

Majorization for Changes in Ritz Values and Canonical Angles Between Subspaces (Part I and Part II)

**Merico Argentati (speaker), Andrew Knyazev,
Ilya Lashuk and Abram Jujunashvili**

*Department of Mathematics and
Center for Computational Mathematics
University of Colorado at Denver*

Supported by the National Science Foundation

Major Sources

- A. V. Knyazev and M. E. Argentati, [Principal Angles between Subspaces in an A-Based Scalar Product: Algorithms and Perturbation Estimates](#), SIAM Journal on Scientific Computing 23 (2002), no. 6, pp. 2009–2041.
- A. V. Knyazev and M. E. Argentati, [On Proximity of Rayleigh Quotients for Different Vectors and Ritz Values Generated by Different Trial Subspaces](#), Linear Algebra and its Applications, accepted.
- A. V. Knyazev and M. E. Argentati, [Majorization for Changes in Principal Angles Between Subspaces](#), in production.

Part I – Canonical Angles Between Subspaces

Outline

1. Angles Between Subspaces in Applications
2. Theory and Algorithms in the Euclidean Norm
3. Generalization to an A -Based Scalar Product
4. Numerical Examples
5. Conclusions

1. Angles Between Subspaces in Applications

- In statistics, the angles are closely related to measures of dependency and covariance of random variables, and can be used to describe canonical correlations of a matrix pair.
- Analyzing perturbations related to eigensolvers, where angles provide information about solution quality and need to be computed with high accuracy (Davis and Kahan [1970]; Li [1999]).
- Information retrieval, web search engines, DNA microarray analysis, and control theory.

2. Definition of Principal Angles

Let us consider two real-valued matrices F and G , each with n rows, and their corresponding column-spaces \mathcal{F} and \mathcal{G} , which are subspaces in R^n , assuming that

$$p = \dim\mathcal{F} \geq \dim\mathcal{G} = q \geq 1.$$

Then the *principal angles* $\theta_1, \dots, \theta_q \in [0, \pi/2]$ between \mathcal{F} and \mathcal{G} may be defined recursively for $k = 1, \dots, q$ by

$$\cos(\theta_k) = \max_{u \in \mathcal{F}} \max_{v \in \mathcal{G}} u^T v = u_k^T v_k$$

subject to

$$\|u\| = \|v\| = 1, \quad u^T u_i = 0, \quad v^T v_i = 0, \quad i = 1, \dots, k-1.$$

The vectors u_1, \dots, u_q and v_1, \dots, v_q are called principal vectors.

Other Representations for the Largest and Smallest Principal Angles

- There is an equivalent, slightly more intuitive, characterization, of the largest principal angle for $p = q$, which is given by

$$\theta_{\max} = \max_{u \in \mathcal{F}} \min_{v \in \mathcal{G}} \angle \{u, v\} \quad u \neq 0, v \neq 0.$$

- Let $P_{\mathcal{F}}$ and $P_{\mathcal{G}}$ be orthogonal projectors onto the subspaces \mathcal{G} and \mathcal{F} , respectively. Then if $p = q$, we have

$$\text{gap}(\mathcal{F}, \mathcal{G}) = \sin(\theta_{\max}) = \|P_{\mathcal{F}} - P_{\mathcal{G}}\|_2,$$

and

$$\cos(\theta_{\min}) = \|P_{\mathcal{F}}P_{\mathcal{G}}\|_2.$$

An SVD Cosine–Based Algorithm

Let columns of matrices $Q_F \in R^{n \times p}$ and $Q_G \in R^{n \times q}$ form orthonormal bases for the subspaces \mathcal{F} and \mathcal{G} respectively. The reduced SVD of $Q_F^T Q_G$ is

$$Y^T Q_F^T Q_G Z = \text{diag}(\sigma_1, \sigma_2, \dots, \sigma_q), \quad \sigma_1 \geq \dots \geq \sigma_q$$

where $Y \in R^{p \times q}$, $Z \in R^{q \times q}$ both have orthonormal columns. Then the principal angles can be computed as

$$\theta_k = \arccos(\sigma_k), \quad k = 1, \dots, q, \quad 0 \leq \theta_1 \leq \dots \leq \theta_q \leq \frac{\pi}{2},$$

while principal vectors are given by

$$u_k = Q_F y_k, \quad v_k = Q_G z_k, \quad k = 1, \dots, q.$$

Inaccuracy in the Cosine-based Algorithm

Let $\mathcal{F} = \text{span} \left\{ (1 \ 0)^T \right\}$, $\mathcal{G} = \text{span} \left\{ (1 \ d)^T \right\}$. Then

$$\cos(\theta) = \frac{1}{\sqrt{1 + d^2}}.$$

The cosine, that is a canonical correlation, is computed accurately and simply equals to one for all $|d| \leq 10^{-8}$ in double precision. Small angle θ cannot be computed accurately from $\cos(\theta)$. However,

$$\theta = \arcsin \left(\frac{|d|}{\sqrt{1 + d^2}} \right),$$

can be accurately computed for small $|d|$.

When $\theta \approx \pi/2$, the sine based formula is inaccurate and the cosine based formula should be used.

The fix: a Sine-Based Algorithm

Theorem 1 (Björck and Golub [1973]) *Singular values $\mu_1 \leq \mu_2 \leq \dots \leq \mu_q$ of matrix $Q_G - Q_F(Q_F^T Q_G)$ are given by $\mu_k = \sqrt{1 - \sigma_k^2}$, $k = 1, \dots, q$. Moreover, the principal angles satisfy the equalities $\theta_k = \arcsin(\mu_k)$.*

The right principal vectors can be computed as

$$v_k = Q_G z_k, \quad k = 1, \dots, q,$$

where z_k are corresponding orthonormal right singular vectors of matrix $Q_G - Q_F(Q_F^T Q_G)$. The left principal vectors are then computed by

$$u_k = Q_F(Q_F^T v_k) / \sigma_k, \quad \text{if } \sigma_k \neq 0, \quad k = 1, \dots, q.$$

3. Generalization to an A -Based Scalar Product

Let $A \in R^{n \times n}$ be a fixed symmetric positive definite (SPD) matrix. Let $(x, y)_A = (x, Ay) = y^T Ax$ be an A -based scalar product, $x, y \in R^n$. Let $\|x\|_A = \sqrt{(x, x)_A}$ be the corresponding vector norm and let $\|B\|_A$ be the corresponding induced matrix norm of a matrix $B \in R^{n \times n}$.

Then

$$\cos(\theta_k) = \max_{u \in \mathcal{F}} \max_{v \in \mathcal{G}} (u, v)_A = (u_k, v_k)_A$$

subject to

$$\|u\|_A = \|v\|_A = 1, (u, u_i)_A = 0, (v, v_i)_A = 0, i = 1, \dots, k-1.$$

The vectors u_1, \dots, u_q and v_1, \dots, v_q are called principal vectors relative to the A -based scalar product.

Algorithms in the A -Based Scalar Product

Theorem 2 *Let $A = K^T K$. The principal angles between subspaces \mathcal{F} and \mathcal{G} relative to the scalar product $(\cdot, \cdot)_A$ coincide with the principal angles between subspaces $K\mathcal{F}$ and $K\mathcal{G}$ relative to the original scalar product (\cdot, \cdot) .*

- If K is available, we can use the representation where the columns of matrices KQ_F and KQ_G are orthonormal with respect to the standard Euclidean inner product to compute the principal angles between the subspaces \mathcal{F} and \mathcal{G} relative to the scalar product $(\cdot, \cdot)_A$.
- In the rest of the talk we do not assume that a factorization of A (such as $K^T K$) is available. We assume A is available through the matrix–vector product function.

