A Priori Error Bounds for Eigenvalues Approximated by Ritz Values

Merico Argentati (speaker) and Andrew Knyazev
Department of Mathematical Sciences
University of Colorado at Denver and Health Sciences Center

Supported by the National Science Foundation

Outline

1. Definition of Principal Angles Between Subspaces
2. Majorization, Ritz Values and Principal Angles
3. New Weak Majorization Bound on the Change in Ritz Values
4. Improvement When one Subspace is Invariant
5. Application to Subspace Iteration
6. Conclusions

1. Definition of Principal Angles Between Subspaces

- Let $P_{\mathcal{X}}$ and $P_{\mathcal{Y}}$ be orthogonal projectors onto the subspaces $\mathcal{X}$ and $\mathcal{Y}$, respectively, of the space $\mathbf{C^n}$. Define the set of cosines of principal angles between subspaces $\mathcal{X}$ and $\mathcal{Y}$ by

$$\cos\Theta(\mathcal{X},\mathcal{Y}) = [s_1(P_{\mathcal{X}}P_{\mathcal{Y}}),\ldots,s_m(P_{\mathcal{X}}P_{\mathcal{Y}})],\; m = \min\{\dim\mathcal{X};\dim\mathcal{Y}\}\,.$$

- Definition is symmetric: $\Theta(\mathcal{X},\mathcal{Y}) = \Theta(\mathcal{Y},\mathcal{X})$.
- Cosines are arranged in nonincreasing order, i.e. $\cos(\Theta(\mathcal{X},\mathcal{Y})) = (\cos(\Theta(\mathcal{X},\mathcal{Y})))^{\downarrow}$.
- Definition of the distance between subspaces: $\mathrm{gap}(\mathcal{X},\mathcal{Y}) = \|P_{\mathcal{X}} - P_{\mathcal{Y}}\|_2 = \sin(\theta_{\max}(\mathcal{X},\mathcal{Y}))$.

  See e.g., Knyazev and Argentati [2002].

2. Definition of Weak Majorization

- 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

$$\sum_{i=1}^{k} x_{[i]} \leq \sum_{i=1}^{k} y_{[i]}, \qquad k = 1, \ldots, n. \tag{1}$$

- We use notation $[x_1, \cdots, x_n] \prec_w [y_1, \cdots, y_n]$ or $x \prec_w y$.
- We have majorization (strong) if, in addition to (1), $\sum_{i=1}^{n} x_i = \sum_{i=1}^{n} y_i$.
- We denote $[|x_1|, \cdots, |x_n|]$ by $|x|$.
- Importance: the following two conditions are equivalent: $x \prec_w y$ and $\sum_{i=1}^{n} \phi(x_i) \leq \sum_{i=1}^{n} \phi(y_i)$ for all nondecreasing convex functions ϕ.

3. 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 and majorization.

- Let $A \in \mathbf{C}^{\mathbf{n}\times\mathbf{n}}$ be a Hermitian matrix and let $\mathcal{X}$ be an m–dimensional subspace of $\mathbf{C}^{\mathbf{n}}$. 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, $\alpha_1 \geq \cdots \geq \alpha_m$.

- The Ritz values are also the eigenvalues of $Q_{\mathcal{X}}^* A Q_{\mathcal{X}}$ where $Q_{\mathcal{X}}$ is a matrix with orthonormal columns that span $\mathcal{X}$. Also the nonzero Ritz values are the nonzero eigenvalues of $P_{\mathcal{X}}AP_{\mathcal{X}}$.

Motivation: 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.

Let $S(A)$ denote the vector of the singular values of A in nonincreasing order.

Theorem 1 (Mirsky [1960]) *Let* $A, B \in \mathbf{C^{n\times n}}$. *Then*

$$|S(A) - S(B)| \prec_w S(A - B).$$

Bounding Eigenvalues

For Hermitian A let $\Lambda(A)$ denote the vector of all eigenvalues of A in nonincreasing order, i.e. $\Lambda(A) = \Lambda^{\downarrow}(A)$, while individual eigenvalues of A, enumerated in nonincreasing order, are denoted by $\lambda_i(A)$. A generalization of Weyl's perturbation theorem for Hermitian matrices is the following:

Corollary 1 *Let $A, B \in \mathbf{C^{n\times n}}$ be Hermitian matrices. Then*

$$|\Lambda(A) - \Lambda(B)| \prec_w S(A - B),$$

which is the same as

$$[|\lambda_1(A) - \lambda_1(B)|, \dots, |\lambda_n(A) - \lambda_n(B)|] \prec_w [s_1(A - B), \dots, s_n(A - B)],$$

where individual singular values of $A - B$, enumerated in nonincreasing order, are denoted by $s_i(A - B)$.

