Krylov sequence methods

Tsung-Ming Huang

Department of Mathematics
National Taiwan Normal University, Taiwan

March 22, 2009

Outline

1. Krylov decompositions
   - Householder transformation
   - Arnoldi decompositions
   - Lanczos decompositions
   - Krylov decompositions
   - Computation of refined and Harmonic Ritz vectors
2. Restarted Arnoldi method
   - The implicitly restarted Arnoldi method
   - Krylov-Schur restarting
3. The Lanczos Algorithm

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. 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 $Q_1 a_1^{(0)} = c e_1$. Then

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

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 $A = Q_1 R_1 = Q_2 R_2$, then there is a unitary diagonal matrix $D = diag(d_i)$ with $|d_i| = 1$ such that $Q_1 = Q_2 D$ and $DR_1 = R_2$.

Proof: Let $A = Q_1 R_1 = Q_2 R_2$. Then $Q_2^* Q_1 = R_2 R_1^{-1} = D$ must be a diagonal unitary matrix. ■

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 independent. Let $U_{k+1} = \begin{bmatrix} u_1 & \cdots & u_{k+1} \end{bmatrix}$ be the Q-factor of K_{k+1}. Then there is a $(k+1) \times k$ unreduced upper Hessenberg matrix $\hat{H}_k$ such that

$$AU_k = U_{k+1} \hat{H}_k. \tag{2}$$

Conversely, if U_{k+1} is orthonormal and satisfies (2), where $\hat{H}_k$ is a $(k+1) \times k$ unreduced upper Hessenberg matrix, 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

$$AU_k = AK_k S_k = K_{k+1} \begin{bmatrix} 0 \\ S_k \end{bmatrix} = U_{k+1} R_{k+1} \begin{bmatrix} 0 \\ S_k \end{bmatrix} = U_{k+1} \hat{H}_k,$$

where

$$\hat{H}_k = R_{k+1} \begin{bmatrix} 0 \\ S_k \end{bmatrix}.$$

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

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

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 (2)

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

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}. ■

Definition

Let $U_{k+1} \in \mathbb{C}^{n\times(k+1)}$ be orthonormal. If there is a $(k+1)\times k$ unreduced upper Hessenberg matrix $\hat{H}_k$ such that

$$AU_k = U_{k+1}\hat{H}_k, \tag{3}$$

then (3) is called an Arnoldi decomposition of order k. If $\hat{H}_k$ is reduced, we say the Arnoldi decomposition is reduced.

Partition

$$\hat{H}_k = \begin{bmatrix} H_k \\ h_{k+1,k}e_k^T \end{bmatrix},$$

and set

$$\beta_k = h_{k+1,k}.$$

Then (3) is equivalent to

$$AU_k = U_kH_k + \beta_k u_{k+1}e_k^T.$$

Theorem

Suppose the Krylov sequence $K_{k+1}(A, u_1)$ does not terminate at $k+1$. Then up to scaling of the columns of U_{k+1}, the Arnoldi decomposition of K_{k+1} is unique.

Proof: Since the Krylov sequence $K_{k+1}(A, u_1)$ does not terminate at $k+1$, the columns of $K_{k+1}(A, u_1)$ are linearly independent. By Theorem 8, there is an unreduced matrix H_k and $\beta_k \neq 0$ such that

$$AU_k = U_k H_k + \beta_k u_{k+1} e_k^T, \tag{4}$$

where $U_{k+1} = [U_k \ u_{k+1}]$ is an orthonormal basis for $\mathcal{K}_{k+1}(A, u_1)$. Suppose there is another orthonormal basis $\tilde{U}_{k+1} = [\tilde{U}_k \ \tilde{u}_{k+1}]$ for $\mathcal{K}_{k+1}(A, u_1)$, unreduced matrix $\tilde{H}_k$ and $\tilde{\beta}_k \neq 0$ such that

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

Then we claim that

$$\tilde{U}_k^H u_{k+1} = 0.$$

For otherwise there is a column $\tilde{u}_j$ of $\tilde{U}_k$ such that

$$\tilde{u}_j = \alpha u_{k+1} + U_k a, \quad \alpha \neq 0.$$

Hence

$$A\tilde{u}_j = \alpha A u_{k+1} + AU_k a$$

which implies that $A\tilde{u}_j$ contains a component along $A^{k+1}u_1$. Since the Krylov sequence $K_{k+1}(A, u_1)$ does not terminate at $k+1$, we have

$$\mathcal{K}_{k+2}(A, u_1) \neq \mathcal{K}_{k+1}(A, u_1).$$

Therefore, $A\tilde{u}_j$ lies in $\mathcal{K}_{k+2}(A, u_1)$ but not in $\mathcal{K}_{k+1}(A, u_1)$ which is a contradiction.

