

Lecture 11: Krylov subspace, Generalized minimal residual method



School of Mathematical Sciences, Xiamen University

1. Krylov subspace

- Given $\mathbf{A} \in \mathbb{C}^{m \times m}$ and nonzero $\mathbf{r} \in \mathbb{C}^m$, the j th Krylov subspace generated by \mathbf{A} and \mathbf{r} is defined by

$$\mathcal{K}_j(\mathbf{A}, \mathbf{r}) := \text{span}\{\mathbf{r}, \mathbf{A}\mathbf{r}, \mathbf{A}^2\mathbf{r}, \dots, \mathbf{A}^{j-1}\mathbf{r}\}.$$

Obviously, $\mathcal{K}_j(\mathbf{A}, \mathbf{r}) \subseteq \mathcal{K}_{j+1}(\mathbf{A}, \mathbf{r})$ and $\dim \mathcal{K}_j(\mathbf{A}, \mathbf{r}) \leq j$.

Proposition 1

Let \mathbb{P}_j denote the set of polynomials of degree $\leq j$. Then

$$\mathcal{K}_j(\mathbf{A}, \mathbf{r}) = \{p(\mathbf{A})\mathbf{r} \mid p \in \mathbb{P}_{j-1}\}.$$

Proposition 2

If the minimal polynomial of the matrix \mathbf{A} has degree n , then for any $j > n$ and any nonzero $\mathbf{r} \in \mathbb{C}^m$, we have

$$\dim \mathcal{K}_j(\mathbf{A}, \mathbf{r}) \leq n.$$

1.1. Arnoldi process

Algorithm: Arnoldi process generating orthonormal basis

Given $\mathbf{A} \in \mathbb{C}^{m \times m}$ and nonzero $\mathbf{r} \in \mathbb{C}^m$

$\mathbf{q}_1 = \mathbf{r} / \|\mathbf{r}\|_2$

for $j = 1, 2, 3, \dots$,

$\mathbf{v} = \mathbf{A}\mathbf{q}_j$

for $i = 1$ **to** j

$h_{ij} = \langle \mathbf{v}, \mathbf{q}_i \rangle = \mathbf{q}_i^* \mathbf{v}$

$\mathbf{v} = \mathbf{v} - h_{ij}\mathbf{q}_i$

end

$h_{j+1,j} = \|\mathbf{v}\|_2$

$\mathbf{q}_{j+1} = \mathbf{v} / h_{j+1,j}$

end

- At the end of step j , we obtain

$$\mathbf{v} = (\mathbf{I} - \mathbf{q}_j \mathbf{q}_j^*) \cdots (\mathbf{I} - \mathbf{q}_2 \mathbf{q}_2^*) (\mathbf{I} - \mathbf{q}_1 \mathbf{q}_1^*) \mathbf{A} \mathbf{q}_j.$$

- We call the Arnoldi process breaks down at step k if $h_{k+1,k} = 0$.

Remark 3

The Arnoldi process is the modified Gram–Schmidt orthogonalization applied to $\{\mathbf{r}, \mathbf{A}\mathbf{q}_1, \mathbf{A}\mathbf{q}_2, \dots, \mathbf{A}\mathbf{q}_k\}$. We have the Arnoldi relation

$$\mathbf{A} [\mathbf{q}_1 \quad \cdots \quad \mathbf{q}_j] = [\mathbf{q}_1 \quad \cdots \quad \mathbf{q}_{j+1}] \begin{bmatrix} h_{11} & \cdots & h_{1j} \\ h_{21} & \ddots & \vdots \\ & \ddots & h_{jj} \\ & & & h_{j+1,j} \end{bmatrix}, \quad \forall 1 \leq j < k,$$

that is $\mathbf{A}\mathbf{Q}_j = \mathbf{Q}_{j+1}\tilde{\mathbf{H}}_j$. ($\tilde{\mathbf{H}}_j$ is upper Hessenberg.) Let

$$\mathbf{H}_j := \begin{bmatrix} h_{11} & \cdots & h_{1j} \\ h_{21} & \ddots & \vdots \\ & \ddots & h_{jj} \end{bmatrix}, \quad \forall 1 \leq j \leq k.$$

We have $\mathbf{A}\mathbf{Q}_k = \mathbf{Q}_k\mathbf{H}_k$ and $\mathbf{H}_j = \mathbf{Q}_j^*\mathbf{A}\mathbf{Q}_j$ for all $1 \leq j \leq k$.

