Majorization for Changes in Principal Angles Between Subspaces

Andrew V. Knyazev and Merico E. Argentati (speaker)

Department of Mathematics and

Center for Computational Mathematics

University of Colorado at Denver

Outline

1. Introduction and some definitions
2. A new inequality for perturbation of principal angles
3. Changes in the trial subspace in the Rayleigh Ritz Method
4. Conclusions

Why Are We Interested in Angles Between Subspaces?

1. Characterize distance or gap between subspaces
2. Statistics (canonical correlations), information retrieval, image processing, signal processing
3. Optimization, robust control, system identification
4. Rayleigh Ritz Method (analysis of the influence of changes in a trial subspace on Ritz values)
5. Grassmannian manifolds (chordal distance metric is the Frobenius norm of the sines of the principal angles)

Definition of Principal Angles Between Subspaces

Let $\mathbf{M^n}$ be a complex n–dimensional vector space. Let $\mathcal{F}$ and $\mathcal{G}$ be m–dimensional subspaces of $\mathbf{M^n}$ with $1 \leq m < n$. Then the *principal angles*

$$\theta_1, \ldots, \theta_m \in [0, \pi/2]$$

between $\mathcal{F}$ and $\mathcal{G}$ may be defined, e.g., [5, 2, 9], recursively for $k = 1, \ldots, m$ by

$$\sigma_k = \cos(\theta_k) = \max_{u\, \in \mathcal{F}} \max_{v\, \in \mathcal{G}} \; |u^* v| \;\; = \;\; |u_k^* v_k|$$

subject to

$$\|u\| = \|v\| = 1, \qquad u^* u_i = 0, \qquad v^* v_i = 0, \qquad i = 1, \ldots, k-1.$$

Definition of Principal Angles Between Subspaces

We use the notation

$$\cos(\angle_k\{\mathcal{F},\mathcal{G}\}) = \sigma_k,\ k = 1,\ldots,m,$$

where $1 \geq \sigma_1 \geq \cdots \geq \sigma_m \geq 0$, and the sines of the principal angles are given by

$$\mu_1 \leq \mu_2 \leq \cdots \leq \mu_m,$$

where $\mu_k = \sqrt{1-\sigma_i^2},\ k = 1,\ldots,m.$

Let the columns of the matrices $Q_{\mathcal{F}}, Q_{\mathcal{G}} \in \mathbf{M^{n\times m}}$ form orthonormal bases for the subspaces $\mathcal{F}$ and $\mathcal{G}$, respectively. Then the σ's are the m singular values of $Q_{\mathcal{F}}^* Q_{\mathcal{G}}$.

Properties of Principal Angles Between Subspaces

Let $P_{\mathcal{F}}$ and $P_{\mathcal{G}}$ be the orthogonal projectors onto the subspaces $\mathcal{F}$ and $\mathcal{G}$, respectively. Then the singular values of $Q_{\mathcal{F}}^{*}Q_{\mathcal{G}}$ are the m largest singular values of $P_{\mathcal{F}}P_{\mathcal{G}}$. Thus the singular values of $P_{\mathcal{F}}P_{\mathcal{G}}$, in decreasing order, are

$$\sigma_1, \sigma_2, \cdots, \sigma_m, 0, \cdots, 0.$$

Let $\mu_1 \leq \mu_2 \leq \cdots \leq \mu_m$ be the sines of the principal angles between $\mathcal{F}$ and $\mathcal{G}$, and let $l = \dim\{\mathcal{F} \bigcap \mathcal{G}\}$. Then the singular values of $P_{\mathcal{F}} - P_{\mathcal{G}}$, in decreasing order, are

$$\mu_m, \mu_m, \cdots, \mu_{l+1}, \mu_{l+1}, 0, \cdots, 0.$$

Properties of Principal Angles Between Subspaces

The singular values of $(I - P_{\mathcal{F}})P_{\mathcal{G}} = P_{\mathcal{F}^\perp} P_{\mathcal{G}}$, in decreasing order, are

$$\mu_m, \cdots, \mu_1, 0, \cdots, 0.$$

$P_{\mathcal{F}}$ and $P_{\mathcal{G}}$ play symmetric roles in all cases.

