Eigenspaces and their Approximation

Tsung-Ming Huang

Department of Mathematics
National Taiwan Normal University, Taiwan

April 8, 2009

Outline

Definition

Let A be of order n and let $\mathcal{X}$ be a subspace of $\mathbb{C}^n$. Then $\mathcal{X}$ is an eigenspace or invariant subspace of A if

$$A\mathcal{X} = \{Ax; x \in \mathcal{X}\} \subset \mathcal{X}.$$

If $(\lambda, x) \equiv (\alpha + \imath\beta, y + \imath z)$ is a complex eigenpair of a real matrix A, i.e.,

$$A(y + \imath z) = (\alpha + \imath\beta)(y + \imath z) = (\alpha y - \beta z) + \imath(\beta y + \alpha z)$$

$$\Rightarrow \quad \begin{cases} Ay = \alpha y - \beta z, \\ Az = \beta y + \alpha z, \end{cases}$$

then

$$A \begin{bmatrix} y & z \end{bmatrix} = \begin{bmatrix} y & z \end{bmatrix} \begin{bmatrix} \alpha & \beta \\ -\beta & \alpha \end{bmatrix}.$$

It implies that $\mathcal{R}\left(\begin{bmatrix} y & z \end{bmatrix}\right)$ is an eigenspace of A.

Theorem

Let $\mathcal{X}$ be an eigenspace of A and let X be a basis for $\mathcal{X}$. Then there is a unique matrix L such that

$$AX = XL.$$

The matrix L is given by

$$L = X^I AX,$$

where X^I is a matrix satisfying $X^I X = I$.
If (λ, x) is an eigenpair of A with $x \in \mathcal{X}$, then $(\lambda, X^I x)$ is an eigenpair of L. Conversely, if (λ, u) is an eigenpair of L, then (λ, Xu) is an eigenpair of A.

Proof: Let

$$X = [x_1 \cdots x_k] \quad \text{and} \quad Y = AX = [y_1 \cdots y_k] \,.$$

Since $y_i \in \mathcal{X}$ and X is a basis for $\mathcal{X}$, there is a unique vector ℓ_i such that

$$y_i = X\ell_i.$$

If we set $L = [\ell_1 \cdots \ell_k]$, then $AX = XL$ and

$$L = X^I XL = X^I AX.$$

Now let (λ, x) be an eigenpair of A with $x \in \mathcal{X}$. Then there is a unique vector u such that $x = Xu$. However, $u = X^I x$. Hence

$$\lambda x = Ax = AXu = XLu \quad \Rightarrow \quad \lambda u = \lambda X^I x = Lu.$$

Conversely, if $Lu = \lambda u$, then

$$A(Xu) = (AX)u = (XL)u = X(Lu) = \lambda(Xu),$$

so that (λ, Xu) is an eigenpair of A. ∎

Definition

Let A be of order n. For $X \in \mathbb{C}^{n\times k}$ and $L \in \mathbb{C}^{k\times k}$, we say that (L, X) is an eigenpair of order k or right eigenpair of order k of A if

$$\begin{array}{ll} 1. & X \text{ is of full rank,} \\ 2. & AX = XL. \end{array}$$

The matrices X and L are called eigenbasis and eigenblock, respectively. If X is orthonormal, we say that the eigenpair (L, X) is orthonormal.
If $Y \in \mathbb{C}^{n\times k}$ has linearly independent columns and $Y^H A = LY^H$, we say that (L, Y) is a left eigenpair of order k of A.

Question

How eigenpairs transform under change of basis and similarities?

Theorem

Let (L, X) be an eigenpair of A. If U is nonsingular, then the pair $(U^{-1}LU, XU)$ is also eigenpair of A. If W is nonsingular, then $(L, W^{-1}X)$ is an eigenpair of $W^{-1}AW$.

proof:

$$A(XU) = (AX)U = (XL)U = (XU)(U^{-1}LU),$$
$$(W^{-1}AW)(W^{-1}X) = W^{-1}AX = (W^{-1}X)L.$$

∎

The eigenvalues of L of an eigenspace with respect to a basis are independent of the choices of the basis.

Theorem

Let $\mathcal{L} = \{\lambda_1, \ldots, \lambda_k\} \subset \Lambda(A)$ be a multisubset of the eigenvalues of A. Then there is an eigenspace $\mathcal{X}$ of A whose eigenvalues are $\lambda_1, \ldots, \lambda_k$.

Proof: Let

$$A[\ U_1 \quad U_2\ ] = [\ U_1 \quad U_2\ ] \begin{bmatrix} T_{11} & T_{12} \\ 0 & T_{22} \end{bmatrix}$$

be a partitioned Schur decomposition of A in which T_{11} is of order k and has the members of $\mathcal{L}$ on its diagonal. Then

$$AU_1 = U_1 T_{11}.$$

Hence the column space of U_1 is an eigenspace of A whose eigenvalues are the members of $\mathcal{L}$. ■

Definition

An eigenvalue whose geometric multiplicity is less than its algebraic multiplicity is defective.

Definition

Let $\mathcal{X}$ be an eigenspace of A with eigenvalues $\mathcal{L}$. Then $\mathcal{X}$ is a simple eigenspace of A if

$$\mathcal{L} \cap [\Lambda(A) \setminus \mathcal{L}] = \emptyset.$$

In the other words, an eigenspace is simple if its eigenvalues are disjoint from the other eigenvalues of A.

Theorem

Let (L_1, X_1) *be a simple orthonormal eigenpairs of* A *and let* (X_1, Y_2) *be unitary so that*

$$\begin{bmatrix} X_1^H \\ Y_2^H \end{bmatrix} A \begin{bmatrix} X_1 & Y_2 \end{bmatrix} = \begin{bmatrix} L_1 & H \\ 0 & L_2 \end{bmatrix}.$$

Then there is a matrix Q *satisfying the Sylvester equation*

$$L_1 Q - Q L_2 = -H$$

such that if we set

$$X = \begin{bmatrix} X_1 & X_2 \end{bmatrix} \quad \text{and} \quad Y = \begin{bmatrix} Y_1 & Y_2 \end{bmatrix},$$

where

$$X_2 = Y_2 + X_1 Q \quad \text{and} \quad Y_1 = X_1 - Y_2 Q^H,$$

then

$$Y^H X = I \quad \text{and} \quad Y^H AX = \text{diag}(L_1, L_2).$$

Proof: Since (L_1, X_1) is a simple eigenpairs of A, it implies that

$$\Lambda(L_1) \cap \lambda(L_2) = \emptyset.$$

By Theorem 1.18 in Chapter 1, there is a unique matrix Q satisfying

$$L_1 Q - Q L_2 = -H$$

such that

$$\begin{bmatrix} I & -Q \\ 0 & I \end{bmatrix} \begin{bmatrix} L_1 & H \\ 0 & L_2 \end{bmatrix} \begin{bmatrix} I & Q \\ 0 & I \end{bmatrix} = \text{diag}(L_1, L_2).$$