Theorem 4

Suppose that the Arnoldi process breaks down at step k . We have

$$\text{span}\{\mathbf{q}_1, \mathbf{q}_2, \dots, \mathbf{q}_j\} = \mathcal{K}_j(\mathbf{A}, \mathbf{r}), \quad j = 1, 2, \dots, k,$$

and the set $\{\mathbf{q}_1, \mathbf{q}_2, \dots, \mathbf{q}_k\}$ is orthonormal.

Corollary 5

The matrices $\{\mathbf{Q}_j\}_{j=1}^k$ generated by the Arnoldi process are Q-factors of reduced QR factorizations of the Krylov matrices,

$$\mathbf{K}_j := [\mathbf{r} \quad \mathbf{A}\mathbf{r} \quad \dots \quad \mathbf{A}^{j-1}\mathbf{r}] = \mathbf{Q}_j\mathbf{R}_j, \quad j = 1, 2, \dots, k.$$

Moreover, $\dim\mathcal{K}_j(\mathbf{A}, \mathbf{r}) = j$ for $1 \leq j \leq k$, and $\dim\mathcal{K}_j(\mathbf{A}, \mathbf{r}) = k$ for $j > k$.

Remark 6

Both \mathbf{K}_j and \mathbf{R}_j are not formed explicitly in the Arnoldi process. We have $r_{11} = \|\mathbf{r}\|_2$. How to obtain \mathbf{R}_j from \mathbf{H}_j ? (Exercise)

2. Generalized minimal residual method (GMRES)

- Idea of GMRES: Consider a nonsingular linear system

$$\mathbf{Ax} = \mathbf{b}, \quad \mathbf{A} \in \mathbb{C}^{m \times m}, \quad \mathbf{b} \in \mathbb{C}^m.$$

For any initial guess \mathbf{x}_0 , at step j , GMRES finds the j th approximate solution

$$\mathbf{x}_j = \operatorname{argmin}_{\mathbf{x} \in \mathbf{x}_0 + \mathcal{K}_j(\mathbf{A}, \mathbf{r}_0)} \|\mathbf{b} - \mathbf{Ax}\|_2,$$

where $\mathbf{r}_0 := \mathbf{b} - \mathbf{Ax}_0$ and

$$\mathcal{K}_j(\mathbf{A}, \mathbf{r}_0) = \operatorname{span}\{\mathbf{r}_0, \mathbf{Ar}_0, \dots, \mathbf{A}^{j-1}\mathbf{r}_0\}.$$

For the residual $\mathbf{r}_j := \mathbf{b} - \mathbf{Ax}_j$, we have

$$\|\mathbf{r}_j\|_2 = \min_{\mathbf{x} \in \mathbf{x}_0 + \mathcal{K}_j(\mathbf{A}, \mathbf{r}_0)} \|\mathbf{b} - \mathbf{Ax}\|_2.$$

Assume that the Arnoldi process for the orthonormal basis of $\mathcal{K}_j(\mathbf{A}, \mathbf{r}_0)$ breaks down at step k . For $1 \leq j < k$, we have

$$\begin{aligned}\|\mathbf{r}_j\|_2 &= \min_{\mathbf{x} \in \mathbf{x}_0 + \mathcal{K}_j(\mathbf{A}, \mathbf{r}_0)} \|\mathbf{b} - \mathbf{A}\mathbf{x}\|_2 = \min_{\mathbf{y} \in \mathbb{C}^j} \|\mathbf{r}_0 - \mathbf{A}\mathbf{Q}_j\mathbf{y}\|_2 \\ &= \min_{\mathbf{y} \in \mathbb{C}^j} \left\| \mathbf{r}_0 - \mathbf{Q}_{j+1}\tilde{\mathbf{H}}_j\mathbf{y} \right\|_2 \quad (\text{by Arnoldi relation}) \\ &= \min_{\mathbf{y} \in \mathbb{C}^j} \left\| \|\mathbf{r}_0\|_2 \mathbf{e}_1 - \tilde{\mathbf{H}}_j\mathbf{y} \right\|_2 > 0.\end{aligned}$$

