Ritz Values of Hermitian Matrices

A nice majorization result.

Chris Paige & Ivo Panayotov
(McGill University);
from work with
M. E. Argentati & A. V. Knyazev
(University of Colorado at Denver).

”Harrachov 2007”, August 19-25, Czech Republic.

Our main result

Theorem: Let $\mathcal{X}$, $\mathcal{Y}$ be subspaces of $\mathbb{C}^n$ having the same dimension k, with orthonormal bases given by the columns of the matrices X and Y respectively. Let $A \in \mathbb{C}^{n \times n}$ be Hermitian, $\mathcal{X}$ be A-invariant and let $\theta \equiv \theta(\mathcal{X}, \mathcal{Y})$ denote the vector of principal angles between the subspaces $\mathcal{X}$ and $\mathcal{Y}$. Then

$$|\lambda(X^H A X) - \lambda(Y^H A Y)| \prec_w \operatorname{spr}(A)(\sin^2\theta + \frac{\sin^4\theta}{2}).$$

Our main result

Theorem: Let $\mathcal{X}$, $\mathcal{Y}$ be subspaces of $\mathbb{C}^n$ having the same dimension k, with orthonormal bases given by the columns of the matrices X and Y respectively. Let $A \in \mathbb{C}^{n\times n}$ be Hermitian, $\mathcal{X}$ be A-invariant and let $\theta \equiv \theta(\mathcal{X}, \mathcal{Y})$ denote the vector of principal angles between the subspaces $\mathcal{X}$ and $\mathcal{Y}$. Then

$$|\lambda(X^H AX) - \lambda(Y^H AY)| \prec_w \operatorname{spr}(A)(\sin^2\theta + \frac{\sin^4\theta}{2}).$$

Moreover, if the A-invariant subspace $\mathcal{X}$ corresponds to the set of k largest or smallest eigenvalues of A then

$$|\lambda(X^H AX) - \lambda(Y^H AY)| \prec_w \operatorname{spr}(A)\sin^2\theta.$$

Basic notation

Hermitian matrix $A = A^H \in \mathbb{C}^{n \times n}$

Eigenvalues $\lambda_1(A) \geq \ldots \geq \lambda_n(A)$

$$\mathrm{spr}(A) \equiv \lambda_1(A) - \lambda_n(A)$$

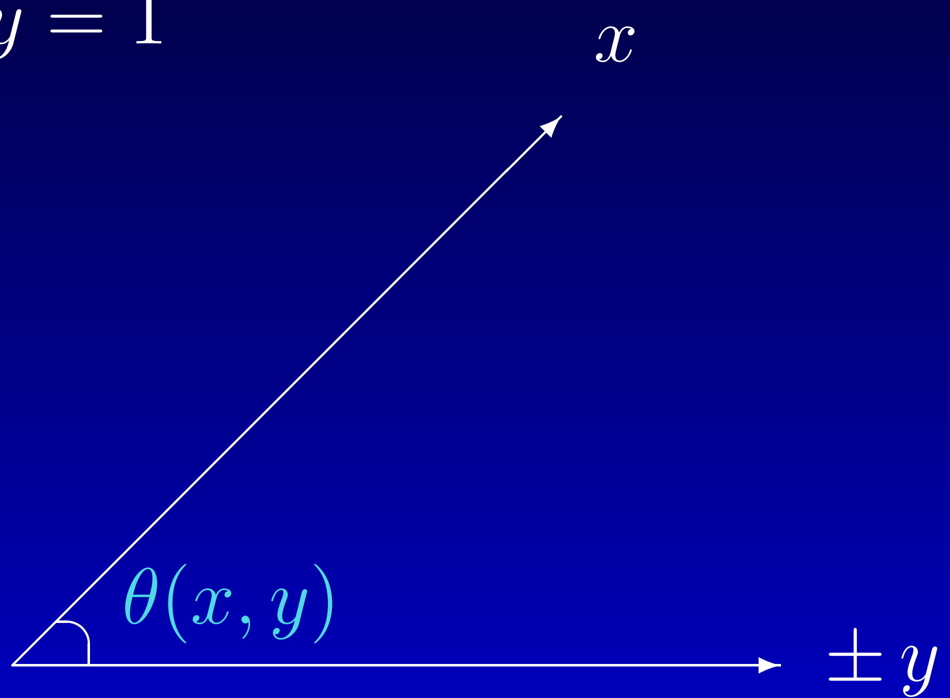
Singular values $\sigma_1(B) \geq \sigma_2(B) \geq \ldots \geq 0$

$$\|B\| \equiv \sigma_1(B)$$

Angles

Acute angle between vectors x and y:

$$x^H x = y^H y = 1$$



$$\cos\theta(x,y) = |x^H y| = \sigma(x^H y)$$

Eigenvalue approximation

Assume $x^H x = y^H y = 1$

Eigenvector x, i.e. $Ax = x.x^H Ax$

y an approximation to x

Rayleigh quotient $y^H Ay \approx x^H Ax$ accuracy?