Theorem 3 *Let columns of $Q_F \in R^{n \times p}$ and $Q_G \in R^{n \times q}$ be now A -orthonormal bases for the subspaces F and G respectively. Let $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_q$ be singular values of $Q_F^T A Q_G$ with corresponding left and right singular vectors y_k and z_k , $k = 1, \dots, q$. Then the principal angles relative to the scalar product $(\cdot, \cdot)_A$ are computed as*

$$\theta_k = \arccos(\sigma_k), \quad k = 1, \dots, q,$$

where

$$0 \leq \theta_1 \leq \dots \leq \theta_q \leq \frac{\pi}{2},$$

while the principal vectors are given by

$$u_k = Q_F y_k, \quad v_k = Q_G z_k, \quad k = 1, \dots, q.$$

Theorem 4 *Let $S = Q_G - Q_F(Q_F^T A Q_G)$ and Q_S be an A -orthonormal basis for the column space of S . Singular values $\mu_1 \leq \mu_2 \leq \dots \leq \mu_q$ of matrix $Q_S^T A S$ are given by $\mu_k = \sqrt{1 - \sigma_k^2}$. Moreover, the principal angles satisfy the equalities $\theta_k = \arcsin(\mu_k)$. The right principal vectors can be computed as*

$$v_k = Q_G z_k, \quad k = 1, \dots, q,$$

where z_k are corresponding orthonormal right singular vectors of matrix $Q_S^T A S$. The left principal vectors are then computed by

$$u_k = Q_F(Q_F^T A v_k) / \sigma_k, \quad \text{if } \sigma_k \neq 0, \quad k = 1, \dots, q.$$

In our code we first use the cosine based approach for large angles, and then recompute only the small angles, using the sine based approach.

Issues for Implementation (SUBSPACEA.m)

1. In our applications $n \gg p$, i.e. the matrices are tall, but not very wide.
2. Our goal is not to store in memory, any matrix larger than $n \times p$.
3. For the A -based scalar product, we allow A to be provided as a function and implement a matrix–vector product.
4. The code for A -orthonormalization, our ORTHA.M is provided.
5. When $n \gg p$, the typical case, the computational costs of the SVDs of a $p \times q$ matrix are negligible; it is the multiplication by A which may be computationally expensive. Therefore we minimize the multiplications by A . The worst case is $2p + q$ matrix–vector multiplications.
6. The well known approach of computing a CS decomposition that requires computation of $n \times n$ matrices is too expensive when $n \gg p$.

Algorithm Implementation - SUBSPACEA.m

Input: matrices F and G with the same number of rows, and a symmetric positive definite matrix A for the scalar product, or a device to compute Ax for a given vector x .

1. Compute A -orthogonal bases $Q_F = \text{ortha}(F)$, $Q_G = \text{ortha}(G)$ of column-spaces of F and G .
2. Compute SVD for cosine $[Y, \Sigma, Z] = \text{svd}(Q_F^T A Q_G)$, $\Sigma = \text{diag}(\sigma_1, \dots, \sigma_q)$.
3. Compute matrices of left $U_{\cos} = Q_F Y$ and right $V_{\cos} = Q_G Z$ principal vectors.
4. Compute large principal angles for $k = 1, \dots, q$:

$$\theta_k = \arccos(\sigma_k) \text{ if } \sigma_k^2 < 1/2.$$
5. Form parts of matrices U and V by picking up corresponding columns of U_{\cos} , V_{\cos} , according to the choice for θ_k above. Put columns of U_{\cos} , V_{\cos} , which are left, in matrices R_F and R_G . Collect the corresponding σ 's in a diagonal matrix Σ_R .

6. Compute the matrix $S = R_G - Q_F(Q_F^T A R_G)$.
7. Compute A -orthogonal basis $Q_S = \text{ortha}(S)$ of the column-space of S .
8. Compute SVD for sine: $[Y, \text{diag}(\mu_1, \dots, \mu_q), Z] = \text{svd}(Q_S^T A S)$.
9. Compute matrices U_{sin} and V_{sin} of left and right principal vectors:

$$V_{\text{sin}} = R_G Z, U_{\text{sin}} = R_F (R_F^T A V_{\text{sin}}) \Sigma_R^{-1}.$$
10. Compute the missing principal angles, for $k = 1, \dots, q$:

$$\theta_k = \arcsin(\mu_k) \text{ if } \mu_k^2 \leq 1/2.$$
11. Complete matrices U and V by adding columns of $U_{\text{sin}}, V_{\text{sin}}$.

Output: Principal angles $\theta_1, \dots, \theta_q$ between column-spaces of matrices F and G , and corresponding matrices U and V of left and right principal vectors, respectively.

For software submitted to MathWorks (subspacea.m and ortha.m) see <http://www.mathworks.com/matlabcentral/fileexchange>.

The Singular Values of $Q_S^T AS$

- With $S = Q_G - Q_F(Q_F^T A Q_G)$, the singular values of $A^{1/2}S$ are the sines of the angles, but we don't want to compute $A^{1/2}$.
- In exact arithmetic, columns of S are A -orthogonal and their A -norms are exactly the sine of principal angles. Thus, if there are several small angles different in orders of magnitude, the columns of S are badly scaled. When we take the norms squared, by explicitly computing the product $S^T AS$, we make the scaling even worse.
- However, we can compute the sines of the angles directly, which are the singular values of $Q_S^T AS$ where Q_S is an A -orthonormal basis for the column space of S .
- When A is the identity we just need to compute the singular values of S .

The Split Between the Small and Large Angles

- The $1/2$ threshold used in above Algorithm to separate small and large principal angles and corresponding vectors seems to be a natural choice.
- However, such an artificial fixed threshold may cause troubles with orthogonality in the resulting choice of vectors if there are several angles close to each other but on different sides of the threshold.
- The problem is that the corresponding principal vectors, picked up from two orthogonal sets computed by different algorithms, may not be orthogonal.
- A more accurate approach would be to identify such possible cluster of principal angles around the original threshold and to make sure that all principal vectors corresponding to the cluster are treated by one algorithm.

Invariant Subspaces Corresponding to Large and Small Angles

- We reduce computational costs of the second SVD by using already computed vectors U_{\cos} and V_{\cos} for the cosines.
- The cosine-based algorithm computes inaccurately individual principal vectors corresponding to small principal angles. It may find accurately the corresponding invariant subspaces spanned by all these vectors.
- Thus, the idea is that, using U_{\cos} and V_{\cos} , we can identify invariant subspaces in \mathcal{F} and \mathcal{G} , which correspond to all small principal angles. Using the second SVD we compute only the columns of U_{\sin} and V_{\sin} that we actually need, which may significantly reduce the size of the matrix in the second SVD.
- This idea is used in the CS decomposition alg. of Van Loan [1985].

4. Numerical Results

The following numerical tests were performed:

- Error Growth with Problem Size - Moderate and Small Angles
- Errors in Individual Angles
- Errors for an Ill-Conditioned Scalar Product

The SUBSPACEA.M code consists of two parts: Alg 3.2 and Alg 6.2 (using enumeration in our paper). Alg 3.2 computes the angles in the standard scalar product, while Alg 6.2 computes the angles in the A -based scalar product.

Error Growth with Problem Size – Figure 1 and Figure 2

According to our analysis an absolute change in cosine and sine of principal angles is bounded by perturbations in matrices F and G , with the constant, proportional to their condition numbers taken after a proper column scaling.