That is

$$\begin{bmatrix} I & -Q \\ 0 & I \end{bmatrix} \begin{bmatrix} X_1^H \\ Y_2^H \end{bmatrix} A \begin{bmatrix} X_1 & Y_2 \end{bmatrix} \begin{bmatrix} I & Q \\ 0 & I \end{bmatrix} = \text{diag}(L_1, L_2).$$

Therefore,

$$\begin{bmatrix} X_1^H - QY_2^H \\ Y_2^H \end{bmatrix} A \begin{bmatrix} X_1 & X_1Q + Y_2 \end{bmatrix} = \text{diag}(L_1, L_2).$$

■

Observations

1. X and Y are said to be biorthogonal.
2. Since

$$A \begin{bmatrix} X_1 & X_2 \end{bmatrix} = \begin{bmatrix} X_1 & X_2 \end{bmatrix} \text{diag}(L_1, L_2),$$

   we see that

$$AX_2 = X_2L_2,$$

   so that (L_2, X_2) is an eigenpair of A. Likewise (L_1, Y_1) is a left eigenpair of A.

Let x and y be nonzero vectors. Then the angle $\angle(x, y)$ of x and y is defined as

$$\cos\angle(x, y) = \frac{|x^H y|}{\| x \|_2 \| y \|_2}.$$

Extend this definition to subspaces in $\mathbb{C}^n$. Let $\mathcal{X}$ and $\mathcal{Y}$ be subspaces of the same dimension. Let X and Y be orthonormal bases for $\mathcal{X}$ and $\mathcal{Y}$, respectively, and define $C = Y^H X$. We have

$$\| C \|_2 \leq \| X \|_2 \| Y \|_2 = 1.$$

Hence all the singular value of C lie in $[0, 1]$ and can be regarded as cosine of angles.

Definition

Let $\mathcal{X}$ and $\mathcal{Y}$ be subspaces of $\mathbb{C}^n$ of dimension p and let X and Y be orthonormal bases for $\mathcal{X}$ and $\mathcal{Y}$, respectively. Then the canonical angles between $\mathcal{X}$ and $\mathcal{Y}$ are

$$\theta_i(\mathcal{X}, \mathcal{Y}) = \cos^{-1} \gamma_i, \tag{1}$$

with

$$\theta_1(\mathcal{X}, \mathcal{Y}) \geq \theta_2(\mathcal{X}, \mathcal{Y}) \geq \cdots \geq \theta_p(\mathcal{X}, \mathcal{Y}),$$

where γ_i are the singular values of $Y^H X$.

- If the canonical angle is small, then the computation of (1) will give inaccurate results.
- For small θ, $\cos(\theta) \cong 1 - \frac{1}{2}\theta^2$. If $\theta \leq 10^{-8}$, then $\cos(\theta)$ will evaluate to 1 in IEEE double-precision arithmetic, and we will conclude that $\theta = 0$.
- The cure for this problem is to compute the sine of the canonical angles.

Theorem

Let X and Y be orthonormal bases for $\mathcal{X}$ and $\mathcal{Y}$, and let $Y_\perp$ be an orthonormal basis for the orthogonal complement of $\mathcal{Y}$. Then the singular values of $Y_\perp^H X$ are the sines of the canonical angles between $\mathcal{X}$ and $\mathcal{Y}$.

Proof: Let

$$\begin{bmatrix} Y^H \\ Y_\perp^H \end{bmatrix} X = \begin{bmatrix} C \\ S \end{bmatrix}.$$

By the orthonormality, we have

$$I = C^H C + S^H S.$$

Let

$$V^H (C^H C) V = \Gamma^2 \equiv \text{diag}(\gamma_1^2, \cdots, \gamma_p^2)$$

be the spectral decomposition of $C^H C$. Then by the definition of canonical angle θ_i in (1), we have

$$\theta_i = \cos^{-1} \gamma_i.$$

But

$$I = V^H (C^H C + S^H S) V = \Gamma^2 + V^H (S^H S) V \equiv \Gamma^2 + \Sigma^2.$$

It follows that

$$\Sigma^2 \equiv \text{diag}(\sigma_1^2, \cdots, \sigma_p^2) = \text{diag}(1-\gamma_1^2, \cdots, 1-\gamma_p^2),$$

where σ_i are singular values of $S = Y_\perp^H X$. Therefore,

$$\sigma_i^2 = 1-\gamma_i^2 = 1-\cos^2\theta_i = \sin^2\theta_i \quad \Rightarrow \quad \theta_i = \sin^{-1}\sigma_i.$$

■

Theorem

Let x be a vector with $\| x \|_2 = 1$ and let $\mathcal{Y}$ be a subspace. Then

$$\sin\angle(x, \mathcal{Y}) = \min_{y\in\mathcal{Y}} \| x - y \|_2 .$$

Proof: Let $(Y, Y_\perp)$ be unitary with $\mathcal{R}(Y) = \mathcal{Y}$. Let $y \in \mathcal{Y}$, then

$$\begin{bmatrix} Y^H \\ Y_\perp^H \end{bmatrix} (x - y) = \begin{bmatrix} \hat{x} \\ \hat{x}_\perp \end{bmatrix} - \begin{bmatrix} \hat{y} \\ 0 \end{bmatrix} = \begin{bmatrix} \hat{x} - \hat{y} \\ \hat{x}_\perp \end{bmatrix}.$$

It implies that

$$\| x - y \|_2 = \left\| \begin{bmatrix} Y^H \\ Y_\perp^H \end{bmatrix} (x - y) \right\|_2 = \left\| \begin{bmatrix} \hat{x} - \hat{y} \\ \hat{x}_\perp \end{bmatrix} \right\|_2$$

and hence

$$\min_{y \in \mathcal{Y}} \|x - y\| = \|\hat{x}_\perp\|_2 = \|Y_\perp^H x\|_2. \tag{2}$$

By Theorem 10 and (2), we have

$$\sin \angle(x, \mathcal{Y}) = \|Y_\perp^H x\|_2 = \min_{y \in \mathcal{Y}} \|x - y\|.$$

Theorem

Let X and Y be orthonormal matrices with $X^H Y = 0$ and let $Z = X + YQ$. Let

$$\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_k > 0 \quad \text{and} \quad \zeta_1 \geq \zeta_2 \geq \cdots \geq \zeta_k > 0$$

denote the nonzero singular values of Z and Q, respectively. Set

$$\theta_1 \geq \theta_2 \geq \cdots \geq \theta_k$$

to be the nonzero canonical angle between $\mathcal{R}(X)$ and $\mathcal{R}(Z)$. Then

$$\sigma_i = \sec\theta_i \quad \text{and} \quad \zeta_i = \tan\theta_i, \quad \text{for } i = 1, \ldots, k.$$

