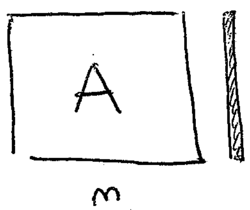
4.3. Krylov Subspaces

Large scale eigenvalue problems $Ax=\lambda x$ are usually solved by constructing a subspace that contains approximations to the desired eigenvectors, and then extracting that information from the subspace.



One particular family of subspaces are Krylov subspaces.

3.1. Krylov sequences and Krylov spaces

Recall the power method: with an initial u, compute $u, Au, A^2u, \ldots A^k u$ tend to the dominant eigenvector. However, in power method we only use the last generated vector $A^k u$, which amounts to throwing away the information contained in the previously generated vectors.

Definition 3.2 Let A be $n\times n$ and $u \neq 0$. Then the sequence $u, Au, A^2u, \ldots$ is a Krylov sequence based on A, u.
The matrix
$$K_k(A,u) = (u \;\; Au \;..\; A^{k-1}u)$$
is the k-th Krylov matrix. The space $\mathcal{K}_k(A,u) = \mathcal{R}(K_k(A,u))$ is called the k-th Krylov subspace.

$$\mathcal{K}_k \equiv \mathcal{K}_k(A,u)$$

Elementary properties

Theorem 3.3 Let A and $u \neq 0$ be given. Then

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

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

2. If $\sigma \neq 0$

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

3. For any $\varkappa$

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

4. If W is nonsingular

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

Proof Easy.

The polynomial connection

typo in the Book γ_{k-1}

Let $v \in \mathcal{K}_k(A,u)$

$$v = \gamma_1 u + \gamma_2 Au + \gamma_3 A^2 u + \dots + \underline{\underline{\gamma_k}} A^{k-1} u$$

$$p(x) = \gamma_1 + \gamma_2 x + \gamma_3 x^2 + \dots + \gamma_k x^{k-1},$$

$$v = p(A)u,$$

$$p(A) = \gamma_1 I + \gamma_2 A + \gamma_3 A^2 + \dots + \gamma_{k-1} A^{k-1}$$

Conversely, if
$v = p(A)$ for any polynomial of degree $\le k-1$
then $v \in \mathcal{K}_k(A, u)$. Thus,

TYPO IN THE BOOK

<u>Theorem 3.4.</u> The space $\mathcal{K}_k(A,u)$ can be represented as

$$\mathcal{K}_k(A,u) = \{ p(A)u \,:\, \deg(p) \le k-1 \}$$

Note that

$$\mathcal{K}_1 \subseteq \mathcal{K}_2 \subseteq \mathcal{K}_3 \subseteq \mathcal{K}_4 \subseteq \dots \subseteq \mathcal{K}_j \subseteq \mathcal{K}_{j+1} \subseteq \dots$$

$\mathcal{K}_1 \to \mathrm{span}(u)$, $\mathcal{K}_2 \to \mathrm{span}(u, Au)$, $\mathcal{K}_3 \to \mathrm{span}(u, Au, A^2u)$

$\Rightarrow$ if u is an eigenvector, $Au = \lambda u \Rightarrow \mathcal{K}_2 = \mathcal{K}_1$, $\mathcal{K}_3 = \mathcal{K}_2$ etc.

We say that a Krylov sequence terminates at l if l is the smallest integer such that

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

<u>Theorem 3.5</u> A Krylov sequence terminates at l if and only if l is the smallest integer such that $\dim \mathcal{K}_{l+1} = \dim \mathcal{K}_l$.

<u>Proof</u> Clear because $\mathcal{K}_l \subseteq \mathcal{K}_{l+1}$.

Note,

if $\mathcal{K}_{l+1} = \mathcal{K}_l$, then

$$A\mathcal{K}_l \subseteq \mathcal{K}_{l+1} = \mathcal{K}_l \Rightarrow A\mathcal{K}_l \subseteq \mathcal{K}_l$$

$\Rightarrow \mathcal{K}_l$ is an invariant subspace of A.

Second claim of Thm 3.5:

If the Krylov sequence terminates at l, then $\mathcal{K}_l$ is an eigenspace of dimension l.

On the other hand, if u lies in an eigenspace of dimension m, then for some $l \leq m$ the sequence terminates at l.

Let $\mathcal{X}$ be an invariant subspace, $\dim \mathcal{X} = m$,

let $u \in \mathcal{X}$.

Then $Au \in A\mathcal{X} \subseteq \mathcal{X} \Rightarrow Au \in \mathcal{X}$

$\ldots A^2u, \ldots \in \mathcal{X}$

If the sequence terminates at $l > m$,

then $l = \dim \mathcal{K}_l > m = \dim \mathcal{K}_m = \dim \mathcal{X}$

which is a contradiction with

$$\mathcal{K}_l \subset \mathcal{X}. \quad \boxtimes$$

Typo IN THE BOOK

Note,
if $\mathcal{K}_{l+1} = \mathcal{K}_l$, then

$$A\mathcal{K}_l \subseteq \mathcal{K}_{l+1} = \mathcal{K}_l \Rightarrow A\mathcal{K}_l \subseteq \mathcal{K}_l$$

$\Rightarrow \mathcal{K}_l$ is an invariant subspace of A.

Second claim of Thm 3.5:
If the Krylov sequence terminates at l,
then $\mathcal{K}_l$ is an eigenspace of dimension l.
On the other hand, if u lies in an eigenspace of dimension m, then for some $l \leq m$ the sequence terminates at l.

Let $\mathcal{X}$ be an invariant subspace, $\dim \mathcal{X} = m$,
let $u \in \mathcal{X}$.
Then $Au \in A\mathcal{X} \subseteq \mathcal{X} \Rightarrow Au \in \mathcal{X}$

$\ldots A^2u, \ldots \in \mathcal{X}$

If the sequence terminates at $l > m$,
then $l = \dim \mathcal{K}_l > m = \dim \mathcal{K}_m = \dim \mathcal{X}$
which is a contradiction with
$\mathcal{K}_l \subset \mathcal{X}$. $\boxtimes$

Typo IN
THE BOOK

Conversely, if
$v = p(A)$ for any polynomial of degree $\leq k-1$
then $v \in \mathcal{K}_k(A, u)$. Thus,

TYPO IN THE BOOK

Theorem 3.4. The space $\mathcal{K}_k(A,u)$ can be represented as

$$\mathcal{K}_k(A,u) = \{ p(A)u : \deg(p) \leq k-1 \}$$

Note that

$$\mathcal{K}_1 \subseteq \mathcal{K}_2 \subseteq \mathcal{K}_3 \subseteq \mathcal{K}_4 \subseteq \dots \subseteq \mathcal{K}_j \subseteq \mathcal{K}_{j+1} \subseteq \dots$$

$\mathcal{K}_1 \to$ span(u), $\mathcal{K}_2 \to$ span(u, Au), $\mathcal{K}_3 \to$ span(u, Au, A^2u)

$\hookrightarrow$ if u is an eigenvector,
$Au = \lambda u \Rightarrow \mathcal{K}_2 = \mathcal{K}_1$, $\mathcal{K}_3 = \mathcal{K}_2$ etc.