Since U_{k+1} and $\tilde{U}_{k+1}$ are orthonormal bases for $\mathcal{K}_{k+1}(A, u_1)$ and $\tilde{U}_k^H u_{k+1} = 0$, it follows that

$$\mathcal{R}(U_k) = \mathcal{R}(\tilde{U}_k) \quad \text{and} \quad U_k^H \tilde{u}_{k+1} = 0,$$

that is

$$U_k = \tilde{U}_k Q$$

for some unitary matrix Q. Hence

$$A(\tilde{U}_k Q) = (\tilde{U}_k Q)(Q^H \tilde{H}_k Q) + \tilde{\beta}_k \tilde{u}_{k+1}(e_k^T Q),$$

or

$$AU_k = U_k(Q^H \tilde{H}_k Q) + \tilde{\beta}_k \tilde{u}_{k+1} e_k^T Q. \tag{5}$$

On premultiplying (4) and (5) by U_k^H, we obtain

$$H_k = U_k^H A U_k = Q^H \tilde{H}_k Q.$$

Similarly, premultiplying by u_{k+1}^H, we obtain

$$\beta_k e_k^T = u_{k+1}^H A U_k = \tilde{\beta}_k (u_{k+1}^H \tilde{u}_{k+1}) e_k^T Q.$$

It follows that the last row of Q is $\omega_k e_k^T$, where $|\omega_k| = 1$. Since the norm of the last column of Q is one, the last column of Q is $\omega_k e_k$. Since H_k is unreduced, it follows from the implicit Q theorem that

$$Q = \text{diag}(\omega_1, \cdots, \omega_k), \quad |\omega_j| = 1, \ j = 1, \ldots, k.$$

Thus up to column scaling $U_k = \tilde{U}_k Q$ is the same as $\tilde{U}_k$. Subtracting (5) from (4), we find that

$$\beta_k u_{k+1} = \omega_k \tilde{\beta}_k \tilde{u}_{k+1}$$

so that up to scaling u_{k+1} and $\tilde{u}_{k+1}$ are the same. ∎

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: ("$\Rightarrow$") 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. ∎

Write the k-th column of the Arnoldi decomposition

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

in the form

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

Then from the orthonormality 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

$$\beta_k = \|Au_k - U_k h_k\|_2$$

and

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

Algorithm (Arnoldi process)

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

$$AU_k = U_k T_k + \beta_k u_{k+1} e_k^T \tag{6}$$

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

$$T_k = \begin{bmatrix} \alpha_1 & \bar{\beta}_1 & & & & \\ \beta_1 & \alpha_2 & \bar{\beta}_2 & & & \\ & \beta_2 & \alpha_3 & \bar{\beta}_3 & & \\ & & \ddots & \ddots & \ddots & \\ & & & \beta_{k-2} & \alpha_{k-1} & \bar{\beta}_{k-1} \\ & & & & \beta_{k-1} & \alpha_k \end{bmatrix}.$$

Equation (6) is called a Lanczos decomposition. The first column of (6) is

$$Au_1 = \alpha_1 u_1 + \beta_1 u_2,$$

or

$$u_2 = \frac{Au_1 - \alpha_1 u_1}{\beta_1}.$$

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 (6) 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

$$AU_k = U_k T_k + \beta_k u_{k+1} e_k^T$$

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

Definition

Let $u_1, u_2, \ldots, u_{k+1}$ be linearly independent and let $U_k = [u_1 \ \cdots \ u_k]$.

$$AU_k = U_k B_k + u_{k+1} b_{k+1}^H$$

is called a Krylov decomposition of order k. $\mathcal{R}(U_{k+1})$ is called the space spanned by the decomposition. Two Krylov decompositions spanning the same spaces are said to be equivalent.

Let $[V \ v]^H$ be any left inverse for U_{k+1}. Then it follows that

$$B_k = V^H AU_k \quad \text{and} \quad b_{k+1}^H = v^H AU_k.$$

In particular, B_k is a Rayleigh quotient of A.

Let

$$AU_k = U_k B_k + u_{k+1} b_{k+1}^H$$

be a Krylov decomposition and Q be nonsingular. That is

$$AU_k = U_{k+1}\hat{B}_k \quad \text{with} \quad \hat{B}_k = \begin{bmatrix} B_k \\ b_{k+1}^H \end{bmatrix}. \tag{7}$$

Then we get an equivalent Krylov decomposition of (7) in the form

$$\begin{aligned} A(U_k Q) &= \left(U_{k+1} \begin{bmatrix} Q & 0 \\ 0 & 1 \end{bmatrix} \right) \left(\begin{bmatrix} Q & 0 \\ 0 & 1 \end{bmatrix}^{-1} \hat{B}_k Q \right) \\ &= \begin{bmatrix} U_k Q & u_{k+1} \end{bmatrix} \begin{bmatrix} Q^{-1} B_k Q \\ b_{k+1}^H Q \end{bmatrix} \\ &= (U_k Q)(Q^{-1} B Q) + u_{k+1}(b_{k+1}^H Q). \end{aligned} \tag{8}$$

The two Krylov decompositions (7) and (8) are said to be similar.

