Majorization, angles, and bounds for Ritz values

Peizhen Zhu (Presenter) and Andrew Knyazev

Department of Mathematical & Statistical Sciences
University of Colorado Denver

July 19, 2010

Peizhen Zhu

Andrew Knyazev

Department of Mathematical & Statistical Sciences

Basic Notation

- Singular values $s_1(A) \geq s_2(A) \geq \cdots \geq 0$
- $S(A) = [s_1(A), \ldots, s_n(A)]$ and $S(A) = S^{\downarrow}(A)$
- $\downarrow$ means the components in vector are in nonincreasing order
- The spectral norm $||A|| = s_1(A)$
- Orthogonal projector $P_{\mathcal{X}}$ on subspace $\mathcal{X}$ or range(X)
- Hermitian matrix $A = A^H \in C^{n \times n}$
- Eigenvalues $\lambda_1(A) \geq \lambda_2(A) \geq \cdots \geq \lambda_n(A)$, eigenvectors $x_1, \ldots, x_n$
- $\Lambda(A) = [\lambda_1(A), \ldots, \lambda_n(A)]$ and $\Lambda(A) = \Lambda^{\downarrow}(A)$
- Unitarily invariant norm $|||\cdot|||$

Principal Angles or Canonical Angles between Subspaces

- Matrices $X \in C^{n\times k}$ and $Y \in C^{n\times l}$ with columns that form orthonormal bases for subspaces $\mathcal{X}$ and $\mathcal{Y}$, respectively. The cosines of principal angles between subspaces are defined by

$$\cos \Theta^{\uparrow}(\mathcal{X},\mathcal{Y}) = S^{\downarrow}(X^H Y) = \left[s_1(X^H Y),\ldots,s_m(X^H Y)\right],$$

  where $m = min(k,l)$ and $s_1 \geq s_2 \geq \cdots \geq s_m$. (see Stewart [2])
- The geometry of principal angles between subspaces

The consines of principal angles can be defined recursively for $j = 1,\ldots,m$:

$$s_j = \cos(\theta_j) = max_{x\in\mathcal{X}} max_{y\in\mathcal{Y}} |x^H y| = |x_j^H y_j|$$

suject to $||x|| = ||y|| = 1$, $x^H x_i = 0$, and $y^H y_i = 0$, $i = 1,\ldots,j-1$.

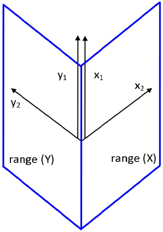

Figure: Angles in 2D

Tangent of Angles in terms of Singular Values

THEOREM 1

For integers $n \geq p$ and $n \geq q$, let the matrix $X \in \mathcal{C}^{n\times p}$ have orthonormal columns and be arbitrarily completed to a unitary matrix $[X, X_\perp]$. Let $Y^H Y = I$ with $Y \in \mathcal{C}^{n\times q}$ and the matrix $X^H Y$ have full rank, then the positive singular values $S_+(T)$ of the matrix $T = X_\perp^H Y \left(X^H Y\right)^+$ satisfy

$$\tan \Theta^{\downarrow}(\mathcal{R}(X), \mathcal{R}(Y)) = [S_+^{\downarrow}(T), 0, \ldots, 0],$$

where $\mathcal{R}(\cdot)$ denotes the matrix column range and $\left(X^H Y\right)^+$ stands for the Moore$-$Penrose inverse of $X^H Y$.

THEOREM 2

Let $[X, X_\perp]$ be unitary and $X \in \mathcal{C}^{n\times p}$. Let $q \leq p$ and $X^H Y$ have full rank with $Y \in \mathcal{C}^{n\times q}$. Let $T = X_\perp^H Y \left(X^H Y\right)^+$, then

$$T = \hat{T},$$

where $\hat{T} = X_\perp^H Z \left(X^H Z\right)^+$ with $Z = Y(Y^H Y)^{-1/2}$.

Theorem 3

Let $[X, X_\perp]$ be unitary and $X \in \mathcal{C}^{n\times p}$. If $\mathit{rank}\left(X^H Y\right)= \mathit{rank}\left(Y\right) \le p$, then the singular values $S(T)$ of the matrix $T = X_\perp^H Y \left(X^H Y\right)^+$satisfy

$$\tan \Theta^{\downarrow}(\mathcal{R}(Y), \mathcal{R}(X)) = [S_+^{\downarrow}(T), 0, \ldots, 0].$$

- We replace the orthonormal bases in subspaces with the corresponding orthogonal projectors in T.

