4.3. \textbf{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.

\[ A \]

3.1. Krylov sequences and Krylov spaces

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

\textbf{Definition 3.2} Let $A$ be an $m \times n$ matrix. Then the sequence $v, Av, A^2v, \ldots$ is a Krylov sequence based on $A$.

The matrix $K_k(A,v) = (v, Av, A^2v, \ldots, A^{k-1}v)$ is the $k$-th Krylov matrix. The space $K_k(A,v) = \mathbb{R}(K_k(A,v))$ is called the $k$-th Krylov subspace.

\[ K_k = K_k(A,v) \]
Elementary properties

**Theorem 3.3** Let $A$ and $n$ be given. Then

1. $J_k(A, n) \subseteq J_{k+1}(A, n)$
   \[ A \cdot J_k(A, n) \subseteq J_{k+1}(A, n) \]

2. If $\delta = 0$
   \[ J_k(A, n) = J_k(\delta A, n) = J_k(A, \delta n) \]

3. For any $x$
   \[ J_k(A, n) = J_k(\lambda - x I, n) \]

4. If $W$ is nonsingular
   \[ J_k(W^{-1}AW, W^{-1}n) = W^{-1}J_k(A, n) \]

**Proof** Easy.

The polynomial connection

Let $n \in J_k(A, n)$

\[ n = y_1n + y_2An + y_3A^2n + \ldots + y_{k-1}A^{k-1}n \]

$p(x) = y_1 + y_2x + y_3x^2 + \ldots + y_{k-1}x^{k-1}$

$q = p(A)n$

$p(A) = y_1I + y_2A + y_3A^2 + \ldots + y_{k-1}A^{k-1}$
Conversely, if $\lambda = p(A)$ for any polynomial of degree $\leq k-1$, then $\lambda \in \text{Ker}(A, n)$. Thus,

**Theorem 3.4.** The space $\text{Ker}(A, n)$ can be represented as

$$\text{Ker}(A, n) = \{ p(A)w : \deg(p) \leq k-1 \}$$

Note that

$$\text{Ker}_1 \subseteq \text{Ker}_2 \subseteq \text{Ker}_3 \subseteq \text{Ker}_4 \subseteq \ldots \subseteq \text{Ker}_j \subseteq \text{Ker}_{j+1} \subseteq \ldots$$

$$\Rightarrow \text{if } n \text{ is an eigenvalue, } A^n w = w \Rightarrow \text{Ker}_2 = \text{Ker}_1, \text{Ker}_3 = \text{Ker}_2 \text{ etc.}.$$

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

$$\text{Ker}_{l+1}(A, n) = \text{Ker}(A, n).$$

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

**Proof.** Clear because $\text{Ker}_l \subseteq \text{Ker}_{l+1}$.  

3
Note, if $K_{L_i} = K_L$, then

$A K_L \subseteq K_{L_i} = K_L \Rightarrow A K_L \subseteq K_L$

$\Rightarrow K_L$ is an invariant subspace of $A$.

Second claim: If $\dim K_L = \infty$,

If the Krylov sequence terminates at $l$,

then $K_L$ is an eigenspace of dimension $l$.

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

Let $X$ be an invariant subspace, thus $X = X$, let $m \in X$. Then $A^n \in AX \subseteq X \Rightarrow A^n \in X$.

If the sequence terminates at $l > M$,

then $\dim K_L > M = \dim K_M = \dim X$.

Which is a contradiction with $K_L \subseteq X$.  

Typo in the book.
Note, \[ \mathbf{K} = \mathbf{K}, \] then
\[ A \mathbf{K} = \mathbf{K} \Rightarrow A \mathbf{K} \subseteq \mathbf{K} \Rightarrow \mathbf{K} \text{ is an invariant subspace of } A. \]

Second claim of Th 3.5:
If the Krylov sequence terminates at \( l \),
then \( \mathbf{K} \) is an eigenspace of dimension \( l \).
On the other hand, if \( M \) lies in an eigenspace of dimension \( m \), then for some \( l \leq m \) the sequence terminates at \( l \).

Let \( X \) be an invariant subspace, \( \dim X = m \),
let \( m \in X \).
Then \( A^m \in AX \subseteq X \Rightarrow A^m \in X \)
... \( A^n \in X \)
If the sequence terminates at \( l > m \),
then \( \mathbf{K} = \dim \mathbf{K} > m = \dim \mathbf{K} \Rightarrow \dim X \)
which is a contradiction since \( \mathbf{K} \subseteq X \).

