Rayleigh-Ritz majorization error bounds with applications to FEM and subspace iterations

Merico Argentati and Andrew Knyazev (speaker)

Department of Applied Mathematical and Statistical Sciences

Center for Computational Mathematics

University Colorado Denver (Downtown Campus)

Householder Symposium XVII Zeuthen, Germany June 5, 2008

Supported by the National Science Foundation

Abstract

Given two subspaces $\mathcal{X}$ and $\mathcal{Y}$ of the same finite dimension, such that $\mathcal{X}$ is A-invariant, the absolute changes in the Ritz values of A with respect to $\mathcal{X}$ compared to the Ritz values with respect to $\mathcal{Y}$ represent the absolute eigenvalue approximation error. A recent paper [1] bounds the error in terms of the principal angles between $\mathcal{X}$ and $\mathcal{Y}$ using weak majorization. We improve and extend this bound, and derive several new related results. We present our Rayleigh-Ritz majorization error bound in the context of the finite element method (FEM). We derive a new majorization-type convergence rate bound of subspace iterations and combine it with the previous result to obtain a similar bound for the block Lanczos method. This presentation is based on [2]. A corresponsing result where neither $\mathcal{X}$ nor $\mathcal{Y}$ is A-invariant can be found in [3]. The case of infinite dimensional subspaces is considered in [4].

References.

[1] M. E. Argentati, A. V. Knyazev, C. C. Paige, and I. Panayotov, Bounds on changes in Ritz values for a perturbed invariant subspace of a Hermitian matrix. SIMAX 30 (2008), 548-559.

[2] A. V. Knyazev and M. E. Argentati, Rayleigh-Ritz majorization error bounds with applications to FEM and subspace iterations, http://arxiv.org/abs/math/0701784.

[3] A. V. Knyazev and M. E. Argentati, Majorization for Changes in Angles Between Subspaces, Ritz values, and graph Laplacian spectra, SIMAX 29 (2006), 15-32.

[4] A. V. Knyazev, A. Jujunashvili, and M. E. Argentati, Angles Between Infinite Dimensional Subspaces with Applications to the Rayleigh-Ritz and Alternating Projectors Methods, http://arxiv.org/abs/0705.1023.

CONTENTS

Majorization

- For vector $x = [x_1, \ldots, x_n]$, we use $x^{\downarrow} \equiv [x_1^{\downarrow}, \ldots, x_n^{\downarrow}]$ to denote x with its elements rearranged in descending order, while $x^{\uparrow} \equiv [x_1^{\uparrow}, \ldots, x_n^{\uparrow}]$ denotes x with its elements rearranged in ascending order. $|x|$ denotes the vector x with the absolute value of its components.

- We say that $x \in \mathbb{R}^n$ is weakly majorized by $y \in \mathbb{R}^n$, written $x \prec_w y$, if

$$\sum_{i=1}^{k} x_i^{\downarrow} \leq \sum_{i=1}^{k} y_i^{\downarrow}, \qquad 1 \leq k \leq n, \tag{1}$$

while x is (strongly) majorized by y, written $x \prec y$, if (1) holds together with

$$\sum_{i=1}^{n} x_i = \sum_{i=1}^{n} y_i. \tag{2}$$

Majorization in Matrix Algebra

Let $\Lambda(A)$ (for Hermitian A) and $S(A)$ be, respectively, vectors of eigenvalues and singular values (both nonincreasing):

- Lidskiĭ theorem: $\Lambda(A+B)-\Lambda(B)\prec\Lambda(A)$ for Hermitian A and B
- Gelfand-Naĭmark theorem: $\log S(AB)-\log S(B)\prec\log S(A)$ for general A and B, where we add zeros to the vectors of singular values if necessary to match the sizes [KA08]
- Generalized pinching inequality [KA08]:

$$[S(A_1^H B C_1), S(A_2^H B C_2)] \prec_w S\left(\sqrt{A_1A_1^H + A_2A_2^H}\,B\,\sqrt{C_1C_1^H + C_2C_2^H}\right)$$

The Rayleigh–Ritz Method

- Let A be *Hermitian* and $\mathcal{X}$ be a subspace
- We define an operator $P_{\mathcal{X}}A|_{\mathcal{X}}$ 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 $\mathcal{X}$. The eigenvalues $\Lambda(P_{\mathcal{X}}A|_{\mathcal{X}})$ are called Ritz values.
- We have $\Lambda(P_{\mathcal{X}}A|_{\mathcal{X}}) = \Lambda(X^H AX)$ where X is a matrix with orthonormal columns that span $\mathcal{X}$