Let

$$\gamma \tilde{u}_{k+1} = u_{k+1} - U_k a.$$

Since $u_1, \ldots, u_k, u_{k+1}$ are linearly independent, we have $\gamma \neq 0$. Then it follows that

$$AU_k = U_k(B_k + ab_{k+1}^H) + \tilde{u}_{k+1}(\gamma b_{k+1}^H).$$

Since $\mathcal{R}([U_k \; u_{k+1}]) = \mathcal{R}([U_k \; \tilde{u}_{k+1}])$, this Krylov decomposition is equivalent to (7).

Theorem

Every Krylov decomposition is equivalent to a (possibly reduced) Arnoldi decomposition.

Proof: Let

$$AU = UB + ub^H$$

be a Krylov decomposition and let

$$U = \tilde{U}R$$

be the QR factorization of U. Then

$$A\tilde{U} = A(UR^{-1}) = (UR^{-1})(RBR^{-1}) + u(b^H R^{-1}) \equiv \tilde{U}\tilde{B} + u\tilde{b}^H$$

is an equivalent decomposition. Let

$$\tilde{u} = \gamma^{-1}(u - Ua)$$

be a vector with $\|\tilde{u}\|_2 = 1$ such that $U^H\tilde{u} = 0$. Then

$$A\tilde{U} = \tilde{U}(\tilde{B} + a\tilde{b}^H) + \tilde{u}(\gamma\tilde{b}^H) \equiv \tilde{U}\hat{B} + \tilde{u}\hat{b}^H$$

is an equivalent orthonormal Krylov decomposition. Let Q be a unitary matrix such that

$$\hat{b}^H Q = \|\hat{b}\|_2 e_k^T$$

and $Q^H\hat{B}Q$ is upper Hessenberg. Then the equivalent decomposition

$$A\hat{U} \equiv A(\tilde{U}Q) = (\tilde{U}Q)(Q^H\hat{B}Q) + \tilde{u}(\hat{b}^HQ) \equiv \hat{U}\bar{B} + \|\hat{b}\|_2\hat{u}e_k^T$$

is a possibly reduced Arnoldi decomposition where

$$\hat{U}^H\hat{u} = Q^H\tilde{U}^H\tilde{u} = Q^HR^{-H}U^H\tilde{u} = 0.$$

■

Reduction to Arnoldi form

Let

$$AU = UB + ub^H$$

be the Krylov decomposition with $B \in \mathbb{C}^{k\times k}$. Let H_1 be a Householder transformation such that

$$b^HH_1 = \beta e_k.$$

Reduce $H_1^H B H_1$ to Hessenberg form as the following illustration:

$$B := \begin{bmatrix} \times & \times & \times & \times \\ \times & \times & \times & \times \\ \times & \times & \times & \times \\ \times & \times & \times & \times \end{bmatrix} \Rightarrow B := BH_2 = \begin{bmatrix} \otimes & \otimes & \otimes & \times \\ \otimes & \otimes & \otimes & \times \\ \otimes & \otimes & \otimes & \times \\ 0 & 0 & \otimes & \times \end{bmatrix}$$

$$\Rightarrow \quad B := H_2^H B = \begin{bmatrix} + & + & + & + \\ + & + & + & + \\ + & + & + & + \\ 0 & 0 & \otimes & \times \end{bmatrix} \Rightarrow B := BH_3 = \begin{bmatrix} \oplus & \oplus & + & + \\ \oplus & \oplus & + & + \\ 0 & \oplus & + & + \\ 0 & 0 & \otimes & \times \end{bmatrix}$$

$$\Rightarrow \quad B := H_3^H B = \begin{bmatrix} * & * & * & * \\ * & * & * & * \\ 0 & \oplus & + & + \\ 0 & 0 & \otimes & \times \end{bmatrix}$$

Let

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

Then $Q^H BQ$ is upper Hessenberg and

$$b^H Q = (b^H H_1)(H_2 \cdots H_{k-1}) = \beta e_k^T (H_2 \cdots H_{k-1}) = \beta e_k^T.$$

Therefore, the Krylov decomposition

$$A(UQ) = (UQ)(Q^H BQ) + \beta u e_k^T \tag{9}$$

is an Arnoldi decomposition.

Assume that

$$AU = UB + ub^H$$

is a n orthonormal Krylov decomposition.

Refined Ritz vectors

If μ is a Ritz value, then the refined Ritz vector associated with μ is the right singular vector of $(A - \mu I)U$ whose singular value is smallest. From (9), we have

$$\begin{aligned}(A-\mu I)U &= U(B-\mu I) + ub^H = \begin{bmatrix} U & u \end{bmatrix} \begin{bmatrix} B - \mu I \\ b^H \end{bmatrix} \\ &\equiv \begin{bmatrix} U & u \end{bmatrix} \hat{B}_\mu.\end{aligned}$$

Since $[U \;\; u]$ is orthonormal, the right singular vectors of $(A - \mu I)U$ are the same as the right singular vectors of $\hat{B}_\mu$. Thus the computation of a refined Ritz vector can be reduced to computing the singular value decomposition of $\hat{B}_\mu$.