The largest principal angle is related to the notion of distance, or a gap, between equidimensional subspaces, and the distance is defined [7] as

$$\operatorname{gap}(\mathcal{F}, \mathcal{G}) = \|P_{\mathcal{F}} - P_{\mathcal{G}}\| = \sin(\angle\{\mathcal{F}, \mathcal{G}\}) = \sin(\angle_m\{\mathcal{F}, \mathcal{G}\}) = \mu_m$$

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

$$0 \leq \cos(\theta) - \cos(\theta + \epsilon) \leq \sin(\theta + \epsilon)\sin(\epsilon) \leq \sin(\epsilon), \tag{1}$$

$$0 \leq \sin(\theta + \epsilon) - \sin(\theta) \leq \cos(\theta)\sin(\epsilon) \leq \sin(\epsilon), \tag{2}$$

$$0 \leq \cos^2(\theta) - \cos^2(\theta + \epsilon) = \sin(2\theta + \epsilon)\sin(\epsilon) \leq \sin(\epsilon), \tag{3}$$

$$0 \leq \sin^2(\theta + \epsilon) - \sin^2(\theta) = \sin(2\theta + \epsilon)\sin(\epsilon) \leq \sin(\epsilon). \tag{4}$$

The One Dimensional Case

Let $x, y, z \in \mathbf{M^n}$ and

$$\cos(\angle\{x, y\}) = \frac{|(x, y)|}{\|x\|\|y\|}. \tag{5}$$

Then

$$|\cos(\angle\{z, x\}) - \cos(\angle\{z, y\})| \leq \sin(\angle\{x, y\}), \tag{6}$$

and

$$|\sin(\angle\{z, x\}) - \sin(\angle\{z, y\})| \leq \sin(\angle\{x, y\}). \tag{7}$$

Also

$$\begin{aligned} |\cos^2(\angle\{z, x\}) - \cos^2(\angle\{z, y\})| &= |\sin^2(\angle\{z, x\}) - \sin^2(\angle\{z, y\})| \\ &\leq \sin(\angle\{x, y\}). \end{aligned} \tag{8}$$

Perturbation of Subspaces

Let $\mathcal{F}$, $\mathcal{G}$ be m–dimensional subspaces of $\mathbf{M^n}$ and let $\tilde{\mathcal{G}}$ be an m–dimensional subspace of $\mathbf{M^n}$ that is a perturbation of $\mathcal{G}$. What can we say about the cosines and sines of the principal angles between $\mathcal{F}$ and $\mathcal{G}$ and between $\mathcal{F}$ and $\tilde{\mathcal{G}}$?

We know from [9] that for $k = 1, \ldots, m$,

$$|\sigma_k - \hat{\sigma}_k| \leq \sin(\angle\{\mathcal{G}, \tilde{\mathcal{G}}\}), \tag{9}$$

and

$$|\mu_k - \hat{\mu}_k| \leq \sin(\angle\{\mathcal{G}, \tilde{\mathcal{G}}\}). \tag{10}$$

Perturbation of Subspaces

QUESTION: Can inequalities be characterized that involve all the angles?

Suppose that

$$|\sigma_{i_1} - \hat{\sigma}_{i_1}| \geq \ldots \geq |\sigma_{i_k} - \hat{\sigma}_{i_k}|$$

Is it true that for $k = 1, \ldots, m$,

$$|\sigma_{i_k} - \hat{\sigma}_{i_k}| \leq \sin(\angle_{m-k+1}\{\mathcal{G}, \tilde{\mathcal{G}}\})?$$

This is not true. However, majorization theory provides a tool that can be used to obtain some very general inequalities that involve unitarily invariant norms.