We say that a Krylov sequence terminates at l if l is the smallest integer such that

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

Theorem 3.5 A Krylov sequence terminates at l if and only if l is the smallest integer such that $\dim \mathcal{K}_{l+1} = \dim \mathcal{K}_l$.

Proof Clear because $\mathcal{K}_l \subseteq \mathcal{K}_{l+1}$.

Conversely, if
$v = p(A)$ for any polynomial of degree $\leq k-1$
then $v \in \mathcal{K}_k(A, u)$. Thus,

TYPO IN THE BOOK

<u>Theorem 3.4.</u> The space $\mathcal{K}_k(A,u)$ can be represented as

$$\mathcal{K}_k(A,u) = \left\{ p(A)u : \deg(p) \leq k-1 \right\}$$

Note that

$$\mathcal{K}_1 \subseteq \mathcal{K}_2 \subseteq \mathcal{K}_3 \subseteq \mathcal{K}_4 \subseteq \dots \subseteq \mathcal{K}_j \subseteq \mathcal{K}_{j+1} \subseteq \dots$$

$\mathcal{K}_1 \to \mathrm{span}(u)$, $\mathcal{K}_2 \to \mathrm{span}(u, Au)$, $\mathcal{K}_3 \to \mathrm{span}(u, Au, A^2u)$

$\Rightarrow$ if u is an eigenvector,
$Au = \lambda u \Rightarrow \mathcal{K}_2 = \mathcal{K}_1$, $\mathcal{K}_3 = \mathcal{K}_2$ etc.

We say that a Krylov sequence terminates at l if l is the smallest integer such that

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

<u>Theorem 3.5</u> A Krylov sequence terminates at l if and only if l is the smallest integer such that $\dim \mathcal{K}_{l+1} = \dim \mathcal{K}_l$.

<u>Proof</u> Clear because $\mathcal{K}_l \subseteq \mathcal{K}_{l+1}$.

~~Found~~
Termination is
- good because we have found an invariant subspace that contains exact information of some eigenpairs
- bad because it stops furnishing info ~~on~~ about other eigenvectors.

3.2. Convergence

We want to discover what mechanism drives Krylov subspace to converge.

For simplicity, we assume that A is Hermitian, with eigenpairs (λ_i, x_i) $i=1\ldots n$, $x_i^* x_j = \delta_{ij}$

Let u be the starting vector for a Krylov sequence.

$u = \alpha_1 x_1 + \alpha_2 x_2 + \dots + \alpha_n x_n$, $\alpha_i = x_i^* u_i$

Any vector in $\mathcal{K}_k$ can be written as $p(A)u$, with polynomial p of degree $\leq k-1$

$$p(A)u = p(A) \sum_{i=1}^{n} \alpha_i x_i = \sum_{i=1}^{n} \alpha_i p(A) x_i$$

$$= \sum_{i=1}^{n} \alpha_i p(\lambda_i) x_i = \alpha_1 p(\lambda_1) x_1 + \alpha_2 p(\lambda_2) x_2 + \dots + \alpha_n p(\lambda_n) x_n$$

$$\underbrace{p(A)u}_{\text{a vector from } \mathcal{K}_k} = \alpha_1 p(\lambda_1) x_1 + \alpha_2 p(\lambda_2) x_2 + \dots + \alpha_n p(\lambda_n) x_n$$

vary over all p, $\deg(p) \le k-1$

NOTE If we can find a polyn. p such that

$$|p(\lambda_i)| \gg \max_{j \ne i} |p(\lambda_j)|$$

then

$$p(A)u = \cancel{\alpha_i p(\lambda_i) x_i + \sum}$$

$$= p(\lambda_i)\left[\alpha_i x_i + \sum_{j \ne i} \alpha_j \frac{p(\lambda_j)}{p(\lambda_i)} x_j\right]$$

will be small in abs value

and $p(A)u$ will be a good approx to x_i

<u>Theorem 3.6</u> If $\alpha_i = x_i^* u \ne 0$ and $p(\lambda_i) \ne 0$

$$\tan \measuredangle (p(A)u, x_i) \le \max_{j \ne i} \left| \frac{p(\lambda_j)}{p(\lambda_i)} \right| \tan \measuredangle (u, x_i)$$

<u>Proof</u>

$$u = \alpha_1 x_1 + \alpha_2 x_2 + \dots + \alpha_i x_i + \dots + \alpha_n x_n$$

$$\cos \measuredangle (u, x_i) = \frac{|x_i^* u|}{\|x_i\|_2 \|u\|_2} = \frac{|\alpha_i|}{\|u\|_2}$$

$$\|u\|_2 = \sqrt{\sum_{i=1}^{n} |\alpha_i|^2}$$

$$\sin^2 \measuredangle(u, x_i) = 1 - \cos^2 \measuredangle(u, x_i) =$$

$$= 1 - \frac{|\alpha_i|^2}{\|u\|_2^2} = \frac{\|u\|_2^2 - |\alpha_i|^2}{\|u\|_2^2} =$$

$$= \frac{\sum_{j \neq i} |\alpha_j|^2}{\|u\|_2^2}$$

$$\Rightarrow \tan^2 \measuredangle(u, x_i) = \sum_{j \neq i} \frac{|\alpha_j|^2}{|\alpha_i|^2}$$

TYPO IN THE BOOK

In the same way

$$\tan^2 \measuredangle(p(A)u, x_i) = \sum_{j \neq i} \frac{|\alpha_j p(\lambda_j)|^2}{|\alpha_i p(\lambda_i)|^2}$$

$$\leq \max_{j \neq i} \frac{|p(\lambda_j)|^2}{|p(\lambda_i)|^2} \sum_{j \neq i} \frac{|\alpha_j|^2}{|\alpha_i|^2}$$

$$= \max_{j \neq i} \frac{|p(\lambda_j)|^2}{|p(\lambda_i)|^2} \cdot \tan^2 \measuredangle(u, x_i)$$