We assume n to be even and $p = q \leq n/2$ with $A = I$. Let D be a diagonal matrix of the size p :

$$D = \text{diag} (d_1, \dots, d_p), \quad d_k > 0, \quad k = 1, \dots, p.$$

We first define n -by- p matrices

$$F_1 = [I \ 0]^T, \quad G_1 = [I \ D \ 0]^T,$$

We multiply matrices F_1 and G_1 by orthogonal random matrices on the left and right. The exact values of sine and cosine of principal angles between

column-spaces of matrices F_1 and G_1 are obviously given by

$$\mu_k = \frac{d_k}{\sqrt{1 + d_k^2}}, \quad \sigma_k = \frac{1}{\sqrt{1 + d_k^2}}, \quad k = 1, \dots, p,$$

The collective error in principal angles is measured as the following sum:

$$\sqrt{(\mu_1 - \tilde{\mu}_1)^2 + \dots + (\mu_p - \tilde{\mu}_p)^2} + \sqrt{(\sigma_1 - \tilde{\sigma}_1)^2 + \dots + (\sigma_p - \tilde{\sigma}_p)^2},$$

where μ 's are the sine and σ 's are the cosine of principal angles, and the tilde sign $\tilde{}$ is used for actual computed values.

In the first two tests, diagonal entries of D are chosen as uniformly distributed random numbers *rand* on the interval $(0, 1)$.

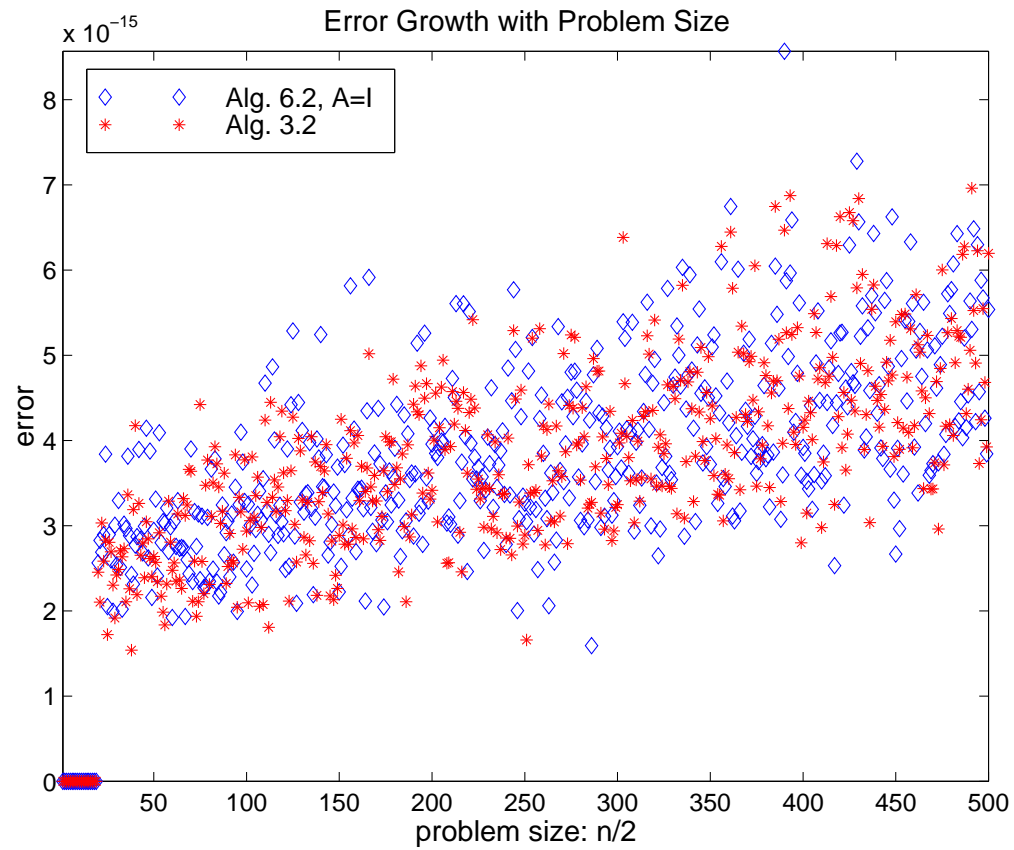


Figure 1: Errors in principal angles as functions of $n/2$; $p = 20$.

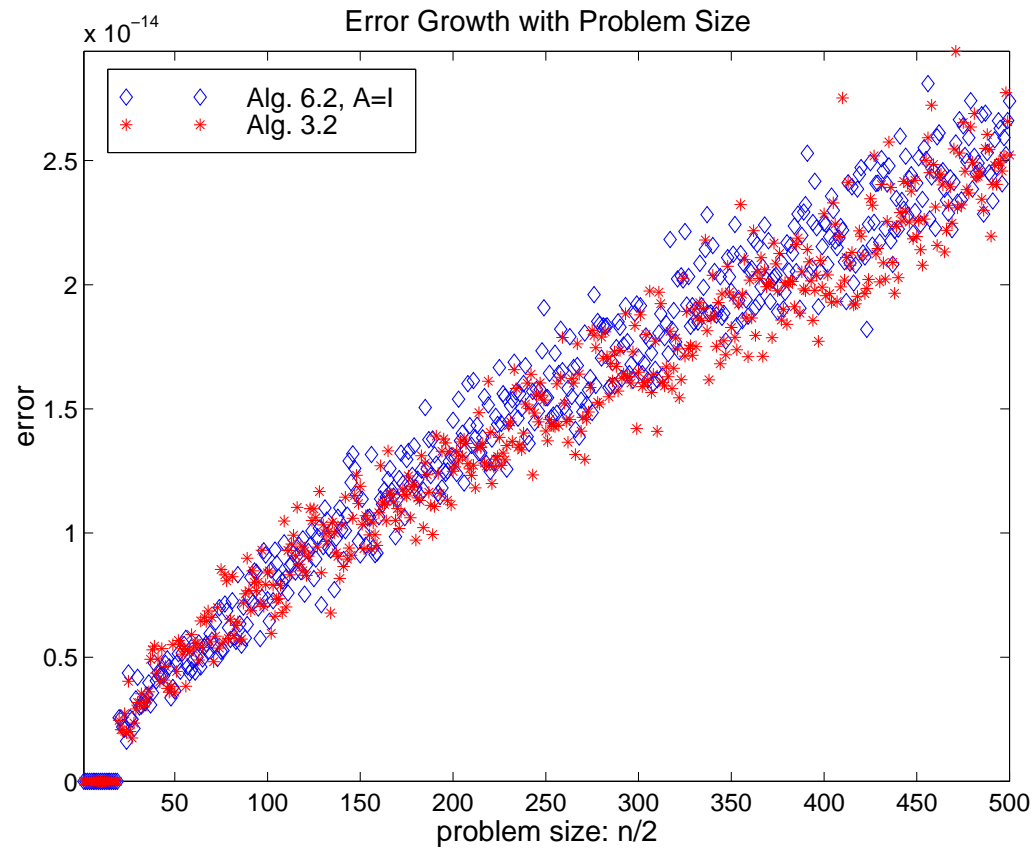


Figure 2: Errors in principal angles as functions of $n/2$; $p = n/2$.

Errors For Individual Angles – Figure 3 and Figure 4.