Principal Angles Between Subspaces

- Let subspaces $\mathcal{X}$ and $\mathcal{Y} \subseteq \mathbb{C}^n$ have orthonormal bases given by the columns of the matrices X and Y The principal angles, arranged in descending order, are denoted by $\Theta(\mathcal{X},\mathcal{Y}) = \Theta^{\downarrow}(\mathcal{X},\mathcal{Y})$, and defined using $\cos\Theta(\mathcal{X},\mathcal{Y}) = S^{\uparrow}(X^H Y)$
- The definition is symmetric: $\Theta(\mathcal{X},\mathcal{Y}) = \Theta(\mathcal{Y},\mathcal{X})$ if $\dim\mathcal{X} = \dim\mathcal{Y}$ and describes the angles *between* subspaces.
- If $\dim\mathcal{X} < \dim\mathcal{Y}$, the angles $\Theta(\mathcal{X},\mathcal{Y})$ are *from* $\mathcal{X}$ *to* $\mathcal{Y}$
- $\text{gap}(\mathcal{X},\mathcal{Y}) = \|P_{\mathcal{X}} - P_{\mathcal{Y}}\|_2 = \sin(\Theta_{\max}(\mathcal{X},\mathcal{Y}))$.

Traditional bounds for Ritz values

Let $\dim\mathcal{X} = \dim\mathcal{Y}$, then [KA06]

$$\max|\Lambda((P_\mathcal{X}A)|_\mathcal{X}) - \Lambda((P_\mathcal{Y}A)|_\mathcal{Y})| \le (\lambda_{\max} - \lambda_{\min}) \max\sin\Theta(\mathcal{X},\mathcal{Y}),$$

where $\lambda_{\min}$ and $\lambda_{\max}$ are the smallest and largest eigenvalues of the A, respectively. If in addition one of the subspaces is A-invariant then [KA08]

$$\max|\Lambda((P_\mathcal{X}A)|_\mathcal{X}) - \Lambda((P_\mathcal{Y}A)|_\mathcal{Y})| \le (\lambda_{\max} - \lambda_{\min}) \max\sin^2\Theta(\mathcal{X},\mathcal{Y}).$$

A similar bound with the min does not hold and the inequality

$$|\Lambda((P_\mathcal{X}A)|_\mathcal{X}) - \Lambda((P_\mathcal{Y}A)|_\mathcal{Y})|^\downarrow \le (\lambda_{\max} - \lambda_{\min}) \sin\Theta(\mathcal{X},\mathcal{Y})$$

is WRONG! Majorization is necessary to analyse low-rank perturbations of the trial subspace, leading to many zero angles in $\Theta(\mathcal{X},\mathcal{Y})$.

Majorization bounds for Ritz values

Let $\dim\mathcal{X} = \dim\mathcal{Y}$, then from the Lidskiǐ theorem we can obtain [KA05]

$$\begin{aligned} \left|\Lambda\left(X^H AX\right) - \Lambda\left(Y^H AY\right)\right| &\prec_w \left|\Lambda\left(X^H AX - Y^H AY\right)\right| \\ &\prec_w \sqrt{2}\left(\lambda_{\max} - \lambda_{\min}\right)\sin\Theta(\mathcal{X},\mathcal{Y}). \end{aligned}$$

The constant multiplier $\sqrt{2}$ is artificial. In reality [KA06]

$$\left|\Lambda\left((P_{\mathcal{X}}A)|_{\mathcal{X}}\right) - \Lambda\left((P_{\mathcal{Y}}A)|_{\mathcal{Y}}\right)\right| \prec_w \left(\lambda_{\max} - \lambda_{\min}\right)\ \sin\Theta(\mathcal{X},\mathcal{Y}).$$

If in addition one of the subspaces is A-invariant then [KA08]

$$\left|\Lambda\left((P_{\mathcal{X}}A)|_{\mathcal{X}}\right) - \Lambda\left((P_{\mathcal{Y}}A)|_{\mathcal{Y}}\right)\right| \prec_w \left(\lambda_{\max} - \lambda_{\min}\right)\ \sin^2\Theta(\mathcal{X},\mathcal{Y}).$$

Conjectured and proved only for a particular case in [AKPP08].