Perturbation of Subspaces

Let

$$C = \mathrm{diag}\ (\sigma_1, \ldots, \sigma_m), \qquad \hat{C} = \mathrm{diag}\ (\hat{\sigma}_1, \ldots, \hat{\sigma}_m),$$

and

$$S = \mathrm{diag}\ (\mu_1, \ldots, \mu_m), \qquad \hat{S} = \mathrm{diag}\ (\hat{\mu}_1, \ldots, \hat{\mu}_m).$$

Then (9) and (10) are equivalent to

$$\|C - \hat{C}\|_2 \leq \|\mathrm{diag}\ (\sin(\angle_1\{\mathcal{G}, \tilde{\mathcal{G}}\}), \ldots, \sin(\angle_m\{\mathcal{G}, \tilde{\mathcal{G}}\}))\|_2$$

and

$$\|S - \hat{S}\|_2 \leq \|\mathrm{diag}\ (\sin(\angle_1\{\mathcal{G}, \tilde{\mathcal{G}}\}), \ldots, \sin(\angle_m\{\mathcal{G}, \tilde{\mathcal{G}}\}))\|_2.$$

Majorization and Unitarily Invariant Norms

Let $x = [x_i],\ y = [y_i] \in \mathbf{R^n}$ be given vectors, and denote their algebraically decreasing ordered entries by $x_{i_1} \geq \cdots \geq x_{i_n}$ and $y_{i_1} \geq \cdots \geq y_{i_n}$. Then we say y weakly majorizes x if

$$\sum_{j=1}^{k} x_{i_j} \leq \sum_{j=1}^{k} y_{i_j}, \qquad k = 1, \ldots, n.$$

If in addition to satisfying the above equation, we have

$$\sum_{i=1}^{n} x_i = \sum_{j=1}^{n} y_i,$$

then we say y (strongly) majorizes x. We use the notation

$$[x_1, \cdots, x_n] \prec_w [y_1, \cdots, y_n]$$

to indicate that y weakly majorizes x.

Majorization and Unitarily Invariant Norms

1. Majorization inequalities [3] are important in characterizing the relationship between the main diagonal entries and eigenvalues of a Hermitian matrix, and have an important connection to unitarily invariant norms.

2. First proved by Schur, the eigenvalues of a Hermitian matrix majorizes its diagonal entries [4]. Also the singular values of a general matrix weakly majorizes the absolute values of its diagonal entries [11].

Majorization and Unitarily Invariant Norms

Let $A \in \mathbf{M^{n\times m}}$, then a unitarily invariant matrix norm $\|\cdot\|$, satisfies $\|UAV\| = \|A\|$, for all unitary matrices $U \in \mathbf{M^{n\times n}}$ and $V \in \mathbf{M^{m\times m}}$.

Some examples of unitarily invariant norms include the 2–norm, the Frobenius norm, the trace norm, p–norms ($p \geq 1$) and Ky Fan k norms [4][Theorem 7.4.45]. If $A \in \mathbf{M^{n\times m}}$ is a matrix, the singular values of A are denoted by $s_1(A) \geq \cdots \geq s_q(A)$, where $q = \min\{n, m\}$. Some of the commonly recognized unitarily invariant norms include

$$\|A\|_2 = s_1(A), \qquad \|A\|_F = \sqrt{\sum_{i=1}^{q} s_i^2(A)}, \qquad \|A\|_{\mathrm{tr}} = \sum_{i=1}^{q} s_i(A),$$

where $\|A\|_2$, $\|A\|_F$ and $\|A\|_{\mathrm{tr}}$ are respectively, the 2–norm, the Frobenius norm and the trace norm.