Type in the book.
Conversely, if
\[ p(A) \in \mathbb{R} \text{ for any polynomial of degree } \leq k-1 \]
then \( N = \ker(A^n) \). Thus,

\[ \ker(A^n) = \left\{ p(A)w : \deg(p) \leq k-1 \right\} \]

Theorem 3.4. The space \( \ker(A^n) \) can be represented as

\[ \ker(A^n) = \left\{ p(A)w : \deg(p) \leq k-1 \right\} \]

Note that

\[ \ker(A) \subseteq \ker(A^2) \subseteq \ker(A^3) \subseteq \ldots \subseteq \ker(A^n) \subseteq \ldots \]

\[ \implies \text{if } n \text{ is an eigenvalue, } \]

\[ A^n v = \lambda^n v \implies \ker(A^n) = \ker(A^{n+1}) = \ker(A^{n+2}) \text{ etc.} \]

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

\[ \ker(A^{l+1}) = \ker(A^l). \]

Theorem 3.5. A Krylov sequence terminates at \( l \)

if and only if \( l \) is the smallest integer

such that \( \dim \ker(A^{l+1}) = \dim \ker(A^l) \).

Proof. Clear because \( \ker(A^l) \subseteq \ker(A^{l+1}). \)
Conversely, if $v = p(A)v$ for any polynomial of degree $\leq k-1$, then $v \in \mathcal{K}_k(A,v)$. Thus,

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

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

Note that

$$\mathcal{K}_1 \subseteq \mathcal{K}_2 \subseteq \mathcal{K}_3 \subseteq \cdots \subseteq \mathcal{K}_k \subseteq \cdots$$

$\Rightarrow$ if $v$ is an eigenvector, $A^n v = Av = \cdots = A v = v$ etc.

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

$$\mathcal{K}_{L+1}(A,v) = \mathcal{K}_L(A,v).$$

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

**Proof:** Clear because $\mathcal{K}_L \subseteq \mathcal{K}_{L+1}$. 

3
Termination is good because we have found an invariant subspace that contains enough information about some eigenvectors. It's bad because it stops furnishing lots of information about other eigenvectors.

3.2. Convergence

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

For simplicity, we assume that \( A \) is Hermitian, with eigenpairs \( (\lambda_i, x_i) \), \( i = 1, \ldots, m \). Let \( n \) be the starting vector for a k-th sequence, \( n = x_1 x_1 + \cdots + x_m x_m \), \( x_i^* n = x_i^* x_i \).

Any vector in \( \mathbb{C}^k \) can be written as \( p(A) n \), with multivariate polynomial \( p \) of degree \( \leq k - 1 \):

\[
p(A) n = p(A) \sum_{i=1}^{m} x_i x_i = \sum_{i=1}^{m} x_i p(A) x_i
\]

\[
= \sum_{i=1}^{m} x_i p(\lambda_1) x_i = x_1 p(\lambda_1) x_1 + x_2 p(\lambda_2) x_2 + \cdots + x_m p(\lambda_m) x_m
\]
\[ p(A)^n = \alpha_1 p(h_1)x_1 + \alpha_2 p(h_2)x_2 + \cdots + \alpha_m p(h_m)x_m \]

Let \( \alpha_i \) be a vector from \( \mathbb{R}^k \)

**Note**: If we can find a polynomial \( p \) such that

\[ |p(h_i)| \geq \max_{j \neq i} |p(h_j)| \]

then

\[ p(A)^n = \alpha_i p(h_i)x_i + \sum_{j \neq i} \alpha_j \frac{p(h_j)}{p(h_i)} x_j \]

and \( p(A)^n \) will be a good approx to \( x_i \)

**Theorem 3.6**: If \( \alpha_i = x_i^* n \neq 0 \)

and \( p(h_i) \neq 0 \)

\[ \tan \chi(p(A)^n, x_i) \leq \max_{j \neq i} \left| \frac{p(h_j)}{p(h_i)} \right| \tan \chi(n, x_i) \]

**Proof**: \( n = \alpha_1 x_1 + \alpha_2 x_2 + \cdots + \alpha_i x_i + \cdots + \alpha_m x_m \)

\[ \cos \chi(n, x_i) = \frac{|x_i^* n|}{\|x_i^* n\| \|n\|} = \frac{\|x_i\|}{\|x_i^* n\|} \|n\| \]

\[ \|n\| = \sqrt{\sum_{i=1}^m |\alpha_i|^2} \]

6
\[
\sin^2 (\mathbf{m}, \mathbf{x}_i) = 1 - \cos^2 (\mathbf{m}, \mathbf{x}_i) = \\
= 1 - \frac{|d_i|^2}{||m||^2} = \frac{||m||^2 - |d_i|^2}{||m||^2} = \\
\sum_{j \neq i} |d_j|^2 \\
= \frac{||m||^2}{||m||^2} \\
\Rightarrow \tan^2 (\mathbf{m}, \mathbf{x}_i) = \sum_{j \neq i} \frac{|d_j|^2}{|d_i|^2}
\]