Harmonic Ritz vectors

Recall: $(\kappa + \delta, Uw)$ is a harmonic Ritz pair if

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

Since

$$(A-\kappa I)U = U(B-\kappa I) + ub^H,$$

we have

$$U^H(A-\kappa I)^H(A-\kappa I)U = (B-\kappa I)^H(B-\kappa I) + bb^H$$

and

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

It follows that

$$\left[(B-\kappa I)^H(B-\kappa I) + bb^H\right] w = \delta(B-\kappa I)^H w$$

which is a small generalized eigenvalue problem.

Let

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

be an Arnoldi decomposition.

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

- Choose a new starting vector for the underlying Krylov sequence
- A natural choice would be a linear combination of Ritz vectors that we are interested in.

Filter polynomials

Assume A has a complete system of eigenpairs (λ_i, x_i) and we are interested in the first k of these eigenpairs. Expand u_1 in the form

$$u_1 = \sum_{i=1}^{k} \gamma_i x_i + \sum_{i=k+1}^{n} \gamma_i x_i.$$

If p is any polynomial, we have

$$p(A)u_1 = \sum_{i=1}^{k} \gamma_i p(\lambda_i) x_i + \sum_{i=k+1}^{n} \gamma_i p(\lambda_i) x_i.$$

- Choose p so that the values $p(\lambda_i)$ $(i = k+1, \ldots, n)$ are small compared to the values $p(\lambda_i)$ $(i = 1, \ldots, k)$.
- Then $p(A)u_1$ is rich in the components of the x_i that we want and deficient in the ones that we do not want.
- p is called a filter polynomial.
- Suppose we have Ritz values $\mu_1, \ldots, \mu_m$ and $\mu_{k+1}, \ldots, \mu_m$ are not interesting. Then take

$$p(t) = (t - \mu_{k+1}) \cdots (t - \mu_m).$$

Implicitly restarted Arnoldi: Let

$$AU_m = U_m H_m + \beta_m u_{m+1} e_m^T \tag{10}$$

be an Arnoldi decomposition with order m. Choose a filter polynomial p of degree $m-k$ and use the implicit restarting process to reduce the decomposition to a decomposition

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

of order k with starting vector $p(A)u_1$.

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. Then take

$$p(t) = (t - \kappa_1)(t - \kappa_2) \cdots (t - \kappa_{m-k}).$$

The starting vector $p(A)u_1$ is equal to

$$\begin{aligned} p(A)u_1 &= (A - \kappa_{m-k}I) \cdots (A - \kappa_2 I)(A - \kappa_1 I)u_1 \\ &= (A - \kappa_{m-k}I) \left[\cdots \left[(A - \kappa_2 I) \left[(A - \kappa_1 I)u_1 \right] \right] \right]. \end{aligned}$$

In the first, we construct an Arnoldi decomposition with starting vector $(A - \kappa_1 I)u_1$. From (10), we have

$$\begin{aligned} (A - \kappa_1 I)U_m &= U_m(H_m - \kappa_1 I) + \beta_m u_{m+1} e_m^T \\ &= U_m Q_1 R_1 + \beta_m u_{m+1} e_m^T, \end{aligned} \qquad (11)$$

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

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

It implies that

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

where

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

($H_m^{(1)}$: one step of single shifted QR algorithm)

Theorem

Let H_m be an unreduced Hessenberg matrix. Then $H_m^{(1)}$ has the form

$$H_m^{(1)} = \begin{bmatrix} \hat{H}_m^{(1)} & \hat{h}_{12} \\ 0 & \kappa_1 \end{bmatrix},$$

where $\hat{H}_m^{(1)}$ is unreduced.

Proof: Let

$$H_m - \kappa_1 I = Q_1 R_1$$

be the QR factorization of $H_m - \kappa_1 I$ with

$$Q_1 = G(1, 2, \theta_1) \cdots G(m-1, m, \theta_{m-1})$$

where $G(i, i+1, \theta_i)$ for $i = 1, \ldots, m-1$ are Given rotations.

Since H_m is unreduced upper Hessenberg, i.e., the subdiagonal elements of H_m are nonzero, we get

$$\theta_i \neq 0 \quad \text{for} \;\; i = 1, \ldots, m-1 \tag{12}$$

and

$$(R_1)_{ii} \neq 0 \quad \text{for} \;\; i = 1, \ldots, m-1. \tag{13}$$

Since κ_1 is an eigenvalue of H_m, we have that $H_m - \kappa_1 I$ is singular and then

$$(R_1)_{mm} = 0. \tag{14}$$

Using the results of (12), (13) and (14), we get

$$\begin{aligned} H_m^{(1)} &= R_1 Q_1 + \kappa_1 I = R_1 G(1,2,\theta_1) \cdots G(m-1, m, \theta_{m-1}) + \kappa_1 I \\ &= \begin{bmatrix} \hat{H}_m^{(1)} & \hat{h}_{12} \\ 0 & \kappa_1 \end{bmatrix}, \end{aligned}$$

where $\hat{H}_m^{(1)}$ is unreduced.