Majorization and Unitarily Invariant Norms

Theorem 1 *Let* $A, B \in \mathbf{M^{n\times m}}$ *be given matrices with respective singular values* $s_1(A) \geq \cdots \geq s_q(A)$ *and* $s_1(B) \geq \cdots \geq s_q(B)$*, where* $q = \min\{m, n\}$*. In order that* $\|A\| \leq \|B\|$ *for every unitarily invariant norm* $\|\cdot\|$ *on* $\mathbf{M^{n\times m}}$*, it is sufficient that*

$$s_i(A) \leq s_i(B), \qquad i = 1, \ldots, q$$

and it is necessary and sufficient that

$$\sum_{i=1}^{k} s_i(A) \leq \sum_{i=1}^{k} s_i(B), \qquad k = 1, \ldots, n, \tag{11}$$

that is the singular values of B *must weakly majorize the singular values of* A*.*

Perturbation Theorems for Angles

Ji-Guang [13] has proven the following theorem.

Theorem 2 *Let $\mathcal{F}$, $\mathcal{G}$ and $\tilde{\mathcal{G}}$ be a m–dimensional subspaces of $\mathbf{M^n}$ with $1 \le m < n$. Let $C = \mathrm{diag}\,(\sigma_1, \cdots, \sigma_m)$ be the cosines of the principal angles between $\mathcal{F}$ and $\mathcal{G}$ and let $S = \mathrm{diag}\,(\mu_1, \cdots, \mu_m)$ be the sines of the principal angles between $\mathcal{F}$ and $\mathcal{G}$. Let $\hat{C} = \mathrm{diag}\,(\hat{\sigma}_1, \cdots, \hat{\sigma}_m)$ be the cosines of the principal angles between $\mathcal{F}$ and $\tilde{\mathcal{G}}$ and let $\hat{S} = \mathrm{diag}\,(\hat{\mu}_1, \cdots, \hat{\mu}_m)$ be the sines of the principal angles between $\mathcal{F}$ and $\tilde{\mathcal{G}}$. Then*

$$\|C - \hat{C}\| \le \|P_{\mathcal{G}} - P_{\tilde{\mathcal{G}}}\|$$

and

$$\|S - \hat{S}\| \le \|P_{\mathcal{G}} - P_{\tilde{\mathcal{G}}}\|$$

in an arbitrary unitarily invariant norm.

Perturbation Theorems for Angles

This theorem implies that

$$\|C - \hat{C}\| \leq 2\|\mathrm{diag}\ (\sin(\angle_1\{\mathcal{G}, \tilde{\mathcal{G}}\}), \ldots, \sin(\angle_m\{\mathcal{G}, \tilde{\mathcal{G}}\}))\|$$

and

$$\|S - \hat{S}\| \leq 2\|\mathrm{diag}\ (\sin(\angle_1\{\mathcal{G}, \tilde{\mathcal{G}}\}), \ldots, \sin(\angle_m\{\mathcal{G}, \tilde{\mathcal{G}}\}))\|$$

in an arbitrary unitarily invariant norm.

Perturbation Theorems for Angles

Knyazev and Argentati [8] have proven the following theorem.

Theorem 3 *Let $\mathcal{F}$, $\mathcal{G}$ and $\tilde{\mathcal{G}}$ be a m–dimensional subspaces of* $\mathbf{M^n}$ *with* $1 \leq m < n$. *Let* $C = \operatorname{diag}\,(\sigma_1, \cdots, \sigma_m)$ *be the cosines of the principal angles between* $\mathcal{F}$ *and* $\mathcal{G}$ *and let* $S = \operatorname{diag}\,(\mu_1, \cdots, \mu_m)$ *be the sines of the principal angles between* $\mathcal{F}$ *and* $\mathcal{G}$. *Let* $\hat{C} = \operatorname{diag}\,(\hat{\sigma}_1, \cdots, \hat{\sigma}_m)$ *be the cosines of the principal angles between* $\mathcal{F}$ *and* $\tilde{\mathcal{G}}$ *and let* $\hat{S} = \operatorname{diag}\,(\hat{\mu}_1, \cdots, \hat{\mu}_m)$ *be the sines of the principal angles between* $\mathcal{F}$ *and* $\tilde{\mathcal{G}}$. *Then*

$$\|C^2 - \hat{C}^2\| = \|S^2 - \hat{S}^2\| \leq \|\operatorname{diag}\,(\sin(\angle_1\{\mathcal{G}, \tilde{\mathcal{G}}\}), \ldots, \sin(\angle_m\{\mathcal{G}, \tilde{\mathcal{G}}\})\|$$

in an arbitrary unitarily invariant norm.