In the same way,

\[
\tan^2 (p(A)\mathbf{m}, \mathbf{x}_i) = \sum_{j \neq i} \frac{|d_j|^2}{|d_i|^2}
\]

\[
\leq \max_{j \neq i} \frac{|P(H_j)|^2}{|P(H_i)|^2} \sum_{j \neq i} \frac{|d_j|^2}{|d_i|^2} \\
= \max_{j \neq i} \frac{|P(H_j)|^2}{|P(H_i)|^2} \Rightarrow \tan^2 (\mathbf{m}, \mathbf{x}_i)
\]

So, if there is a polynomial \( P \) that is "large" at \( d_i \) and "small" at \( d_i, d_j \), then the corresponding \( p(A)\mathbf{m} \) will go to the direction of \( \mathbf{x}_i \).
\begin{equation}
\tan \chi(x_i, \pi, x_i) \leq \max_{i \neq i'} \left| \frac{\hat{p}(i)}{\hat{p}(i')} \right| + \tan \chi(x_i, x_i)
\end{equation}

Independent of reality of \( \hat{p} \)

So we can consider \( \hat{p} \) normalized

So that \( \hat{p}(i') = 1 \)

\begin{equation}
\tan \chi(x_i, \pi, x_i) \leq \max_{j \neq i} \left| \frac{\hat{p}(j)}{\hat{p}(i')} \right| + \tan \chi(x_i, x_i)
\end{equation}

\( \hat{p}(i') = 1 \)

\[ \deg(p) \leq k - 1 \]

\( p(4) \in \mathbb{K}_k \)

\[ \Rightarrow \tan \chi(x_i, \mathbb{K}_k) \leq \min_{j \neq i} \max_{\deg(p) = k - 1} \left| \frac{\hat{p}(j)}{\hat{p}(i')} \right| + \tan \chi(x_i, x_i) \]

This fact determines how well \( \mathbb{K}_k \)
will perform in approximating \( x_i \).

It is enough if we find particular polynomials
that will
produce small factors.
We study \( \min \max |P(d)| \), let \( r = 1 \), i.e. we want \( x_1 \).

Since \( \lambda \) is assumed nonnegative, we can order the exponents \( \lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n \).

\( \lambda_1 \) assume this for the moment.

\[
\begin{array}{c}
\hline
-\lambda_1 & -\lambda_3 & -\lambda_2 & -\lambda_1 \\
\hline
\end{array}
\]

\( \lambda_1 \) larger obviously, includes \( \lambda_2, \ldots, \lambda_n \).

Note:

\[
\min \max |P(d)| \leq \min_{\deg(p) \leq k - 1} \max_{n \in \mathbb{N}, m \in \mathbb{N}} |P(d)|.
\]

So we have

\[
\tan \chi (x_1, x_k) \leq \min_{\deg(p) \leq k - 1} \max_{n \in \mathbb{N}, m \in \mathbb{N}} |P(d)| \cdot \tan \chi (n, x_1).
\]

An elegant solution using Chebyshev polynomials.
Chebyshev polynomials

\[ C_k(t) = \begin{cases} \cos(k \cos^{-1} t), & |t| \leq 1 \\ \cosh(k \cosh^{-1} t), & |t| \geq 1 \end{cases} \]

Theorem

1. For \(|t| > 1\)
   \[ C_k(t) = \frac{2t C_{k-1}(t) - C_{k-2}(t)}{t^2 - 1}, \quad k = 1, 2, \ldots \]

\(\cos(k \gamma) \cos((k-1) \gamma) = 2 \cos k \gamma \cos \gamma\)

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 \(|t| \leq 1\), \(|C_k(t)| \leq 1\)
   \[ b_{ik} = \cos \left(\frac{k-i}{2}\pi\right), \quad i = 0, \ldots, k \]
   \[ C_k(t_{ik}) = (-1)^{k-i} \]

4. For \(k > 1\)
   \[ \text{Max}_{\deg P \leq k} \text{Max}_{\gamma \in [-1, 1]} |p(t)| = \frac{1}{C_k(1)} \quad \text{at} \quad P(t) = \frac{C_k(t)}{C_k(1)} \]

Typo in the book

Typo in the book

Typo in the book
Now, to estimate \( f_\ell (x_1, K_{12}) \), we need

\[
K_k = \min \left\{ \max \left\{ P(x) \right\} \right. \text{ s.t. } \deg(P) = k-1 \text{ and } \begin{cases} P(1) = 1 \\
\end{cases}
\]

First, we transform \([a_1, a_2]\) to \([-1, 1]\):

change of variable

\[
\gamma = a_2 + (a_2 - 1) \frac{a_1 - a_2}{2},
\]

\[
\mu_1 = \mu(a_1) = 1 + 2 \frac{a_1 - a_2}{a_2 - a_1}
\]

\[
\Rightarrow \delta_k = \frac{1}{c_{k-1}(\mu_1)}
\]

Theorem: Let \( A = A^n \), \((x_i, x_i) i = 1..n\), \( x_i^T x_j = \delta_{ij} \).