$\boxtimes$

So, ~~we need a po~~ if there is a polynomial p that is "large" at λ_i and "small" at λ_j, $j \neq i$, then the corresp. vector $p(A)u$ will go to the direction of $\underline{x_i}$.

$$\tan \measuredangle (p(A)r, x_i) \le \max_{j \ne i} \left| \frac{p(\lambda_j)}{p(\lambda_i)} \right| \tan \measuredangle (r, x_i)$$

independent of scaling of p
so we can consider p normalized
so that $p(\lambda_i) = 1$

$$\tan \measuredangle (p(A)r, x_i) \le \max_{j \ne i} |p(\lambda_j)| \tan \measuredangle (r, x_i)$$

$p(\lambda_i) = 1$
$\deg(p) \le k-1$

$$p(A)r \in \mathcal{K}_k$$

$$\Rightarrow \tan \measuredangle (x_i, \mathcal{K}_k) \le \boxed{\min_{\substack{\deg(p) \le k-1 \\ p(\lambda_i) = 1}} \max_{j \ne i} |p(\lambda_j)|} \tan \measuredangle (r, x_i)$$

this factor determines how well $\mathcal{K}_k$ will perform in approximating x_i.

It is enough if we find particular polynomials that will produce small factors.

We study

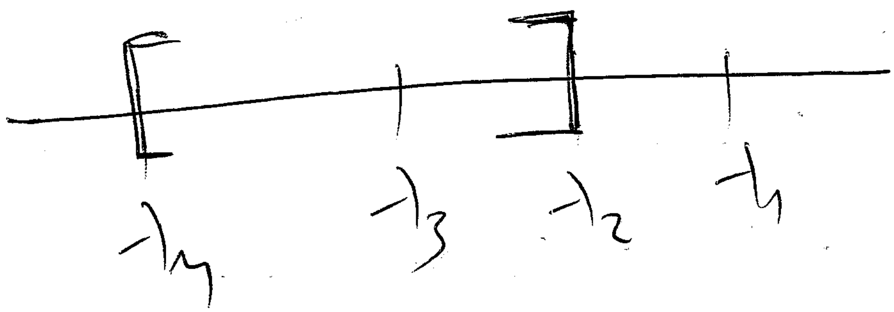
$$\min_{\substack{\deg(p)\le k-1\\ p(\lambda_i)=1}} \max_{j\ne i} |p(\lambda_j)|.$$

Let $i=1$, i.e. we want x_1.

Since A is assumed Hermitian we can order the eigenvalues

$$\lambda_1 > \lambda_2 \ge \lambda_3 \ge \dots \ge \lambda_n$$

assume this for the moment



Note

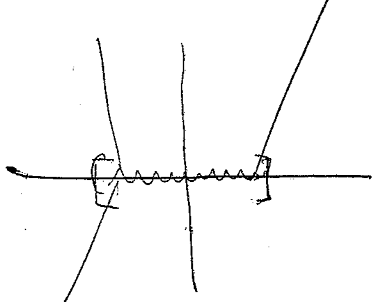
$$\min_{\substack{\deg(p)\le k-1\\ p(\lambda_1)=1}} \max_{j\ne 1} |p(\lambda_j)| \le \min_{\substack{\deg(p)\le k-1\\ p(\lambda_1)=1}} \max_{\lambda\in[\lambda_n,\lambda_2]} |p(\lambda)|$$

larger domain, includes $\lambda_2, \dots \lambda_n$

So we have

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

elegant solution using Chebyshev polynomials



Chebyshev polynomials

$$C_k(t) = \begin{cases} \cos(k \cos^{-1} t), & |t| \le 1 \\ \cosh(k \cosh^{-1} t) & |t| \ge 1 \end{cases}$$

Theorem

TYPO IN THE BOOK

1. $C_0(t) = 1$, $C_1(t) = t$

   $C_{k+1}(t) = 2t C_k(t) - C_{k-1}(t)$, $k = 1, 2, \dots$

   $\left(\cos(k+1)\varphi + \cos(k-1)\varphi = 2 \cos k\varphi \cos \varphi \right)$

2. For $|t| > 1$

   $$C_k(t) = \left(1 + \sqrt{t^2 - 1}\right)^k + \left(1 + \sqrt{t^2 - 1}\right)^{-k}$$

3. For $|t| \le 1$, $|C_k(t)| \le 1$

   $$t_{ik} = \cos \frac{(k-i)\pi}{k} \qquad i = 0 \dots k$$

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

4. For $s > 1$

   $$\min_{\substack{\deg(p) \le k \\ p(s) = 1}} \max_{t \in [-1,1]} |p(t)| = \frac{1}{C_k(s)} \quad \text{at} \quad p(t) = \frac{C_k(t)}{C_k(s)}$$

   TYPO IN THE BOOK

Now, to estimate $\tan \measuredangle (x_1, \mathcal{K}_k)$

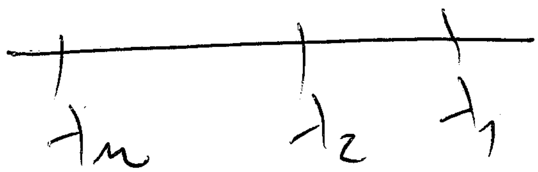
we need

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

First, we transform $[\lambda_n, \lambda_2]$ to $[-1,1]$:

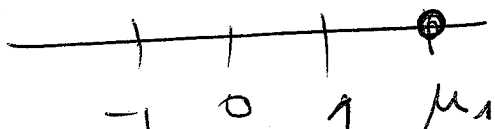
Change of variable

$$\lambda = \lambda_2 + (\mu - 1)\frac{\lambda_2 - \lambda_n}{2}$$



$$\mu_1 = \mu(\lambda_1) = 1 + 2\,\frac{\lambda_1 - \lambda_2}{\lambda_2 - \lambda_n}$$

$$\Rightarrow \quad \sigma_k = \frac{1}{C_{k-1}(\mu_1)}$$



<u>Theorem</u> Let $A = A^*$, (λ_i, x_i) $i = 1 \dots n$, $x_i^* x_j = \delta_{ij}$