Eigenvalue approximation

Assume $x^H x = y^H y = 1$

Eigenvector x, i.e. $Ax = x.x^H Ax$

y an approximation to x

Rayleigh quotient $y^H Ay \approx x^H Ax$ accuracy?

Classical *a priori* result: ?Axel Ruhe 1976?

$$|x^H Ax - y^H Ay| \leq \operatorname{spr}(A).\sin^2 \theta(x,y)$$

$$\sin^2 \theta(x,y) = 1 - \cos^2 \theta(x,y) = 1 - |x^H y|^2.$$

Eigenvalue approximation

Assume $x^H x = y^H y = 1$

Eigenvector x, i.e. $Ax = x.x^H Ax$

y an approximation to x

Rayleigh quotient $y^H Ay \approx x^H Ax$ accuracy?

Classical *a priori* result: ?Axel Ruhe 1976?

$$\boxed{|x^H Ax - y^H Ay| \leq \mathrm{spr}(A).\sin^2 \theta(x,y)}$$

Generalize this to subspaces $\mathcal{X}$ and $\mathcal{Y}$?

Vector notation

$$\boldsymbol{\lambda}(A) \equiv \begin{bmatrix} \lambda_1(A) \\ \cdot \\ \lambda_n(A) \end{bmatrix}, \qquad \boldsymbol{\sigma}(B) \equiv \begin{bmatrix} \sigma_1(B) \\ \sigma_2(B) \\ \cdot \end{bmatrix},$$

$$\boldsymbol{\lambda}(A) = \boldsymbol{\lambda}^{\downarrow}(A), \qquad \boldsymbol{\sigma}(B) = \boldsymbol{\sigma}^{\downarrow}(B),$$

$\boldsymbol{\lambda}^{\downarrow}$ elements ordered "downwards",

$\boldsymbol{\lambda}^{\uparrow}$ elements ordered "upwards".

Angles between subspaces

$$X^H X = Y^H Y = I_k$$

$$\mathcal{X} = \mathrm{Range}(X), \quad \mathcal{Y} = \mathrm{Range}(Y)$$

$\boldsymbol{\theta}(\mathcal{X}, \mathcal{Y})$ vector of angles between $\mathcal{X}$ & $\mathcal{Y}$

$$\boldsymbol{\theta}(\mathcal{X}, \mathcal{Y}) = \boldsymbol{\theta}^{\downarrow}(\mathcal{X}, \mathcal{Y}) = [\theta_1(\mathcal{X}, \mathcal{Y}), \ldots, \theta_k(\mathcal{X}, \mathcal{Y})]^T$$

Angles between subspaces

$$X^H X = Y^H Y = I_k$$

$$\mathcal{X} = \mathrm{Range}(X), \quad \mathcal{Y} = \mathrm{Range}(Y)$$

$\boldsymbol{\theta}(\mathcal{X}, \mathcal{Y})$ vector of angles between $\mathcal{X}$ & $\mathcal{Y}$

$$\boldsymbol{\theta}(\mathcal{X}, \mathcal{Y}) = \boldsymbol{\theta}^{\downarrow}(\mathcal{X}, \mathcal{Y}) = [\theta_1(\mathcal{X}, \mathcal{Y}), \ldots, \theta_k(\mathcal{X}, \mathcal{Y})]^T$$

These angles are related to singular values:

$$\begin{aligned} \cos \boldsymbol{\theta}(\mathcal{X}, \mathcal{Y}) &\equiv \boldsymbol{\sigma}^{\uparrow}(X^H Y) \\ &= [\sigma_k(X^H Y), \ldots, \sigma_1(X^H Y)]^T \end{aligned}$$

Eigenproblem for $A = A^H \in \mathbb{C}^{n\times n}$

Assume $\quad X, Y \in \mathbb{C}^{n\times k}, \quad X^H X = Y^H Y = I_k$

Subspaces $\quad \mathcal{X} = \mathrm{Range}(X), \quad \mathcal{Y} = \mathrm{Range}(Y)$

Invariant subspace $\quad \mathcal{X},$ i.e. $\quad AX = X.X^H AX$

Eigenproblem for $A = A^H \in \mathbb{C}^{n\times n}$

Assume $\quad X, Y \in \mathbb{C}^{n\times k}, \quad X^H X = Y^H Y = I_k$

Subspaces $\quad \mathcal{X} = \mathrm{Range}(X), \quad \mathcal{Y} = \mathrm{Range}(Y)$

Invariant subspace $\quad \mathcal{X}$, i.e. $\quad AX = X.X^H AX$

Problem: $\quad \mathcal{Y}$ some approximation to $\mathcal{X}$,

"Ritz values" $\quad \boldsymbol{\lambda}(Y^H AY)$,

Bound $\quad d \equiv |\boldsymbol{\lambda}(X^H AX) - \boldsymbol{\lambda}(Y^H AY)|$?