Then, \( A_{12} \geq \ldots \geq A_{n,n} \) and \( \gamma = \frac{a_1 - a_2}{a_2 - a_1} \).

Then \( f_\ell (x_1, K_{12}(A_{1,m})) \leq \frac{\tan \gamma (x_1, y)}{c_{k-1}(1 + 2 \gamma)} \)

\[
= \frac{\tan \gamma (x_1, y)}{(1 + \sqrt{(1 + 2 \gamma)^2 - 1} )^{k-1}} \left( 1 + \sqrt{(1 + 2 \gamma)^2 - 1} \right)^{-k-1}
\]
\[
\begin{align*}
\frac{\tan \phi (x_1, n)}{(1 + 2 \sqrt{\gamma_1^2 + \gamma_2^2})^{k-1}} + \frac{\tan \phi (x_1, n)}{(1 + 2 \sqrt{\gamma_1^2 + \gamma_2^2})^{k-1}} & \quad \text{(K-1)} \\
\end{align*}
\]

- It is possible to give bounds for the remaining eigenvectors $x_2, -x_3$
Let $\lambda_1 > \lambda_2 > \ldots > \lambda_m$

$m = x_1 x_1 + d_2 x_2 + \ldots + d_n x_n$

$A^k m = (x_1 x_1 + d_2 x_2) \lambda_1^k + d_3 x_3 \lambda_3^k + \ldots + d_n x_n \lambda_n^k$

This will approximate $x_1 x_1 + d_2 x_2$

and "see" $(\lambda_1, x_1 x_1 + d_2 x_2)$ as a

simple eigenpair.

We can try with another vec

$m' = \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \ldots + \beta_n x_n$

$A^k m' = (\beta_1 x_1 + \beta_2 x_2) \lambda_1^k + \beta_3 x_3 \lambda_3^k + \ldots + \beta_n x_n \lambda_n^k$

unless $x_1 x_1 + d_2 x_2$ and $\beta_1 x_1 + \beta_2 x_2$

are collinear, we will have two orthogonal

subspace bases for $\lambda_1 = \lambda_2$.

So we can use

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

$K_{x_1} (A, U)$ block Krylov subspace
Theorem 3.1: \( x_{\geq 1} \geq l_n \geq l_m \) 

multiple of \( l_1 \leq l_m 

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

n = UB^{-1}e_1 \implies \tan \phi(x_1, y_k | A_1 U) \leq \frac{\tan (x_1, y)}{C_{k-1} (1+2\gamma)} 

\gamma = \frac{l_{n-1} - l_{m+1}}{l_{m+1} - l_m} 

\begin{array}{c|c|c|c} \hline \lambda & l_{m+1} & l_2 & l_1 \\ \hline \end{array}
4.4. Rayleigh Ritz

→ suppose now that we have an approximate subspace containing approximate approximations to an eigenvalue of A

4.1. Rayleigh–Ritz methods

\[ A \begin{bmatrix} U \end{bmatrix} \]

\[ U = R(U) \] contains an approximate subspace of A

\[ U \] has been computed by some method, e.g. as a Fourier subspace

**Theorem 4.1:** \( U = R(U) \). Let \( V^* \) be a left inverse of \( U \).

Set \( B = V^* A U \).

If \( X \subseteq U \) is an eigenvalue of \( A \), then there is an eigenpair \((L, W)\) of \( B \) such that \((L, U W)\) is an eigenpair of \( A \), with \( R(UW) = X \).

**Proof**

\[ BW = WL \iff (LW) \text{ eigenpairs of } B \]

Let \( AX = XL = 1 \)

\( X = R(X) \)

\[ A U W = U W L \]

\[ V^* A U W = V^* U W L = WL \]

\[ V^* (A U W - U W L) = 0 \]

\[ BW = WL \]

Hence, to identify the exact eigenvalue by examining the eigenpairs of \( B \).
Effect: When \( \mathbf{U} \) contains only an approximate eigenspace \( \tilde{X} \) of \( \mathbf{A} \), there would be an eigenspace \( (\tilde{X}, \tilde{W}) \) of \( \mathbf{B} \) such that \( (\tilde{X}, \tilde{W}) \) is an approx eigenspace of \( \mathbf{A} \).