$\lambda_1 > \lambda_2 \ge \dots \ge \lambda_n$, $\quad \eta = \dfrac{\lambda_1 - \lambda_2}{\lambda_2 - \lambda_n}$

Then

$$\tan \measuredangle (x_1, \mathcal{K}_k(A, u)) \le \frac{\tan \measuredangle (x_1, u)}{C_{k-1}(1 + 2\eta)}$$

$$= \frac{\tan \measuredangle (x_1, u)}{\left(1 + \sqrt{(1+2\eta)^2 - 1}\right)^{k-1} + \left(1 + \sqrt{(1+2\eta)^2 - 1}\right)^{-(k-1)}}$$

$$= \frac{\tan \measuredangle (x_1, u)}{\left(1 + 2\sqrt{\eta + \eta^2}\right)^{k-1} + \left(1 + 2\sqrt{\eta + \eta^2}\right)^{-(k-1)}}$$

$$\approx \frac{\tan \measuredangle (x_1, u)}{\left(1 + 2\sqrt{\eta + \eta^2}\right)^{k-1}}$$

$\boxtimes$

- It is possible to give bounds for the remaining eigenvectors $x_2, \dots, x_n$

Let now $\lambda_1=\lambda_2>\lambda_3\ge\dots\ge\lambda_n$

$u=\alpha_1x_1+\alpha_2x_2+\dots+\alpha_nx_n$

$A^ku=(\alpha_1x_1+\alpha_2x_2)\lambda_1^k+\alpha_3x_3\lambda_3^k+\dots+\alpha_nx_n\lambda_n^k$

$\mathcal{K}_k$ will approximate $\alpha_1x_1+\alpha_2x_2$ and "see" $(\lambda_1,\alpha_1x_1+\alpha_2x_2)$ as a simple eigenpair.

We can try with another vector

$v=\beta_1x_1+\beta_2x_2+\beta_3x_3+\dots+\beta_nx_n$

$A^kv=(\beta_1x_1+\beta_2x_2)\lambda_1^k+\beta_3x_3\lambda_3^k+\dots+\beta_nx_n^k$

unless $\alpha_1x_1+\alpha_2x_2$ and $\beta_1x_1+\beta_2x_2$ are colinear, we will have two-dim. subspace belong. to $\lambda_1=\lambda_2$.

So we can use

$U,\ AU,\ A^2U,\ \dots,$

$\mathcal{K}_k(A,U)$ block-Krylov subspace

Thm 3.11. $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n$

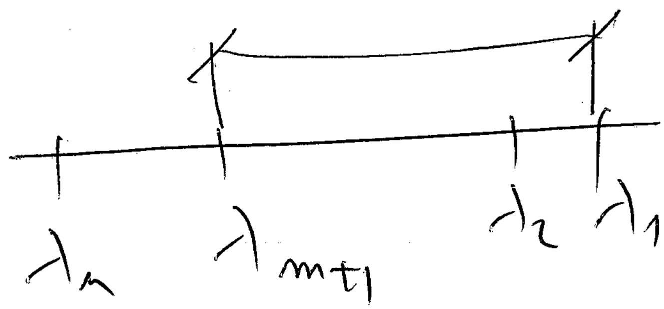
multipl. of $\lambda_1 \leq m$

$$B = \begin{pmatrix} x_1^* \\ \vdots \\ x_m^* \end{pmatrix} U$$

$$v = U B^{-1} e_1 \quad \Rightarrow$$

$$\tan \measuredangle (x_1, \mathcal{K}_k(A, U)) \leq \frac{\tan (x_1, v)}{c_{k-1}(1+2\eta)} .$$

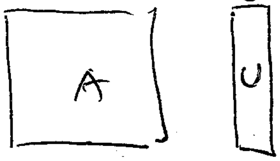
$$\eta = \frac{\lambda_1 - \lambda_{m+1}}{\lambda_{m+1} - \lambda_n}$$



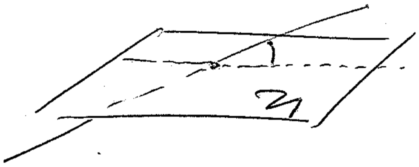
4.4. Rayleigh Ritz

→ suppose now that we have ~~an~~ a subspace containing accurate approximations to an eigenspace of A

4.1. Rayleigh-Ritz methods



$\mathcal{U} = \mathcal{R}(U)$ contains an approximate subspace of A

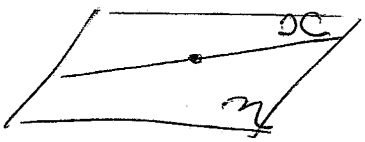


$\mathcal{U}$ has been comp. by some method, e.g. as a Krylov subspace

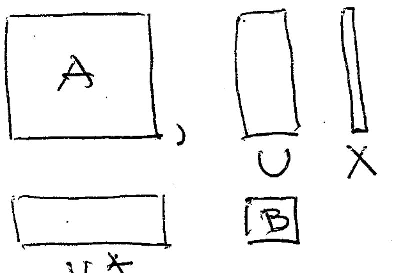
<u>Theorem 4.1</u> $\mathcal{U} = \mathcal{R}(U)$. Let V^* be a left inverse of U. Set $B = V^*AU$.
If $\mathcal{X} \subseteq \mathcal{U}$ is an eigenspace of A, then there is an eigenpair (L, W) of B such that (L, UW) is an eigenpair of A, with $\mathcal{R}(UW) = \mathcal{X}$.

<u>Proof</u>

$BW = WL \iff (L, W)$ eigenpair of B



Let $AX = XL$
$\mathcal{X} = \mathcal{R}(X)$



$X = UW$

$$\to A\,UW = UWL \quad \Big|\; V^*$$

$$\underbrace{V^*AU}_{B}W = \underbrace{V^*U}_{I}WL = WL$$

$(V^*(AUW - UWL) = 0)$

$$\boxed{BW = WL} \checkmark$$

Hence, to identify the exact ~~sta~~ eigenspace by examining the eigenpairs of B.

Expect

Hope: When $\mathcal{U}$ contains only an approximate eigenspace $\tilde{\mathcal{X}}$ of A, there would be an eigenpair $(\tilde{L}, \tilde{W})$ of B such that $(\tilde{L}, U\hat{W})$ is an approx eigenpair of A.