Asymptotic Inequalities

Theorem 3 implies the following results. As all principal angles approach zero, we have

$$\|C - \hat{C}\| \lesssim \frac{1}{2}\|\mathrm{diag}\ (\sin(\angle_1\{\mathcal{G}, \tilde{\mathcal{G}}\}), \ldots, \sin(\angle_m\{\mathcal{G}, \tilde{\mathcal{G}}\}))\|,$$

and as all principal angles approach ninety degrees, we have

$$\|S - \hat{S}\| \lesssim \frac{1}{2}\|\mathrm{diag}\ (\sin(\angle_1\{\mathcal{G}, \tilde{\mathcal{G}}\}), \ldots, \sin(\angle_m\{\mathcal{G}, \tilde{\mathcal{G}}\}))\|.$$

Changes in the Trial Subspace in the Rayleigh–Ritz Method

Concerning the perturbation of a subspace, an application is the analysis of the influence of changes in a trial subspace in the Rayleigh–Ritz method.

Let $A \in \mathbf{M^{n\times n}}$ be a Hermitian matrix and let $\mathcal{X}$ be an m–dimensional subspace of $\mathbf{M^n}$. 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 [12]. The eigenvalues of $\tilde{A}$ are called Ritz values, $\alpha_1 \geq \cdots \geq \alpha_m$.

The Ritz values are also the eigenvalues of

$$Q_{\mathcal{X}}^* A Q_{\mathcal{X}},$$

and the nonzero Ritz values are the nonzero eigenvalues of

$$P_{\mathcal{X}} A P_{\mathcal{X}}.$$

Changes in the Trial Subspace in the Rayleigh–Ritz Method

One of the main results of [1, 10] is the following theorem.

Theorem 4 *Let $\mathcal{X}$ and $\mathcal{Y}$ both be m–dimensional subspaces of $\mathbf{R^n}$, and $\alpha_1 \geq \cdots \geq \alpha_m$ and $\beta_1 \geq \cdots \geq \beta_m$ 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*

$$\max_{j=1,\ldots,m}|\alpha_j - \beta_j| \leq (\lambda_{\max} - \lambda_{\min}) \sin(\angle\{\mathcal{X}, \mathcal{Y}\}). \tag{12}$$

Changes in the Trial Subspace in the Rayleigh–Ritz Method

One of the key results of [10][Theorem 10], that also involves unitarily invariant norms, is given below.

Theorem 5 *Let* $A \in \mathbf{M^{n\times n}}$ *be a symmetric real-valued matrix and let* $\mathcal{X}$ *and* $\mathcal{Y}$ *both be* m*–dimensional subspaces of* $\mathbf{M^n}$*, and let* $\alpha_1 \geq \cdots \geq \alpha_m$ *and* $\beta_1 \geq \cdots \geq \beta_m$ *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*

$$\|\mathrm{diag}\ (\alpha_1, \ldots, \alpha_m) - \mathrm{diag}\ (\beta_1, \ldots, \beta_m)\| \leq$$
$$C\ \|\mathrm{diag}\ (\sin(\angle_1\{\mathcal{X}, \mathcal{Y}\}), \ldots, \sin(\angle_m\{\mathcal{X}, \mathcal{Y}\}))\|$$

in an arbitrary unitarily invariant norm, where $C = \sqrt{2}(\lambda_{\max} - \lambda_{\min})$.

Changes in the Trial Subspace in the Rayleigh–Ritz Method

Interestingly, we can use Theorem 3 to prove Theorem 5 with a reduced constant, if A is an orthogonal projector.

Theorem 6 *Under the assumptions of Theorem 5, assuming that A is an orthogonal projector, the $\sqrt{2}$ factor in the constant C can be eliminated. Thus $C = \lambda_{\max} - \lambda_{\min} = 1$ for this case.*

In fact, Theorem 3 is equivalent to Theorem 6.