COROLLARY 1

Let $P_{\mathcal{X}}$,$P_{\mathcal{X}_\perp}$, and $P_{\mathcal{Y}}$ be orthogonal projectors on subspaces $\mathcal{R}(X)$, $\mathcal{R}(X_\perp)$, and $\mathcal{R}(Y)$, respectively. If $\Theta(\mathcal{R}(X), \mathcal{R}(Y)) < \frac{\pi}{2}$, then the positive singular values $S_+(T)$ of $T = P_{\mathcal{X}_\perp} P_{\mathcal{Y}} \left(P_{\mathcal{X}} P_Y\right)^+$ satisfy

$$\tan \Theta^{\downarrow}(\mathcal{R}(X), \mathcal{R}(Y)) = [S_+^{\downarrow}(T), 0, \ldots, 0].$$

The Geometry of the Operator $T = P_{\mathcal{X}_\perp} P_{\mathcal{Y}} \left(P_{\mathcal{X}} P_{Y}\right)^{+}$

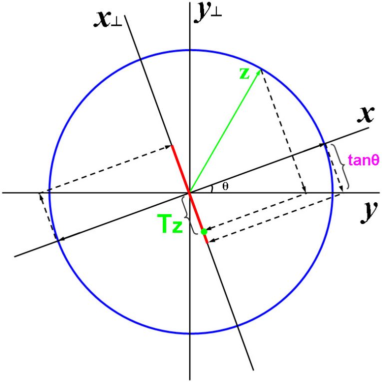

Figure: Geometry of $T = P_{\mathcal{X}_\perp} P_{\mathcal{Y}} \left(P_{\mathcal{X}} P_{Y}\right)^{+}$

Choose a unit vector z and project z on $\mathcal{Y}$ along $\mathcal{X}_\perp$, then project on $\mathcal{X}_\perp$, which is exactly Tz. The red segment is the image of T applied to all unit vectors. So, $s(T) = ||T|| = \tan(\theta)$.

Majorization (Weak and Strong)

Let $x^{\downarrow}$ and $y^{\downarrow} \in R^n$ be the vectors obtained by rearranging the entries of vectors x and y in algebraically descending order. We denote $x^{\downarrow} = \left[x_{[1]}, \ldots, x_{[n]}\right]$ and $y^{\downarrow} = \left[y_{[1]}, \ldots, y_{[n]}\right]$.

- Weak Majorization

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

  using notation $x \prec_w y$
- Majorization or Strong Majorization
  If in addition

  $$\sum_{i=1}^{n} x_{[i]} = \sum_{i=1}^{n} y_{[i]},$$

  using notation $x \prec y$

For example: Strong Majorization

$$\left(\tfrac{1}{n}, \ldots, \tfrac{1}{n}\right) \prec \left(\frac{1}{n-1}, \ldots, \frac{1}{n-1}, 0\right) \prec \cdots \prec (1, 0, \ldots, 0)$$

Interestingly, $x \leq y$ implies $x \prec_w y$, but $x \prec_w y$ does not imply $x \leq y$.

Geometry of Majorization

For $x, y \in \Re^n$, $x \prec y$ if and only if x is in the convex hull of all vectors obtained by permuting the coordinate of y.

1. Orbits of y under permutations for the case n=2
   - For fixed y, we have

$$(\frac{y_1+y_2}{2}, \frac{y_1+y_2}{2}) \prec \cdots \prec x \prec y$$

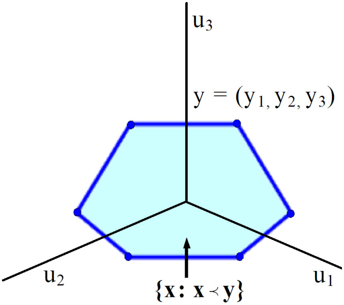

Figure: Geometry of majorization

2. Orbits of y under permutations for the case n=3
   - For fixed y, we have

$$(\frac{y_1+y_2+y_3}{3}, \frac{y_1+y_2+y_3}{3}, \frac{y_1+y_2+y_3}{3}) \prec \cdots \prec x \prec y$$

Majorization in Matrix Theory

- Majorization inequalities appear naturally, when describing the eigenvalues or singular values of differences, sums, and products of matrices.
  - Schur's Theorem: if A is Hermitian, then $diag(A) \prec \Lambda(A)$.
  - Lidskii's Theorem: if A and B are Hermitian matrices, then $$\Lambda(A) - \Lambda(B) \prec \Lambda(A - B).$$
  - Mirsky's Theorem: for any two matrices A and B, $$|S(A) - S(B)| \prec_w S(A - B).$$
- Unitarily invariant norm: a norm on $m \times n$ matrices satisfies $$|||UAV||| = |||A|||,$$ for all unitary matrices U and V.
  - $|||A||| = g(S(A))$ where g is a symmetric gauge function.
  - The 2-norm, Ky-Fan k norms and Forbenius norm are unitarily invariant norms.
- Inequality of unitarily invariant norms is equivalent to weak majorization for singular values, $$S(A) \prec_w S(B) \Leftrightarrow |||A||| \leq |||B|||,$$ for every unitarily invariant norm.

