Krylov Subspace Methods for Large/Sparse Eigenvalue Problems

Wen-Wei Lin

Department of Applied Mathematics
National Chiao Tung University, Taiwan

Outline

Reference

1. G. H. Golub and C. F. Van Loan, Matrix Computations
2. W.-W. Lin, Lecture Notes of Matrix Computations
3. G. W. Stewart, Matrix Algorithms, Volume II: Eigensystems

Orthogonal projection methods

Let $A \in \mathbb{C}^{n \times n}$ and $\mathcal{K} \subset \mathbb{C}^n$, $\dim(\mathcal{K}) = m < n$. The eigenvalue problem is to find $0 \neq x \in \mathbb{C}^n, \lambda \in \mathbb{C}$ such that

$$Ax = \lambda x.$$

In an orthogonal projection technique onto $\mathcal{K}$, we want to find an approximate eigenpair $\tilde{\lambda} \in \mathbb{C}, \tilde{x} \in \mathcal{K}$ satisfying Galerkin condition

$$(A\tilde{x} - \tilde{\lambda}\tilde{x}, v) = 0, \quad \forall\, v \in \mathcal{K}. \tag{1}$$

Let $V = [v_1, v_2, \ldots, v_m]$ be an orthonormal basis for $\mathcal{K}$. Letting $\tilde{x} = Vy$, (1) becomes

$$(AVy - \tilde{\lambda}Vy, v_j) = 0, \quad j = 1, 2, \ldots, m.$$

Therefore, y and $\tilde{\lambda}$ satisfy

$$A_m y = \tilde{\lambda} y \quad \text{with} \quad A_m = V^H AV.$$

Theorem

$$\|AV - VA_m\|_2 \leq \|AV - VB\|_2\,, \quad \forall B \in \mathbb{C}^{m \times m}.$$

$\Rightarrow A_m$ can be seen to be a best approximation of A in $\mathcal{K}$.

Algorithm (Rayleigh-Ritz Procedure)

1. *Compute* $V = [v_1, ..., v_m]$ *forms an orthonormal basis for* $\mathcal{K}$.
2. *Compute* $A_m = V^H AV$.
3. *Compute eigenvalues* $\tilde{\lambda}_i$ *of* A_m *(Ritz values)*
4. *Compute eigenvector* y_i *of* A_m*,* $\tilde{x}_i = V y_i$ *(Ritz vector).*

- No guarantee that the result approximates the desired eigenpair of A.

Example

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

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

Then

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

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

- Note that the approximate eigenpairs $(\tilde{\lambda}_i, \tilde{x}_i)$, $i = 1, \ldots, m$, are exact eigenpairs of A provided that $\mathcal{K}$ is an invariant subspace of A.
- **Question**: Which subspace $\mathcal{K}$ is meaningful and effective?

Suppose the eigenvalue with maximum module is wanted.

Power method

Compute the dominant eigenpair

Disadvantage

At each step it considers only the single vector $A^k u$, which throws away the information contained in the previously generated vectors $u, Au, A^2u, \ldots, A^{k-1}u$.

Krylov Subspaces

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.

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

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

where

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

Assume that $A^\top = A$ and $A x_i = \lambda_i x_i$ for $i = 1, \ldots, n$. Write u in the form

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

Since $p(A) x_i = p(\lambda_i) x_i$, we have

$$p(A) u = \alpha_1 p(\lambda_1) x_1 + \alpha_2 p(\lambda_2) x_2 + \cdots + \alpha_n p(\lambda_n) x_n. \tag{2}$$

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 (2), 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

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

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

(ii) *For* $|t| > 1$, $c_k(t) = (1 + \sqrt{t^2 - 1})^k + (1 + \sqrt{t^2 - 1})^{-k}$.

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

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

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

(iv) *For* $s > 1$,

$$\min_{deg(p) \leq k, p(s)=1} \max_{t \in [0,1]} |p(t)| = \frac{1}{c_k(s)}, \tag{3}$$

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

For applying (3), 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

$$\begin{aligned} & \min_{deg(p) \leq k-1, p(\lambda_1)=1} \max_{\lambda \in [\lambda_n, \lambda_2]} |p(\lambda)| \\ = & \min_{deg(p) \leq k-1, p(\mu_1)=1} \max_{\mu \in [0,1]} |p(\mu)| = \frac{1}{c_{k-1}(\mu_1)} \end{aligned}$$

Theorem

Let $A^\top = A$ *and* $Ax_i = \lambda_i x_i,\ i = 1, \cdots, n$ *with* $\lambda_1 > \lambda_2 \geq \cdots \geq \lambda_n$. *Let* $\eta = \frac{\lambda_1 - \lambda_2}{\lambda_2 - \lambda_n}$. *Then*

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

- 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 conv. rate 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 Ews of $A \in \mathbb{R}^{100\times 100}$. Then $\eta = 0.0530$ and the bound on the conv. 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.

Householder transformation

Definition

A Householder transformation or elementary reflector is a matrix of

$$H = I - uu^*$$

where $\|u\|_2 = \sqrt{2}$.

Note that H is Hermitian and unitary.