Numerical Tests

We use a 4-dimensional vector space and 2-dimensional subspaces. There are two principal angles between each pair of subspaces. We compute the Ky–Fan N_k norms for $k = 1, 2$ and form the N_1 ratio

$$N_1 = \frac{\max\{|\alpha_1 - \beta_1|, |\alpha_2 - \beta_2|\}}{(\lambda_{\max} - \lambda_{\min}) \sin(\angle_2\{\mathcal{X}, \mathcal{Y}\})},$$

and the N_2 ratio

$$N_2 = \frac{|\alpha_1 - \beta_1| + |\alpha_2 - \beta_2|}{(\lambda_{\max} - \lambda_{\min})(\sin(\angle_1\{\mathcal{X}, \mathcal{Y}\}) + \sin(\angle_2\{\mathcal{X}, \mathcal{Y}\}))}.$$

We plot the largest N_1 (left) and N_2 (right) ratios for 400,000 trials, each trial involving a different random symmetric 4×4 matrix. For each trial (matrix), we vary the two principal angles from 0 to 90 degrees by 2 degree increments.

Numerical Tests

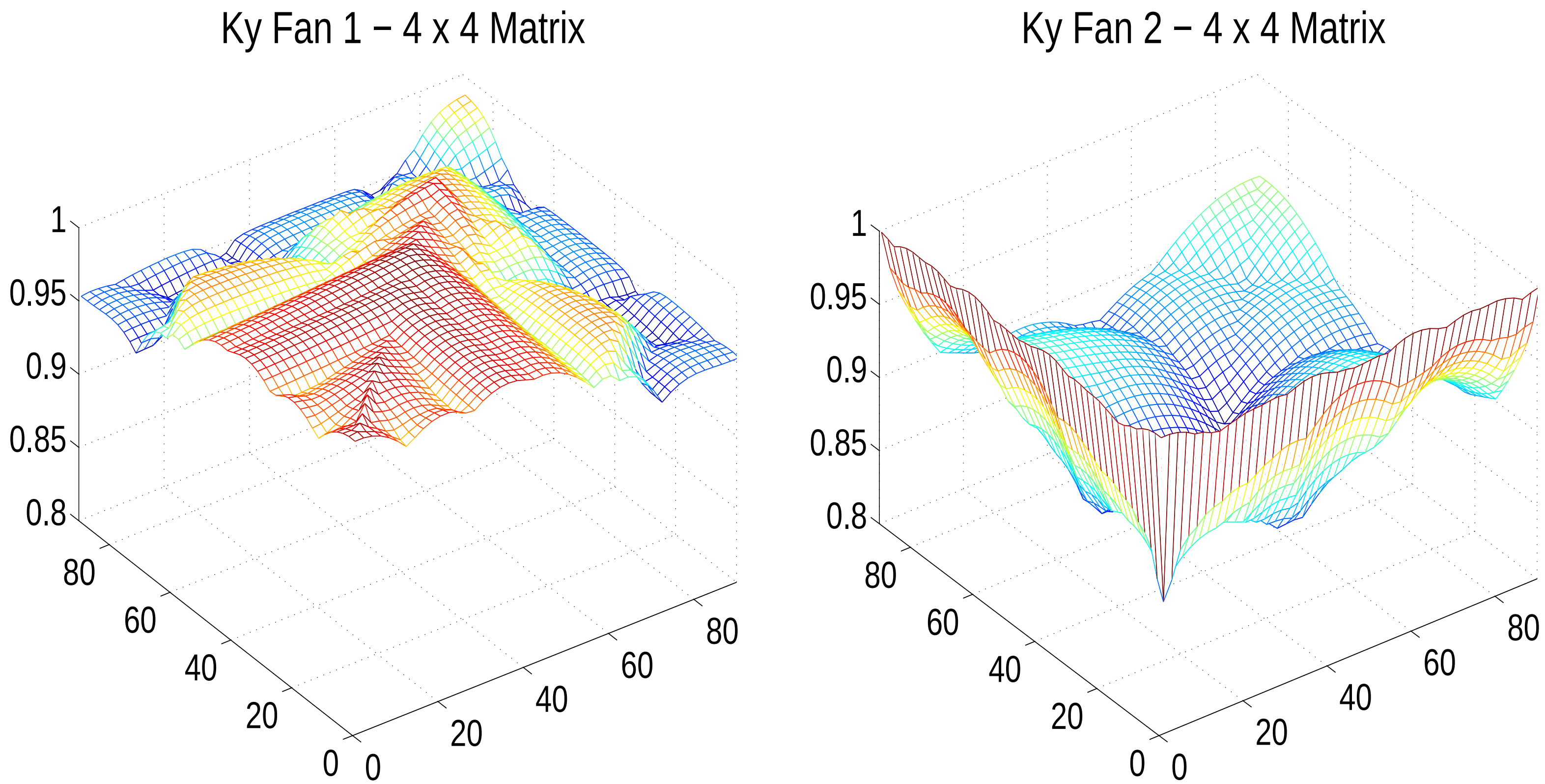

Figure 1: Ky–Fan N_1 (left) and N_2 (right) Ratios

Conclusions

1. We present a new result concerning perturbations of principal angles where the absolute value of the difference of the squares of the cosines/sines are majorized by the sines of the angles between the perturbed subspaces, with a constant of one.

2. We prove a new result that the absolute value of the perturbations in the Ritz values is bounded by a constant times the gap between the original trial subspace and its perturbation, and the constant is sharp.

3. We generalize this result to unitarily invariant norms, but we have to increase the constant by a factor $\sqrt{2}$.

4. Numerical results are consistent with our theorems and support our hypothesis that the $\sqrt{2}$ factor is artificial.

Conclusions

5. In general, utilizing all angles provides finer detail and information concerning smaller changes, and is more natural for some applications, e.g., in the analysis of Grassmannian manifolds [6], where the chordal distance is the Frobenius norm of the sine of the principal angles.