Bounding Eigenvalues (Continued)

It is worth noting that Weyl's theorem

$$\max_{j=1,\ldots,n}|\lambda_j(A)-\lambda_j(B)| \leq \|A-B\|_2, \tag{2}$$

and a Hermitian analog of the Hoffman–Wielandt theorem, e.g. Stewart and Sun [1990],

$$\sqrt{\sum_{j=1}^{n}(\lambda_j(A)-\lambda_j(B))^2} \leq \|A-B\|_F,$$

both follow from Corollary 1.

Changes in the Trial Subspace in the Rayleigh–Ritz Method

The closing result of (Argentati [2003]; Knyazev and Argentati [2006b]):

Theorem 2 *Let* $A \in \mathbf{C^{n \times n}}$ *be a Hermitian matrix and let* $\mathcal{X}$ *and* $\mathcal{Y}$ *both be* m*–dimensional subspaces of* $\mathbf{C^n}$*. 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* $\mathcal{Y}$*. Then*

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

where $\lambda_{\min}$ *and* $\lambda_{\max}$ *are the smallest and largest eigenvalues of* A*.*

Weak Majorization Bound on the Change in Ritz Values

Previous result generalized in Knyazev and Argentati [2006a]:

Theorem 3 *Let* $A \in \mathbf{C^{n \times n}}$ *be a Hermitian matrix and let* $\mathcal{X}$ *and* $\mathcal{Y}$ *both be* m*–dimensional subspaces of* $\mathbf{C^n}$*, 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* $\mathcal{Y}$*. Then*

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

Remark 1 *Pioneering results using angles between subspaces in the framework of unitarily invariant norms and symmetric gauge functions, equivalent to majorization, appear in Davis and Kahan [1970], which introduces many of the tools that we use here.*

Implications of Theorem 3

The weak majorization inequality directly implies

$$\sum_{i=1}^{k} |\alpha_i - \beta_i|^{\downarrow} \leq (\lambda_{\max} - \lambda_{\min}) \sum_{i=1}^{k} \sin(\Theta_i(\mathcal{X}, \mathcal{Y}))^{\downarrow}, \qquad k = 1, \ldots, m,$$

e.g., for $k = m$ we obtain

$$\sum_{i=1}^{m} |\alpha_i - \beta_i| \leq (\lambda_{\max} - \lambda_{\min}) \sum_{i=1}^{m} \sin(\Theta_i(\mathcal{X}, \mathcal{Y})),$$

and for $k = 1$ we have (as in Theorem 2)

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

Implications of Theorem 3 (Continued)

For real vectors x and y the weak majorization $x \prec_w y$ is equivalent to the inequality $\sum_{i=1}^{n} \phi(x_i) \leq \sum_{i=1}^{n} \phi(y_i)$ for any continuous nondecreasing convex real valued function ϕ, e.g., Marshall and Olkin [1979]. Taking, e.g., $\phi(t) = t^p$ with $p \geq 1$, in Theorem 3, we obtain

$$\left(\sum_{i=1}^{m} |\alpha_i - \beta_i|^p\right)^{\frac{1}{p}} \leq (\lambda_{\max} - \lambda_{\min})\left(\sum_{i=1}^{m} \sin(\Theta_i(\mathcal{X}, \mathcal{Y}))^p\right)^{\frac{1}{p}}, \quad 1 \leq p < \infty.$$