We fix a small $p = q = 10$, $n = 100$ with $A = I$, and compute angles 500 times, changing only the random matrices used in the construction of our F_3 and G_3 . We now compute errors for individual principal angles as

$$|\mu_k - \tilde{\mu}_k| + |\sigma_k - \tilde{\sigma}_k|, \quad k = 1, \dots, p.$$

The worst-case scenario found numerically corresponds to (Figure 3)

$$D = \text{diag} \{1, 0.5, 10^{-11}, 10^{-12}, 10^{-13}, 5 \cdot 10^{-15}, 2 \cdot 10^{-15}, 10^{-15}, 10^{-16}, 0\}$$

To test our code for an ill-conditioned case, we add two large values to the previous choice of D to obtain (Figure 4)

$$D = \text{diag} \{10^{10}, 10^8, 1, 0.5, 10^{-11}, 10^{-12}, 10^{-13}, 5 \cdot 10^{-15}, 2 \cdot 10^{-15}, 10^{-15}, 10^{-16}, 0\}.$$

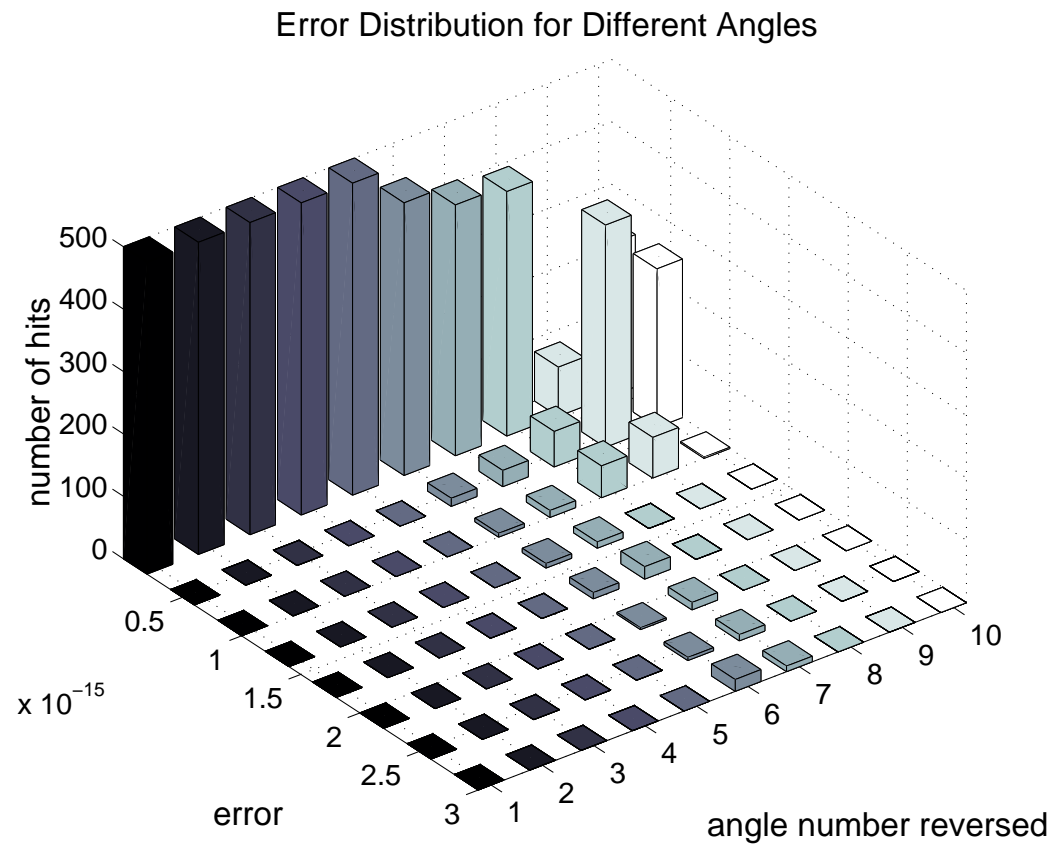


Figure 3: Errors for individual angles ($n = 100, p = 10$).

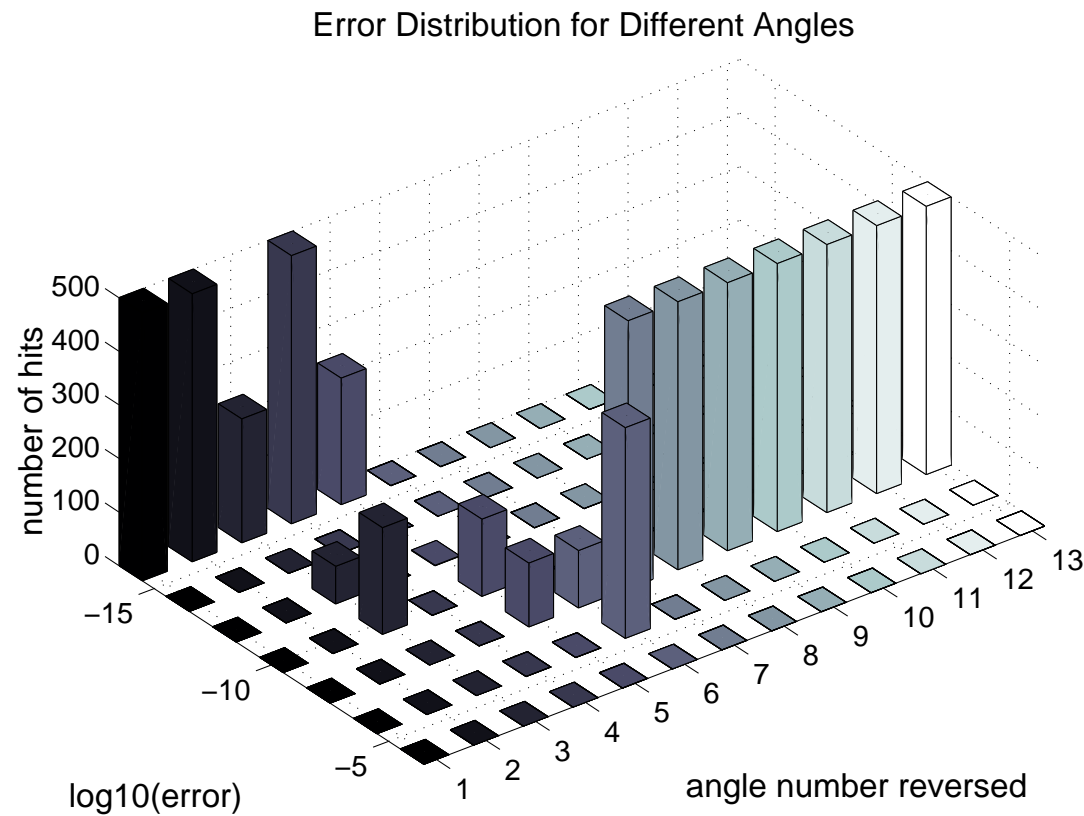


Figure 4: Errors in individual angles for ill-conditioned case.

Ill-Conditioned Scalar Products – Figure 5.

We take G to be the first ten columns of the identity matrix of size twenty, and F to be the last ten columns of the Vandermonde matrix of size twenty with elements $v_{i,j} = i^{20-j}$, $i, j = 1, \dots, 20$. Matrix F is ill-conditioned, $\text{cond}F \approx 10^{13}$. We compute principal angles and vectors between F and G in an A -based scalar product for the following family of matrices:

$$A = A_l = 10^{-l}I + H, \quad l = 1, \dots, 16,$$

where I is identity and H is the Hilbert matrix of the order twenty, whose elements are given by $h_{i,j} = 1/(i + j - 1)$, $i, j = 1, \dots, 20$.

MATLAB was apparently compiled on LINUX to take advantage of extended 80 bit precision of FPU registers of PIII, while Microsoft compilers apparently set the FPU to 64 bit operations.