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

$$b_{m+1}^{(1)H} = \begin{bmatrix} 0 & \cdots & 0 & q_{m-1,m}^{(1)} & q_{m,m}^{(1)} \end{bmatrix};$$

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

- For on postmultiplying (11) by e_1, we get

$$(A - \kappa_1 I)u_1 = (A - \kappa_1 I)(U_m e_1) = U_m^{(1)} R_1 e_1 = r_{11}^{(1)} u_1^{(1)}.$$

  Since H_m is unreduced, $r_{11}^{(1)}$ is nonzero. Therefore, the first column of $U_m^{(1)}$ is a multiple of $(A - \kappa_1 I)u_1$.
- 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 $H_m^{(1)}$.

Similarly,

$$\begin{aligned}(A-\kappa_2 I)U_m^{(1)} &= U_m^{(1)}(H_m^{(1)}-\kappa_2 I)+\beta_m u_{m+1}b_{m+1}^{(1)H} \\ &= U_m^{(1)}Q_2R_2+\beta_m u_{m+1}b_{m+1}^{(1)H},\end{aligned} \qquad (15)$$

where

$$H_m^{(1)}-\kappa_2 I = Q_2R_2$$

is the QR factorization of $H_m^{(1)}-\kappa_2 I$ with upper Hessenberg matrix Q_2. Postmultiplying by Q_2, we get

$$(A-\kappa_2 I)(U_m^{(1)}Q_2) = (U_m^{(1)}Q_2)(R_2Q_2)+\beta_m u_{m+1}(b_{m+1}^{(1)H}Q_2).$$

It implies that

$$AU_m^{(2)} = U_m^{(2)}H_m^{(2)}+\beta_m u_{m+1}b_{m+1}^{(2)H},$$

where

$$U_m^{(2)} \equiv U_m^{(1)}Q_2$$

is orthonormal,

$$H_m^{(2)} \equiv R_2 Q_2 + \kappa_2 I = \left[\begin{array}{c|cc} H_{m-2}^{(2)} & * & * \\ \hline & \kappa_2 & * \\ & & \kappa_1 \end{array}\right]$$

is upper Hessenberg with unreduced matrix $H_{m-2}^{(2)}$ and

$$\begin{aligned} b_{m+1}^{(2)H} &\equiv b_{m+1}^{(1)H} Q_2 = q_{m-1,m}^{(1)} e_{m-1}^H Q_2 + q_{m,m}^{(1)} e_m^T Q_2 \\ &= \left[\begin{array}{cccccc} 0 & \cdots & 0 & \times & \times & \times \end{array}\right]. \end{aligned}$$

For on postmultiplying (15) by e_1, we get

$$(A - \kappa_2 I) u_1^{(1)} = (A - \kappa_2 I)(U_m^{(1)} e_1) = U_m^{(2)} R_2 e_1 = r_{11}^{(2)} u_1^{(2)}.$$

Since $H_m^{(1)}$ is unreduced, $r_{11}^{(2)}$ is nonzero. Therefore, the first column of $U_m^{(2)}$ is a multiple of

$$(A - \kappa_2 I) u_1^{(1)} = 1/r_{11}^{(1)} (A - \kappa_2 I)(A - \kappa_1 I) u_1.$$

Repeating this process with $\kappa_3, \ldots, \kappa_{m-k}$, 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 $\kappa_1, \ldots, \kappa_m$ be eigenvalues of H_m. If the implicitly restarted QR step is performed with shifts $\kappa_1, \ldots, \kappa_{m-k}$, 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 $\kappa_1, \ldots, \kappa_{m-k}$ 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

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

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

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

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

- Avoid any matrix-vector multiplications in forming the new starting vector.
- Get its Arnoldi decomposition of order k for free.
- For large n the major cost will be in computing UQ.

If a Krylov decomposition can be partitioned in the form

$$A\left[\begin{array}{cc} U_1 & U_2 \end{array}\right] = \left[\begin{array}{cc} U_1 & U_2 \end{array}\right]\left[\begin{array}{cc} B_{11} & B_{12} \\ 0 & B_{22} \end{array}\right] + u\left[\begin{array}{cc} b_1^H & b_2^H \end{array}\right],$$

then

$$AU_1 = U_1 B_{11} + u b_1^H$$

is also a Krylov decomposition.
The process of Krylov-Schur restarting:

- Compute the Schur decomposition of the Rayleigh quotient
- Move the desired eigenvalues to the beginning
- Throw away the rest of the decomposition

Exchanging eigenvalues and eigenblocks

- Move an eigenvalue from one place to another.

Let a triangular matrix be partitioned in the form

$$R \equiv \begin{bmatrix} A & B & C \\ 0 & S & D \\ 0 & 0 & E \end{bmatrix},$$

where

$$S = \begin{bmatrix} s_{11} & s_{12} \\ 0 & s_{22} \end{bmatrix}.$$

Suppose that Q is a unitary matrix such that

$$Q^H S Q = \begin{bmatrix} s_{22} & \hat{s}_{12} \\ 0 & s_{11} \end{bmatrix},$$

then the eigenvalues s_{11} and s_{22} in the matrix

$$\text{diag}\left(\begin{array}{ccc} I & Q^H & I \end{array}\right) R \, \text{diag}\left(\begin{array}{ccc} I & Q & I \end{array}\right) = \begin{bmatrix} A & BQ & C \\ 0 & Q^H S Q & Q^H D \\ 0 & 0 & E \end{bmatrix}$$

will have traded places.