Theorem

Let x be a vector such that $\|x\|_2 = 1$ and x_1 is real and nonnegative. Let

$$u = (x + e_1)/\sqrt{1 + x_1}.$$

Then

$$Hx = (I - uu^*)x = -e_1.$$

Proof:

$$\begin{aligned}(I - uu^*)\,x &= x - (u^*x)u = x - \frac{x^*x + x_1}{\sqrt{1+x_1}} \cdot \frac{x + e_1}{\sqrt{1+x_1}} \\ &= x - (x + e_1) = -e_1\end{aligned}$$

□

Theorem

Let x *be a vector with* $x_1 \neq 0$. *Let*

$$u = \frac{\rho \frac{x}{\|x\|_2} + e_1}{\sqrt{1 + \rho \frac{x_1}{\|x\|_2}}},$$

where $\rho = \bar{x}_1 / |x_1|$. *Then*

$$Hx = -\bar{\rho} \|x\|_2 e_1.$$

Proof: Since

$$\begin{aligned} & [\bar{\rho}x^*/\|x\|_2 + e_1^T][\rho x/\|x\|_2 + e_1] \\ = \;& \bar{\rho}\rho + \rho x_1/\|x\|_2 + \bar{\rho}\bar{x}_1/\|x\|_2 + 1 \\ = \;& 2[1 + \rho x_1/\|x\|_2], \end{aligned}$$

it follows that

$$u^*u = 2 \quad \Rightarrow \quad \|u\|_2 = \sqrt{2}$$

and

$$u^*x = \frac{\bar{\rho}\|x\|_2 + x_1}{\sqrt{1 + \rho\frac{x_1}{\|x\|_2}}}.$$

Hence,

$$
\begin{aligned}
Hx &= x - (u^* x) u = x - \frac{\bar{\rho}\|x\|_2 + x_1}{\sqrt{1 + \rho \frac{x_1}{\|x\|_2}}} \frac{\rho \frac{x}{\|x\|_2} + e_1}{\sqrt{1 + \rho \frac{x_1}{\|x\|_2}}} \\
&= \left[1 - \frac{(\bar{\rho}\|x\|_2 + x_1) \frac{\rho}{\|x\|_2}}{1 + \rho \frac{x_1}{\|x\|_2}} \right] x - \frac{\bar{\rho}\|x\|_2 + x_1}{1 + \rho \frac{x_1}{\|x\|_2}} e_1 \\
&= -\frac{\bar{\rho}\|x\|_2 + x_1}{1 + \rho \frac{x_1}{\|x\|_2}} e_1 \\
&= -\bar{\rho}\|x\|_2 e_1 .
\end{aligned}
$$

□

Definition

A complex $m \times n$-matrix $R = [r_{ij}]$ is called an upper (lower) triangular matrix, if $r_{ij} = 0$ for $i > j$ $(i < j)$.

Definition

Given $A \in \mathbb{C}^{m\times n}, Q \in \mathbb{C}^{m\times m}$ unitary and $R \in \mathbb{C}^{m\times n}$ upper triangular such that $A = QR$. Then the product is called a QR-factorization of A.

Theorem

Any complex $m \times n$ matrix A can be factorized by the product $A = QR$, where Q is $m \times m$-unitary and R is $m \times n$ upper triangular.

Proof: Let $A^{(0)} = A = [a_1^{(0)} | a_2^{(0)} | \cdots | a_n^{(0)}]$. Find $Q_1 = (I - 2w_1 w_1^*)$ such that $Q_1 a_1^{(0)} = ce_1$. Then

$$
\begin{aligned}
A^{(1)} &= Q_1 A^{(0)} = [Q_1 a_1^{(0)}, Q_1 a_2^{(0)}, \cdots, Q_1 a_n^{(0)}] \\
&= \left[\begin{array}{c|c|c|c} c_1 & * & \cdots & * \\ \hline 0 & & & \\ \vdots & a_2^{(1)} & \cdots & a_n^{(1)} \\ 0 & & & \end{array}\right].
\end{aligned} \tag{4}
$$

Find $Q_2 = \left[\begin{array}{c|c} 1 & 0 \\ \hline 0 & I - w_2 w_2^* \end{array}\right]$ such that $(I - 2w_2 w_2^*)a_2^{(1)} = c_2 e_1$.

Then

$$A^{(2)} = Q_2 A^{(1)} = \left[\begin{array}{cc|ccc} c_1 & * & * & \cdots & * \\ 0 & c_2 & * & \cdots & * \\ \hline 0 & 0 & & & \\ \vdots & \vdots & a_3^{(2)} & \cdots & a_n^{(2)} \\ 0 & 0 & & & \end{array}\right].$$

We continue this process. Then after $l = \min(m, n)$ steps $A^{(l)}$ is an upper triangular matrix satisfying

$$A^{(l-1)} = R = Q_{l-1} \cdots Q_1 A.$$

Then $A = QR$, where $Q = Q_1^* \cdots Q_{l-1}^*$. □

Theorem

Let A be a nonsingular $n \times n$ matrix. Then the QR- factorization is essentially unique. That is, if $A = Q_1R_1 = Q_2R_2$, then there is a unitary diagonal matrix $D = diag(d_i)$ with $|d_i| = 1$ such that $Q_1 = Q_2D$ and $DR_1 = R_2$.