Rayleigh-Ritz procedure

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

(M, UW) → Ritz pair

UW Ritz basis

(λ, Uw)

primitive Ritz vector

Ritz vector

Ritz value

<u>Note</u> For a matrix V such that V^*U is nonsingular, $\hat{V} = V(V^*U)^{-*}$ satisfies

$$\hat{V}^* U = (V^*U)^{-1} V^* U = I$$

$$(V^*U)^{-1} V^* A U w = \mu w \iff$$

$$(V^* A U) w = \mu (V^*U) w$$

generalized eigen-problem

Example $A = \begin{bmatrix} 0 & & \\ & 1 & \\ & & -1 \end{bmatrix}$ target: $(0, e_1)$

$U = \begin{bmatrix} 1 & 0 \\ 0 & 1/\sqrt{2} \\ 0 & 1/\sqrt{2} \end{bmatrix}$, $U^*U = I$

$B = U^*AU = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$ no clue

$Bw = 0 \cdot w$ for any $w \neq 0$

U contains e_1 exactly

The orthogonal RR procedure

U orthonormal, $V = U$, take W also orthonormal so UW is orthonormal.

e.g. for $A = A^*$, $B = U^*AU = B^*$

Theorem 4.3. Let $(M, \hat{X})$ be an orthogonal RR pair with respect to U. Then

$R = A\hat{X} - \hat{X}M$ is minimal in any unit inv. norm. (over all $A\hat{X} - \hat{X}N$)

(see Thm 2.6)

4.2. Convergence

goal: an eigenpair (λ, x) of A.

Suppose we obtain orthonormal basis U_ϑ for which $\measuredangle(x, U_\vartheta) \equiv \vartheta$ is small.

We apply RR and obtain a Ritz pair $(\mu_\vartheta, U_\vartheta p_\vartheta)$

If satisfied, STOP. Else, get another U_ϑ with smaller ϑ. As $\vartheta \to 0$, we need to show convergence.

$B_\vartheta = U_\vartheta^* A U_\vartheta$, we will show <u>first</u> that B_ϑ has Ritz value μ_ϑ that converges to λ.

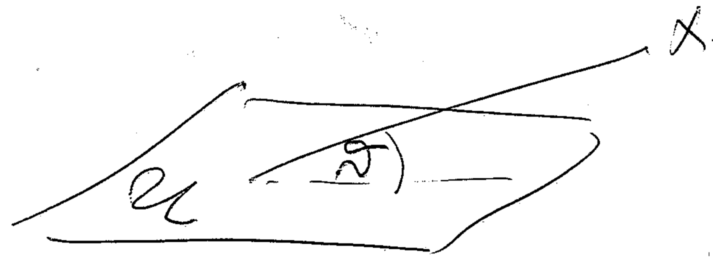
<u>Convergence of the Ritz value</u>

<u>Theorem 4.4</u> Let B_ϑ ~~be the~~ $= U_\vartheta^* A U_\vartheta$.

Then there is a matrix E_ϑ satisfying

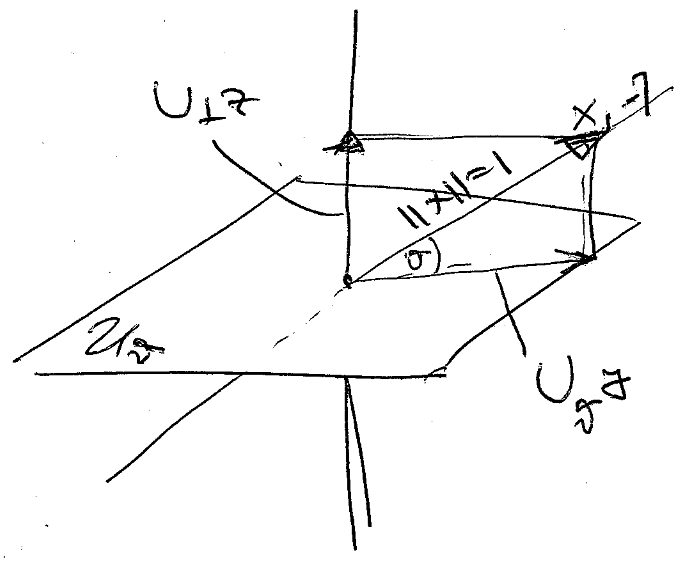
$$\|E_\vartheta\|_2 \le \frac{\sin\vartheta}{\sqrt{1-\sin^2\vartheta}} \|A\|_2$$

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



Proof Let $(U_\vartheta, U_\perp)$ be unitary

Set $y = U_\vartheta^* x, \quad z = U_\perp^* x$



$$x = U_\vartheta U_\vartheta^* x + U_\perp U_\perp^* x$$

$$x = U_\vartheta y + U_\perp z$$

$$\|z\|_2 = \sin\vartheta$$

$$\|y\|_2 = \cos\vartheta$$

$$Ax - \lambda x = 0$$

$$U_\vartheta^* A (U_\vartheta \; U_\perp) \begin{pmatrix} U_\vartheta^* \\ U_\perp^* \end{pmatrix} x - \lambda \underbrace{U_\vartheta^* x}_{y} = 0$$

$$\underbrace{U_\vartheta^* A U_\vartheta}_{B_\vartheta} \underbrace{U_\vartheta^* x}_{y} + U_\vartheta^* A U_\perp \underbrace{U_\perp^* x}_{z} - \lambda y = 0$$

$$B_\vartheta y + U_\vartheta^* A U_\perp z - \lambda y = 0$$

$$\hat{y} = \frac{y}{\|y\|_2} \; ; \quad r = B_\vartheta \hat{y} - \lambda \hat{y}$$

$$\|r\|_2 \le \frac{\|U_\vartheta^* A U_\perp z\|_2}{\|y\|_2} \le$$

$$\le \|A\|_2 \cdot \frac{\sin\vartheta}{\sqrt{1-\sin^2\vartheta}}$$

Define $E_\vartheta = -r\hat{y}^*$

$\|E_\vartheta\|_2 = \|r\|_2$

$(B_\vartheta + E_\vartheta)\hat{y} =$

$= B_\vartheta \hat{y} - r\hat{y}^*\hat{y} = \lambda\hat{y}$

$\boxtimes$

By perturbation theory (Chapter 1, Th 3.1)

$$|\mu_\vartheta - \lambda| \le 4\left(2\|A\|_2 + \|E_\vartheta\|_2\right)^{1-\frac{1}{m}} \|E_\vartheta\|_2^{1/m} \longrightarrow 0 \quad (\text{as } \vartheta \to 0)$$