For $j = k$, we have

$$\begin{aligned}\|\mathbf{r}_k\|_2 &= \min_{\mathbf{x} \in \mathbf{x}_0 + \mathcal{K}_k(\mathbf{A}, \mathbf{r}_0)} \|\mathbf{b} - \mathbf{A}\mathbf{x}\|_2 = \min_{\mathbf{y} \in \mathbb{C}^k} \|\mathbf{r}_0 - \mathbf{A}\mathbf{Q}_k\mathbf{y}\|_2 \\ &= \min_{\mathbf{y} \in \mathbb{C}^k} \|\mathbf{r}_0 - \mathbf{Q}_k\mathbf{H}_k\mathbf{y}\|_2 \quad (\text{by Arnoldi relation}) \\ &= \min_{\mathbf{y} \in \mathbb{C}^k} \left\| \|\mathbf{r}_0\|_2 \mathbf{e}_1 - \mathbf{H}_k\mathbf{y} \right\|_2 = 0.\end{aligned}$$

Once \mathbf{y}_j is found, set $\mathbf{x}_j = \mathbf{x}_0 + \mathbf{Q}_j\mathbf{y}_j$.

- The least squares problem about \mathbf{y} can be solved inexpensively with Givens rotations, exploiting the upper Hessenberg structure of $\tilde{\mathbf{H}}_j$, costing just $\mathcal{O}(j^2)$ or $\mathcal{O}(j)$ instead of $\mathcal{O}(j^3)$.

2.1. Convergence of GMRES

Theorem 7

Assume that the Arnoldi process for the orthonormal basis of $\mathcal{K}_j(\mathbf{A}, \mathbf{r}_0)$ breaks down at step k .

- (1) *For $1 \leq j < k$, the residual \mathbf{r}_j satisfies $(\mathbf{A}\mathbf{Q}_j)^*\mathbf{r}_j = \mathbf{0}$, i.e.,*

$$\mathbf{r}_j \perp \mathbf{A}\mathcal{K}_j.$$

- (2) *For $0 \leq j \leq k$, the residual \mathbf{r}_j satisfies*

$$\|\mathbf{r}_0\|_2 \geq \|\mathbf{r}_1\|_2 \geq \cdots \geq \|\mathbf{r}_{k-1}\|_2 > \|\mathbf{r}_k\|_2 = 0.$$

That is to say GMRES converges monotonically and finds the exact solution at step k .

Theorem 8

Suppose \mathbf{A} is diagonalizable, i.e., $\mathbf{A} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^{-1}$ for some nonsingular matrix \mathbf{V} and diagonal matrix $\mathbf{\Lambda}$. At step j of the GMRES iteration, the residual \mathbf{r}_j satisfies

$$\frac{\|\mathbf{r}_j\|_2}{\|\mathbf{r}_0\|_2} \leq \min_{p \in \mathbb{P}_j, p(0)=1} \|p(\mathbf{A})\|_2 \leq \kappa(\mathbf{V}) \min_{p \in \mathbb{P}_j, p(0)=1} \max_{\lambda \in \Lambda(\mathbf{A})} |p(\lambda)|,$$

where $\Lambda(\mathbf{A})$ is the set of eigenvalues of \mathbf{A} , and $\kappa(\mathbf{V}) = \|\mathbf{V}\|_2 \|\mathbf{V}^{-1}\|_2$.

- Y. Saad and M.H. Schultz

GMRES: A generalized minimal residual algorithm for solving non-symmetric linear systems

SIAM J. Sci. Stat. Comput., 7: 856–869, 1986.

- Y. Saad

A flexible inner-outer preconditioned GMRES algorithm

SIAM J. Sci. Comput., 14: 461–469, 1993.