Proof: Let $A = Q_1R_1 = Q_2R_2$. Then $Q_2^*Q_1 = R_2R_1^{-1} = D$ must be a diagonal unitary matrix. □

Arnoldi Method

Suppose that the columns of K_{k+1} are linearly independent and let

$$K_{k+1} = U_{k+1} R_{k+1}$$

be the QR factorization of K_{k+1}. Then the columns of U_{k+1} are results of successively orthogonalizing the columns of K_{k+1}.

Theorem

Let $\|u_1\|_2 = 1$ *and the columns of* $K_{k+1}(A, u_1)$ *be linearly indep. Let* $U_{k+1} = [u_1 \ \cdots \ u_{k+1}]$ *be the* Q*-factor of* K_{k+1}*. Then there is a* $(k+1) \times k$ *unreduced upper Hessenberg matrix*

$$\hat{H}_k \equiv \left[\begin{array}{cccc} \hat{h}_{11} & \cdots & \cdots & \hat{h}_{1k} \\ \hat{h}_{21} & \hat{h}_{22} & \cdots & \hat{h}_{2k} \\ & \ddots & \ddots & \vdots \\ & & \hat{h}_{k,k-1} & \hat{h}_{kk} \\ \hline & & & \hat{h}_{k+1,k} \end{array}\right] \quad \text{with} \quad \hat{h}_{i+1,i} \neq 0 \qquad (5)$$

such that

$$AU_k = U_{k+1}\hat{H}_k. \text{ (Arnoldi decomp.)} \qquad (6)$$

Conversely, if U_{k+1} *is orthonormal and satisfies (6), where* $\hat{H}_k$ *is defined in (5), then* U_{k+1} *is the* Q*-factor of* $K_{k+1}(A, u_1)$*.*

Proof: ("$\Rightarrow$") Let $K_k = U_k R_k$ be the QR factorization and $S_k = R_k^{-1}$. Then

$$AU_k = AK_kS_k = K_{k+1}\left[\begin{array}{c} 0 \\ S_k \end{array}\right] = U_{k+1}R_{k+1}\left[\begin{array}{c} 0 \\ S_k \end{array}\right] = U_{k+1}\hat{H}_k,$$

where

$$\hat{H}_k = R_{k+1}\left[\begin{array}{c} 0 \\ S_k \end{array}\right].$$

It implies that $\hat{H}_k$ is a $(k+1)\times k$ Hessenberg matrix and

$$h_{i+1,i} = r_{i+1,i+1}s_{ii} = \frac{r_{i+1,i+1}}{r_{ii}}.$$

Thus by the nonsingularity of R_k, $\hat{H}_k$ is unreduced.
("$\Leftarrow$") If $k=1$, then

$$Au_1 = h_{11}u_1 + h_{21}u_2 \quad \Rightarrow \quad u_2 = \frac{-h_{11}}{h_{21}}u_1 + \frac{1}{h_{21}}Au_1.$$

Since $\begin{bmatrix} u_1 & u_2 \end{bmatrix}$ is orthonormal and u_2 is a linear combination of u_1 and Au_1, $\begin{bmatrix} u_1 & u_2 \end{bmatrix}$ is the Q-factor of K_2.
Assume U_k is the Q-factor of K_k. If we partition

$$\hat{H}_k = \begin{bmatrix} \hat{H}_{k-1} & h_k \\ 0 & h_{k+1,k} \end{bmatrix},$$

then from (6)

$$Au_k = U_k h_k + h_{k+1,k} u_{k+1}.$$

Thus u_{k+1} is a linear combination of Au_k and the columns of U_k. Hence U_{k+1} is the Q-factor of K_{k+1}. □

Theorem

Let the orthonormal matrix U_{k+1} *satisfy*

$$AU_k = U_{k+1}\hat{H}_k,$$

where $\hat{H}_k$ *is Hessenberg. Then* $\hat{H}_k$ *is reduced if and only if* $\mathcal{R}(U_k)$ *contains an eigenspace of* A*.*

Proof: ("⇒") Suppose that $\hat{H}_k$ is reduced, say that $h_{j+1,j} = 0$. Partition

$$\hat{H}_k = \begin{bmatrix} H_{11} & H_{12} \\ 0 & H_{22} \end{bmatrix} \quad \text{and} \quad U_k = \begin{bmatrix} U_{11} & U_{12} \end{bmatrix},$$

where H_{11} is an $j \times j$ matrix and U_{11} is consisted the first j columns of U_{k+1}. Then

$$A\begin{bmatrix} U_{11} & U_{12} \end{bmatrix} = \begin{bmatrix} U_{11} & U_{12} & u_{k+1} \end{bmatrix} \begin{bmatrix} H_{11} & H_{12} \\ 0 & H_{22} \end{bmatrix}.$$

It implies that

$$AU_{11} = U_{11}H_{11}$$

so that U_{11} is an eigenbasis of A.
("$\Leftarrow$") Suppose that A has an eigenspace that is a subset of $\mathcal{R}(U_k)$ and $\hat{H}_k$ is unreduced. Let $(\lambda, U_k w)$ for some w be an eigenpair of A. Then

$$\begin{aligned} 0 &= (A - \lambda I)U_k w = (U_{k+1}\hat{H}_k - \lambda U_k)w \\ &= \left(U_{k+1}\hat{H}_k - \lambda U_{k+1} \begin{bmatrix} I \\ 0 \end{bmatrix} \right) w = U_{k+1}\hat{H}_\lambda w, \end{aligned}$$

where

$$\hat{H}_\lambda = \begin{bmatrix} H_k - \lambda I \\ h_{k+1,k} e_k^T \end{bmatrix}.$$

Since $\hat{H}_\lambda$ is unreduced, the matrix $U_{k+1}\hat{H}_\lambda$ is of full column rank. It follows that $w = 0$ which is a contradiction.