Improved sine-based bounds

Substitutions of the space $\mathcal{X}+\mathcal{Y}$ and the operator $(P_{\mathcal{X}+\mathcal{Y}}A)|_{\mathcal{X}+\mathcal{Y}}$ for $\mathcal{H}$ and A improve the constant $\lambda_{\max}-\lambda_{\min}\geq\lambda_{\max(\mathcal{X}+\mathcal{Y})}-\lambda_{\min(\mathcal{X}+\mathcal{Y})}$, using the spread of the spectrum of the operator $(P_{\mathcal{X}+\mathcal{Y}}A)|_{\mathcal{X}+\mathcal{Y}}$. *Spread vector*

$$\operatorname{Spr}_{(\mathcal{X}+\mathcal{Y})}=\left[\lambda_{i(\mathcal{X}+\mathcal{Y})}-\lambda_{-i(\mathcal{X}+\mathcal{Y})},\, i=1,\ldots,\dim\mathcal{X}\right]$$

where $\lambda_{1(\mathcal{X}+\mathcal{Y})}\geq\ldots$ and $\lambda_{-1(\mathcal{X}+\mathcal{Y})}\leq\ldots$ are the $\dim\mathcal{X}$ largest and smallest eigenvalues of $(P_{\mathcal{X}+\mathcal{Y}}A)|_{\mathcal{X}+\mathcal{Y}}$. Our *CONJECTURES*:

$$|\Lambda\left((P_{\mathcal{X}}A)|_{\mathcal{X}}\right)-\Lambda\left((P_{\mathcal{Y}}A)|_{\mathcal{Y}}\right)|\prec_w \operatorname{Spr}_{(\mathcal{X}+\mathcal{Y})}\sin\Theta(\mathcal{X},\mathcal{Y});$$

if in addition one of the subspaces is A-invariant then

$$|\Lambda\left((P_{\mathcal{X}}A)|_{\mathcal{X}}\right)-\Lambda\left((P_{\mathcal{Y}}A)|_{\mathcal{Y}}\right)|\prec_w \operatorname{Spr}_{(\mathcal{X}+\mathcal{Y})}\sin^2\Theta(\mathcal{X},\mathcal{Y}) \qquad (3)$$

We *PROVE* (3) for the largest (or smallest) eigenvalues only [KA08].

Bounds for the largest/smallest eigenvalues

$\mathcal{X}$ represents the largest eigenvalues of $A > 0$ then $\mathrm{Spr}_{(\mathcal{X}+\mathcal{Y})} \leq \Lambda\left((P_{\mathcal{X}}A)|_{\mathcal{X}}\right)$ so $0 \leq \Lambda\left((P_{\mathcal{X}}A)|_{\mathcal{X}}\right) - \Lambda\left((P_{\mathcal{Y}}A)|_{\mathcal{Y}}\right) \prec_w \Lambda\left((P_{\mathcal{X}}A)|_{\mathcal{X}}\right) \sin^2\Theta(\mathcal{X},\mathcal{Y})$. New:

$$0 \leq \log\Lambda\left((P_{\mathcal{X}}A)|_{\mathcal{X}}\right) - \log\Lambda\left((P_{\mathcal{Y}}A)|_{\mathcal{Y}}\right) \prec_w \log\cos^{-2}\Theta(\mathcal{X},\mathcal{Y}),$$

—a *relative multiplicative* error bound, which implies

$$0 \leq \frac{\Lambda\left((P_{\mathcal{X}}A)|_{\mathcal{X}}\right)}{\Lambda\left((P_{\mathcal{Y}}A)|_{\mathcal{Y}}\right)} - 1 \prec_w \tan^2\Theta(\mathcal{X},\mathcal{Y}).$$

For $\mathcal{X}$ representing the smallest eigenvalues of $A > 0$ this gives

$$0 \leq \frac{\Lambda\left((P_{\mathcal{Y}}A)|_{\mathcal{Y}}\right)}{\Lambda\left((P_{\mathcal{X}}A)|_{\mathcal{X}}\right)} - 1 \prec_w \tan^2\Theta_A(\mathcal{X},\mathcal{Y}).$$

The sine- and tangent-based bounds do not follow from each other!

Application to the FEM

The results are generalized to $\dim\mathcal{X} \le \dim\mathcal{Y}$ with truncated $\Lambda\left((P_{\mathcal{Y}}A)|_{\mathcal{Y}}\right)$ in infinite dimensional Hilbert spaces.