μ_ϑ: e.g. B_ϑ $\quad$ λ: e.g. $B_\vartheta + E_\vartheta$

If A is Hermitian

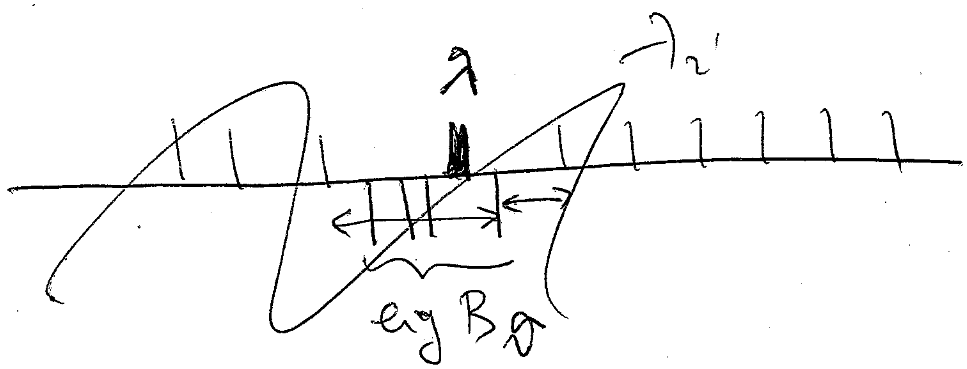
$$|\mu_\vartheta - \lambda| \le \|E_\vartheta\|_2$$

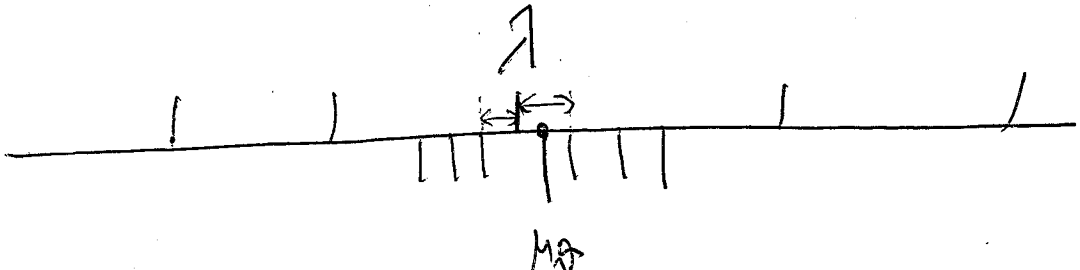
Convergence of Ritz vectors

Theorem 4.6 Let $(\mu_\vartheta, w_\vartheta)$ be a primitive Ritz pair ($B w_\vartheta = \mu_\vartheta w_\vartheta$) and let $(w_\vartheta\ W_\vartheta)$ be unitary so that

$$\begin{pmatrix} w_\vartheta^* \\ W_\vartheta^* \end{pmatrix} B_\vartheta (w_\vartheta\ W_\vartheta) = \left[\begin{array}{c|c} \mu_\vartheta & h_\vartheta^* \\ \hline 0 & N_\vartheta \end{array}\right]. \text{ Then}$$

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





$\| h_\vartheta \|_2 \le \|A\|_2$ ✓

Let $\sigma(N_\vartheta)$: $\nu_1 \dots \nu_\ell$

$$|\lambda - \nu_i| \ge \underbrace{|\mu_\vartheta - \nu_i|}_{\ge \operatorname{sep}(\mu_\vartheta, N_\vartheta)} - \underbrace{|\mu_\vartheta - \lambda|}_{\to 0}$$

<u>Corollary</u> 4.7. Let $(\mu_\vartheta, U_\vartheta w_\vartheta)$ be a Ritz pair for which $\mu_\vartheta \to \lambda$. If there is $\alpha > 0$ ∩ $\operatorname{sep}(\mu_\vartheta, N_\vartheta) \ge \alpha$ then asymptotically

$$\sin \measuredangle (x, U_\vartheta w_\vartheta) \lessapprox \sin\vartheta \sqrt{1 + \frac{\|A\|^2}{\alpha^2}}$$

NOW RECALL THE EXAMPLE WITH $B = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$

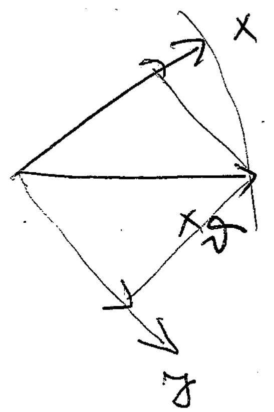
Convergence of Ritz values revisited

$(\mu_\vartheta, x_\vartheta) \sim (\lambda, x)$

$x_\vartheta = U_\vartheta w_\vartheta \quad , \quad B_\vartheta w_\vartheta = \mu_\vartheta w_\vartheta$

$B_\vartheta = U_\vartheta^* A U_\vartheta$

$\mu_\vartheta = x_\vartheta^* A x_\vartheta$



$y^* x = 0$

$y^* \underbrace{Ax}_{\lambda x} = 0$

$x_\vartheta = \gamma x + \sigma y \quad , \quad y \perp x$

$|\gamma| = \cos \angle (x_\vartheta, x)$

$|\sigma| = \sin \angle (x_\vartheta, x)$

$|\sigma| \sim O(\vartheta)$ by Corollary 4.7

$$\mu_\vartheta = (\gamma x + \sigma y)^* A (\gamma x + \sigma y)$$
$$= |\gamma|^2 \lambda + \bar{\gamma} \sigma x^* A y + |\sigma|^2 y^* A y$$

$$|\mu_\vartheta - \lambda| = \left| (|\gamma|^2 - 1)\lambda + \bar{\gamma} \sigma x^* A y + |\sigma|^2 y^* A y \right|$$

$$\leq |\sigma|^2 \underbrace{|\lambda|}_{\leq \|A\|} + \underbrace{|\gamma|}_{\leq 1} \underbrace{|\sigma|}_{\leq 1} \underbrace{|x^* A y|}_{\leq \|A\|} + |\sigma|^2 \underbrace{|y^* A y|}_{\leq \|A\|}$$

$$\leq \|A\| \cdot |\sigma| \left(2|\sigma| + 1 \right) = O(\vartheta)$$

4.4 Rayleigh-Ritz part II

Theo 4.3

Orthogonal RR. is optimal → usual approach

→ $A\hat{X} - \hat{X}M = R$ min. residual in any unit. invar. norm.