Rayleigh-Ritz procedure

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

\[
\begin{align*}
\text{Ritz pair} & \quad \text{Ritz basis} \quad \text{primitive Ritz vectors} \\
\mathbf{UW} & \quad (\mathbf{UW})^\dagger & & (\mathbf{UW})^\dagger
\end{align*}
\]

Note: For a matrix \( \mathbf{V} \) such that \( \mathbf{V}^\dagger \mathbf{U} \) is nonsingular, \( \tilde{\mathbf{V}} = \mathbf{V} (\mathbf{V}^\dagger \mathbf{U})^{-1} \) satisfies

\[
\tilde{\mathbf{V}}^\star \mathbf{U} = (\mathbf{V}^\dagger \mathbf{U})^{-1} \mathbf{V}^\dagger \mathbf{U} = \mathbf{I}
\]

\[
(\mathbf{V}^\dagger \mathbf{U})^{-1} \mathbf{V}^\dagger \mathbf{A} \mathbf{U} \mathbf{w} = \mu \mathbf{w} \iff (\mathbf{V}^\dagger \mathbf{A} \mathbf{U}) \mathbf{w} = \mu (\mathbf{V}^\dagger \mathbf{U}) \mathbf{w}
\]

generalized eigen-problem
Example \[ A = \begin{bmatrix} 0 & 1 \\ 0 & -1 \end{bmatrix} \] target: \((0, e_1)\)

\[ U = \begin{bmatrix} 1 & 0 \\ 0 & \sqrt{2} \end{bmatrix}, \quad U^* U = I \]

\[ B = U^* A U = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} \]

case

\[ B w = 0 \] for any \( w \neq 0 \)

The orthogonal RR procedure

U orthogonal, \( V = U \), take \( W \) also orthogonal, so \( UW \) is orthogonal.

e.g. \( A = A^* \), \( B = U^* A U = B^* \)

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

\[ R = A \hat{x} - X M \] is minimal in any unit invar. norm. (over all \( A \hat{x} - X N \))

(see Th 2.6)
4.2. Convergence

Goal: an eigepair \((\lambda, x)\) of \(A\).

Suppose we obtain orthonormal basis

\[ U \]

for which \(4(x, U_\lambda, U_\lambda) = \delta \) is small.

We apply RR and obtain a Ritz pair \((\lambda_m, U_\lambda, p_\lambda)\).

If satisfied, stop. Else, get another \(U_\lambda\) with

smaller \( \delta \). As \( \delta \to 0 \), we need to show convergence.

\[
B_\lambda = U_\lambda^* A U_\lambda, \quad \text{we will show first that}
\]

\[ B_\lambda \] has Ritz value \(\lambda_m\) that converges to \(\lambda\).

Convergence of the Ritz value

Theorem 4.4 Let \( B_\lambda = U_\lambda^* A U_\lambda \),

Then there is a matrix \( E_\lambda \) satisfying

\[
\| E_\lambda \|_2 \leq \frac{\sin \beta}{\sqrt{1 - \sin^2 \beta}} \| A \|_2
\]

such that \( \lambda \) is an eigenvalue of

\[ B_\lambda + E_\lambda. \]
Proof  Let \((U^c, U^c)\) be unitary

Set \(y = U^c x, z = U^c y\)

\[
x = U^c x + U^c y + U^c z
\]

\[
\|x\|_2 = \sqrt{\|y\|_2^2 + \|z\|_2^2}
\]

\[
A x - 2x = 0
\]

\[
U^c A \left(U^c U^c\right)^* x - 2 U^c x = 0
\]

\[
U^c A U^c U^c x + U^c A U^c z - 1 y = 0
\]

\[
B y + U^c A U^c z - 1 y = 0
\]

\[
\hat{y} = \frac{y}{\|y\|_2}
\]

\[
\hat{z} = \frac{z}{\|z\|_2}
\]

Define \(E_y = -\hat{y} \hat{y}^*\)

\[
\|E_y\|_2 = \|\hat{y}\|_2
\]

\[
(B^* + E_y) \hat{y} = 0
\]

\[
= B^* y - \hat{y} \hat{y}^* y = 0
\]
By perturbation theory (Chapter 1, Th 3.1)

\[ |\lambda_0 - \lambda| \leq 4 \left( 2 \| A \|_2 + \| E_0 \|_2 \right)^{\frac{1}{2m}} \| E_0 \|_2^{\frac{1}{m}} \xrightarrow{\text{as }\delta \to 0} 0 \]

eg. $B_0$  eg. $B_0 + E_0$

If $A$ is Hermitian

\[ |\lambda_0 - \lambda| \leq \| E_0 \|_2 \]

---

Convergence of Ritz vectors

Theorem 4.6  Let $(\mu_0, W_0)$ be a primitive Ritz pair ($B W_0 = \mu_0 W_0$) and let $(\mu_0 W_0)$ be unitary so that

\[
\begin{bmatrix}
W_0^* \\
W_0^* \\
\end{bmatrix}
B
\begin{bmatrix}
W_0 \\
W_0 \\
\end{bmatrix} =
\begin{bmatrix}
\mu_0 & 0 \\
0 & N_0 \\
\end{bmatrix}
\]

