENTIFIC COMPUTING & COMPUTATIONAL MATHEMATICS, STANFORD UNIVERSITY, STANFORD, CA 94305-2140
Current address: Computer Science Department, 309 Whitmore Laboratory, Penn State University, University Park, PA 16802-6103
E-mail *address:* zha@cholesky.stanford.edu

FIGURE 1. Convergence behavior of the modified Lanczos method