- How to find such unitary matrix Q?

Let

$$S = \begin{bmatrix} S_{11} & S_{12} \\ 0 & S_{22} \end{bmatrix},$$

where S_{ii} is of order n_i $(i = 1, 2)$. Therefore are four cases to consider.

1. $n_1 = 1, \; n_2 = 1.$
2. $n_1 = 2, \; n_2 = 1.$
3. $n_1 = 1, \; n_2 = 2.$
4. $n_1 = 2, \; n_2 = 2.$

- For the first two cases ($n_1 = 1,\ n_2 = 1$ or $n_1 = 2,\ n_2 = 1$):

Let

$$S = \begin{bmatrix} S_{11} & s_{12} \\ 0 & s_{22} \end{bmatrix},$$

where S_{11} is of order one or two. Let x be a normalized eigenvector corresponding to s_{22} and let $Q = [x\ Y]$ be orthogonal. Then

$$Q^T SQ = \begin{bmatrix} x^T \\ Y^T \end{bmatrix} S \begin{bmatrix} x & Y \end{bmatrix} = \begin{bmatrix} x^T Sx & x^T SY \\ Y^T Sx & Y^T SY \end{bmatrix} = \begin{bmatrix} s_{22} & \hat{s}_{12}^T \\ 0 & \hat{S}_{11} \end{bmatrix}.$$

Note that $\hat{S}_{11}$ and S_{11} have the same eigenvalues.

- For the third case ($n_1 = 1,\ n_2 = 2$):

Let

$$S = \begin{bmatrix} s_{11} & s_{12}^T \\ 0 & S_{22} \end{bmatrix},$$

where S_{22} is of order two. Let y be a normalized left eigenvector corresponding to s_{11} and let $Q = [X\ y]$ be orthogonal. Then

$$Q^T S Q = \begin{bmatrix} X^T \\ y^T \end{bmatrix} S \begin{bmatrix} X & y \end{bmatrix} = \begin{bmatrix} X^T S X & X^T S y \\ y^T S X & y^T S y \end{bmatrix} = \begin{bmatrix} \hat{S}_{22} & \hat{s}_{12} \\ 0 & s_{11} \end{bmatrix}.$$

- For the last case ($n_1 = 2,\ n_2 = 2$):

Let

$$S = \begin{bmatrix} S_{11} & S_{12} \\ 0 & S_{22} \end{bmatrix}.$$

Let (S_{22}, X) be an orthonormal eigenpair, i.e.,

$$SX = X\left(US_{22}U^{-1}\right)$$

for some nonsingular U, and let $Q = [X\ Y]$ be orthogonal. Then

$$\begin{aligned} Q^T SQ &= \begin{bmatrix} X^T SX & X^T SY \\ Y^T SX & Y^T SY \end{bmatrix} = \begin{bmatrix} X^T XUS_{22}U^{-1} & X^T SY \\ Y^T XUS_{22}U^{-1} & Y^T SY \end{bmatrix} \\ &= \begin{bmatrix} US_{22}U^{-1} & \hat{S}_{12} \\ 0 & \hat{S}_{11} \end{bmatrix}. \end{aligned}$$

Question

How to compute the orthonormal eigenbasis X?

Let the eigenbasis be $\begin{bmatrix} P \\ I \end{bmatrix}$, where P is to be determined.

Then

$$\begin{bmatrix} S_{11} & S_{12} \\ 0 & S_{22} \end{bmatrix} \begin{bmatrix} P \\ I \end{bmatrix} = \begin{bmatrix} P \\ I \end{bmatrix} S_{22}.$$

Hence P can be solved from the Sylvester equation

$$S_{11}P - PS_{22} = -S_{12}.$$

The orthonormal eigenbasis X can be computed by the QR factorization

$$\begin{bmatrix} P \\ I \end{bmatrix} = \begin{bmatrix} X & Y \end{bmatrix} \begin{bmatrix} R \\ 0 \end{bmatrix}.$$

The Krylov-Schur cycle

Assume $A \in \mathbb{C}^{n\times n}$.