Partition $\hat{H}_k = \begin{bmatrix} H_k \\ \hat{h}_{k+1,k} e_k^\top \end{bmatrix}$, and set $\beta_k = \hat{h}_{k+1,k}$. Then (6) is equivalent to the Arnoldi decomp.

$$AU_k = U_k H_k + \beta_k u_{k+1} e_k^\top. \tag{7}$$

Write (7) in the form

$$Au_k = U_k h_k + \beta_k u_{k+1}.$$

Then from the orthogonality of U_{k+1}, we have

$$h_k = U_k^H A u_k.$$

Since $\beta_k u_{k+1} = Au_k - U_k h_k$ and $\|u_{k+1}\|_2 = 1$, we must have

$$\beta_k = \|Au_k - U_k h_k\|_2, \quad u_{k+1} = \beta_k^{-1}(Au_k - U_k h_k).$$

Algorithm (Arnoldi process)

1. *For* $k = 1, 2, \dots$
2. $h_k = U_k^H A u_k$.
3. $v = A u_k - U_k h_k$
4. $\beta_k = h_{k+1,k} = \|v\|_2$
5. $u_{k+1} = v/\beta_k$
6. $\hat{H}_k = \begin{bmatrix} \hat{H}_{k-1} & h_k \\ 0 & h_{k+1,k} \end{bmatrix}$
7. *end for* k

- The computation of u_{k+1} is actually a form of the well-known Gram-Schmidt algorithm.
- In the presence of inexact arithmetic cancelation in statement 3 can cause it to fail to produce orthogonal vectors.
- The cure is process called reorthogonalization.

Algorithm (Reorthogonalized Arnoldi process)

For $k = 1, 2, \ldots$

$h_k = U_k^H A u_k.$

$v = A u_k - U_k h_k.$

$w = U_k^H v.$

$h_k = h_k + w.$

$v = v - U_k w.$

$\beta_k = h_{k+1,k} = \|v\|_2$

$u_{k+1} = v / \beta_k$

$$\hat{H}_k = \begin{bmatrix} \hat{H}_{k-1} & h_k \\ 0 & h_{k+1,k} \end{bmatrix}$$

end for k

Let $y_i^{(k)}$ be an eigenvector of H_k associated with the Ew $\lambda_i^{(k)}$ and $x_i^{(k)} = U_k y_i^{(k)}$ the Ritz approximate eigenvector.

Theorem

$$(A - \lambda_i^{(k)} I) x_i^{(k)} = h_{k+1,k} e_k^T y_i^{(k)} u_{k+1}.$$

and therefore,

$$\|(A - \lambda_i^{(k)} I) x_i^{(k)}\|_2 = |h_{k+1,k}| \, | e_k^\top y_i^{(k)} |.$$

Lanczos Method

Let A be Hermitian and let

$$AU_k = U_k T_k + \beta_k u_{k+1} e_k^\top \qquad (8)$$

be an Arnoldi decomposition. Since T_k is upper Hessenberg and $T_k = U_k^H A U_k$ is Hermitian, it follows that T_k is tridiagonal and can be written in the form

$$T_k = \begin{bmatrix} \alpha_1 & \beta_1 & & & & \\ \beta_1 & \alpha_2 & \beta_2 & & & \\ & \beta_2 & \alpha_3 & \beta_3 & & \\ & & \ddots & \ddots & \ddots & \\ & & & \beta_{k-2} & \alpha_{k-1} & \beta_{k-1} \\ & & & & \beta_{k-1} & \alpha_k \end{bmatrix}.$$

Equation (8) is called a Lanczos decomposition.

The first column of (8) is

$$Au_1 = \alpha_1 u_1 + \beta_1 u_2 \quad \text{or} \quad u_2 = \frac{Au_1 - \alpha_1 u_1}{\beta_1}.$$

From the orthonormality of u_1 and u_2, it follows that

$$\alpha_1 = u_1^H A u_1$$

and

$$\beta_1 = \|Au_1 - \alpha_1 u_1\|_2.$$

More generality, from the j-th column of (8) we get the relation

$$u_{j+1} = \frac{Au_j - \alpha_j u_j - \bar{\beta}_{j-1} u_{j-1}}{\beta_j}$$

where

$$\alpha_j = u_j^H A u_j \quad \text{and} \quad \beta_j = \|Au_j - \alpha_j u_j - \bar{\beta}_{j-1} u_{j-1}\|_2.$$

This is the Lanczos three-term recurrence.

Algorithm (Lanczos recurrence)

Let u_1 be given. This algorithm generates the Lanczos decomposition

$$AU_k = U_kT_k + \beta_k u_{k+1} e_k^\top$$

where T_k is Hermitian tridiagonal.