For the clamped membrane vibration problem, denoting the two smallest eigenvalues $\Lambda\left((P_{\mathcal{X}}A)|_{\mathcal{X}}\right) = [\lambda_2, \lambda_1]$ of $A = -\Delta$, let the corresponding eigenfunctions v_1 and v_2 in $\mathcal{X}$ be approximated by the FEM subspace $\mathcal{Y}$ such that $\sin\Theta(v_1 + v_2, \mathcal{Y}) = h$, but $\sin\Theta(v_1 - v_2, \mathcal{Y}) = h^2$, both in $\dot{H}^1(\Omega)$, where Ω denotes the polygonal membrane. Denoting the eigenvalues of the FEM discrete negative Laplacian by $\Lambda_{\dim\mathcal{X}}\left((P_{\mathcal{Y}}A)|_{\mathcal{Y}}\right) = [\lambda_2^h, \lambda_1^h]$, we get error bounds for the trace: $0 \le \lambda_1^h + \lambda_2^h - \lambda_1 - \lambda_2 \le \lambda_1 h^2$, and the product

$$1 \le \frac{\lambda_1^h}{\lambda_1}\frac{\lambda_2^h}{\lambda_2} \le 1 + h^2.$$

Compared to the standard bound, the gain in the constant is the factor of 2.

Convergence bounds of subspace iterations

Let F be invariant on both $\mathcal{X}$ and on $\mathcal{X}^\perp$, and $(P_\mathcal{X} F)\,|_\mathcal{X}$ be invertible. Assume $\dim\mathcal{X} = \dim\mathcal{Y}$ and $\Theta(\mathcal{X},\mathcal{Y}) < \pi/2$. Then $\dim(F\mathcal{Y}) = \dim\mathcal{Y}$ and

$$\log\left(\frac{\tan\Theta(\mathcal{X},F\mathcal{Y})}{\tan\Theta(\mathcal{X},\mathcal{Y})}\right) \prec \log\left(S\left(\left((P_\mathcal{X} F)\,|_\mathcal{X}\right)^{-1}\right) S\left((P_{\mathcal{X}^\perp} F)\,|_{\mathcal{X}^\perp}\right)\right),$$

which implies, by applying the exponential function to both sides,

$$\tan\Theta(\mathcal{X},F\mathcal{Y}) \prec_w S\left(\left((P_\mathcal{X} F)\,|_\mathcal{X}\right)^{-1}\right) S\left((P_{\mathcal{X}^\perp} F)\,|_{\mathcal{X}^\perp}\right) \tan\Theta(\mathcal{X},\mathcal{Y}).$$

If $F = f(A)$, combining with our RR error bounds, we obtain

$$\left|\Lambda\left((P_\mathcal{X} A)\,|_\mathcal{X}\right) - \Lambda\left((P_{F\mathcal{Y}} A)\,|_{F\mathcal{Y}}\right)\right|$$

$$\prec_w (\lambda_{\max} - \lambda_{\min}) \left(\frac{\left|f\left(\Lambda\left((P_{\mathcal{X}^\perp} A)\,|_{\mathcal{X}^\perp}\right)\right)\right|^\downarrow}{\left|f\left(\Lambda\left((P_\mathcal{X} A)\,|_\mathcal{X}\right)\right)\right|^\uparrow}\right)^2 \tan^2\Theta(\mathcal{X},\mathcal{Y}).$$

Convergence bounds of the block Lanczos

Let $\dim\mathcal{X} = \dim\mathcal{X}_0 = m$, operator A be Hermitian, and let the A-invariant subspace X correspond to the contiguous set of the largest eigenvalues of A. Let $\mathcal{Y} = \mathcal{X}_0 + A\mathcal{X}_0 + \cdots + A^k\mathcal{X}_0$, then

$$0 \leq \frac{\Lambda\left((P_{\mathcal{X}}A)|_{\mathcal{X}}\right) - \Lambda_{\dim\mathcal{X}}\left((P_{\mathcal{Y}}A)|_{\mathcal{Y}}\right)}{\Lambda_{\dim\mathcal{X}}\left((P_{\mathcal{Y}}A)|_{\mathcal{Y}}\right) - \lambda_{\min}} \prec_w [\sigma_m^*, \ldots, \sigma_1^*] \tan^2\Theta(\mathcal{X}, \mathcal{X}_0).$$

where

$$\sigma_i^* = \left| T_k\left(\frac{\lambda_{m+1} + \lambda_{\min} - 2\lambda_i}{\lambda_{m+1} - \lambda_{\min}}\right)\right|^{-2} \quad i = 1, \ldots, m,$$

and where T_k is the kth Chebyshev polynomial of the first kind.

Ideally, we would like to have the σ's on the left-hand side, as this would imply all previously known bounds!

Conclusions

- Majorization is a powerful tool that gives elegant and general error bounds for eigenvalues approximated by the Rayleigh-Ritz method.
- We discover several new results of this kind, including multiplicative bounds for relative errors.
- We apply majorization, apparently for first time, in the contexts of FEM error bounds and convergence rate bounds for subspace iterations and the block Lanczos method.
- Our initial results are promising and expected to lead to further development of the majorization technique for the theory of eigenvalue computations