Then

\[
\min \left\{ (x, U_\mu W_0) \leq \min \left( 1 + \frac{\| h_0 \|_2^2}{\text{sep}(\lambda_1, N_0)^2} \right) \right. 
\]

6
\[ \|x - \nu_i\| \geq |\mu_\theta - \nu_i| - |\mu_\theta - \nu| \]

\[ \sum_{\text{sep}(\mu_\theta, \nu_\theta)} \rightarrow 0 \]

**Corollary 4.7.** Let \((\mu_\theta, U_{\text{seg}})\) be a finite pair for which \(\mu_\theta \rightarrow \mathcal{A}\). If there is \(\alpha > 0\) with

\[ \text{sep}(\mu_\theta, \nu_\theta) > \alpha \]

then asymptotically

\[ \lim_{x \rightarrow \infty} x \cdot (x, U_{\text{seg}}) \leq \min \sqrt{1 + \|A\|^2} \]

**Now recall the example with \(B = (0,0)\)**
Convergence of Fitz values revised

\((\mu, x) \sim (\lambda, x)\)

\(x = U\phi^m\sigma, \quad B_\phi = \mu \phi^m\sigma\)

\(B_\phi = U\phi^* A U\phi\)

\(\mu = x^\phi A x\phi\)

\(x = \Phi x + \sigma x, \quad \sigma \perp x\)

\(A = \cos A(x, x)\)

\(\sigma = \sin A(x, x)\)

\(\sigma \sim O(\varphi)\) by Corollary 4.7

\(M_\phi = (\Phi x + \sigma x)^\phi A (\Phi x + \sigma x)\)

\(= |\sigma|^2 I + \overline{\Phi} \sigma x^A \sigma^* A \sigma + |\sigma|^2 y^A \sigma^* A \sigma\)

\(|\mu - \lambda| = |(1 + |\sigma|^2) \sigma + \overline{\Phi} \sigma x^A \sigma^* A \sigma + |\sigma|^2 y^A \sigma|\)

\(\leq |\sigma|^2 |I| + \|\overline{\Phi} \sigma x^A \sigma^* A \sigma\| + |\sigma|^2 |y^A \sigma|^2\)

\(\leq \|A\| \|\sigma\|^2 (2|\sigma| + 1) = O(\varphi)\)
Orthogonal RR is optimal \( \rightarrow \) usual approach
\[
\rightarrow AX - \hat{X}M = R \quad \text{min. residual in any unit. norm.}
\]

Convergence
\[
A \rightarrow \text{eigenpair } \lambda, x
\]

As \( \nu^2 = \Delta(U, x) \rightarrow 0 \), how do approx of \( \lambda \) and \( x \) behave.

Rayleigh quotient \( B_\nu = U_\nu^H A U_\nu \)

\( B \) as above. Then \( E_\nu \) exists s.t.
\[
\|E_\nu\|_2 \leq \frac{\sin \nu}{\sqrt{1 - \sin^2 \nu}} \|A\|_2 \quad \text{and}
\]
\[
d \in \Lambda(B_\nu + E_\nu)
\]

Expand \( x = U_\nu y + U_\perp z \) (\( U_\nu U_\nu^H \) unitary)
\[
A(U_\nu U_\nu^H)(U_\nu^H) x - \lambda x = 0
\]
\[
\sum_{\nu=1}^{\infty} U_\nu^H A(U_\nu U_\nu^H)(U_\nu^H) y - \lambda U_\nu^H x = 0
\]

\[
B_\nu y + U_\nu^H A U_\perp z - \lambda y = 0 \quad \Rightarrow
\]

\[
B_\nu y - \lambda y = r = U_\nu^H A U_\perp \frac{z}{\sqrt{1 - \sin^2 \nu}} \quad \Rightarrow
\]
\[
\|r\|_2 \leq \frac{\sin \nu}{\sqrt{1 - \sin^2 \nu}} \|A\|_2
\]

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

\[
\|u_\nu - \hat{u}\| \leq \epsilon (\|A\|_2 + \|E_\nu\|_2)^{-\frac{1}{m}} \|E_\nu\|_2^\frac{1}{m}
\]

\( m \) is \( \dim(U) \)
In practice, we're not interested in large \( n \). So, we need iteration that expands and prunes \( U_0 \) while keeping \( \alpha \) modest.

It can be shown that Ritz vector converges \( H \) (for \( \theta \to 0 \)) if \( \alpha \) bounded away from \( \Delta(\theta) \setminus \{ \mu \} \).