1. $u_0 = 0; \beta_0 = 0;$
2. *for* $j = 1$ *to* k
3. $\quad u_{j+1} = Au_j$
4. $\quad \alpha_j = u_j^H u_{j+1}$
5. $\quad v = u_{j+1} - \alpha_j u_j - \beta_{j-1} u_{j-1}$
6. $\quad \beta_j = \|v\|_2$
7. $\quad u_{j+1} = v/\beta_j$
8. *end for* j

Theorem (Stop criterion)

Suppose that j steps of the Lanczos algorithm have been performed and that

$$S_j^H T_j S_j = diag(\theta_1, \cdots, \theta_j)$$

is the Schur decomposition of the tridiagonal matrix T_j, if $Y_j \in \mathbb{C}^{n\times j}$ is defined by

$$Y_j \equiv \begin{bmatrix} y_1 & \cdots & y_j \end{bmatrix} = U_j S_j$$

then for $i = 1, \cdots, j$ we have

$$\|Ay_i - \theta_i y_i\|_2 = |\beta_j||s_{ji}|$$

where $S_j = [s_{pq}]$.

Proof : Post-multiplying

$$AU_j = U_j T_j + \beta_j u_{j+1} e_j^\top$$

by S_j gives

$$AY_j = Y_j diag(\theta_1, \cdots, \theta_j) + \beta_j u_{j+1} e_j^T S_j,$$

i.e.,

$$Ay_i = \theta_i y_i + \beta_j u_{j+1} (e_j^T S_j e_i) \ , \quad i = 1, \cdots, j.$$

The proof is complete by taking norms. □

Remark

- *Stop criterion* = $|\beta_j||s_{ji}|$. *Do not need to compute* $\|Ay_i - \theta_i y_i\|_2$.
- *In general,* $|\beta_j|$ *is not small. It is possible that* $|\beta_j||s_{ji}|$ *is small.*

Theorem

Let A be $n \times n$ symmetric matrix with eigenvalues $\lambda_1 \geq \cdots \geq \lambda_n$ and corresponding orthonormal eigenvectors $z_1, \cdots, z_n$. If $\theta_1 \geq \cdots \geq \theta_j$ are the eigenvalues of T_j obtained after j steps of the Lanczos iteration, then

$$\lambda_1 \geq \theta_1 \geq \lambda_1 - \frac{(\lambda_1 - \lambda_n)(\tan \phi_1)^2}{[c_{j-1}(1 + 2\rho_1)]^2},$$

where $\cos \phi_1 = |u_1^\top z_1|$, c_{j-1} is a Chebychev polynomal of degree $j - 1$ and

$$\rho_1 = \frac{\lambda_1 - \lambda_2}{\lambda_2 - \lambda_n}.$$

Proof: From Courant-Fischer theorem we have

$$\theta_1 = \max_{y \neq 0} \frac{y^T T_j y}{y^T y} = \max_{y \neq 0} \frac{(U_j y)^T A (U_j y)}{(U_j y)^T (U_j y)} = \max_{0 \neq w \in \mathcal{K}(q_1, A, j)} \frac{w^T A w}{w^T w}.$$

Since λ_1 is the maximum of $w^T Aw/w^T w$ over all nonzero w, it follows that $\lambda_1 \geq \theta_1$. To obtain the lower bound for θ_1, note that

$$\theta_1 = \max_{p \in P_{j-1}} \frac{q_1^T p(A) A p(A) q_1}{q_1^T p(A)^2 q_1},$$

where P_{j-1} is the set of all $j-1$ degree polynomials. If $q_1 = \sum_{i=1}^n d_i z_i$, then

$$\begin{aligned} \frac{q_1^T p(A) A p(A) q_1}{q_1^T p(A)^2 q_1} &= \frac{\sum_{i=1}^n d_i^2 p(\lambda_i)^2 \lambda_i}{\sum_{i=1}^n d_i^2 p(\lambda_i)^2} \\ &\geq \lambda_1 - (\lambda_1 - \lambda_n) \frac{\sum_{i=2}^n d_i^2 p(\lambda_i)^2}{d_1^2 p(\lambda_1)^2 + \sum_{i=2}^n d_i^2 p(\lambda_i)^2}. \end{aligned}$$

We can make the lower bound tight by selecting a polynomial $p(x)$ that is large at $x = \lambda_1$ in comparison to its value at the remaining eigenvalues. Set

$$p(x) = c_{j-1}\left(-1 + 2\frac{x - \lambda_n}{\lambda_2 - \lambda_n}\right),$$

where $c_{j-1}(z)$ is the $(j-1)$-th Chebychev polynomial generated by

$$c_j(z) = 2zc_{j-1}(z) - c_{j-2}(z), \quad c_0 = 1, c_1 = z.$$

These polynomials are bounded by unity on [-1,1]. It follows that $|p(\lambda_i)|$ is bounded by unity for $i = 2, \cdots, n$ while $p(\lambda_1) = c_{j-1}(1 + 2\rho_1)$. Thus,

$$\theta_1 \geq \lambda_1 - (\lambda_1 - \lambda_n)\frac{(1 - d_1^2)}{d_1^2}\frac{1}{c_{j-1}^2(1 + 2\rho_1)}.$$

The desired lower bound is obtained by noting that $\tan(\phi_1)^2 = (1 - d_1^2)/d_1^2$.