Proof: Since

$$X^H X = I, \quad Y^H Y = I, \quad X^H Y = 0 \quad \text{and} \quad Z = X + YQ,$$

we have

$$Z^H Z = (X^H + Q^H Y^H)(X + YQ) = I + Q^H Q.$$

This implies that

$$\sigma_i^2 = 1 + \zeta_i^2, \ \text{for} \ i = 1, \ldots, k. \tag{3}$$

Define

$$\hat{Z} \equiv Z(Z^H Z)^{-1/2} = (X + YQ)(I + Q^H Q)^{-1/2},$$

where $(I + Q^H Q)^{-1/2}$ is the inverse of the positive definite square root of $I + Q^H Q$. Then $\hat{Z}$ is an orthonormal basis for $\mathcal{R}(Z)$ and

$$X^H \hat{Z} = (I + QQ^H)^{-1/2}.$$

Hence the singular values γ_i of $X^H \hat{Z}$ are

$$\gamma_i = \left(\sqrt{1 + \zeta_i^2} \right)^{-1}$$

for $i = 1, \ldots, k$. Using (3) and the definition of canonical angles θ_i between $\mathcal{R}(X)$ and $\mathcal{R}(Z)$, we have

$$\cos \theta_i = \gamma_i = \sigma_i^{-1}.$$

That is

$$\sigma_i = \frac{1}{\cos \theta_i} = \sec \theta_i.$$

The relation $\tan \theta_i = \zeta_i$ now follows from (3). ∎

Let (L_1, X_1) be a simple right orthonormal eigenpair of A and let (X_1, Y_2) be unitary. From Theorem 8, $(L_1, Y_1 \equiv X_1 - Y_2 Q^H)$ is left eigenpair of A and $Y_1^H X_1 = I$. By Theorem 12, we obtain the following corollary.

Corollary

Let X be an orthonormal basis for a simple eigenspace $\mathcal{X}$ of A and let Y be a basis for the corresponding left eigenspace $\mathcal{Y}$ of A normalized so that $Y^H X = I$. Then the singular values of Y are the secants of the canonical angles between $\mathcal{X}$ and $\mathcal{Y}$. In particular,

$$\|Y\|_2 = \sec \theta_1(\mathcal{X}, \mathcal{Y}).$$

Theorem

Let $[X \ \ X_\perp]$ *be unitary. Let* $R = AX - XL$ *and* $S^H = X^H A - LX^H$. *Then* $\|R\|$ *and* $\|S\|$ *are minimized when*

$$L = X^H AX,$$

in which case

$$\|R\| = \|X_\perp^H AX\| \quad \textit{and} \quad \|S\| = \|X^H AX_\perp\|.$$

Proof: Set

$$\begin{bmatrix} X^H \\ X_\perp^H \end{bmatrix} A \begin{bmatrix} X & X_\perp \end{bmatrix} = \begin{bmatrix} \hat{L} & H \\ G & M \end{bmatrix}.$$

Then

$$\begin{bmatrix} X^H \\ X_\perp^H \end{bmatrix} R = \begin{bmatrix} \hat{L} & H \\ G & M \end{bmatrix} \begin{bmatrix} X^H \\ X_\perp^H \end{bmatrix} X - \begin{bmatrix} X^H \\ X_\perp^H \end{bmatrix} XL = \begin{bmatrix} \hat{L} - L \\ G \end{bmatrix}.$$

It implies that

$$\|R\| = \left\| \begin{bmatrix} X^H \\ X_\perp^H \end{bmatrix} R \right\| = \left\| \begin{bmatrix} \hat{L} - L \\ G \end{bmatrix} \right\|,$$

which is minimized when $L = X^H A X$ and

$$\min \|R\| = \|G\| = \|X_\perp^H A X\|.$$

The proof for S is similar. ■

Definition

Let X be of full column rank and let X^I be a left inverse of X. Then $X^I A X$ is a Rayleigh quotient of A.

Theorem

Let X be orthonormal and let

$$R = AX - XL.$$

Let $\ell_1, \ldots, \ell_k$ be the eigenvalues of L. Then there are eigenvalues $\lambda_{j_1}, \ldots, \lambda_{j_k}$ of A such that

$$|\ell_i - \lambda_{j_i}| \leq \|R\|_2$$

and

$$\sqrt{\sum_{i=1}^{k} (\ell_i - \lambda_{j_i})^2} \leq \sqrt{2}\|R\|_F.$$

- Power method: compute the dominant eigenpair
- Disadvantage: at each step it considers only the single vector $A^k u$, which amounts to throwing away the information contained in $u, Au, A^2u, \ldots, A^{k-1}u$.

Definition

Let A be of order n and let $u \neq 0$ be an n vector. Then

$$\{u, Au, A^2u, A^3u, \ldots\}$$

is a Krylov sequence based on A and u. We call the matrix

$$K_k(A,u) = \begin{bmatrix} u & Au & A^2u & \cdots & A^{k-1}u \end{bmatrix}$$

the kth Krylov matrix. The space

$$\mathcal{K}_k(A,u) = \mathcal{R}[K_k(A,u)]$$

is called the kth Krylov subspace.

Theorem

Let A and $u \neq 0$ be given. Then

1. *The sequence of Krylov subspaces satisfies*

$$\mathcal{K}_k(A,u) \subset \mathcal{K}_{k+1}(A,u), \quad A\mathcal{K}_k(A,u) \subset \mathcal{K}_{k+1}(A,u).$$

2. *If $\sigma \neq 0$, then*

$$\mathcal{K}_k(A,u) = \mathcal{K}_k(\sigma A,u) = \mathcal{K}_k(A,\sigma u).$$

3. *For any κ,*

$$\mathcal{K}_k(A,u) = \mathcal{K}_k(A-\kappa I,u).$$

4. *If W is nonsingular, then*

$$\mathcal{K}_k(W^{-1}AW, W^{-1}u) = W^{-1}\mathcal{K}_k(A,u).$$

A Krylov sequence terminates at ℓ if ℓ is the smallest integer such that

$$\mathcal{K}_{\ell+1}(A,u) = \mathcal{K}_{\ell}(A,u).$$

Theorem

A Krylov sequence terminates based on A and u at ℓ if and only if ℓ is the smallest integer for which

$$\textit{dim}[\mathcal{K}_{\ell+1}] = \textit{dim}[\mathcal{K}_{\ell}].$$

If the Krylov sequence terminates at ℓ, then $\mathcal{K}_{\ell}$ is an eigenspace of A of dimension ℓ. On the other hand, if u lies in an eigenspace of dimension m, then for some $\ell \le m$, the sequence terminates at ℓ.