References

[1] 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.

[2] Gene H. Golub and Charles F. Van Loan. *Matrix Computations*. Johns Hopkins University Press, Baltimore, MD, third edition, 1996.

[3] G. H. Hardy, J. E. Littlewood, and G. Pólya. *Inequalities*. Cambridge University Press, Cambridge, 1959.

[4] R. A. Horn and C. R. Johnson. *Matrix analysis*. Cambridge University Press, Cambridge, 1990. Corrected reprint of the 1985 original.

[5] H. Hotelling. Relation between two sets of variables. *Biometrica*, 28:322–377, 1936.

[6] R. H. Hardin J. H. Conway and N. J. A. Sloane. Packing lines, planes, etc.: packings in Grassmannian spaces. *Experimental Mathematics*, 5(2):139–159, 1996.

[7] T. Kato. *Perturbation Theory for Linear Operators*. Springer–Verlag, New York, NY, corrected printing of the second edition edition, 1980.

[8] A. V. Knyazev and M. E. Argentati. Majorization for changes in principal angles between subspaces. *Draft Paper*.

[9] 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(6):2009–2041, 2002.

[10] A. V. Knyazev and M. E. Argentati. On proximity of Rayleigh quotients for different vectors and Ritz values generated by different trial subspaces. *Submitted to Linear Algebra and its Applications*, August 2004.

[11] A. W. Marshall and I. Olkin. *Inequalities: theory of majorization and its applications*, volume 143 of *Mathematics in Science and Engineering*. Academic Press Inc. [Harcourt Brace Jovanovich Publishers], New York, 1979.

[12] B. N. Parlett. *The symmetric eigenvalue problem*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1998. Corrected reprint of the 1980 original.

[13] Ji Guang Sun. Perturbation of angles between linear subspaces. *J. Comput. Math.*, 5(1):58–61, 1987.