Theorem

Using the same notation as Theorem 19,

$$\lambda_n \le \theta_j \le \lambda_n + \frac{(\lambda_1 - \lambda_n)\tan^2 \varphi_n}{[c_{j-1}(1 + 2\rho_n)]^2},$$

where

$$\rho_n = \frac{\lambda_{n-1} - \lambda_n}{\lambda_1 - \lambda_{n-1}}, \quad \cos \varphi_n = |u_1^\top z_n|.$$

Proof : Apply Theorem 19 with A replaced by $-A$. □

- Rounding errors greatly affect the behavior of the Lanczos iteration.
- The basic difficulty is caused by loss of orthogonality among the Lanczos vectors.
- To avoid these difficulties we can reorthogonalize the Lanczos vectors.

For details of complete reorthogonalization and selective reorthogonalization see the books:

- Parlett: "Symmetric Eigenvalue problem" (1980) pp.257–
- Golub & Van Loan: "Matrix computation" (1981) pp.332–

Generalized eigenvalue problem

Consider the generalized eigenvalue problem

$$Ax = \lambda Bx,$$

where B is symmetric positive definite. Let

$$C = B^{-1}A.$$

Applying Arnoldi process to matrix C, we get

$$CU_k = U_k H_k + \beta_k u_{k+1} e_k^\top,$$

or

$$AU_k = BU_k H_k + \beta_k B u_{k+1} e_k^\top. \tag{9}$$

Write the k-th column of (9) in the form

$$Au_k = BU_k h_k + \beta_k B u_{k+1}. \tag{10}$$

Let U_k satisfy that

$$U_k^\top B U_k = I_k.$$

Then

$$h_k = U_k^\top A u_k$$

and

$$\beta_k B u_{k+1} = Au_k - BU_k h_k \equiv t_k \;\Rightarrow\; \beta_k u_{k+1} = B^{-1} t_k$$

Since u_{k+1} satisfies $u_{k+1}^\top B u_{k+1} = 1$, it implies that

$$\beta_k^2 = (\beta_k u_{k+1})^\top B(\beta_k u_{k+1}) = t_k^\top B^{-1} t_k,$$

which implies that

$$\beta_k = \sqrt{t_k^\top B^{-1} t_k}$$

and

$$u_{k+1} = \beta_k^{-1} B^{-1} t_k.$$

Algorithm (Arnoldi process for GEP)

1. For $k = 1, 2, \ldots$
2. $\quad h_k = U_k^\top A u_k$.
3. $\quad t = Au_k - BU_k h_k$
4. $\quad$ Solve linear system $Bv = t$
4. $\quad \beta_k = h_{k+1,k} = \sqrt{t^\top v}$
5. $\quad u_{k+1} = v/\beta_k$
6. $\quad H_k = \begin{bmatrix} H_{k-1} & h_k \\ 0 & h_{k+1,k} \end{bmatrix}$
7. end for k

Shift-and-invert Lanczos for GEP

Consider the generalized eigenvalue problem

$$Ax = \lambda Bx, \tag{11}$$

where A is symmetric and B is symmetric positive definite.

Shift-and-invert

- Compute the eigenvalues which are closest to a given shift value σ.
- Transform (11) into

$$(A - \sigma B)^{-1} Bx = (\lambda - \sigma)^{-1} x. \tag{12}$$

The basic recursion for applying Lanczos method to (12) is

$$(A - \sigma B)^{-1} B V_j = V_j T_j + \beta_j v_{j+1} e_j^\top, \tag{13}$$

where the basis V_j is B-orthogonal and T_j is a real symmetric tridiagonal matrix defined by

$$T_j = \begin{bmatrix} \alpha_1 & \beta_1 & & \\ \beta_1 & \alpha_2 & \ddots & \\ & \ddots & \ddots & \beta_{j-1} \\ & & \beta_{j-1} & \alpha_j \end{bmatrix}$$

or equivalent to

$$(A - \sigma B)^{-1} B v_j = \alpha_j v_j + \beta_{j-1} v_{j-1} + \beta_j v_{j+1}.$$

By the condition $V_j^* B V_j = I_j$, it holds that

$$\alpha_j = v_j^* B (A - \sigma B)^{-1} B v_j, \quad \beta_j^2 = t_j^* B t_j,$$

where

$$t_j \equiv (A - \sigma B)^{-1} B v_j - \alpha_j v_j - \beta_{j-1} v_{j-1} = \beta_j v_{j+1}.$$

An eigenpair (θ_k, s_k) of T_j is used to get an approximate eigenpair (λ_k, x_k) of (A, B) by

$$\lambda_k = \sigma + \frac{1}{\theta_k}, \quad x_k = V_j s_k.$$

The corresponding residual is

$$\begin{aligned} r_k &= A V_j s_k - \lambda_k B V_j s_k = (A - \sigma B) V_j s_k - \theta_k^{-1} B V_j s_k \\ &= -\theta_k^{-1} \left[B V_j - (A - \sigma B) V_j T_j \right] s_k \\ &= -\theta_k^{-1} (A - \sigma B) \left[(A - \sigma B)^{-1} B V_j - V_j T_j \right] \\ &= -\theta_k^{-1} \beta_j (e_j^\top s_k) (A - \sigma B) v_{j+1} \end{aligned}$$

which implies $\|r_k\|$ is small whenever $|\beta_j (e_j^\top s_k)/\theta_k|$ is small.