Theo 4.6 and papers by Stewart [163] and Jia & Stewart [133]

Since this separation is not guaranteed, it may need 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 [18])

...uniform separation condition.

Later

\( \Rightarrow \) If Ritz vec come \( \Rightarrow \) better bound \( (\mu_0 - 1) = O(\theta) \)

\( \Rightarrow \) result from earlier analysis

If eig. vec. \( O(\theta) \) \( \Rightarrow \beta_0 = O(\theta^2) \) for Herm. matrix.

If \( \beta_0 \) remains "well-conditioned" then the Ritz vector \( \frac{u_0}{\omega} \) converges to eigenvector \( x_0 \).
Corr. 4.7 (based on Theo 4.6)

let \((\mu_0, U_0, w_0)\) be a Ritz pair for which \(\mu_0 \to 1\). If there is a constant \(\alpha > 0\) s.t. \(\text{sep}(\mu_0, N_0) \geq \alpha > 0\) then (asymptotically)

\[
\sin \Delta(x, U_0, w_0) \leq \sin \theta \sqrt{1 + \frac{\|A_1\|^2}{\alpha^2}}
\]

let \(B_0 = U_0^H A\) and let \((\mu_0, w_0)\) be a primitive Ritz pair. Furthermore, let \((w_0, W_0)\) be unitary and

\[
(w_0, W_0)^H B_0 (w_0, W_0) = \begin{pmatrix} \mu_0 & w_0^H \\ 0 & N_0 \end{pmatrix}
\]

Clearly, if \(\text{sep}(\mu_0, N_0)\) is very small, there is a small perturbation of \(B_0\) that has double eigenvalue \(\mu_0\) and \(w_0, w_0\) cannot be expected to be computed (or to have been computed) accurately.

If \(\mu_0 \to 1\) and \(d \in \Delta(A)\) where \(|d - \hat{d}|\) 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 \( \mu_0 \) is not. The latter is just an "artifact" of the projection \( U_0^H A U_0 \). In this case we must do extra work to get accurate eigenvector.

If (or once) we have a converging eigenvector, the Rayleigh quotient converges at least as fast as the eigenvector. Let \( x_0 \) be approx. to \( x \) and

\[ \mu_0 = x_0^H A x_0. \]

Let \( x_0 = y x + 0 y \) where \( \| y \|_2 = 1 \) and \( y \perp x \) (also let \( \| x \|_2 = \| x_0 \|_2 = 1 \) ). Then

\[ y_1 = \cos \Delta(x_0, x), \quad 1 \| y_1 \|_2 = \sin \Delta(x_0, x) \]

\[ x_0^H A x_0 = y_1^2 x_0^H A x + y_0^2 x_0^H A y + \]

\[ \overline{y}_1 y_1^H A x + 161^2 y_1^H A y \]

\[ = d 1 y_1^2 + y_0^2 x_0^H A y + 161^2 y_1^H A y = \mu_0 \]

\[ \| x_0 - d \| = \| (y_1^2 - 1) + \cdots \| \leq \]

\[ 10^2 |d| + 10^1 |y| \| x_0^H A y \| + 10^2 |y^H A y| \leq \]

\[ 10 \| A \|_2 \| (10^1 + 10^1 + 1^1) \| \leq (1 + 2 10^1) |10^1 \| A \|_2| \]
Since $|\theta| = \sin \Delta(x_0, x) \leq \sin \theta \cdot k$
we have $|\mu_0 - \mu| = O(\theta)$

If $A$ is Hermitian, use the term
\[ y^H \sigma x^H A y = 0 \] (drops out) and
we get $|\mu_0 - \mu| \leq 2|\theta|^2 \|A\|_2 \rightarrow$
$|\mu_0 - \mu| = O(\theta^2)$. See also notes on "Overview of Methods" discussing
the Rayleigh quotient.

Unfortunately, we have no way to know
$\Delta(x, u_0)$ or $\Delta(x_0, x)$ or $\theta$.

What can we learn from residual?

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

Then
\[
\sin \Delta(\tilde{x}, x) \leq \frac{p}{\text{sep}(\mu, \lambda)} \leq \frac{p}{\text{sep}(\lambda, \mu) - |\mu - \lambda|}
\]

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

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

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

$\sin \Delta(\tilde{x}, x) \leq \text{sep}(\mu, \lambda) \|r\|_2$
(For simple $d$)

Last step follows from continuity of sep.

So, if $\mu \to 1$ and $\|\| \to 0$ the eig.

pair converges.

If $\mu \to 1$ but $\mu$ nearly multiple eig.
val., although $d$ simple, eig. vector may not

converge.

Near "multiple" copy" of $\mu$ is called

spurious eigenval. as it does not corr.
to eigenval. of $A$ (it's artifact of

projection).