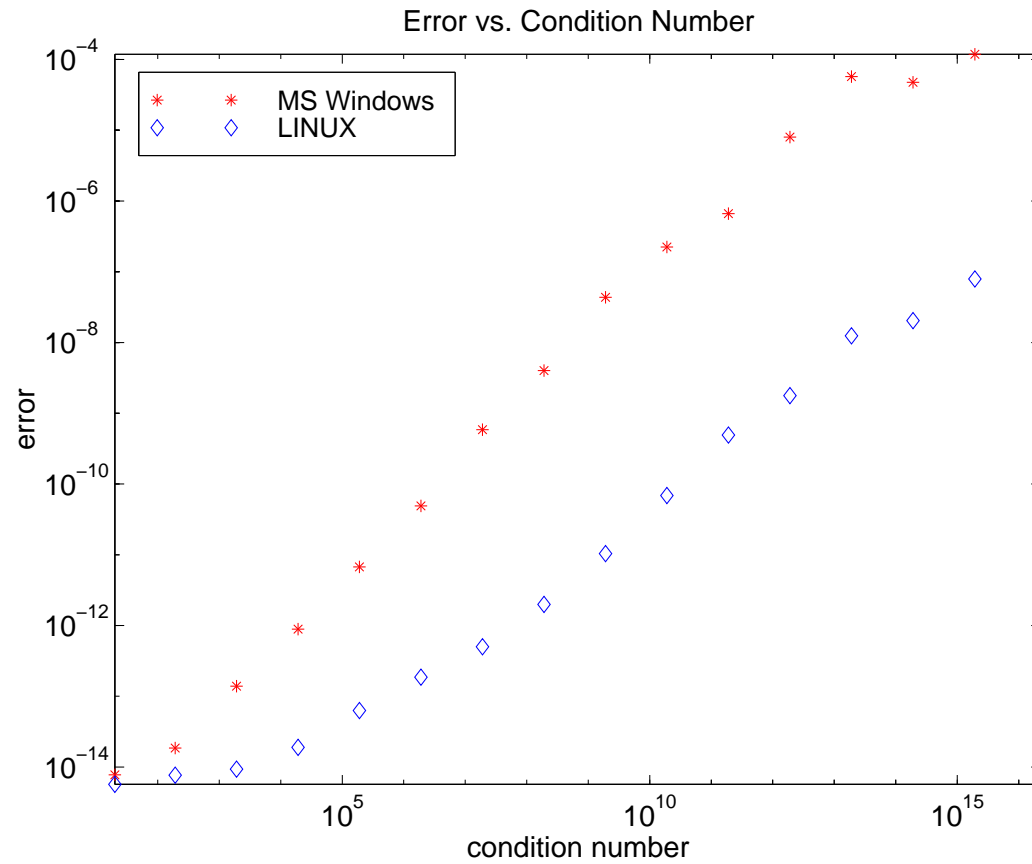


Figure 5: Ill-Conditioned Scalar Products .

5. Conclusions (Part I)

- An algorithm is presented that computes all principal angles accurately in computer arithmetic and is proved to be equivalent to the standard algorithm in exact arithmetic.
- A generalization of the algorithm to an arbitrary scalar product given by a symmetric positive definite matrix is suggested and justified.
- Code is provided as well as results of numerical tests. The code is robust in practice and provides accurate angles for large-scale and ill-conditioned cases we tested numerically. It is also reasonably efficient for large-scale applications with $n \gg p$.

Part II – Majorization for Changes in Angles/Ritz Values Outline

1. Basic Perturbation Theorems
2. Majorization of principal angles and unitarily invariant norms
3. Majorization inequalities and perturbation of principal angles
4. Perturbation of Ritz values in the Rayleigh-Ritz method with the change of the trial subspace
5. Principal angles in a Hilbert Space
6. Conclusions

1. Perturbation Analysis – Some Motivation

Trigonometric Inequalities

If an angle $\theta \in [0, \pi/2]$ is perturbed by $\epsilon \in [0, \pi/2]$ such that $\theta + \epsilon \in [0, \pi/2]$, then

$$0 \leq \cos(\theta) - \cos(\theta + \epsilon) \leq \sin(\theta + \epsilon) \sin(\epsilon) \leq \sin(\epsilon),$$

$$0 \leq \sin(\theta + \epsilon) - \sin(\theta) \leq \cos(\theta) \sin(\epsilon) \leq \sin(\epsilon),$$

$$0 \leq \cos^2(\theta) - \cos^2(\theta + \epsilon) = \sin(2\theta + \epsilon) \sin(\epsilon) \leq \sin(\epsilon),$$

$$0 \leq \sin^2(\theta + \epsilon) - \sin^2(\theta) = \sin(2\theta + \epsilon) \sin(\epsilon) \leq \sin(\epsilon).$$

The One Dimensional Case

Let $x, y, z \in \mathbf{C}^n$. Then

$$|\angle\{z, x\} - \angle\{z, y\}| \leq \angle\{x, y\}, \quad (1)$$

$$|\sin(\angle\{z, x\}) - \sin(\angle\{z, y\})| \leq \sin(\angle\{x, y\}), \quad (2)$$

and

$$|\cos(\angle\{z, x\}) - \cos(\angle\{z, y\})| \leq \sin(\angle\{x, y\}).$$

Also

$$|\cos^2(\angle\{z, x\}) - \cos^2(\angle\{z, y\})| = |\sin^2(\angle\{z, x\}) - \sin^2(\angle\{z, y\})| \leq \sin(\angle\{x, y\}).$$

Inequality (1) implies inequality (2), since the sine function is increasing and sub-additive.

The Gap – Notion of Distance Between Subspaces

Let \mathcal{F} and \mathcal{G} be p -dimensional subspaces of \mathbf{C}^n with $1 \leq p < n$. Then the largest principal angle is related to the notion of distance, or a gap, between equidimensional subspaces. Let

$$0 \leq \theta_1 \leq \cdots \leq \theta_p \leq \frac{\pi}{2},$$

be the principal angles between equidimensional subspaces \mathcal{F} and \mathcal{G} . Then the distance is defined as

$$\text{gap}(\mathcal{F}, \mathcal{G}) = \|P_{\mathcal{F}} - P_{\mathcal{G}}\|_2 = \sin(\theta_p) = \sqrt{1 - (\cos(\theta_p))^2},$$

where $P_{\mathcal{F}}$ and $P_{\mathcal{G}}$ are orthogonal projectors onto \mathcal{F} and \mathcal{G} , respectively.

Perturbation Analysis – Sine

Lemma 1 *Let $\mu_1 \leq \mu_2 \leq \dots \leq \mu_q$ and $\tilde{\mu}_1 \leq \tilde{\mu}_2 \leq \dots \leq \tilde{\mu}_q$ be sine of principal angles between subspaces \mathcal{F} , \mathcal{G} , and \mathcal{F} , $\tilde{\mathcal{G}}$, respectively, computed in the A -based scalar product. Then, for $k = 1, \dots, q$,*

$$|\mu_k - \tilde{\mu}_k| \leq \max\{\sin(\theta_{\max}\{(\mathcal{G} + \tilde{\mathcal{G}}) \ominus \mathcal{G}, \mathcal{F}\}); \sin(\theta_{\max}\{(\mathcal{G} + \tilde{\mathcal{G}}) \ominus \tilde{\mathcal{G}}, \mathcal{F}\})\} \text{gap}_A(\mathcal{G}, \tilde{\mathcal{G}}),$$

where θ_{\max} is the largest angle between corresponding subspaces, measured in the A -based scalar product.

We use the notation \ominus for example:

$$(\mathcal{G} + \tilde{\mathcal{G}}) \ominus \mathcal{G} = (\mathcal{G} + \tilde{\mathcal{G}}) \cap \mathcal{G}^\perp,$$

where \ominus and the orthogonal complement to \mathcal{G} are understood in the A -based scalar product.