Shift-and-invert Lanczos method for symmetric GEP

```
Given starting vector $t$, compute $q = Bt$ and $\beta_0 = \sqrt{|q^* t|}$.
for $j = 1, 2, \ldots,$ do
    Compute $w_j = q/\beta_{j-1}$ and $v_j = t/\beta_{j-1}$.
    Solve linear system $(A - \sigma B)t = w_j$.
    Set $t := t - \beta_{j-1} v_{j-1}$; compute $\alpha_j = w_j^* t$ and reset
    $t := t - \alpha_j v_j$.
    B-reorthogonalize $t$ to $v_1, \ldots, v_j$ if necessary.
    Compute $q = Bt$ and $\beta_j = \sqrt{|q^* t|}$.
    Compute approximate Ews $T_j = S_j \Theta_j S_j^*$.
    Test for convergence.
end for
Compute approximate Evs $X = V_j S_j$.
```

Restarting method

Let

$$AU_k = U_k H_k + \beta_k u_{k+1} e_k^\top$$

be an Arnoldi decomposition.

- In principle, we can keep expanding the Arnoldi decomposition until the Ritz pairs have converged.
- Unfortunately, it is limited by the amount of memory to storage of U_k.
- Restarted the Arnoldi process once k becomes so large that we cannot store U_k.
  - Implicitly restarting method
  - Krylov-Schur restarting method

Implicitly restarting method

- Let
$$AU_m = U_m H_m + \beta_m u_{m+1} e_m^\top \tag{14}$$
be an Arnoldi decomposition with order m.
- Let $\kappa_1, \ldots, \kappa_m$ be eigenvalues of H_m and suppose that $\kappa_1, \ldots, \kappa_{m-k}$ correspond to the part of the spectrum we are not interested in.
- From (14), we have
$$\begin{aligned}(A-\kappa_1 I)U_m &= U_m(H_m - \kappa_1 I) + \beta_m u_{m+1} e_m^\top \\ &= U_m Q_1 R_1 + \beta_m u_{m+1} e_m^\top,\end{aligned}$$
where
$$H_m - \kappa_1 I = Q_1 R_1$$
is the QR factorization of $H_m - \kappa_1 I$.

Postmultiplying by Q_1, we get

$$(A - \kappa_1 I)(U_m Q_1) = (U_m Q_1)(R_1 Q_1) + \beta_m u_{m+1}(e_m^\top Q_1).$$

It implies that

$$AU_m^{(1)} = U_m^{(1)} H_m^{(1)} + \beta_m u_{m+1} b_{m+1}^{(1)T},$$

where

$$U_m^{(1)} = U_m Q_1, \quad H_m^{(1)} = R_1 Q_1 + \kappa_1 I, \quad b_{m+1}^{(1)T} = e_m^\top Q_1.$$

Remark

- *$U_m^{(1)}$ is orthonormal.*
- *Since H_m is upper Hessenberg and Q_1 is the Q-factor of the QR factorization of $H_m - \kappa_1 I$, it implies that Q_1 and $H_m^{(1)}$ are also upper Hessenberg.*

Remark (continue)

- *The vector* $b_{m+1}^{(1)H} = e_m^\top Q_1$ *has the form*

$$b_{m+1}^{(1)H} = \left[\begin{array}{ccccc} 0 & \cdots & 0 & q_{m-1,m}^{(1)} & q_{m,m}^{(1)} \end{array} \right];$$

i.e., only the last two components of $b_{m+1}^{(1)}$ *are nonzero.*

- *By the definition of* $H_m^{(1)}$, *we get*

$$Q_1 H_m^{(1)} Q_1^H = Q_1(R_1 Q_1 + \kappa_1 I) Q_1^H = Q_1 R_1 + \kappa_1 I = H_m.$$

Therefore, $\kappa_1, \kappa_2, \ldots, \kappa_m$ *are also eigenvalues of* $H_m^{(1)}$.

Repeating this process with $\kappa_2, \kappa_3, \ldots, \kappa_{m-k}$, the result will be a Krylov decomposition

$$AU_m^{(m-k)} = U_m^{(m-k)} H_m^{(m-k)} + \beta_m u_{m+1} b_{m+1}^{(m-k)H}$$

with the following properties

1. $U_m^{(m-k)}$ is orthonormal.
2. $H_m^{(m-k)}$ is upper Hessenberg.
3. The first $k-1$ components of $b_{m+1}^{(m-k)H}$ are zero.
4. The first column of $U_m^{(m-k)}$ is a multiple of $(A - \kappa_1 I) \cdots (A - \kappa_{m-k} I) u_1$.

Corollary

Let $\kappa_1, \dots, \kappa_m$ *be eigenvalues of* H_m*. If the implicitly restarted* QR *step is performed with shifts* $\kappa_1, \dots, \kappa_{m-k}$*, then the matrix* $H_m^{(m-k)}$ *has the form*

$$H_m^{(m-k)} = \begin{bmatrix} H_{kk}^{(m-k)} & H_{k,m-k}^{(m-k)} \\ 0 & T^{(m-k)} \end{bmatrix},$$

where $T^{(m-k)}$ *is an upper triangular matrix with Ritz value* $\kappa_1, \dots, \kappa_{m-k}$ *on its diagonal.*