Outline of Proof for Theorem 3

1. First prove that for $\dim\mathcal{X} = \dim\mathcal{Y}$, that $|\cos^2\Theta(\mathcal{X},\mathcal{Z}) - \cos^2\Theta(\mathcal{Y},\mathcal{Z})| \prec_w \sin\Theta(\mathcal{X},\mathcal{Y})$.
2. Observe that if $A = P_{\mathcal{Z}}$ is an orthogonal projector, then the set of the Ritz values of A are $\Lambda\left(P_{\mathcal{X}}P_{\mathcal{Z}}|_{\mathcal{X}}\right) = [\cos^2\Theta(\mathcal{X},\mathcal{Z}), 0, \ldots, 0]$.
3. For Hermitian A let $\hat{A} = (A - \lambda_{\min} I)/(\lambda_{\max} - \lambda_{\min})$, so $\Lambda(\hat{A}) \in [0,1]$.
4. Let $P_{\hat{\mathcal{Z}}} = \begin{bmatrix} \hat{A} & \sqrt{\hat{A}(I-\hat{A})} \\ \sqrt{\hat{A}(I-\hat{A})} & I-\hat{A} \end{bmatrix}$, $\hat{\mathcal{X}} = \begin{bmatrix} \mathcal{X} \\ 0 \end{bmatrix}$, $\hat{\mathcal{Y}} = \begin{bmatrix} \mathcal{Y} \\ 0 \end{bmatrix}$.
5. Observe that $\Lambda(P_{\mathcal{X}}\hat{A}|_{\mathcal{X}}) = \Lambda(P_{\hat{\mathcal{X}}}P_{\hat{\mathcal{Z}}}|_{\hat{\mathcal{X}}})$ and $\Lambda(P_{\mathcal{Y}}\hat{A}|_{\mathcal{Y}}) = \Lambda(P_{\hat{\mathcal{Y}}}P_{\hat{\mathcal{Z}}}|_{\hat{\mathcal{Y}}})$, and principal angles between $\hat{\mathcal{X}}$ and $\hat{\mathcal{Y}}$ are the same as between $\mathcal{X}$ and $\mathcal{Y}$.
6. Then we have $|\Lambda(P_{\mathcal{X}}\hat{A}|_{\mathcal{X}}) - \Lambda(P_{\mathcal{Y}}\hat{A}|_{\mathcal{Y}})| = |\Lambda(P_{\hat{\mathcal{X}}}P_{\hat{\mathcal{Z}}}|_{\hat{\mathcal{X}}}) - \Lambda(P_{\hat{\mathcal{Y}}}P_{\hat{\mathcal{Z}}}|_{\hat{\mathcal{Y}}})|$ $\prec_w \sin\Theta(\hat{\mathcal{X}},\hat{\mathcal{Y}}) = \sin\Theta(\mathcal{X},\mathcal{Y})$, by item 1.

4. Improvement When one Subspace is Invariant

In Theorem 3 no assumptions are made concerning the subspaces $\mathcal{X}$ and $\mathcal{Y}$ (other than that they are of the same dimension). What happens if one of the subspaces is invariant w.r.t. to the Hermitian operator A? Of course in this case the Ritz values are actually eigenvalues of A, so we now estimate eigenvalues.

It is natural to expect a much better bound that involves the squares of the sine of the angles. The following is well known.

Theorem 4 *Let $A \in \mathbf{C^{n\times n}}$ be Hermitian and let x or y be an eigenvector of A, then*

$$|\lambda(x) - \lambda(y)| \leq (\lambda_{\max} - \lambda_{\min}) \sin^2(\angle\{x, y\}),$$

where $\lambda(x) = \dfrac{(Ax, x)}{(x, x)}$, is the Rayleigh Quotient.

Improvement When one Subspace is Invariant (Continued)

We also have the following stronger result, e.g., Knyazev [1986].

Theorem 5 *Let* $A \in \mathbf{C^{n\times n}}$ *be a Hermitian matrix and let* $\mathcal{X}$ *and* $\mathcal{Y}$ *both be* m*–dimensional subspaces of* $\mathbf{C^n}$*, with* $\mathcal{X}$ *invariant w.r.t* A*, where the Ritz values of* A *with respect to* $\mathcal{X}$ *are the largest* m *contiguous eigenvalues of* A*. 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* $\mathcal{Y}$*. Then*

$$0 \leq \alpha_j - \beta_j \leq (\lambda_{\max} - \lambda_{\min}) \sin^2(\Theta_{\max}(\mathcal{X}, \mathcal{Y})), \quad j = 1, \ldots, m.$$