Convergence

A → eigenpair λ, x

As $\vartheta = \measuredangle(\mathcal{U}, x) \to 0$ how do approx. to λ and x behave.

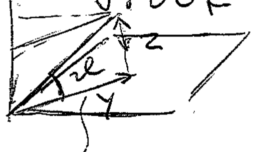
Rayleigh quotient $B_\vartheta = \mathcal{U}_\vartheta^H A \mathcal{U}_\vartheta$

Theo 4.4. B_ϑ as above. Then E_ϑ exists s.t.

$$\|E_\vartheta\|_2 \leq \frac{\sin\vartheta}{\sqrt{1-\sin^2\vartheta}} \|A\|_2 \quad \text{and}$$

$$\lambda \in \Lambda(B_\vartheta + E_\vartheta)$$

Proof



$\|x\| = 1 \to \|y\| = \cos\vartheta$

$\|z\| = \sin\vartheta$

Expand $x = \mathcal{U}_\theta y + \mathcal{U}_\perp z$ $\quad (\mathcal{U}_\theta \, \mathcal{U}_\perp)$ unit.

$$A\underbrace{(\mathcal{U}_\theta \, \mathcal{U}_\perp)\begin{pmatrix} \mathcal{U}_\theta^H \\ \mathcal{U}_\perp^H \end{pmatrix}}_{I} x - \lambda x = 0$$

$$\mathcal{U}_\theta^H A (\mathcal{U}_\theta \, \mathcal{U}_\perp)\begin{pmatrix} y \\ z \end{pmatrix} - \lambda \mathcal{U}_\theta^H x =$$

$$B_\theta y + \mathcal{U}_\theta^H A \mathcal{U}_\perp z - \lambda y = 0 \implies$$

$$\hat{y} = \frac{y}{\|y\|_2}$$

$$B_\theta \hat{y} - \lambda \hat{y} = r = \mathcal{U}_\theta^H A \mathcal{U}_\perp \frac{z}{\sqrt{1-\sin^2\vartheta}} \implies$$

$$\|r\|_2 \leq \frac{\sin\vartheta}{\sqrt{1-\sin^2\vartheta}} \|A\|_2$$

Take $E_\theta = r\hat{y}$ (see earlier notes)

By Theo 3.1

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

(m is $\dim(\mathcal{U})$)

In practice, we're not interested ~~in~~ in large m. So, we need iteration that expands and prunes U_θ while keeping dim modest

It can be shown that Ritz vector converges ~~if~~ (for $\theta \to 0$) if λ bounded away from $\Lambda(B_\theta) \setminus \{\mu_\theta\}$.

Theo 4.6 and papers by Stewart [268] and Jia & Stewart [133]

Since this separation is not guaranteed, ~~it may have to be~~ Ritz vectors may not converge to eigenvectors unless additional work is done. This led to work by Jia and Jia & Stewart on refined Ritz vectors. (and work E&S)

(Pirsl) → ~~unif~~ uniform separation condition.

later

Corr. 4.7. → If Ritz vec conv → better bnd
$|\mu_\theta - \lambda| = O(\theta)$
↳ result from earlier analysis

If eig. vec. $O(\epsilon) \to RQ = O(\theta^2)$ for Herm. matrix.

~~If the eigenvalue problem for the Ritz block B_θ~~ → prim. Ritz vector

If the eigenvector corr. to μ_θ for B_θ remains "well-conditioned" then the Ritz vector $U w_\theta$ converges to eigenvector x.

Corr. 4.7 (based on theo 4.6)

Let $(\mu_\Theta, U_\Theta w_\Theta)$ be a Ritz pair for which $\mu_\Theta \to \lambda$. If there is a constant $\alpha > 0$ s.t. $\text{sep}(\mu_\Theta, N_\Theta) \geq \alpha > 0$ then (asymptotically)

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

swap order

Let $B_\Theta = U_\Theta^H A U_\Theta$ and let (μ_Θ, w_Θ) be a primitive Ritz pair. Furthermore, let $(w_\Theta \; W_\Theta)$ be unitary and $(w_\Theta \; W_\Theta)^H B_\Theta (w_\Theta \; W_\Theta) = \begin{pmatrix} \mu_\Theta & h_\Theta^H \\ 0 & N_\Theta \end{pmatrix}$

Clearly, if $\text{sep}(\mu_\Theta, N_\Theta)$ is very small, there is a small perturbation of B_Θ that has double eigenvalue μ_Θ and w_Θ cannot be expected to be computed (or to have been computed) accurately.

If $\mu_\Theta \to \lambda$ and $\lambda, \hat{\lambda} \in \Lambda(A)$ where $|\lambda - \hat{\lambda}|$ very small (compared to $\|A\|$) then x is ill-conditioned and there is little we can do. However, it is quite possible that x is well-conditioned but μ_θ is not. The latter is just an "artifact" of the projection $U_\theta^H A U_\theta$. In this case we must do extra work to get accurate eigenvector.

If (or once) we have a converging eig. vector, the Rayleigh quotient converges at least as fast as the eig. vector. Let x_θ be approx. to x and $\mu_\theta = x_\theta^H A x_\theta$. Let $x_\theta = \gamma x + \sigma y$ where $\|y\|_2 = 1$ and $y \perp x$ (also let $\|x\|_2 = \|x_\theta\|_2 = 1$). ~~Thus~~ Then

$|\gamma| = \cos\angle(x_\theta, x)$, $|\sigma| = \sin\angle(x_\theta, x)$

$$x_\theta^H A x_\theta = |\gamma|^2 x^H A x + \bar{\gamma}\sigma x^H A y + \bar{\sigma}\gamma y^H A x + |\sigma|^2 y^H A y$$

$$= \lambda|\gamma|^2 + \bar{\gamma}\sigma x^H A y + |\sigma|^2 y^H A y = \mu_\theta$$

$$|\mu_\theta - \lambda| = \left|\lambda(|\gamma|^2 - 1) + \cdots\right| \le$$

$$|\sigma|^2|\lambda| + |\gamma||\sigma|\,|x^H A y| + |\sigma^2|\,|y^H A y| \le$$

$$|\sigma|\,\|A\|_2\left(|\sigma| + |\gamma| + |\sigma|\right) \le (1 + 2|\sigma|)\,|\sigma|\,\|A\|_2$$

Since $|\sigma| = \sin \Delta(x_\theta, x) \lesssim \sin\theta \cdot K$

we have $|\mu_\theta - \lambda| = O(\theta)$

If A is Hermitian ~~so~~ the term

$\bar{\gamma}\sigma x^H A y = 0$ (drops out) and

we get $|\mu_\theta - \lambda| \leq 2|\sigma|^2 \|A\|_2 \rightarrow$

$|\mu_\theta - \lambda| = O(\theta^2)$. See also notes on "Overview of Methods" discussing the Rayleigh quotient.