1. Write the corresponding Krylov decomposition in the form
$$AU_m = U_m T_m + \beta_m u_{m+1} e_m^T.$$
2. Compute the Schur decomposition of T_m,
$$S_m = Q^H T_m Q$$
where S_m is upper triangular.
3. Transform the decomposition to the form
$$A\hat{U}_m = \hat{U}_m S_m + u_{m+1} b_{m+1}^H.$$
4. Select $m-k$ Ritz values and move them to the end of S_m, accumulating the transformations in Q_1.
5. Truncate the decomposition, i.e.,
$$S_k := S_m[1:k, 1:k],\ b_k^H := b_{m+1}^H Q_1[:, 1:k],\ U_k := \hat{U}_m Q_1[:, 1:k].$$

Deflation

We say a Krylov decomposition has been deflated if it can be partitioned in the form

$$A\begin{bmatrix} U_1 & U_2 \end{bmatrix} = \begin{bmatrix} U_1 & U_2 \end{bmatrix}\begin{bmatrix} B_{11} & B_{12} \\ 0 & B_{22} \end{bmatrix} + u\begin{bmatrix} 0 & b_2^H \end{bmatrix}.$$

It implies that

$$AU_1 = U_1 B_{11},$$

so that U_{11} spans an eigenspace of A.

Criterion of Deflation:

Theorem

Let

$$AU = UB + ub^H$$

be an orthonormal Krylov decomposition, and let $[M, \tilde{U}] = [M, UW]$ *be an orthonormal pair. Let* $[W, W_\perp]$ *be unitary, and set*

$$\tilde{B} = \begin{bmatrix} W^H \\ W_\perp^H \end{bmatrix} B \begin{bmatrix} W & W_\perp \end{bmatrix} \equiv \begin{bmatrix} \tilde{B}_{11} & \tilde{B}_{12} \\ \tilde{B}_{21} & \tilde{B}_{22} \end{bmatrix}$$

and

$$\tilde{b}^H = b^H \begin{bmatrix} W & W_\perp \end{bmatrix} = \begin{bmatrix} \tilde{b}_1^H & \tilde{b}_2^H \end{bmatrix}.$$

Then

$$\|A\tilde{U} - \tilde{U}M\|_F^2 = \|\tilde{B}_{21}\|_F^2 + \|\tilde{b}_1\|_F^2 + \|\tilde{B}_{11} - M\|_F^2.$$

Proof: Let

$$\left[\begin{array}{cc} \tilde{U} & \tilde{U}_\perp \end{array} \right] = U \left[\begin{array}{cc} W & W_\perp \end{array} \right].$$

Then

$$\begin{aligned} & A\tilde{U} - \tilde{U}M = UBW + ub^H W - UWM \\ = \; & U \left[\begin{array}{cc} W & W_\perp \end{array} \right] \left(\left[\begin{array}{c} W^H \\ W_\perp^H \end{array} \right] B \left[\begin{array}{cc} W & W_\perp \end{array} \right] \left[\begin{array}{c} I \\ 0 \end{array} \right] - \left[\begin{array}{c} I \\ 0 \end{array} \right] M \right) \\ & + ub^H \left[\begin{array}{cc} W & W_\perp \end{array} \right] \left[\begin{array}{c} I \\ 0 \end{array} \right] \\ = \; & \left[\begin{array}{cc} \tilde{U} & \tilde{U}_\perp \end{array} \right] \left[\begin{array}{c} \tilde{B}_{11} - M \\ \tilde{B}_{21} \end{array} \right] + u\tilde{b}_1^H = \left[\begin{array}{ccc} \tilde{U} & \tilde{U}_\perp & u \end{array} \right] \left[\begin{array}{c} \tilde{B}_{11} - M \\ \tilde{B}_{21} \\ \tilde{b}_1^H \end{array} \right]. \end{aligned}$$

Since $u^H U = 0$, we have

$$u^H \begin{bmatrix} \tilde{U} & \tilde{U}_\perp \end{bmatrix} = u^H U \begin{bmatrix} W & W_\perp \end{bmatrix} = 0.$$

It implies that $[\tilde{U}, \tilde{U}_\perp, u]$ is an orthonormal matrix. Therefore,

$$\begin{aligned} \|A\tilde{U} - \tilde{U}M\|_F^2 &= \left\| \begin{bmatrix} \tilde{B}_{11} - M \\ \tilde{B}_{21} \\ \tilde{b}_1^H \end{bmatrix} \right\|_F^2 \\ &= \|\tilde{B}_{21}\|_F^2 + \|\tilde{b}_1\|_F^2 + \|\tilde{B}_{11} - M\|_F^2. \end{aligned}$$

■

Suppose that $A\tilde{U} - \tilde{U}M$ is small. Transform the Krylov decomposition to the form

$$A \begin{bmatrix} \tilde{U} & \tilde{U}_\perp \end{bmatrix} = \begin{bmatrix} \tilde{U} & \tilde{U}_\perp \end{bmatrix} \begin{bmatrix} \tilde{B}_{11} & \tilde{B}_{12} \\ \tilde{B}_{21} & \tilde{B}_{22} \end{bmatrix} + u \begin{bmatrix} \tilde{b}_1^H & \tilde{b}_2^H \end{bmatrix}$$

$$= \begin{bmatrix} \tilde{U} & \tilde{U}_\perp \end{bmatrix} \begin{bmatrix} \tilde{B}_{11} & \tilde{B}_{12} \\ 0 & \tilde{B}_{22} \end{bmatrix} + u \begin{bmatrix} 0 & \tilde{b}_2^H \end{bmatrix} + \left(\tilde{U}_\perp \tilde{B}_{21} + u \tilde{b}_1^H \right).$$