For $k = 3$ and $m = 6$,

$$
\begin{aligned}
& A \left[\begin{array}{ccc|ccc} u & u & u & u & u & u \end{array}\right] \\
= \; & \left[\begin{array}{ccc|ccc} u & u & u & u & u & u \end{array}\right] \left[\begin{array}{ccc|ccc} \times & \times & \times & \times & \times & \times \\ \times & \times & \times & \times & \times & \times \\ 0 & \times & \times & \times & \times & \times \\ \hline 0 & 0 & \times & \times & \times & \times \\ 0 & 0 & 0 & \times & \times & \times \\ 0 & 0 & 0 & 0 & \times & \times \end{array}\right] \\
& + u \left[\begin{array}{ccc|ccc} 0 & 0 & q & q & q & q \end{array}\right].
\end{aligned}
$$

Therefore, the first k columns of the decomposition can be written in the form

$$
AU_k^{(m-k)} = U_k^{(m-k)} H_{kk}^{(m-k)} + h_{k+1,k} u_{k+1}^{(m-k)} e_k^\top + \beta_k q_{mk} u_{m+1} e_k^\top,
$$

where $U_k^{(m-k)}$ consists of the first k columns of $U_m^{(m-k)}$, $H_{kk}^{(m-k)}$ is the leading principal submatrix of order k of $H_m^{(m-k)}$, and q_{km} is from the matrix $Q = Q_1 \cdots Q_{m-k}$. Hence if we set

$$
\begin{aligned}
\tilde{U}_k &= U_k^{(m-k)}, \\
\tilde{H}_k &= H_{kk}^{(m-k)}, \\
\tilde{\beta}_k &= \|h_{k+1,k} u_{k+1}^{(m-k)} + \beta_k q_{mk} u_{m+1}\|_2, \\
\tilde{u}_{k+1} &= \tilde{\beta}_k^{-1} (h_{k+1,k} u_{k+1}^{(m-k)} + \beta_k q_{mk} u_{m+1}),
\end{aligned}
$$

then

$$
A\tilde{U}_k = \tilde{U}_k \tilde{H}_k + \tilde{\beta}_k \tilde{u}_{k+1} e_k^\top
$$

is an Arnoldi decomposition whose starting vector is proportional to $(A - \kappa_1 I) \cdots (A - \kappa_{m-k} I) u_1$.

Krylov-Schur restarting method (Locking and Purging)

Use Arnoldi process to generate the Arnoldi decomposition of order $j+p$

$$AV_{j+p} = V_{j+p}H_{j+p} + \beta_{j+p}v_{j+p+1}e_{j+p}^T. \tag{15}$$

Let

$$H_{j+p} = U_{j+p}R_{j+p}U_{j+p}^\top \equiv \begin{bmatrix} U_j & U_p \end{bmatrix} \begin{bmatrix} R_j & \star \\ 0 & R_p \end{bmatrix} \begin{bmatrix} U_j^\top \\ U_p^\top \end{bmatrix} \tag{16}$$

be a Schur decomposition of H_{j+p} where the diagonal elements of R_j and R_p are the j wanted and p unwanted Ritz values, respectively. Substituting (16) into (15), it holds that

$$\begin{aligned} & A(V_{j+p}U_{j+p}) \\ = \; & (V_{j+p}U_{j+p})(U_{j+p}^\top H_{j+p}U_{j+p}) + \beta_{j+p}v_{j+p+1}(e_{j+p}^T U_{j+p}) \end{aligned}$$

which implies that

$$A\tilde{V}_j = \tilde{V}_j R_j + \beta_{j+p} v_{j+p+1} t_j^T$$

is a Krylov decomposition of order j where $\tilde{V}_j \equiv V_{j+p} U_j$ and $e_{j+p}^T U_{j+p} \equiv \left[t_j^T, t_p^T \right]$. Let H_1 be a Householder transformation such that

$$t_j^T H_1 = \beta e_j^\top.$$

Reduce $H_1^T R_j H_1$ to Hessenberg form by using Householder transformations H_i for $i = 2, \ldots, j-1$ as follows:

$$SQ_j := H_1^\top R_j H_1 = \begin{bmatrix} \times & \times & \times & \times \\ \times & \times & \times & \times \\ \times & \times & \times & \times \\ \times & \times & \times & \times \end{bmatrix}$$

$$\Rightarrow \quad S_j := H_2^\top S_j H_2 = \begin{bmatrix} \times & \times & \times & \times \\ \times & \times & \times & \times \\ \times & \times & \times & \times \\ 0 & 0 & \times & \times \end{bmatrix}$$

$$\Rightarrow \quad S_j := H_3^\top S_j H_3 = \begin{bmatrix} \times & \times & \times & \times \\ \times & \times & \times & \times \\ 0 & \times & \times & \times \\ 0 & 0 & \times & \times \end{bmatrix}$$

Let

$$Q = H_1 H_2 \cdots H_{j-1}.$$

Then $\tilde{H}_j \equiv Q^T R_j Q$ is Hessenberg and

$$t_j^T Q = (t_j^T H_1)(H_2 \cdots H_{j-1}) = \beta e_j^\top (H_2 \cdots H_{j-1}) = \beta e_j^\top.$$

Therefore, the Krylov decomposition

$$A(\tilde{V}_j Q) = (\tilde{V}_j Q)\tilde{H}_j + (\beta_{j+p}\beta) v_{j+p+1} e_j^\top$$

is a Arnoldi decomposition of order j and we can use it to generate a new Arnoldi decomposition of order $j+p$ if the j eigenpairs of A do not converge.