Proof:

- If $\mathcal{K}_{\ell+1} = \mathcal{K}_\ell$, then dim$[\mathcal{K}_{\ell+1}] =dim[\mathcal{K}_\ell]$. On the other hand, if dim$[\mathcal{K}_{\ell+1}] =$dim$[\mathcal{K}_\ell]$, then $\mathcal{K}_{\ell+1} = \mathcal{K}_\ell$ because $\mathcal{K}_\ell \subset \mathcal{K}_{\ell+1}$.
- If the sequence terminates at ℓ, then

$$A\mathcal{K}_\ell \subset \mathcal{K}_{\ell+1} = \mathcal{K}_\ell,$$

  so that $\mathcal{K}_\ell$ is an invariant subspace of A.
- Let $\mathcal{X}$ be an invariant subspace with dimension m. If $u \in \mathcal{X}$, then $A^i u \in \mathcal{X}$ for all i. That is $\mathcal{K}_i \subset \mathcal{X}$ and dim$(\mathcal{K}_i) \leq m$ for all i. If the sequence terminates at $\ell > m$, then $\mathcal{K}_\ell$ is an invariant subspace and dim$(\mathcal{K}_\ell) >$ dim$(\mathcal{K}_m) =$ dim$(\mathcal{X})$, which is impossible. ■

By the definition of $\mathcal{K}_k(A, u)$, for any vector $v \in \mathcal{K}_k(A, u)$ can be written in the form

$$v = \gamma_1 u + \gamma_2 Au + \cdots + \gamma_k A^{k-1} u \equiv p(A)u,$$

where

$$p(A) = \gamma_1 I + \gamma_2 A + \gamma_3 A^2 + \cdots + \gamma_k A^{k-1}.$$

Assume that A is Hermitian and has an orthonormal eigenpairs (λ_i, x_i) for $i = 1, \ldots, n$. Write u in the form

$$u = \alpha_1 x_1 + \alpha_2 x_2 + \cdots + \alpha_n x_n,$$

where $\alpha_i = x_i^H u$. Since $p(A)x_i = p(\lambda_i)x_i$, we have

$$p(A)u = \alpha_1 p(\lambda_1)x_1 + \alpha_2 p(\lambda_2)x_2 + \cdots + \alpha_n p(\lambda_n)x_n. \tag{4}$$

If $p(\lambda_i)$ is large compared with $p(\lambda_j)$ for $j \neq i$, then $p(A)u$ is a good approximation to x_i.

Theorem

If $x_i^H u \neq 0$ *and* $p(\lambda_i) \neq 0$*, then*

$$\tan\angle(p(A)u, x_i) \leq \max_{j \neq i} \frac{|p(\lambda_j)|}{|p(\lambda_i)|} \tan\angle(u, x_i).$$

Proof: From (4), we have

$$\cos\angle(p(A)u, x_i) = \frac{|x_i^H p(A)u|}{\|p(A)u\|_2 \|x_i\|_2} = \frac{|\alpha_i p(\lambda_i)|}{\sqrt{\sum_{j=1}^n |\alpha_j p(\lambda_j)|^2}}$$

and

$$\sin\angle(p(A)u, x_i) = \frac{\sqrt{\sum_{j \neq i} |\alpha_j p(\lambda_j)|^2}}{\sqrt{\sum_{j=1}^n |\alpha_j p(\lambda_j)|^2}}$$

Hence

$$\begin{aligned}
\tan^2 \angle(p(A)u, x_i) &= \sum_{j\neq i} \frac{|\alpha_j p(\lambda_j)|^2}{|\alpha_i p(\lambda_i)|^2} \\
&\leq \max_{j\neq i} \frac{|p(\lambda_j)|^2}{|p(\lambda_i)|^2} \sum_{j\neq i} \frac{|\alpha_j|^2}{|\alpha_i|^2} \\
&= \max_{j\neq i} \frac{|p(\lambda_j)|^2}{|p(\lambda_i)|^2} \tan^2 \angle(u, x_i).
\end{aligned}$$

■

Assume that $p(\lambda_i) = 1$, then

$$\tan \angle(p(A)u, x_i) \leq \max_{j\neq i, p(\lambda_i)=1} |p(\lambda_j)| \tan \angle(u, x_i) \quad \forall \quad p(A)u \in \mathcal{K}_k.$$

Hence

$$\tan \angle(x_i, \mathcal{K}_k) \leq \min_{deg(p)\leq k-1, p(\lambda_i)=1} \max_{j\neq i} |p(\lambda_j)| \tan \angle(u, x_i).$$

Assume that

$$\lambda_1 > \lambda_2 \geq \cdots \geq \lambda_n$$

and that our interest is in the eigenvector x_1. Then

$$\tan \angle(x_1, \mathcal{K}_k) \leq \min_{deg(p)\leq k-1, p(\lambda_1)=1} \max_{\lambda\in[\lambda_n,\lambda_2]} |p(\lambda)| \tan \angle(u, x_1).$$

Question

How to compute

$$\min_{deg(p)\leq k-1, p(\lambda_1)=1} \max_{\lambda\in[\lambda_n,\lambda_2]} |p(\lambda)|?$$

Definition

The Chebyshev polynomials are defined by

$$c_k(t) = \begin{cases} \cos(k \cos^{-1} t), & |t| \leq 1, \\ \cosh(k \cosh^{-1} t), & |t| \geq 1. \end{cases}$$

Theorem

(a) $c_0(t) = 1,\ c_1(t) = t$ *and*

$$c_{k+1}(t) = 2c_k(t) - c_{k-1}(t), \quad k = 1, 2, \ldots.$$

(b) *For* $|t| > 1$,

$$c_k(t) = (1 + \sqrt{t^2 - 1})^k + (1 + \sqrt{t^2 - 1})^{-k}.$$

(c) *For* $t \in [-1, 1], \quad |c_k(t)| \le 1$. *Moreover, if*

$$t_{ik} = \cos \frac{(k-i)\pi}{k}, \quad i = 0, 1, \ldots, k,$$

then

$$c_k(t_{ik}) = (-1)^{k-i}.$$

(d) For $s > 1$,

$$\min_{deg(p)\leq k, p(s)=1} \max_{t\in[0,1]} |p(t)| = \frac{1}{c_k(s)}, \tag{5}$$

and the minimum is obtained only for $p(t) = c_k(t)/c_k(s)$.