Eigenproblem for $A = A^H \in \mathbb{C}^{n\times n}$

Assume $\quad X, Y \in \mathbb{C}^{n\times k}, \quad X^H X = Y^H Y = I_k$

Subspaces $\quad \mathcal{X} = \mathrm{Range}(X), \quad \mathcal{Y} = \mathrm{Range}(Y)$

Invariant subspace $\quad \mathcal{X},$ i.e. $\quad AX = X.X^H AX$

Problem: $\quad \mathcal{Y}$ some approximation to $\quad \mathcal{X},$

“Ritz values” $\quad \boldsymbol{\lambda}(Y^H AY),$

Bound $\quad d \;\equiv\; |\boldsymbol{\lambda}(X^H AX) - \boldsymbol{\lambda}(Y^H AY)| \quad$?

$d^{\downarrow} \leq \mathrm{spr}(A)\sin^2\theta(\mathcal{X},\mathcal{Y})$ is <u>false</u>!

Majorization (weak and strong)

A comparison relation between real vectors.

Majorization (weak and strong)

A comparison relation between real vectors.

Majorization inequalities appear naturally, e.g., when describing the spectrum, or singular values of sums and products of matrices.

Majorization (weak and strong)

A comparison relation between real vectors.

Majorization inequalities appear naturally, e.g., when describing the spectrum, or singular values of sums and products of matrices.

A well developed theoretical field applied extensively in Matrix Analysis, e.g.

Marshall & Olkin 1979, (revised Nov 2007?)
Horn & Johnson 1991,
Bhatia 1997.

Majorization (weak and strong)

A comparison relation between real vectors.

Majorization inequalities appear naturally, e.g., when describing the spectrum, or singular values of sums and products of matrices.

A well developed theoretical field applied extensively in Matrix Analysis, e.g.

Marshall & Olkin 1979, (revised Nov 2007?)
Horn & Johnson 1991,
Bhatia 1997.

It has recently been used in the analysis of numerical algorithms.

Majorization

$y \in \mathbb{R}^n$ is weakly majorized by $x \in \mathbb{R}^n$, written $y \prec_w x$,

iff

$$\begin{aligned} y_1^\downarrow &\leq x_1^\downarrow \\ y_1^\downarrow + y_2^\downarrow &\leq x_1^\downarrow + x_2^\downarrow \\ \cdots &\leq \cdots \\ y_1^\downarrow + \cdots + y_n^\downarrow &\leq x_1^\downarrow + \cdots + x_n^\downarrow . \end{aligned}$$

Majorization

$y \in \mathbb{R}^n$ is weakly majorized by $x \in \mathbb{R}^n$, written $y \prec_w x$,

iff

$$\begin{aligned} y_1^\downarrow &\leq x_1^\downarrow \\ y_1^\downarrow + y_2^\downarrow &\leq x_1^\downarrow + x_2^\downarrow \\ \cdots &\leq \cdots \\ y_1^\downarrow + \cdots + y_n^\downarrow &\leq x_1^\downarrow + \cdots + x_n^\downarrow. \end{aligned}$$

$y \in \mathbb{R}^n$ is strongly majorized by $x \in \mathbb{R}^n$, written $y \prec x$, if also $\sum_{i=1}^n y_i = \sum_{i=1}^n x_i$.

Geometry of majorization

The linear inequalities of weak and strong majorization define convex sets in $\mathbb{R}^n$.

Geometry of majorization

The linear inequalities of weak and strong majorization define convex sets in $\mathbb{R}^n$.

For a fixed $x \in \mathbb{R}^n$, the set of all $y \in \mathbb{R}^n$ satisfying $y \prec x$ is the convex hull of all Px , where P runs over the permutation matrices.

Geometry of majorization

The linear inequalities of weak and strong majorization define convex sets in $\mathbb{R}^n$.

For a fixed $x \in \mathbb{R}^n$, the set of all $y \in \mathbb{R}^n$ satisfying $y \prec x$ is the convex hull of all Px , where P runs over the permutation matrices.

The set of all $y \in \mathbb{R}^n$ satisfying $y \prec_w x$ is an unbounded convex set. However if x and y are positive vectors, then we also obtain a convex polygon.