From Theorem 16, we have

$$\left\| \begin{bmatrix} \tilde{B}_{21} \\ \tilde{b}_1^H \end{bmatrix} \right\|_F \leq \|A\tilde{U} - \tilde{U}M\|_F,$$

with equality if and only if $M = W^H BW$. Therefore, if the residual norm $\|A\tilde{U} - \tilde{U}M\|_F$ is sufficiently small, we may set $\tilde{B}_{21}$ and $\tilde{b}_1$ to zero to get the approximate decomposition

$$A \begin{bmatrix} \tilde{U} & \tilde{U}_\perp \end{bmatrix} \approx \begin{bmatrix} \tilde{U} & \tilde{U}_\perp \end{bmatrix} \begin{bmatrix} \tilde{B}_{11} & \tilde{B}_{12} \\ 0 & \tilde{B}_{22} \end{bmatrix} + u \begin{bmatrix} 0 & \tilde{b}_2^H \end{bmatrix}.$$

Rational Krylov transformations

- Shift-and-invert transformations in Arnoldi's method is to focus the algorithm on the eigenvalues near the shift κ.
- How to do when it needs to use more than one shift?
  - Restart with a new shift and a new vector
  - Change a Krylov decomposition from one in $(A - \kappa_1 I)^{-1}$ to one in $(A - \kappa_2 I)^{-1}$.

Suppose we have a Krylov sequence

$$u, (A-\kappa_1 I)^{-1}u, (A-\kappa_1 I)^{-2}u, \cdots, (A-\kappa_1 I)^{1-k}u.$$

Set $v = (A-\kappa_1 I)^{1-k}u$, then the sequence with its terms in reverse order is

$$v, (A-\kappa_1 I)v, \cdots, (A-\kappa_1 I)^{k-1}v,$$

so that

$$\mathcal{K}_k[(A-\kappa_1 I)^{-1}, u] = \mathcal{K}_k[A-\kappa_1 I, v].$$

By the shift invariance of a Krylov sequence

$$\mathcal{K}_k[A-\kappa_1 I, v] = \mathcal{K}_k[A-\kappa_2 I, v].$$

Set

$$w = (A-\kappa_2 I)^{k-1}v,$$

we have

$$\mathcal{K}_k[A-\kappa_2 I, v] = \mathcal{K}_k[(A-\kappa_2 I)^{-1}, w].$$

It follows that

$$\mathcal{K}_k[(A-\kappa_1 I)^{-1}, u] = \mathcal{K}_k[(A-\kappa_2 I)^{-1}, w].$$

That is the Krylov subspace in $(A-\kappa_1 I)^{-1}$ with starting vector u is exactly the same as the Krylov subspace in $(A-\kappa_2 I)^{-1}$ with a different starting vector w.

Let

$$(A-\kappa_1 I)^{-1} U = \hat{U}\hat{H}$$

be an orthonormal Arnoldi decomposition. Then

$$U = (A-\kappa_1 I)\hat{U}\hat{H} = (A-\kappa_2 I)\hat{U}\hat{H} - (\kappa_1-\kappa_2)\hat{U}\hat{H}.$$

Hence

$$(A-\kappa_2 I)^{-1}\hat{U}\left[\hat{I} + (\kappa_1-\kappa_2)\hat{H}\right] = \hat{U}\hat{H},$$

where

$$\hat{I} = \begin{bmatrix} I & 0 \end{bmatrix}^T.$$

Let

$$\hat{I} + (\kappa_1 - \kappa_2)\hat{H} = Q(\hat{I}R)$$

be a QR factorization of $\hat{I} + (\kappa_1 - \kappa_2)\hat{H}$. Then

$$(A - \kappa_2 I)^{-1}(\hat{U}Q\hat{I}) = \hat{U}\hat{H}R^{-1} = (\hat{U}Q)(Q^H\hat{H}R^{-1}).$$

Set

$$\hat{V} = \hat{U}Q, \; V = \hat{V}\hat{I} \;\text{ and }\; \hat{B} = Q^H\hat{H}R^{-1},$$

then

$$(A - \kappa_2 I)^{-1}V = \hat{V}\hat{B}.$$

This is a Krylov decomposition.

If A is symmetric and

$$AU_k = U_k T_k + \beta_k u_{k+1} e_k^T$$

is an Arnoldi decomposition, then T_k is a tridiagonal matrix of the form

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

It implies that u_k can be generated by a three-term recurrence

$$\beta_j u_{j+1} = Au_j - \alpha_j u_j - \beta_{j-1} u_{j-1},$$

where

$$\alpha_j = u_j^T A u_j, \quad \beta_j = \|Au_j - \alpha_j u_j - \beta_{j-1} u_{j-1}\|_2.$$

Problem

- Mathematically, u_j must be orthogonal.
- In practice, they can lose orthogonality.

Solutions

Reorthogonalize the vectors at each step and restart when it becomes impossible to store $\{u_i\}$ in main memory.