Unfortunately, we have no way to know $\Delta(x, U_\theta)$ or $\Delta(x_\theta, x)$ or θ.

What can we learn from residual?

Theorem 4.8 and $\|x\|_2 = 1$

Let $A = \lambda x y^H + X L Y^H$, where $\|\tilde{x}\|_2 = 1$ and Y orthonorm. Let $(\mu, \tilde{x})$ be an approx. to (λ, x) and $\rho = \|A\tilde{x} - \mu\tilde{x}\|_2$.

Then

$$\sin\Delta(\tilde{x}, x) \leq \frac{\rho}{sep(\mu, L)} \leq \frac{\rho}{sep(\lambda, L) - |\mu - \lambda|}$$

Proof.

$\|Y^H\tilde{x}\|_2 = \sin\Delta(\tilde{x}, x)$. $r = A\tilde{x} - \mu\tilde{x} \Rightarrow$

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

$\|(L - \mu I)^{-1} Y^H r\|_2 = \sin\Delta(\tilde{x}, x) \Rightarrow$

$\sin\Delta(\tilde{x}, x) \leq sep(\mu, L)\|r\|_2$

last step follows from continuity of sep

(For simple λ)

So, if $\mu \to \lambda$ and $\|r\| \to 0$ the eig. pair converges.

If $\mu \to \lambda$ but μ nearly multiple eig. val., although λ simple, eig. vector may not converge.

Near "multiple"/"copy" of μ is called spurious eig. val. as it does not corr. to eigenval. of A (it's artefact of projection).

Alternative for Ritz vector $\to$ Refined Ritz vector:

Def. 4.9

$$\hat{x} = \arg\min \| A\tilde{x} - \mu_\theta \tilde{x} \|_2 \quad \text{s.t.}$$

$$\tilde{x} \in \mathcal{U}_\theta, \quad \|\tilde{x}\|_2 = 1$$

Theorem 4.10

Let A have spectral representation

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

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

$$\sin \angle (x, x_\theta) \leq \frac{\|A - \mu_\theta I\|_2 \sin\theta + |\lambda - \mu_\theta|}{\sqrt{1-\sin^2\theta}\,(sep(\lambda, L) - |\lambda - \mu_\theta|)}$$

<u>Proof</u>

Let U be orthonormal basis for $\mathcal{U}$, and $x = w + z$, where $w = UU^H x$.
Then $\|w\|_2 = \sqrt{1-\sin^2\theta}$ $(= \cos\theta)$

$$\|z\|_2 = \sin\theta.$$

Let $\hat{w} = w / \|w\|_2$. Then

$$(A - \mu_\theta I)\hat{w} = \frac{(A-\mu_\theta I) w}{\sqrt{1-\sin^2\theta}} = \frac{(A-\mu_\theta I)(x - z)}{\sqrt{1-\sin^2\theta}} = \frac{(\lambda - \mu_\theta) x - (A - \mu_\theta I) z}{(1-\sin^2\theta)^{1/2}}$$

by def. of Ref. Ritz vector:

$$\|(A-\mu_\theta I)\hat{x}_\theta\|_2 \leq \|(A-\mu_\theta I)\hat{w}\|_2 \leq \frac{|\lambda - \mu_\theta| + \|(A - \mu_\theta I)\| \sin\theta}{(1-\sin^2\theta)^{1/2}}.$$

~~Using~~ Using (theo 4.8)

$$\sin \angle (x, \hat{x}_\theta) \leq \frac{\|(A - \mu_\theta I)\hat{x}_\theta\|_2}{sep(\lambda, L) - |\mu - \lambda|}$$

completes the proof.

For $\theta \to 0$, $\mu_\theta \to \lambda$ (and $\sin\theta \to 0$),

hence $\Delta(\hat{x}, x) \to 0$. So, refined Ritz vectors converge unconditionally.

Note that once the refined Ritz vector is computed, it is generally best to replace the Ritz value μ_θ by the Rayleigh quotient

$$\hat{\mu}_\theta = \hat{x}_\theta^H A \hat{x}_\theta$$

residual $\|A\hat{x}_\theta - \hat{\mu}_\theta \hat{x}_\theta\|_2$ optimal.

Let U be orthonormal basis for $\mathcal{U}$. Then refined Ritz vector $\hat{x} = Uz$ for some z, $\|z\|_2 = 1$. We want

$$\|(A - \mu I) U z\|_2 \text{ minimal}$$

So, we must compute the right singular vector corresponding to the smallest singular value of

$$AU - \mu U \to z$$

Then

$$\hat{x} = Uz.$$

It turns out that for some algorithms this can be done ~~very cheaply~~ efficiently. Otherwise, one can use the "cross-product" algorithm (Chap. 3, ~~Section~~ Alg 3.1)

Instead of computing the singular values and vectors of $AU - \mu U$ we can compute the eigendecomposition of

$$B_\mu = (AU - \mu U)^H (AU - \mu U) \text{ for any } \mu$$

After precomputing

$V = AU$, $C_0 = V^H V$ and $C_1 = V^H U + U^H V$

computing $B_\mu = \mu^2 I - \mu C_1 + C_0$ is cheap.

This approach may lead to some loss of accuracy, although it will often be good if μ sufficiently close to λ and

$\sigma_{p-1}(AU - \mu U)$ not too close to $\sigma_p(AU - \mu U)$.

An alternative for refined Ritz vectors are harmonic Ritz vectors.

Def 4.11

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

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

($\to$ harmonic Rayleigh-Ritz method)

Multiplying by w^H from the right, we see that

$$\cancel{(Uw)^H [(A-\kappa I)Uw]^H [(A-\kappa I)Uw] = \delta (Uw)^H}$$

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

$$\|(A-\kappa I)Uw\|_2^2 \le |\delta| \, \|(A-\kappa I)Uw\|_2 \iff$$

$$\|(A-\kappa I)Uw\|_2 \le |\delta|$$

So, if $\kappa \to \lambda$ and $\{\delta\} \to 0$, the residual must go to zero and convergence is guaranteed.