Perturbation Analysis – Cosine

Lemma 2 *Let $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_q$ and $\tilde{\sigma}_1 \geq \tilde{\sigma}_2 \geq \dots \geq \tilde{\sigma}_q$ be cosine of principal angles between subspaces \mathcal{F}, \mathcal{G} and $\mathcal{F}, \tilde{\mathcal{G}}$, respectively, computed in the A -based scalar product. Then, for $k = 1, \dots, q$,*

$$|\sigma_k - \tilde{\sigma}_k| \leq \max\{\cos(\theta_{\min}\{(\mathcal{G} + \tilde{\mathcal{G}}) \ominus \mathcal{G}, \mathcal{F}\}); \cos(\theta_{\min}\{(\mathcal{G} + \tilde{\mathcal{G}}) \ominus \tilde{\mathcal{G}}, \mathcal{F}\})\} \text{gap}_A(\mathcal{G}, \tilde{\mathcal{G}}),$$

where θ_{\min} is the smallest angle between corresponding subspaces, measured in the A -based scalar product.

Perturbation Analysis – Known Results

Let us also highlight that simpler estimates,

$$|\mu_k - \tilde{\mu}_k| \leq \text{gap}_A(\mathcal{G}, \tilde{\mathcal{G}}), \quad |\sigma_k - \tilde{\sigma}_k| \leq \text{gap}_A(\mathcal{G}, \tilde{\mathcal{G}}), \quad k = 1, \dots, q,$$

which are not as sharp as those we prove in Lemmas 2 and 1, can be derived almost trivially using orthoprojectors (see Wedin [1983]; Sun [1987]; Golub and Zha [1994]), where this approach is used for the case $A = I$.

2. Majorization, Unitarily Invariant Norms and Symmetric Gauge Functions

Let $x, y \in \mathbf{R}^n$ be given real vectors, and denote their algebraically decreasing ordered entries by $x_{[1]} \geq \cdots \geq x_{[n]}$ and $y_{[1]} \geq \cdots \geq y_{[n]}$. Then we say that y weakly majorizes x if

$$\sum_{i=1}^k x_{[i]} \leq \sum_{i=1}^k y_{[i]}, \quad k = 1, \dots, n,$$

and we use the notation $[x_1, \dots, x_n] \prec_w [y_1, \dots, y_n]$ or $x \prec_w y$. We denote $[|x_1|, \dots, |x_n|]$ by $|x|$.

Majorization, Unitarily Invariant Norms and Symmetric Gauge Functions (Continued)

There is a relationship between symmetric gauge functions and unitarily invariant norms. For $A \in \mathbf{C}^{n \times n}$, $\|A\|_{\Phi} = \|\text{diag}(\sigma(A))\|_{\Phi} = \Phi(\sigma(A))$.

Let $x, y \in \mathbf{R}^n$ be given real vectors. Then the following are equivalent:

$$|x| \prec_w |y|,$$

$$\Phi(x) \leq \Phi(y)$$

for every symmetric gauge function,

$$\|\text{diag}(x)\|_{\Phi} \leq \|\text{diag}(y)\|_{\Phi}$$

for every unitarily invariant norm.

3. Majorization and Perturbation of Principal Angles

Some Notation

Let \mathcal{G} and $\tilde{\mathcal{G}}$ be p -dimensional subspaces of \mathbf{C}^n with $1 \leq p < n$. Then

$$\Theta(\mathcal{G}, \tilde{\mathcal{G}}) = \{\angle_1\{\mathcal{G}, \tilde{\mathcal{G}}\}, \dots, \angle_p\{\mathcal{G}, \tilde{\mathcal{G}}\}\},$$

with $\angle_1\{\mathcal{G}, \tilde{\mathcal{G}}\} \leq \dots \leq \angle_p\{\mathcal{G}, \tilde{\mathcal{G}}\}$, e.g.,

$$\cos \Theta(\mathcal{G}, \tilde{\mathcal{G}}) = \{\cos \angle_1\{\mathcal{G}, \tilde{\mathcal{G}}\}, \dots, \cos \angle_p\{\mathcal{G}, \tilde{\mathcal{G}}\}\}.$$

Perturbation of Subspaces

Known Gap Bounds

Let \mathcal{F} , \mathcal{G} and $\tilde{\mathcal{G}}$ be p -dimensional subspaces of \mathbf{C}^n with $1 \leq p < n$. Then

$$\|\text{diag} (\cos \Theta(\mathcal{F}, \mathcal{G}) - \cos \Theta(\mathcal{F}, \tilde{\mathcal{G}}))\|_2 \leq \|\text{diag} (\sin \Theta(\mathcal{G}, \tilde{\mathcal{G}}))\|_2$$

and

$$\|\text{diag} (\sin \Theta(\mathcal{F}, \mathcal{G}) - \sin \Theta(\mathcal{F}, \tilde{\mathcal{G}}))\|_2 \leq \|\text{diag} (\sin \Theta(\mathcal{G}, \tilde{\mathcal{G}}))\|_2.$$

The gap bound: $\|\text{diag} (\sin \Theta(\mathcal{G}, \tilde{\mathcal{G}}))\|_2 = \text{gap}(\mathcal{G}, \tilde{\mathcal{G}})$.

Known Perturbation Theorems for Angles

Theta Theorem

This result involves the angles themselves:

Theorem 5 (Qiu et al. [2004]) *Let \mathcal{F} , \mathcal{G} and $\tilde{\mathcal{G}}$ be p -dimensional subspaces of \mathbf{C}^n with $1 \leq p < n$. Then*

$$\Phi(\Theta(\mathcal{F}, \mathcal{G}) - \Theta(\mathcal{F}, \tilde{\mathcal{G}})) \leq \Phi(\Theta(\mathcal{G}, \tilde{\mathcal{G}}))$$

for all symmetric gauge functions.

This is equivalent to

$$|\Theta(\mathcal{F}, \mathcal{G}) - \Theta(\mathcal{F}, \tilde{\mathcal{G}})| \prec_w \Theta(\mathcal{G}, \tilde{\mathcal{G}}).$$

For $p = 1$ this theorem implies the sine perturbation inequality (2).

Known Perturbation Theorems for Angles

Cosine/Sine Theorem

Theorem 6 (Sun [1987]) *Let \mathcal{F} , \mathcal{G} and $\tilde{\mathcal{G}}$ be p -dimensional subspaces of \mathbf{C}^n with $1 \leq p < n$, and let $P_{\mathcal{G}}$ and $P_{\tilde{\mathcal{G}}}$ be the corresponding orthogonal projectors, respectively. Then*

$$\|\text{diag} (\cos \Theta(\mathcal{F}, \mathcal{G}) - \cos \Theta(\mathcal{F}, \tilde{\mathcal{G}}))\| \leq \|P_{\mathcal{G}} - P_{\tilde{\mathcal{G}}}\|$$

and

$$\|\text{diag} (\sin \Theta(\mathcal{F}, \mathcal{G}) - \sin \Theta(\mathcal{F}, \tilde{\mathcal{G}}))\| \leq \|P_{\mathcal{G}} - P_{\tilde{\mathcal{G}}}\|$$

for every unitarily invariant norm.

This Theorem will lead to cosine and sine majorization results with a constant equal to 2.

Singular Values of $P_{\mathcal{G}} - P_{\tilde{\mathcal{G}}}$

We state the following Theorem in a slightly different form than in Stewart and Sun [1990]:

Theorem 7 (Stewart and Sun [1990]) *Let \mathcal{G} and $\tilde{\mathcal{G}}$ both be p -dimensional subspaces of \mathbf{C}^n . Let $P_{\mathcal{G}}$ and $P_{\tilde{\mathcal{G}}}$ be the corresponding orthogonal projectors, respectively. Let $l = \dim\{\mathcal{G} \cap \tilde{\mathcal{G}}\}$. Then the singular values of $P_{\mathcal{G}} - P_{\tilde{\mathcal{G}}}$ are*

$$\sin \angle_p\{\mathcal{G}, \tilde{\mathcal{G}}\}, \sin \angle_p\{\mathcal{G}, \tilde{\mathcal{G}}\}, \dots, \sin \angle_{l+1}\{\mathcal{G}, \tilde{\mathcal{G}}\}, \sin \angle_{l+1}\{\mathcal{G}, \tilde{\mathcal{G}}\}, 0, \dots, 0.$$

Corollary 1 *Under the assumptions of Theorem 7, we have*

$$\|P_{\mathcal{F}} - P_{\mathcal{G}}\| \leq 2 \|\text{diag}(\sin \Theta(\mathcal{G}, \tilde{\mathcal{G}}))\|$$

for every unitarily invariant norm.