- **Exercise:** Assume the Arnoldi process for $\{\mathbf{A}, \mathbf{r}_0\}$ breaks down at step $k > 1$. For $1 \leq j < k$, we have the Arnoldi relation

$$\mathbf{A}\mathbf{Q}_j = \mathbf{Q}_{j+1}\tilde{\mathbf{H}}_j.$$

For $1 \leq j < k$, prove the following:

- (a) The j th residual \mathbf{r}_j of GMRES can be *uniquely* expressed as

$$\mathbf{r}_j = p_j(\mathbf{A})\mathbf{r}_0, \quad \deg(p_j) \leq j, \quad p_j(0) = 1.$$

- (b) Let $\mathbf{H}_j = \mathbf{Q}_j^* \mathbf{A} \mathbf{Q}_j$. The unique polynomial p_j in (a) is given by

$$p_j(z) = \prod_{i=1}^j \left(1 - \theta_i^{(j)} z\right),$$

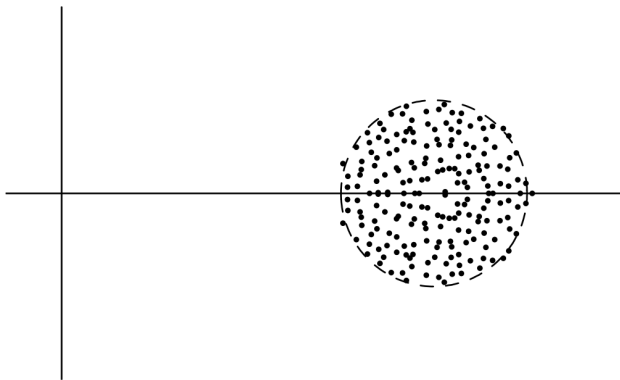
where $\theta_i^{(j)}$, $i = 1, 2, \dots, j$, are the eigenvalues of $(\tilde{\mathbf{H}}_j^* \tilde{\mathbf{H}}_j)^{-1} \mathbf{H}_j^*$.

2.2. Numerical examples

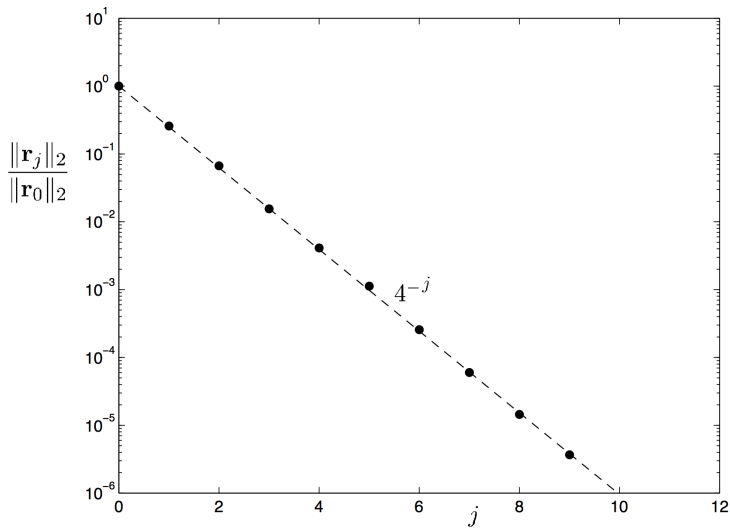
- Example 1: \mathbf{A} , 200×200 entries from real normal distribution of mean 2 and standard deviation $0.5/\sqrt{200}$

$m = 200$; $\mathbf{A} = 2*\text{eye}(m)+0.5*\text{randn}(m)/\text{sqrt}(m)$;

$\mathbf{Ax} = \mathbf{b}$, $\mathbf{x}_0 = \mathbf{0}$, $\mathbf{b} = [1 \ 1 \ \dots \ 1]^\top$



Convergence history of Example 1

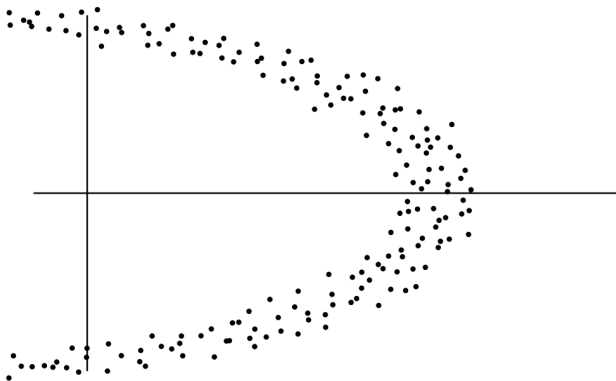


• Example 2:

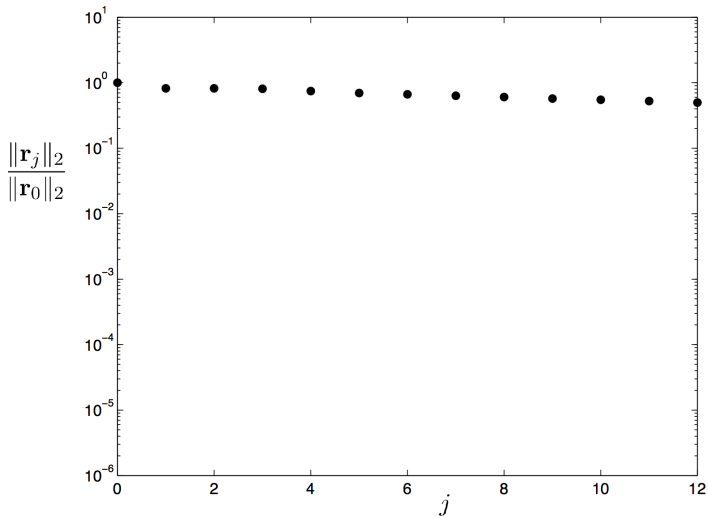
$m = 200$; $B = 2 \cdot \text{eye}(m) + 0.5 \cdot \text{randn}(m) / \text{sqrt}(m)$;

$A = B + D$, D is the diagonal matrix with complex entries

$$d_i = (-2 + 2 \sin \theta_i) + i \cos \theta_i, \quad \theta_i = \frac{(i-1)\pi}{m-1}, \quad 1 \leq i \leq m.$$



Convergence history of Example 2



2.3. Preconditioning (see Lecture 40 of NLA)

- To improve the convergence of Krylov subspace methods, it is important to have a preconditioner (suitable approximation for the original coefficient matrix \mathbf{A}), denoted by \mathbf{M} .
- Left preconditioning, i.e.,

$$\mathbf{M}^{-1}\mathbf{A}\mathbf{x} = \mathbf{M}^{-1}\mathbf{b}.$$

- Right preconditioning is often used, i.e.,

$$\mathbf{A}\mathbf{M}^{-1}\mathbf{z} = \mathbf{b}, \quad \mathbf{x} = \mathbf{M}^{-1}\mathbf{z},$$

because it produces the same residual as that of the original system in exact precision arithmetic.

- Note that we never explicitly form \mathbf{M}^{-1} . Only the action of applying the preconditioner solve operation \mathbf{M}^{-1} to a given vector is computed in iterative methods. So $\mathbf{M}^{-1}\mathbf{z}$ must be cheap.

- How to find a good preconditioner? **It's problem dependent.**

Example. Let $\mathcal{A} = \begin{bmatrix} \mathbf{A} & \mathbf{B}^* \\ \mathbf{C} & \mathbf{0} \end{bmatrix}$ and $\mathcal{M} = \begin{bmatrix} \mathbf{A} & \mathbf{0} \\ \mathbf{0} & \mathbf{CA}^{-1}\mathbf{B}^* \end{bmatrix}$, where $\mathbf{A} \in \mathbb{C}^{m \times m}$ is invertible, and $\mathbf{B}, \mathbf{C} \in \mathbb{C}^{n \times m}$ with $m \geq n$. Assume that $-\mathbf{CA}^{-1}\mathbf{B}^*$ is invertible.

The preconditioned matrix $\mathcal{M}^{-1}\mathcal{A}$ is diagonalizable and has at most three distinct eigenvalues 1, $(1 + \sqrt{5})/2$, and $(1 - \sqrt{5})/2$.

2.4. Restarted GMRES

- For larger values of j , the cost of GMRES in operations and storage may be prohibitive. In such circumstances a method called l -step restarted GMRES or GMRES(l) is often employed.
- GMRES(l): After l steps, the GMRES iteration is started anew with the current vector \mathbf{x}_l as an initial guess.
- Note that GMRES(l) can be expected fail to converge, whereas GMRES always succeeds for exact arithmetic. (**Embree's paper**)
- **GMRES-IR, GMRES-DR, FGMRES-DR**, etc.