Classical majorization results

Theorem (Shur): Let $A \in \mathbb{R}^{n \times n}$ be Hermitian.
Then $\mathrm{diag}(A) \prec \lambda(A)$.

Classical majorization results

Theorem (Shur): Let $A \in \mathbb{R}^{n\times n}$ be Hermitian.
Then $\mathrm{diag}(A) \prec \lambda(A)$.

Theorem (Lidskii): Let $A, B \in \mathbb{R}^{n\times n}$ be Hermitian.
Then $\lambda(A) - \lambda(B) \prec \lambda(A - B)$.

Classical majorization results

Theorem (Shur): Let $A \in \mathbb{R}^{n \times n}$ be Hermitian.
Then $\mathrm{diag}(A) \prec \lambda(A)$.

Theorem (Lidskii): Let $A, B \in \mathbb{R}^{n \times n}$ be Hermitian.
Then $\lambda(A) - \lambda(B) \prec \lambda(A - B)$.

Theorem (Ky-Fan): Let $A, B \in \mathbb{R}^{n \times n}$.
Then $\sigma(A + B) \prec_w \sigma(A) + \sigma(B)$.

Our main result again

Theorem: Let $\mathcal{X}$, $\mathcal{Y}$ be subspaces of $\mathbb{C}^n$ having the same dimension k, with orthonormal bases given by the columns of the matrices X and Y respectively. Let $A \in \mathbb{C}^{n\times n}$ be Hermitian, $\mathcal{X}$ be A-invariant and let $\theta \equiv \theta(\mathcal{X}, \mathcal{Y})$ denote the vector of principal angles between the subspaces $\mathcal{X}$ and $\mathcal{Y}$. Then

$$|\lambda(X^H AX) - \lambda(Y^H AY)| \prec_w \operatorname{spr}(A)(\sin^2\theta + \frac{\sin^4\theta}{2}).$$

Moreover, if the A-invariant subspace $\mathcal{X}$ corresponds to the set of k largest or smallest eigenvalues of A then

$$|\lambda(X^H AX) - \lambda(Y^H AY)| \prec_w \operatorname{spr}(A)\sin^2\theta.$$

Motivation

Study the convergence of numerical methods which use Ritz approximations for the symmetric (Hermitian) eigenvalue problem

Motivation

Study the convergence of numerical methods which use Ritz approximations for the symmetric (Hermitian) eigenvalue problem, e.g.

- Lanczos algorithm 1950
- block Lanczos algorithm (Golub & Underwood, 1977)

a priori bounds

Our bound depends on the unknown $\boldsymbol{\theta}(\mathcal{X},\mathcal{Y})$, and so is an *a priori* result:

- it can help our understanding, and
- suggest relative performance of algorithms.

a priori bounds

Our bound depends on the unknown $\boldsymbol{\theta}(\mathcal{X}, \mathcal{Y})$, and so is an *a priori* result:

- it can help our understanding, and
- suggest relative performance of algorithms.

Three *a priori* approaches for the (sym.) EVP:

- Classical: angles between eigen-vectors and Ritz vectors.
- Saad: angles between eigen-vectors and the trial (Krylov) subspace $\mathcal{Y}$.
- AKPP: angles between the invariant subspace $\mathcal{X}$, and $\mathcal{Y}$.

Questions

- Can we remove the $\sin^4 \theta(\mathcal{X}, \mathcal{Y})/2$ term from our general bound?

Questions

- Can we remove the $\sin^4\theta(\mathcal{X},\mathcal{Y})/2$ term from our general bound?
- Are our bounds useful for studying the convergence analysis of the algorithms which use Ritz approximations?

Questions

- Can we remove the $\sin^4 \theta(\mathcal{X}, \mathcal{Y})/2$ term from our general bound?
- Are our bounds useful for studying the convergence analysis of the algorithms which use Ritz approximations?
- Can majorization be usefully applied in the convergence analysis of other numerical algorithms?

Questions

- Can we remove the $\sin^4\theta(\mathcal{X},\mathcal{Y})/2$ term from our general bound?
- Are our bounds useful for studying the convergence analysis of the algorithms which use Ritz approximations?
- Can majorization be usefully applied in the convergence analysis of other numerical algorithms?

This talk is dedicated to my mother.

Questions

- Can we remove the $\sin^4 \theta(\mathcal{X}, \mathcal{Y})/2$ term from our general bound?
- Are our bounds useful for studying the convergence analysis of the algorithms which use Ritz approximations?
- Can majorization be usefully applied in the convergence analysis of other numerical algorithms?

This talk is dedicated to my mother.

THANK YOU!