Perturbation Theorems for Angles

Cosine/Sine Theorem

By Theorem 6 and Corollary 1 we have

$$|\cos \Theta(\mathcal{F}, \mathcal{G}) - \cos \Theta(\mathcal{F}, \tilde{\mathcal{G}})| \prec_w 2 \sin \Theta(\mathcal{G}, \tilde{\mathcal{G}})$$

and

$$|\sin \Theta(\mathcal{F}, \mathcal{G}) - \sin \Theta(\mathcal{F}, \tilde{\mathcal{G}})| \prec_w 2 \sin \Theta(\mathcal{G}, \tilde{\mathcal{G}}).$$

New Perturbation Theorems for Angles

Cosine/Sine Theorem

We prove a theorem with an improved constant (Knyazev and Argentati [2004]).

Theorem 8

$$|\cos \Theta(\mathcal{F}, \mathcal{G}) - \cos \Theta(\mathcal{F}, \tilde{\mathcal{G}})| \prec_w 2 \sin \left(\frac{\Theta(\mathcal{G}, \tilde{\mathcal{G}})}{2} \right),$$

$$|\sin \Theta(\mathcal{F}, \mathcal{G}) - \sin \Theta(\mathcal{F}, \tilde{\mathcal{G}})| \prec_w 2 \sin \left(\frac{\Theta(\mathcal{G}, \tilde{\mathcal{G}})}{2} \right).$$

Corollary 2

$$|\cos \Theta(\mathcal{F}, \mathcal{G}) - \cos \Theta(\mathcal{F}, \tilde{\mathcal{G}})| \prec_w \sqrt{2} \sin \Theta(\mathcal{G}, \tilde{\mathcal{G}}),$$

$$|\sin \Theta(\mathcal{F}, \mathcal{G}) - \sin \Theta(\mathcal{F}, \tilde{\mathcal{G}})| \prec_w \sqrt{2} \sin \Theta(\mathcal{G}, \tilde{\mathcal{G}}).$$

New Perturbation Theorems for Angles Cosine/Sine Squared Theorem

Knyazev and Argentati [2005] have proven the following theorem.

Theorem 9 *Let \mathcal{F} be a p -dimensional subspace of \mathbf{C}^n and let \mathcal{G} and $\tilde{\mathcal{G}}$ be q -dimensional subspaces of \mathbf{C}^n , with $1 \leq p < n$ and $1 \leq q < n$. Let $m = \min\{p, q\}$. Then*

$$|\cos^2 \Theta(\mathcal{F}, \mathcal{G}) - \cos^2 \Theta(\mathcal{F}, \tilde{\mathcal{G}})| \prec_w \sin \Theta(\mathcal{G}, \tilde{\mathcal{G}})$$

and

$$|\sin^2 \Theta(\mathcal{F}, \mathcal{G}) - \sin^2 \Theta(\mathcal{F}, \tilde{\mathcal{G}})| \prec_w \sin \Theta(\mathcal{G}, \tilde{\mathcal{G}}).$$

4. Changes in the Trial Subspace in the Rayleigh–Ritz Method

The analysis of the influence of changes in a trial subspace in the Rayleigh–Ritz method provides a natural application of the theory concerning principal angles.

Let $A \in \mathbf{C}^{n \times n}$ be a Hermitian matrix and let \mathcal{X} be an q -dimensional subspace of \mathbf{C}^n . We can define an operator $\tilde{A} = P_{\mathcal{X}}A|_{\mathcal{X}}$ on \mathcal{X} , where $P_{\mathcal{X}}$ is the orthogonal projection onto \mathcal{X} and $P_{\mathcal{X}}A|_{\mathcal{X}}$ denotes the restriction of $P_{\mathcal{X}}A$ to \mathcal{X} , as discussed in Parlett [1998]. The eigenvalues of \tilde{A} are called Ritz values, $\alpha_1 \geq \cdots \geq \alpha_q$.

The Ritz values are also the eigenvalues of $Q_{\mathcal{X}}^*AQ_{\mathcal{X}}$, and the nonzero Ritz values are the nonzero eigenvalues of $P_{\mathcal{X}}AP_{\mathcal{X}}$.

Mirsky's Theorem

- For a fixed Hermitian matrix we vary the subspace and see how the Ritz values change.
- An analogous problem is to vary the Hermitian matrix and see how the eigenvalues change.

A generalization of Weyl's perturbation theorem for Hermitian matrices is the following:

Theorem 10 (Mirsky [1960]) *Let $A, B \in \mathbf{C}^{n \times n}$ be Hermitian matrices and let $\alpha_1 \geq \dots \geq \alpha_n$ and $\beta_1 \geq \dots \geq \beta_n$ denote the eigenvalues for A and B , respectively. Then*

$$\|\text{diag}(\alpha_1 - \beta_1, \dots, \alpha_n - \beta_n)\| \leq \|A - B\|$$

for every unitarily invariant norm.

Ritz Values of an Orthogonal Projector

Lemma 3 *Let $A \in \mathbf{C}^{n \times n}$ be an orthogonal projector onto the p -dimensional subspace \mathcal{Z} of \mathbf{C}^n , and let \mathcal{X} be a q -dimensional subspace of \mathbf{C}^n . Let $\alpha_1 \geq \dots \geq \alpha_q$ be the Ritz values for A with respect to \mathcal{X} , and let $m = \min\{p, q\}$. Then $\alpha_i = \sigma_i^2$, $i = 1, \dots, m$, with the remaining $q - m$ Ritz values being zero, where σ_i , $i = 1, \dots, m$ are the cosines of the principal angles between \mathcal{X} and \mathcal{Z} .*

The key point is that the Ritz values are the eigenvalues of

$$Q_{\mathcal{X}}^* A Q_{\mathcal{X}} = Q_{\mathcal{X}}^* P_{\mathcal{Z}} Q_{\mathcal{X}} = Q_{\mathcal{X}}^* Q_{\mathcal{Z}} (Q_{\mathcal{X}}^* Q_{\mathcal{Z}})^*$$

Proximity Theorem for an Orthogonal Projector

Lemma 4 *Let $A \in \mathbf{C}^{n \times n}$ be an orthogonal projector and let \mathcal{X} and \mathcal{Y} both be q -dimensional subspaces of \mathbf{C}^n , and let $\alpha_1 \geq \dots \geq \alpha_q$ and $\beta_1 \geq \dots \geq \beta_q$ denote the Ritz values for A with respect to \mathcal{X} and \mathcal{Y} . Then*

$$\|\text{diag}(\alpha_1 - \beta_1, \dots, \alpha_q - \beta_q)\| \leq \|\text{diag}(\sin \Theta(\mathcal{X}, \mathcal{Y}))\|$$

for every unitarily invariant norm.

This follows from the cosine squared proximity inequality – Theorem 9.

Changes in the Trial Subspace in the Rayleigh–Ritz Method

One of the main results of (Argentati [2003]; Knyazev and Argentati [2004]) is the following theorem.