Rayleigh-Ritz Method

- Hermitian matrix A and nonzero vector $x \in \mathcal{C}^n$, the vector Rayleigh quotient $\rho(x)$ is defined as:

  $$\rho(x) = \frac{(x,Ax)}{(x,x)},$$

  where $(\cdot,\cdot)$ is the standard inner product $(x,y) = x^H y$.
- Given a subspace $\mathcal{X}$ spanned by the orthonormal columns of an $n \times k$ matrix X $(k \leq n)$, the matrix Rayleigh quotient (Stewart[2]) is

  $$\rho(X) = X^H AX.$$
- The eigenvalues of $X^H AX$ are called Ritz values of A corresponding to X.
- If the subspace $\mathcal{X}$ is A-invariant, i.e. $A\mathcal{X} \subset \mathcal{X}$, then the Ritz values are exactly some k eigenvalues of A.

Perturbation of the Rayleigh Quotient

1. (Knyazev and Argentati [3])for nonzero vectors x and y,

$$|\rho(x) - \rho(y)| \leq (\lambda_1 - \lambda_n) \sin(\theta(x,y)).$$

   In addition, if x is an eigenvector of A,

$$\text{A Priori:} \quad |\rho(x) - \rho(y)| \leq (\lambda_1 - \lambda_n) \sin^2(\theta(x,y)).$$

2. - (Parlett [8])for any unit vector y, let x be normalized eigenvector corresponding to the eigenvalue of A closest to $\rho(y)$. The gap $\delta = min|\lambda_i(A) - \rho(y)|$ over all $\lambda_i \neq \rho(x)$. Then

$$\text{A Posteriori:} \quad |\rho(x) - \rho(y)| \leq \frac{||r(y)||^2}{\delta}.$$

   - (Kahan[7])for any unit vector y, $\exists$ some eigenvector x of A

$$\text{A Posteriori:} \quad |\rho(x) - \rho(y)| \leq ||r(y)||,$$

   where $r(y) = Ay - \rho(y)y$.

3. (Argentati[9])if $\cos(x,y) \neq 0$ with unit vectors x and y,

$$|\rho(x) - \rho(y)| \leq (||r(x)|| + ||r(y)||) \tan(\theta(x,y)).$$

   (Knayazev[4] and Sun[6])for any unit vector y and any eigenvector x,

$$\text{Mixed:} \quad |\rho(x) - \rho(y)| \leq ||r(y)|| \tan(\theta(x,y)).$$

A Priori Majorization Bounds

For a fixed Hermitian matrix, if a subspace $\mathcal{X}$ is perturbed to give rise to another subspace $\mathcal{Y}$, then how to bound

$$|\Lambda(X^H AX) - \Lambda(Y^H AY)|?$$

THEOREM 4 (Knyazev and Argentati [3])

Let $A = A^H$ and $X^H X = I, Y^H Y = I$. $\Lambda(A)$ denotes a vector consisting of eigenvalues of A. Then

$$|\Lambda^{\downarrow}(X^H AX) - \Lambda^{\downarrow}(Y^H AY)| \prec_w (\lambda_1 - \lambda_n) \sin \Theta^{\downarrow}(\mathcal{X}, \mathcal{Y}).$$

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

$$|\Lambda^{\downarrow}(X^H AX) - \Lambda^{\downarrow}(Y^H AY)| \prec_w (\lambda_1 - \lambda_n) \sin^2 \Theta^{\downarrow}(\mathcal{X}, \mathcal{Y}).$$

Let $\Lambda(X^H AX) = [\alpha_1, \alpha_2, \ldots, \alpha_m]$ and $\Lambda(Y^H AY) = [\beta_1, \beta_2, \ldots, \beta_m]$. For $k = 1, \ldots, m$, the weak majorization inequality directly implies

$$\sum_{i=1}^k |\alpha_i - \beta_i|^{\downarrow} \leq (\lambda_1 - \lambda_n) \sum_{i=1}^k \sin \theta_i(\mathcal{X}, \mathcal{Y})^{\downarrow}.$$