For applying (5), we define

$$\lambda = \lambda_2 + (\mu - 1)(\lambda_2 - \lambda_n)$$

to transform interval $[\lambda_n, \lambda_2]$ to $[0, 1]$. Then the value of μ at λ_1 is

$$\mu_1 = 1 + \frac{\lambda_1 - \lambda_2}{\lambda_2 - \lambda_n}$$

and

$$\min_{deg(p)\leq k-1, p(\lambda_1)=1} \max_{\lambda\in[\lambda_n,\lambda_2]} |p(\lambda)|$$

$$= \min_{deg(p)\leq k-1, p(\mu_1)=1} \max_{\mu\in[0,1]} |p(\mu)| = \frac{1}{c_{k-1}(\mu_1)}.$$

Theorem

Let the Hermitian matrix A have an orthonormal eigenpair (λ_i, x_i) with

$$\lambda_1 > \lambda_2 \geq \cdots \geq \lambda_n.$$

Let

$$\eta = \frac{\lambda_1 - \lambda_2}{\lambda_2 - \lambda_n}.$$

Then

$$\begin{aligned}\tan\angle[x_1, \mathcal{K}_k(A,u)] &\leq \frac{\tan\angle(x_1,u)}{c_{k-1}(1+\eta)} \\ &= \frac{\tan\angle(x_1,u)}{(1+\sqrt{2\eta+\eta^2})^{k-1} + (1+\sqrt{2\eta+\eta^2})^{1-k}}.\end{aligned}$$

Remark

- *For k large, we have*

$$\tan \angle[x_1, \mathcal{K}_k(A,u)] \lesssim \frac{\tan \angle(x_1, u)}{(1+\sqrt{2\eta+\eta^2})^{k-1}}.$$

- *For k large and if η is small, then the bound becomes*

$$\tan \angle[x_1, \mathcal{K}_k(A,u)] \lesssim \frac{\tan \angle(x_1, u)}{(1+\sqrt{2\eta})^{k-1}}.$$

- *Compare it with power method:*
  If $|\lambda_1| > |\lambda_2| \geq \cdots \geq |\lambda_n|$, then the convergence of the power method is $|\lambda_2/\lambda_1|^k$.

- For example, let

$$\lambda_1 = 1, \lambda_2 = 0.95, \lambda_3 = 0.95^2, \cdots, \lambda_{100} = 0.95^{99}$$

be the eigenvalues of 100-by-100 matrix A. Then $\eta = 0.0530$ and the bound on the convergence rate is $1/(1+\sqrt{2\eta}) = 0.7544$. Thus the square root effect gives a great improvement over the rate of 0.95 for the power method.

- Replaced A by $-A$, then the Krylov sequence converges to the eigenvector corresponding to the smallest eigenvalue of A. However, the smallest eigenvalues of a matrix – particularly a positive definite matrix – often tend to cluster together, so that the bound will be unfavorable.

- The hypothesis $\lambda_1 > \lambda_2$ can be relaxed. Suppose that $\lambda_1 = \lambda_2 > \lambda_3$. Expand u in the form

$$u = \alpha_1 x_1 + \alpha_2 x_2 + \alpha_3 x_3 + \cdots + \alpha_n x_n.$$

Then

$$A^k u = \lambda_1^k(\alpha_1 x_1 + \alpha_2 x_2) + \alpha_3 \lambda_3^k x_3 + \cdots + \alpha_n \lambda_n^k x_n.$$

This shows that the spaces $\mathcal{K}_k(A, u)$ contain only approximations to $\alpha_1 x_1 + \alpha_2 x_2$.

Theorem

Let λ be a simple eigenvalue of A and let

$$A = \begin{bmatrix} x & X \end{bmatrix} \begin{bmatrix} \lambda & 0 \\ 0 & L \end{bmatrix} \begin{bmatrix} y^H \\ Y^H \end{bmatrix} = \lambda x y^H + XLY^H$$

be a spectral representation. Let

$$u = \alpha x + Xa,$$

where

$$\alpha = y^H u \quad \text{and} \quad a = Y^H u.$$

Then

$$\sin \angle [x, \mathcal{K}_k(A, u)] \leq |\alpha|^{-1} \min_{deg(p) \leq k-1, p(\lambda)=1} \|Xp(L)a\|_2.$$

Proof: From Theorem 11,

$$\begin{aligned}\sin\angle[x,\mathcal{K}_k(A,u)] &= |\alpha|^{-1}\min_{y\in\mathcal{K}_k(A,u)}\|\alpha x-y\|_2\\ &= |\alpha|^{-1}\min_{deg(p)\le k-1}\|\alpha x-p(A)u\|_2\\ &\le |\alpha|^{-1}\min_{deg(p)\le k-1,p(\lambda)=1}\|\alpha x-p(A)u\|_2.\end{aligned}$$

Since

$$p(\lambda)=1 \quad \text{and} \quad AX=XL,$$

we have

$$p(A)u=p(A)(\alpha x+Xa)=\alpha p(\lambda)x+Xp(L)a=\alpha x+Xp(L)a.$$

Hence

$$\begin{aligned}\sin\angle[x,\mathcal{K}_k(A,u)] &\le |\alpha|^{-1}\min_{deg(p)\le k-1,p(\lambda)=1}\|\alpha x-(\alpha x+Xp(L)a)\|_2\\ &= |\alpha|^{-1}\min_{deg(p)\le k-1,p(\lambda)=1}\|Xp(L)a\|_2.\end{aligned}$$

Let (λ_i, x_i) be an eigenpair of A for $i = 1, \ldots, n$. Write vector u in the form

$$u = \alpha_1 x_1 + \alpha_2 x_2 + \cdots + \alpha_n x_n.$$

Assume that λ_1 is double, i.e., $\lambda_1 = \lambda_2$. Then

$$A^k u = \lambda_1^k(\alpha_1 x_1 + \alpha_2 x_2) + \lambda_3^k \alpha_3 x_3 + \cdots + \lambda_n^k \alpha_n x_n.$$

Hence the Krylov sequence can only produce the approximation $\alpha_1 x_1 + \alpha_2 x_2$ to a vector in the eigenspace of λ_1. Let U be a matrix with linearly independent columns. Then the sequence

$$\{U, AU, A^2U, \ldots\}$$

is called a block Krylov sequence and the space $\mathcal{K}_k(A, U)$ is called the k-th block Krylov space.

Gaol: passing to block Krylov sequence improves the convergence bound.