Alternative for Ritz vector $\to$

Refined Ritz vector:

$\hat{x} = \arg \min_{x} \| A\hat{x} - \mu_0 \hat{x} \|_2$ s.t.

$\hat{x} \in U_0, \| \hat{x} \|_2 = 1$
Theorem 4.10

Let \( A \) have spectral representation

\[
A = \lambda xy^H + XLy^H,
\]

where \( \|x\|_2 = 1 \), \( y \) orthonormal. Let \( \mu_0 \) be Ritz value and \( \hat{x}_0 \) be corr. refined Ritz vector. If \( \text{sep}(d,L) = |\mu_0 - d| > 0 \), then

\[
\sin \Delta(x, x_0) \leq \frac{\|A - \mu_0 I\|_2 \sin \Theta + |d - \mu_0|}{\sqrt{1 - \sin^2 \Theta} \left( \text{sep}(d,L) - |d - \mu_0| \right)}
\]

Proof

Let \( U \) be orthonormal basis for \( U_1 \), and \( x = w + z \), where \( w = UX^1x \).

Then \( \|w\|_2 = \sqrt{1 - \sin^2 \Theta} (= \cos \Theta) \)

\[
\|z\|_2 = \sin \Theta.
\]

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

\[
(A - \mu_0 I) \hat{w} = \frac{(A - \mu_0 I)w}{\sqrt{1 - \sin^2 \Theta}} = \frac{(A - \mu_0 I)(x - z)}{\sqrt{1 - \sin^2 \Theta}} = \frac{(A - \mu_0 I)x - (A - \mu_0 I)z}{(1 - \sin^2 \Theta)^{1/2}}
\]

\[
\|A - \mu_0 I\|_2 \|\hat{w}\|_2 \leq \|A - \mu_0 I\|_2 \|\hat{x}_0\|_2 \leq |d - \mu_0| + \|A - \mu_0 I\|_2 \sin \Theta
\]

\[
(1 - \sin^2 \Theta)^{1/2}.
\]

Using (Theo 4.8)

\[
\sin \Delta(x, x_0) \leq \frac{\|A - \mu_0 I\|_2 \|\hat{x}_0\|_2}{\text{sep}(d,L) - |d - \mu_0|}
\]

completes the proof.
For $\theta \to 0$, $\mu_0 \to 1$ (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 $\mu_0$ by the Rayleigh quotient

$$\hat{\mu}_0 = \hat{x}_0^\dagger \hat{A} \hat{x}_0$$

residual $\| \hat{A} \hat{x}_0 - \hat{\mu}_0 \hat{x}_0 \|_2$ optimal.

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

$$\| (A - \mu I) U z \|_2$$

minimal

So, we must compute the right singular vector corresponding to the smallest singular value of $A - \mu I$.

Then

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

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

Instead of computing the singular values and vectors of $A - \mu I$, we can compute the eigenvalue decomposition of

$$B \mu = (A - \mu U) \hat{U}^\dagger (A - \mu U)$$

for any $\mu$.

After precomputing
\( V = \mathbb{R}^n, \quad C_0 = V^H V \quad \text{and} \quad C_1 = V^H U U^H V \)

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 \( \mu \) sufficiently close to \( 1 \) and \( \sigma_p(\mathbb{R}^n - \mu U) \) not too close to \( \sigma_p(\mathbb{R}^n - \mu U) \).

An alternative for refined Ritz vectors are harmonic Ritz vectors.

Let \( U \) be subspace and \( \mathcal{U} \) orthonormal basis for \( U \). Then \((\mathbb{R}^n, U \mathcal{U})\) is a harmonic Ritz pair with shift \( \kappa \) of

\[
U^H (A - \kappa \mathcal{I})^H (A - \kappa \mathcal{I}) U \mathcal{U} = \delta U^H (A - \kappa \mathcal{I}) U \mathcal{U}
\]

(\( \rightarrow \) harmonic Rayleigh-Ritz method)

Multiplying by \( U^H \) from the right, we see that

\[
\begin{align*}
U^H ((A - \kappa \mathcal{I}) U \mathcal{U})^H (A - \kappa \mathcal{I}) (U \mathcal{U}) &= \delta (U \mathcal{U})^H (A - \kappa \mathcal{I}) (U \mathcal{U}) \\
\| (A - \kappa \mathcal{I}) U \mathcal{U} \|^2_2 &\leq |\delta| \| (A - \kappa \mathcal{I}) U \mathcal{U} \|^2_2 \rightarrow \\
\| (A - \kappa \mathcal{I}) U \mathcal{U} \|^2_2 &\leq |\delta|
\end{align*}
\]

So, if \( \kappa \to 1 \) and \( |\delta| \to 0 \), the residual must go to zero and convergence is guaranteed.