Weak Majorization Rayleigh–Ritz Method

Eigenvalue Error Estimate

Theorem 6 *Using the same assumptions as in Theorem 5, we have*

$$|\alpha - \beta| \prec_w (\lambda_{\text{max}} - \lambda_{\text{min}}) \ \sin^2\Theta(\mathcal{X}, \mathcal{Y}). \tag{3}$$

Implications of Theorem 6

The weak majorization inequality (3) directly implies

$$\sum_{i=1}^{k} |\alpha_i - \beta_i|^{\downarrow} \le (\lambda_{\text{max}} - \lambda_{\text{min}}) \sum_{i=1}^{k} \sin^2(\Theta_i(\mathcal{X},\mathcal{Y}))^{\downarrow}, \qquad k = 1, \ldots, m,$$

and we also have

$$\left(\sum_{i=1}^{m} |\alpha_i - \beta_i|^p \right)^{\frac{1}{p}} \le (\lambda_{\text{max}} - \lambda_{\text{min}}) \left(\sum_{i=1}^{m} \sin(\Theta_i(\mathcal{X},\mathcal{Y}))^{2p} \right)^{\frac{1}{p}}, \qquad 1 \le p < \infty,$$

for example, if the Ritz values of A with respect to $\mathcal{X}$ are the largest (or smallest) m contiguous eigenvalues of A.

Weak Majorization Rayleigh–Ritz Method

Eigenvalue Error Estimate (Continued)

Theorem 7 *Let* $A \in \mathbf{C^{n\times n}}$ *be a Hermitian matrix and let* $\mathcal{X}$ *and* $\mathcal{Y}$ *both be* m*–dimensional subspaces of* $\mathbf{C^n}$*, with* X *invariant w.r.t.* A*. 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* $\mathcal{Y}$*. Then*

$$|\alpha - \beta| \prec_w \frac{3}{2}(\lambda_{\max} - \lambda_{\min})\ \sin^2\Theta(\mathcal{X},\mathcal{Y}). \tag{4}$$

Our numerical tests support the conjecture that the constant factor is $\lambda_{\max} - \lambda_{\min}$ *in this general case.*

Outline of Proof for Theorem 7

1. First choose $Q_{\mathcal{X}}$ and $Q_{\mathcal{Y}}$ such that $Q_{\mathcal{X}}^* Q_{\mathcal{Y}} = \Sigma$. Do a change of basis by letting $U = [Q_{\mathcal{X}} \; Q_{\mathcal{X}^\perp}]$ which is a unitary matrix.
2. Then $\hat{A} = U^* A U = \begin{bmatrix} B & 0 \\ 0 & C \end{bmatrix}$, $\hat{Q}_{\mathcal{X}} = U^* Q_{\mathcal{X}} = \begin{bmatrix} I \\ 0 \end{bmatrix}$, $\hat{Q}_{\mathcal{Y}} = U^* Q_{\mathcal{X}} = \begin{bmatrix} \Sigma \\ Z \end{bmatrix}$.
3. Then
$$|\Lambda(P_{\mathcal{X}} A|_{\mathcal{X}}) - \Lambda(P_{\mathcal{Y}} A|_{\mathcal{Y}})| \prec_w [|\Lambda(B) - \Lambda(\Sigma B \Sigma)|^{\downarrow} + S(Z^* C Z)].$$
4. Observe that $[S(Z), 0, \ldots, 0] = [(\sin \Theta(\mathcal{X}, \mathcal{Y}))^{\downarrow}, 0, \ldots, 0]$.
5. Using the Corollary to Mirsky's theorem we obtain
$$|\Lambda(P_{\mathcal{X}} A|_{\mathcal{X}}) - \Lambda(P_{\mathcal{Y}} A|_{\mathcal{Y}})| \prec_w 3\|A\| \sin^2 \Theta(\mathcal{X}, \mathcal{Y}).$$
6. Replace A with $A - (\lambda_{\max} + \lambda_{\min})/2$, i.e., shifted A, to obtain (4).