Theorem

Let A be Hermitian and let (λ_i, x_i) be a complete system (n eigenvectors are linearly independent) of orthonormal eigenpairs of A with

$$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n,$$

and assume that the multiplicity of λ_1 is not greater than m. If

$$B = \begin{bmatrix} x_1^H \\ \vdots \\ x_m^H \end{bmatrix} U$$

is nonsingular and we set

$$v = UB^{-1}e_1,$$

then

$$\begin{aligned} \tan\angle[x_1, \mathcal{K}_k(A,U)] &\leq \frac{\tan\angle(x_1, v)}{c_{k-1}(1+2\eta)} \\ &= \frac{\tan\angle(x_1, v)}{(1+2\sqrt{\eta+\eta^2})^{k-1} + (1+2\sqrt{\eta+\eta^2})^{1-k}} \end{aligned} \quad (6)$$

where

$$\eta = \frac{\lambda_1 - \lambda_{m+1}}{\lambda_{m+1} - \lambda_n}.$$

Proof: Since $v \in \mathcal{R}(U)$, we have $\mathcal{K}_k(A,v) \subset \mathcal{K}_k(A,U)$. By Theorem 11,

$$\begin{aligned} \sin\angle[x_1, \mathcal{K}_k(A,U)] &= \min_{y \in \mathcal{K}_k(A,U)} \|x_1 - y\|_2 \\ &\leq \min_{y \in \mathcal{K}_k(A,v)} \|x_1 - y\|_2 = \sin\angle[x_1, \mathcal{K}_k(A,v)]. \end{aligned}$$

This implies that

$$\angle[x_1, \mathcal{K}_k(A, U)] \leq \angle[x_1, \mathcal{K}_k(A, v)].$$

By the definition of B, we have

$$\begin{bmatrix} x_1^H \\ \vdots \\ x_n^H \end{bmatrix} UB^{-1} = I \quad \Rightarrow \quad x_i^H UB^{-1} = e_i^T \text{ for } i = 1, \ldots, m.$$

By the definition of v,

$$x_i^H v = x_i^H UB^{-1} e_1 = 0 \text{ for } i = 2, \ldots, m.$$

On the other hand, (λ_i, x_i) is an eigenpair of Hermitian A, i.e., $x_i^H A = \lambda_i x_i^H$. Hence

$$x_i^H A^j v = \lambda_i^j x_i^H v = 0 \text{ for } i = 2, \ldots, m.$$

This implies that $x_2, \ldots, x_m$ are not contained in $\mathcal{K}_k(A, v)$. That is

$$A^j v = \alpha_1 \lambda_1^j x_1 + \alpha_{m+1} \lambda_{m+1}^j x_{m+1} + \cdots + \alpha_n \lambda_n^j x_n$$

for $j = 1, \ldots, k-1$. We may now apply Theorem 23 to get (6). ∎

Theorem

Let $\mathcal{U}$ be a subspace and let U be a basis for $\mathcal{U}$. Let V be a left inverse of U and set

$$B = V^H AU.$$

If $\mathcal{X} \subset \mathcal{U}$ is an eigenspace of A, then there is an eigenpair (L, W) of B such that (L, UW) is an eigenpair of A with $\mathcal{R}(UW) = \mathcal{X}$.

Proof: Let (L, X) be an eigenpair of A and let $X = UW$. Then from the relation

$$AUW = UWL$$

we obtain

$$BW = V^H AUW = V^H UWL = WL,$$

so that (L, W) is an eigenpair of B. ∎

We can find exact eigenspaces contained in $\mathcal{U}$ by looking at eigenpairs of the Rayleigh quotient B.

Algorithm (Rayleigh-Ritz procedure)

1. *Let U be a basis for $\mathcal{U}$ and let V^H be a left inverse of U.*
2. *Form the Rayleigh quotient $B = V^H AU$.*
3. *Let (M, W) be a suitable eigenpair of B.*
4. *Return (M, UW) as an approximate eigenpair of A.*

- (M, UW) is called a Ritz pair. Written Ritz pair in the form (λ, Uw), we will call λ a Ritz value and Uw a Ritz vector.
- Two difficulties for Rayleigh-Ritz procedure: (i) how to choose the eigenpair (M, W) in statement 3. (ii) no guarantee that the result approximates the desired eigenpair of A.

Example

Let $A = \text{diag}(0, 1, -1)$ and suppose we interested in approximating the eigenpair $(0, e_1)$. Assume

$$U = \begin{bmatrix} 1 & 0 \\ 0 & 1/\sqrt{2} \\ 0 & 1/\sqrt{2} \end{bmatrix}.$$

Then

$$B = U^H A U = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$$

and any nonzero vector p is an eigenvector of B. If we take $p = [1, 1]^T$, then $Up = [1, 1/\sqrt{2}, 1/\sqrt{2}]$ is an approximate eigenvector of A, which is completely wrong. Thus the method can fail, even though the space $\mathcal{U}$ contains the desired eigenvector. ∎

- The matrices U and V in Algorithm 1 satisfy the condition $V^H U = I$ and they can differ. Hence Algorithm 1 is called an oblique Rayleigh-Ritz method.
- If the matrix U is taken to be orthonormal and $V = U$. In addition, W is taken to be orthonormal, so that $\hat{X} \equiv UW$ is also orthonormal. We call this procedure the orthogonal Rayleigh-Rite method.

Theorem

Let $(M, \hat{X} \equiv UW)$ be an orthogonal Rayleigh-Rite pair. Then

$$R = A\hat{X} - \hat{X}M$$

is minimal in any unitarily invariant norm.

Proof: By Theorem 14, we need to show $M = \hat{X}^H A \hat{X}$. Since (M, W) is an eigenpair of B and W is orthonormal, we have

$$M = W^H B W$$

and

$$\hat{X}^H A \hat{X} = W^H U^H A U W = W^H B W = M.$$

∎

Let (λ, x) be the desired eigenpair of A and U_θ be an orthonormal basis for which $\theta = \angle(x, U_\theta)$ is small.

Theorem

Let

$$B_\theta = U_\theta^H A U_\theta.$$

Then there is a matrix E_θ satisfying

$$\|E_\theta\|_2 \leq \frac{\sin\theta}{\sqrt{1-\sin^2\theta}}\|A\|_2$$

such that λ is an eigenvalue of $B_\theta + E_\theta$.

Proof. Let $(U_\theta, U_\perp)$ be unitary and set

$$y = U_\theta^H x \quad \text{and} \quad z = U_\perp^H x.$$

From Theorem 10, we have

$$\|z\|_2 = \sin\theta \quad \text{and} \quad \|y\|_2 = \sqrt{1-\sin^2\theta}.$$

Since $Ax - \lambda x = 0$, we have

$$U_\theta^H A[U_\theta, U_\perp] \begin{bmatrix} U_\theta^H \\ U_\perp^H \end{bmatrix} x - \lambda U_\theta^H x = 0,$$

or

$$B_\theta y + U_\theta^H A U_\perp z - \lambda y = 0.$$

Let $\hat{y} = y/\|y\|_2 = y/\sqrt{1-\sin^2\theta}$. If

$$r \equiv B_\theta \hat{y} - \lambda \hat{y} = \frac{-1}{\sqrt{1-\sin^2\theta}} U_\theta^H A U_\perp z,$$

it follows that

$$\|r\|_2 \le \frac{\sin\theta}{\sqrt{1-\sin^2\theta}} \|A\|_2.$$

Now define

$$E_\theta = -r\hat{y}^H.$$

Then

$$\|E_\theta\|_2 = \sqrt{\lambda_1((r\hat{y}^H)(\hat{y}r^H))} = \sqrt{\lambda_1(rr^H)} = \|r\|_2 \leq \frac{\sin\theta}{\sqrt{1-\sin^2\theta}}\|A\|_2$$

and

$$(B_\theta + E_\theta)\hat{y} = B_\theta\hat{y} - (r\hat{y}^H)\hat{y} = B_\theta\hat{y} - r = \lambda\hat{y}.$$

Therefore, $(\lambda, \hat{y})$ is an eigenpair of $B_\theta + E_\theta$. ■

Corollary

There is an eigenvalue μ_θ of B_θ such that

$$|\mu_\theta - \lambda| \leq 4(2\|A\|_2 + \|E_\theta\|_2)^{1-1/m}\|E_\theta\|_2^{1/m},$$

where m is the order of B_θ.