Theorem 11 *Let \mathcal{X} and \mathcal{Y} both be q -dimensional subspaces of \mathbf{R}^n , and $\alpha_1 \geq \dots \geq \alpha_q$ and $\beta_1 \geq \dots \geq \beta_q$ denote the Ritz values for A with respect to \mathcal{X} and \mathcal{Y} , i.e. α 's and β 's are the stationary values of the the Rayleigh quotient on subspaces \mathcal{X} and \mathcal{Y} , correspondingly. Then*

$$\max_{j=1,\dots,q} |\alpha_j - \beta_j| \leq (\lambda_{\max} - \lambda_{\min}) \text{gap}(\mathcal{X}, \mathcal{Y}).$$

Changes in the Trial Subspace in the Rayleigh–Ritz Method

One of the key results of Knyazev and Argentati [2005], that also involves unitarily invariant norms, is given below.

Theorem 12 *Let $A \in \mathbf{C}^{n \times n}$ be a Hermitian matrix and let \mathcal{X} and \mathcal{Y} both be q -dimensional subspaces of \mathbf{C}^n , and let $\alpha_1 \geq \dots \geq \alpha_q$ and $\beta_1 \geq \dots \geq \beta_q$ denote the Ritz values for A with respect to \mathcal{X} and \mathcal{Y} . Then*

$$\|\text{diag}(\alpha_1 - \beta_1, \dots, \alpha_q - \beta_q)\| \leq (\lambda_{\max} - \lambda_{\min}) \|\text{diag}(\sin \Theta(\mathcal{X}, \mathcal{Y}))\|$$

for every unitarily invariant norm, and this is equivalent to

$$|\alpha - \beta| \prec_w (\lambda_{\max} - \lambda_{\min}) \sin \Theta(\mathcal{X}, \mathcal{Y}).$$

5. Principal Angles in a Hilbert Space

Let \mathcal{F} and \mathcal{G} be subspaces in a Hilbert space H . Then the (*smallest discrete*) *principal angles* $\theta_1, \dots, \theta_q \in [0, \pi/2]$ between \mathcal{F} and \mathcal{G} may be defined recursively for $k = 1, \dots, q$, where $q = \min\{\dim\mathcal{F}; \dim\mathcal{G}\}$, by

$$\cos(\theta_k) = \sup_{u \in \mathcal{F}} \sup_{v \in \mathcal{G}} |(u, v)| = |(u_k, v_k)|$$

subject to

$$\|u\| = \|v\| = 1, \quad u^T u_i = 0, \quad v^T v_i = 0, \quad i = 1, \dots, k-1,$$

provided the existence of the pair of vectors u_k, v_k . If for some $k = n$ the pair u_k, v_k does not exist, we set $\theta_k = \theta_n$ for all $k > n$.

An Equivalent Definition Based on s-Numbers

s-numbers (generalization of singular values) of a bounded operator A are defined as

$$s_{k+1}(A) = \min_{A_k, \text{rank}(A_k)=k} \|A - A_k\|.$$

Let P_F and P_G be orthoprojectors on subspaces \mathcal{F} and \mathcal{G} . Then

$$s_k(P_F P_G) = s_k(P_G P_F) = \cos(\theta_k).$$

Perturbation Analysis – Cosine

Lemma 5 For $k = 1, \dots, q$,

$$|\cos(\theta_k) - \cos(\tilde{\theta}_k)| \leq \max\{\cos(\theta_{\min}\{\overline{(\mathcal{G} + \tilde{\mathcal{G}})} \ominus \mathcal{G}, \mathcal{F}\}); \cos(\theta_{\min}\{\overline{(\mathcal{G} + \tilde{\mathcal{G}})} \ominus \tilde{\mathcal{G}}, \mathcal{F}\})\} \|P_{\mathcal{G}} - P_{\tilde{\mathcal{G}}}\|.$$

It is known that the sum $\mathcal{F} + \mathcal{G}$ of subspaces \mathcal{F} and \mathcal{G} is closed if $\theta_{\min}\{\mathcal{F}, \mathcal{G}\} > 0$, thus we have

Corollary 3 Let $\theta_{\min}\{\mathcal{F}, \mathcal{G}\} > 0$ so that $\mathcal{F} + \mathcal{G}$ is closed and let $\|P_{\mathcal{G}} - P_{\tilde{\mathcal{G}}}\| < 1 - \cos(\theta_{\min}\{\mathcal{F}, \mathcal{G}\})$. Then $\theta_{\min}\{\mathcal{F}, \tilde{\mathcal{G}}\} > 0$ and $\mathcal{F} + \tilde{\mathcal{G}}$ is also closed.

6. Conclusions (Part II)

- Perturbation estimates for absolute errors in cosine and sine of principal angles, with improved constants and for an arbitrary scalar product, are derived.
- We present a new perturbation where the absolute value of the difference of the cosines/sines are majorized by the sines of the angles between the perturbed subspaces, with a constant equal to $\sqrt{2}$. We prove a similar theorem for squares of the cosines/sines, with a constant of one.
- The cosine squared formula is equivalent to a majorization inequality for Ritz values for an orthogonal projector. We then prove the theorem for a Hermitian operator, with a factor $\lambda_{\max} - \lambda_{\min}$.

References

- Merico E. Argentati. *Principal Angles Between Subspaces as Related to Rayleigh Quotient and Rayleigh Ritz Inequalities with Applications to Eigenvalue Accuracy and an Eigenvalue Solver*. PhD thesis, University of Colorado at Denver, 2003.
- Åke Björck and Gene H. Golub. Numerical methods for computing angles between linear subspaces. *Math. Comp.*, 27:579–594, 1973.
- C. Davis and W. M. Kahan. The rotation of eigenvectors by a perturbation. III. *SIAM J. Numer. Anal.*, 7(1):1–46, 1970.
- Gene H. Golub and Hong Yuan Zha. Perturbation analysis of the canonical correlations of matrix pairs. *Linear Algebra Appl.*, 210:3–28, 1994. ISSN 0024-3795.
- A. V. Knyazev and M. E. Argentati. On proximity of Rayleigh quotients for different vectors and Ritz values generated by different trial subspaces. *Linear Algebra and its Applications*, *accepted*, 2004.
- A. V. Knyazev and M. E. Argentati. Majorization for changes in principal angles between subspaces. *Draft Paper*, 2005.

- Ren-Cang Li. Relative perturbation theory. II. Eigenspace and singular subspace variations. *SIAM J. Matrix Anal. Appl.*, 20(2):471–492 (electronic), 1999. ISSN 1095-7162.
- L. Mirsky. Symmetric gauge functions and unitarily invariant norms. *Quarterly Journal of Mathematics. Oxford Series(2)*, 11:50–59, 1960.
- B. N. Parlett. *The symmetric eigenvalue problem*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1998. ISBN 0-89871-402-8. Corrected reprint of the 1980 original.
- Li Qiu, Yanxia Zhang, and Chi-Kwong Li. Unitarily invariant metrics on the Grassmann space. *SIAM Journal on Matrix Analysis and Applications*, submitted, 2004.
- G. W. Stewart and J. G. Sun. *Matrix perturbation theory*. Academic Press Inc., Boston, MA, 1990. ISBN 0-12-670230-6.
- Ji Guang Sun. Perturbation of angles between linear subspaces. *J. Comput. Math.*, 5(1): 58–61, 1987. ISSN 0254-9409.
- Charles Van Loan. Computing the CS and the generalized singular value decompositions. *Numer. Math.*, 46(4):479–491, 1985. ISSN 0029-599X.

P. A. Wedin. On angles between subspaces of a finite-dimensional inner product space. In Bo Kågström and Axel Ruhe, editors, *Matrix Pencils. Proceedings of the Conference Held at Pite Havsbud, March 22–24, 1982*, pages 263–285. Springer-Verlag, Berlin, 1983. ISBN 3-540-11983-3.