We Can Replace $\lambda_{\max} - \lambda_{\min}$ With a Reduced Constant

- As in Knyazev and Argentati [2006b, Remark 7], the constant $\lambda_{\max} - \lambda_{\min}$ can be replaced with

$$\max_{x \in \mathcal{X}+\mathcal{Y},\, \|x\|=1}(x, Ax) - \min_{x \in \mathcal{X}+\mathcal{Y},\, \|x\|=1}(x, Ax),$$

  which for some subspaces $\mathcal{X}$ and $\mathcal{Y}$ can provide a significant improvement.

- This effectively replaces A with $P_{(\mathcal{X}+\mathcal{Y})}AP_{(\mathcal{X}+\mathcal{Y})}$, which has the same action on $\mathcal{X}$ and $\mathcal{Y}$.

Theorem 6 is Sharp in a Strong Sense

If the Ritz values of A with respect to $\mathcal{X}$ are the largest (or smallest) contiguous eigenvalues of A then the weak majorization inequality is sharp. Consider the following example: Let I be an $m \times m$ identity matrix and let $A = \begin{bmatrix} I & 0 \\ 0 & -I \end{bmatrix}$. Let Σ be a diagonal of given cosines and let Γ be the diagonal of sines, so $\Sigma^2 + \Gamma^2 = I$. Let $Q_{\mathcal{X}} = \begin{bmatrix} I \\ 0 \end{bmatrix}$, and $Q_{\mathcal{Y}} = \begin{bmatrix} \Sigma \\ \Gamma \end{bmatrix}$.

Then $Q_{\mathcal{X}}^* A Q_{\mathcal{X}} = I$ and $Q_{\mathcal{Y}}^* A Q_{\mathcal{Y}} = 2\Sigma^2 - I$ and

$$|\alpha - \beta|^{\downarrow} = 2[\sin\Theta(\mathcal{X},\mathcal{Y}))^{\downarrow}] = (\lambda_{\max} - \lambda_{\min})[\sin\Theta(\mathcal{X},\mathcal{Y}))^{\downarrow}].$$

5. Application: Subspace Iteration

- Concerning the Hermitian eigenvalue problem and methods that access the matrix only as matrix-vector operations (e.g. Lanczos method) we work with a sequence of subspaces and compute Ritz values.
- Consider the subspace iteration method where $U^k = F_k(A)U^0$ and U^0 and U^k are both m–dimensional subspaces.
- Let U be the m–dimensional invariant subspace consisting of eigenvectors corresponding to the the m largest eigenvalues of A.
- We are attempting to estimate the convergence rate using weak majorization.

6. Conclusions

- The absolute value of the difference of Ritz values for a Hermitian operator are majorized by $\lambda_{\max} - \lambda_{\min}$ times the sines of the angles between the perturbed subspaces.

- When one of the subspaces is invariant w.r.t. A and if the Ritz values are the largest (smallest) m contiguous eigenvalues of A, the absolute value of the difference of the Ritz values is majorized by $\lambda_{\max} - \lambda_{\min}$ times the squares of the sines of the angles between the perturbed subspaces, and this majorization inequality is strongly sharp.

- When one of the subspaces is invariant w.r.t. A, the absolute value of the difference of the Ritz values is majorized by $3/2(\lambda_{\max} - \lambda_{\min})$ times the squares of the sines of the angles between the perturbed subspaces.

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.

C. Davis and W. M. Kahan. The rotation of eigenvectors by a perturbation. III. *SIAM Journal on Numerical Analysis*, 7(1):1–46, 1970.

A. V. Knyazev. *Computation of Eigenvalues and Eigenvectors for Mesh Problems: Algorithms and Error Estimates*. Dept. Numerical Math. USSR Academy of Sciences, Moscow, 1986. (In Russian).

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.

A. V. Knyazev and M. E. Argentati. Majorization for changes in angles between subspaces, Ritz values, and graph laplacian spectra. *SIAM J. Matrix Analysis, accepted*, 2006a. Published as a Technical Report math.NA/0508591 at http://arxiv.org/pdf/math.NA/0508591.

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*, 415:82–95, 2006b.

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. ISBN 0-12-473750-1.

L. Mirsky. Symmetric gauge functions and unitarily invariant norms. *Quarterly Journal of Mathematics. Oxford Series(2)*, 11:50–59, 1960.

Beresford 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.

G. W. Stewart and J. G. Sun. *Matrix perturbation theory*. Academic Press Inc., Boston, MA, 1990. ISBN 0-12-670230-6.