Theorem

Let (μ_θ, w_θ) *be an eigenpair of* B_θ *and let* $\begin{bmatrix} w_\theta & W_\theta \end{bmatrix}$ *be unitary, so that*

$$\begin{bmatrix} w_\theta^H \\ W_\theta^H \end{bmatrix} B_\theta \begin{bmatrix} w_\theta & W_\theta \end{bmatrix} = \begin{bmatrix} \mu_\theta & h_\theta^H \\ 0 & N_\theta \end{bmatrix}.$$

Then

$$\sin \angle(x, U_\theta w_\theta) \leq \sin\theta \sqrt{1 + \frac{\|h_\theta\|_2^2}{\textit{sep}(\lambda, N_\theta)^2}},$$

where sep$(\lambda, N_\theta) = \|(\lambda I - N_\theta)^{-1}\|^{-1}$.

By the continuity of sep, we have

$$
\begin{aligned}
& |\text{sep}(\lambda, N_\theta) - \text{sep}(\mu_\theta, N_\theta)| \leq |\mu_\theta - \lambda| \\
\Rightarrow \quad & \text{sep}(\lambda, N_\theta) \geq \text{sep}(\mu_\theta, N_\theta) - |\mu_\theta - \lambda|.
\end{aligned}
$$

Suppose $\mu_\theta \to \lambda$ and $\text{sep}(\mu_\theta, N_\theta)$ is bounded below. Then $\text{sep}(\lambda, N_\theta)$ is also bounded below. Since $\|h_\theta\|_2 \leq \|A\|_2$, we have $\sin \angle(x, U_\theta w_\theta) \to 0$ along with θ.

Corollary

Let $(\mu_\theta, U_\theta w_\theta)$ be a Ritz pair for which $\mu_\theta \to \lambda$. If there is a constant $\alpha > 0$ such that

$$\text{sep}(\mu_\theta, N_\theta) \geq \alpha > 0, \tag{7}$$

then

$$\sin \angle(x, U_\theta w_\theta) \lesssim \sin\theta \sqrt{1 + \frac{\|A\|_2}{\alpha^2}}.$$

This corollary justifies that eigenvalue convergence plus separation equals eigenvector convergence.
The condition (7) is called the uniform separation condition.
Let (μ_θ, x_θ) with $\|x_\theta\|_2 = 1$ be the Ritz approximation to (λ, x).
Then by construction, we have

$$\mu_\theta = x_\theta^H A x_\theta. \tag{8}$$

Write

$$x_\theta = \gamma x + \sigma y, \tag{9}$$

where $y \perp x$ and $\|y\|_2 = 1$. Then

$$|\gamma| = |x^H x_\theta| = \cos\angle(x_\theta, x) \quad \text{and} \quad |\sigma| = |y^H x_\theta| = \sin\angle(x_\theta, x).$$

If the uniform separation is satisfied, we have

$$|\sigma| = \sin\angle(x_\theta, x) = O(\theta).$$

Substituting (9) into (8) and using the facts of $Ax = \lambda x$ and $y^H x = 0$, we find that

$$\begin{aligned} \mu_\theta &= (\bar{\gamma} x^H + \bar{\sigma} y^H)(\gamma \lambda x + \sigma A y) \\ &= |\gamma|^2 \lambda + \sigma x_\theta^H A y. \end{aligned}$$

Hence

$$\begin{aligned} |\mu_\theta - \lambda| &= |(|\gamma|^2 - 1)\lambda + \sigma x_\theta^H A y| \\ &\leq |\sigma|^2 |\lambda| + |\sigma| \|x_\theta\|_2 \|A\|_2 \|y\|_2 \\ &\leq |\sigma|(1 + |\sigma|)\|A\|_2 \\ &= O(\theta). \end{aligned}$$

Thus the Ritz value converges at least as fast as the eigenvector approximation of x in $\mathcal{U}(U_\theta)$.

If A is Hermitian, then

$$\begin{aligned}\mu_\theta &= (\bar{\gamma}x^H + \bar{\sigma}y^H)(\gamma\lambda x + \sigma Ay)\\ &= |\gamma|^2\lambda + \bar{r}\sigma x^H Ay + |\sigma|^2 y^H Ay = |\gamma|^2\lambda + \bar{r}\sigma\bar{\lambda}x^H y + |\sigma|^2 y^H Ay\\ &= |\gamma|^2\lambda + |\sigma|^2 y^H Ay\end{aligned}$$

and

$$\begin{aligned}|\mu_\theta - \lambda| &= |(|\gamma|^2 - 1)\lambda + |\sigma|^2 y^H Ay|\\ &\le |\sigma|^2|\lambda| + |\sigma|^2\|y\|_2\|A\|_2\|y\|_2\\ &\le 2|\sigma|^2\|A\|_2\\ &= O(\theta^2).\end{aligned}$$

Since the angle $\theta = \angle(x, U_\theta)$ cannot be known and hence cannot compute error bounds. Thus, we must look to the residual as an indication of convergence.

Theorem

Let A have the spectral representation

$$A = \lambda x y^H + X L Y^H,$$

where $\|x\|_2 = 1$ and Y is orthonormal. Let $(\mu, \tilde{x})$ be an approximation to (λ, x) and let

$$\rho = \|A\tilde{x} - \mu\tilde{x}\|_2.$$

Then

$$\sin\angle(\tilde{x}, x) \le \frac{\rho}{\text{sep}(\mu, L)} \le \frac{\rho}{\text{sep}(\lambda, L) - |\mu - \lambda|}.$$

Proof: Since $Y^H x = 0$, we have that Y is an orthonormal basis for the orthogonal complement of the space span$\{x\}$ and then $\sin \angle(\tilde{x}, x) = \|Y^H \tilde{x}\|_2$. Let $r = A\tilde{x} - \mu\tilde{x}$. Then

$$Y^H r = Y^H A\tilde{x} - \mu Y^H \tilde{x} = (L - \mu I) Y^H \tilde{x}.$$

It follows that

$$\sin \angle(\tilde{x}, x) = \|(L - \mu I)^{-1} Y^H r\|_2 \leq \frac{\|r\|_2}{\text{sep}(\mu, L)}.$$