A Posteriori Majorization Bounds

Let the columns of $X \in C^{n\times m}$ be orthonormal and $A = A^H$. For any m by m matrix B, there is an associated residual matrix

$$R\left(B\right) = AX - XB$$

THEOREM 5 (Stewart and Sun [1], also see Kahan [7])

Let Y be any orthonormal n by m matrix and let $A = A^H$. Associated with it are H_y and $R(H_y)$, where $H_y = Y^H AY$ and $R(H_y) = AY - YH_y$.There exists a set of indices $1 \leq i_1 < i_2 < \cdots < i_m \leq n$, such that

$$|\Lambda_I^{\downarrow}(A) - \Lambda^{\downarrow}(Y^H AY)| \prec_w [s_1, s_1, s_2, s_2, \ldots] \prec_w 2S^{\downarrow}\left[R(H_y)\right],$$

where $\Lambda_I(A) = (\lambda_{i_1}, \ldots, \lambda_{i_m})$ and $S\left[R(H_y)\right] = [s_1, s_2, \ldots, s_n]$ denotes the singular values of $R(H_y)$.

Bhatia in [10] shows that the multiplier 2 cannot be removed.

Mixed Type Majorization Bounds

THEOREM 6

Under the assumptions of Theorem 4, in addition, if $\Theta(\mathcal{X},\mathcal{Y}) < \pi/2$, then $|\Lambda^{\downarrow}(X^HAX) - \Lambda^{\downarrow}(Y^HAY)| \prec_w$

$$\frac{1}{\cos\Theta_{max}(\mathcal{X},\mathcal{Y})}\left\{S^{\downarrow}[R(H_x)] + S^{\downarrow}[R(H_y)]\right\}\sin\Theta^{\downarrow}(\mathcal{X},\mathcal{Y})$$

where $R(H_x) = AX - XX^HAX$. If the subspace $\mathcal{X}$ is A-invariant, then

$$|\Lambda^{\downarrow}(X^HAX) - \Lambda^{\downarrow}(Y^HAY)| \quad \prec_w \quad \frac{1}{\cos\Theta_{max}(\mathcal{X},\mathcal{Y})} S^{\downarrow}[R(H_y)]\sin^{\downarrow}\Theta(\mathcal{X},\mathcal{Y})$$

Under the assumptions of Theorem 6, we also have

$$|\Lambda(X^HAX) - \Lambda(Y^HAY)|^2 \quad \prec_w \quad \left\{S[R(H_x)] + S[R(H_y)]\right\}^2\tan^2\Theta(\mathcal{X},\mathcal{Y})$$

If the subspace $\mathcal{X}$ is A-invariant, then

$$|\Lambda(X^HAX) - \Lambda(Y^HAY)|^2 \quad \prec_w \quad S^2[R(H_y)]\tan^2\Theta(\mathcal{X},\mathcal{Y})$$

Our conjecture: The square on both sides can be removed.

Mixed Type Majorization Bounds VS Known Bounds

Comparing to the a posteriori majorization bound

- The mixed type majorization bound is more general than the a posteriori bound, since the a posteriori bound holds only if one subspace is A-invariant.
- We can choose any A-invariant subspace in mixed type majorization. However, in the a posteriori majorization bound, there exists an A-invariant subspace.
- If the principal angles between subspaces are small, such as $\tan\Theta \leq 1$, then the mixed type majorization bound is sharper.

Comparing to the a priori majorization bound

- One of the main advantages of mixed type majorization bound is in the case where both subspaces $\mathcal{X}$ and $\mathcal{Y}$ approximate the same A-invariant subspace, so that the principal angles between the subspaces are small and the singular values of the residual matrix are small.

- G. W. Stewart and J.-G Sun, *Matrix perturbation theory*, 1990.
- G. W. Stewart, *Matrix algorithms. Vol. II*, SIAM, 2001.
- A. V. Knyazev and M. E. Argentati, *Rayleigh-ritz majorization error bounds with applications to FEM*, SIMAX, 2010.
- A. V. Knyazev, *New estimates for Ritz vectors*, Math. Comp., 1997.
- C. Davis and W. M. Kahan, *The rotation of eigenvectors by a perturbation. III*, SINUM, 1970.
- J.-G Sun, *Eigenvalue of Rayleigh quotient matrices*, NUMAT, 1991.
- W. M. Kahan, *Inclusion theorems for clusters of eigenvalues of Hermitian matrices*, Tech Report NO.CS42, 1967.
- B. N. Parlett, *The symmetric eigenvalue problem*, SIAM, 1998.
- M. E. Argentati, Ph.D. thesis, UC Denver, 2003.
- R. Bhatia, *Matrix analysis*, 1997.