By the fact that sep$(\mu, L) \geq$ sep$(\lambda, L) - |\mu - \lambda|$, the second inequality is obtained. ■

Since λ is assumed to be simple, this theorem says that: *Sufficient condition for $\tilde{x}$ to converge to x is for μ to converge to λ and for the residual to converge to zero.*

Definition

Let μ_θ be a Ritz value associated with $\mathcal{U}_\theta$. A refined Ritz vector is a solution of the problem

$$\begin{array}{ll} \min & \|A\hat{x}_\theta - \mu_\theta \hat{x}_\theta\|_2 \\ \text{subject to} & \hat{x}_\theta \in \mathcal{U}_\theta, \ \|\hat{x}_\theta\|_2 = 1. \end{array}$$

Theorem

Let A have the spectral representation

$$A = \lambda x y^H + XLY^H,$$

where $\|x\|_2 = 1$ and Y is orthonormal. Let μ_θ be a Ritz value and $\hat{x}_\theta$ the corresponding refined Ritz vector. If $\textit{sep}(\lambda, L) - |\mu_\theta - \lambda| > 0$, then

$$\sin \angle(x, \hat{x}_\theta) \leq \frac{\|A - \mu_\theta I\|_2 \sin\theta + |\lambda - \mu_\theta|}{\sqrt{1 - \sin^2\theta}\,[\textit{sep}(\lambda, L) - |\lambda - \mu_\theta|]}.$$

Proof: Let U be an orthonormal basis for $\mathcal{U}_\theta$ and let $x = y + z$, where $z = UU^H x$. Then

$$\|z\|_2 = \|U^H x\|_2 = \sin\theta.$$

Moreover, since y and x are orthogonal,

$$\begin{aligned} \|z\|_2^2 &= \|x - y\|_2^2 = (x^H - y^H)(x - y) \\ &= \|x\|_2^2 + \|y\|_2^2 = 1 + \|y\|_2^2 \\ \Rightarrow &\ \|y\|_2^2 = 1 - \|z\|_2^2 = 1 - \sin^2\theta. \end{aligned}$$

Let

$$\hat{y} = \frac{y}{\sqrt{1 - \sin^2\theta}},$$

we have

$$\begin{aligned} (A - \mu_\theta I)\hat{y} &= \frac{(A - \mu_\theta I)y}{\sqrt{1 - \sin^2\theta}} = \frac{(A - \mu_\theta I)(x - z)}{\sqrt{1 - \sin^2\theta}} \\ &= \frac{(\lambda - \mu_\theta)x - (A - \mu_\theta I)z}{\sqrt{1 - \sin^2\theta}}. \end{aligned}$$

Hence

$$\|(A-\mu_\theta I)\hat{y}\|_2 \le \frac{|\lambda-\mu_\theta| + \|A-\mu_\theta I\|\sin\theta}{\sqrt{1-\sin^2\theta}}.$$

By the definition of a refined Ritz vector we have

$$\|(A-\mu_\theta I)\hat{x}\|_2 \le \frac{|\lambda-\mu_\theta| + \|A-\mu_\theta I\|\sin\theta}{\sqrt{1-\sin^2\theta}}.$$

The result now follows from Theorem 35. ∎

Remark

- *By Corollary 30,* $\mu_\theta \to \lambda$. *It follows that* $\sin\angle(x,\hat{x}_\theta) \to 0$. *In other words, refined Ritz vectors are guaranteed to converge.*
- $\hat{\mu}_\theta = \hat{x}_\theta^H A\hat{x}_\theta$ *is more accurate than* μ_θ *and* $\|A\hat{x}_\theta - \hat{\mu}_\theta\hat{x}_\theta\|_2$ *is optimal.*

The computation of a refined Ritz vector amounts to solve

$$\begin{array}{ll} \min & \|A\hat{x} - \mu\hat{x}\|_2 \\ \text{subject to} & \hat{x} \in \mathcal{U}, \ \|\hat{x}\|_2 = 1. \end{array} \tag{10}$$

Let U be an orthonormal basis for $\mathcal{U}$. Then (10) is equivalent to

$$\begin{array}{ll} \min & \|(A - \mu I)Uz\|_2 \\ \text{subject to} & \|z\|_2 = 1. \end{array}$$

The solution of this problem is the right singular vector of $(A - \mu I)U$ corresponding to its smallest singular value. Thus refined Ritz vector can be computed by the following algorithm.

1. $V = AU$
2. $W = V - \mu U$
3. Compute the smallest singular value of W and its right singular vector z
4. $\hat{x} = Uz$

Exterior eigenvalues are easily convergent than interior eigenvalues by Rayleigh quotient. The quality of the refined Ritz vector depends on the accuracy of the Ritz value μ and each refined Ritz vector must be calculated independently from its own distinct value of μ.

Definition

Let U be an orthonormal basis for subspace $\mathcal{U}$. Then $(\kappa+\delta, Uw)$ is a Harmonic Ritz pair with shift κ if

$$U^H(A-\kappa I)^H(A-\kappa I)Uw = \delta U^H(A-\kappa I)^H Uw. \tag{11}$$

Given shift κ, (11) is a generalized eigenvalue problem with eigenvalue δ.

Theorem

Let (λ, x) be an eigenpair of A with $x = Uw$. Then (λ, Uw) is a harmonic Ritz pair.

Proof: Since (λ, x) is an eigenpair of A with $x = Uw$, we have

$$Ax = \lambda x \quad \Rightarrow \quad AUw = \lambda Uw.$$

It implies that

$$U^H(A - \kappa I)^H(A - \kappa I)Uw = (\lambda - \kappa)U^H(A - \kappa I)^H Uw.$$

Taking eigenvalue $\delta = \lambda - \kappa$, we obtain

$$U^H(A - \kappa I)^H(A - \kappa I)Uw = \delta U^H(A - \kappa I)^H Uw.$$

That is $(\kappa + \delta, Uw) = (\lambda, Uw)$ is a harmonic Ritz pair. ■

Given a shift κ, if we want to compute the eigenvalue λ of A which is closest to κ, then we need to compute the eigenvalue δ of (11) such that $|\delta|$ is the smallest value of all of the absolute values for the eigenvalues of (11).

Expect

If x is approximately represented in $\mathcal{U}$, then the harmonic Rayleigh-Ritz will produce an approximation to x.

Question

How to compute the eigenpair (δ, w) of (11)?

Let

$$(A - \kappa I)U = QR$$

be the QR factorization of $(A - \kappa I)U$. Then (11) can be rewritten as

$$R^H R w = \delta R^H Q^H U w.$$

That is

$$(Q^H U) w = \delta^{-1} R w.$$

This eigenvalue can be solved by the QZ algorithm. The harmonic Ritz vector $\hat{x} = Uw$ and the corresponding harmonic Ritz value is $\mu = \hat{x}^H A \hat{x}$.