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 jth Krylov subspace generated by $\mathbf{A}$ and $\mathbf{r}$ is defined by

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

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 $\mathbf{r} \in \mathbb{C}^m$,

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

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 $j = 1, 2, 3, \ldots,$

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

  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–Schimidt orthogonalization applied to $\{\mathbf{r}, \mathbf{A}\mathbf{q}_1, \mathbf{A}\mathbf{q}_2, \cdots, \mathbf{A}\mathbf{q}_k\}$. *We have the Arnoldi relation*

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

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

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

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

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

and the set $\{\mathbf{q}_1, \mathbf{q}_2, \cdots, \mathbf{q}_k\}$ 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,*

$$\mathbf{K}_j := \begin{bmatrix} \mathbf{r} & \mathbf{A}\mathbf{r} & \cdots & \mathbf{A}^{j-1}\mathbf{r} \end{bmatrix} = \mathbf{Q}_j \mathbf{R}_j, \quad j = 1, 2, \ldots, k.$$

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

Remark 6

Both $\mathbf{K}_j$ and $\mathbf{R}_j$ are not formed explicitly in the Arnoldi process.

2. Generalized minimal residual method (GMRES)

- Idea of GMRES: Consider a nonsingular linear system

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

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

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

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

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

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

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

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

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

For $j = k$, we have

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

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 $\widetilde{\mathbf{H}}_j$, 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 $(\mathbf{AQ}_j)^* \mathbf{r}_j = \mathbf{0}$, 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 nonsymmetric 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 $1 \le j < k$, we have the Arnoldi relation

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

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

(a) The jth residual $\mathbf{r}_j$ of GMRES can be *uniquely* expressed as

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

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

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

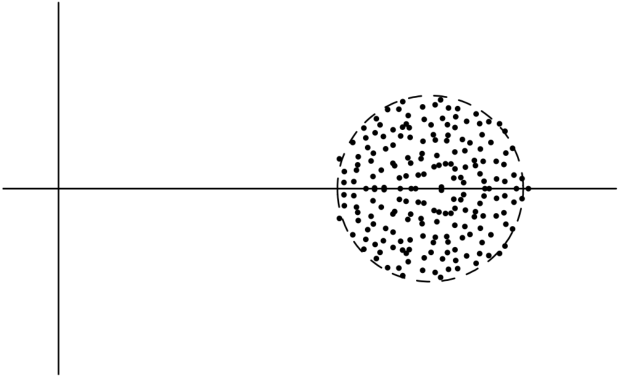
where $\theta_i^{(j)}$, $i = 1, 2, \ldots, j$, are the eigenvalues of $(\widetilde{\mathbf{H}}_j^* \widetilde{\mathbf{H}}_j)^{-1} \mathbf{H}_j^*$.

2.2. Numerical examples

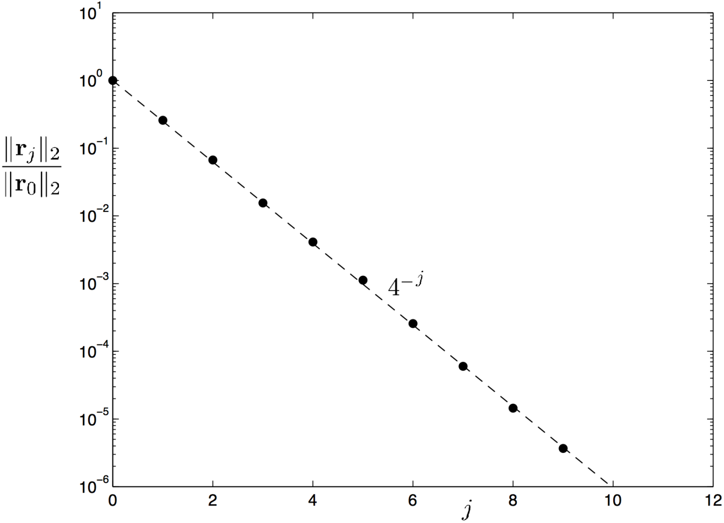
- Example 1: $\mathbf{A}$, 200×200 entries from real normal distribution of mean 2 and standard deviation $0.5/\sqrt{200}$

```
m = 200; A = 2*eye(m)+0.5*randn(m)/sqrt(m);
```

$$\mathbf{A}\mathbf{x} = \mathbf{b}, \quad \mathbf{x}_0 = \mathbf{0}, \quad \mathbf{b} = \begin{bmatrix} 1 & 1 & \cdots & 1 \end{bmatrix}^\top$$



Convergence history of Example 1

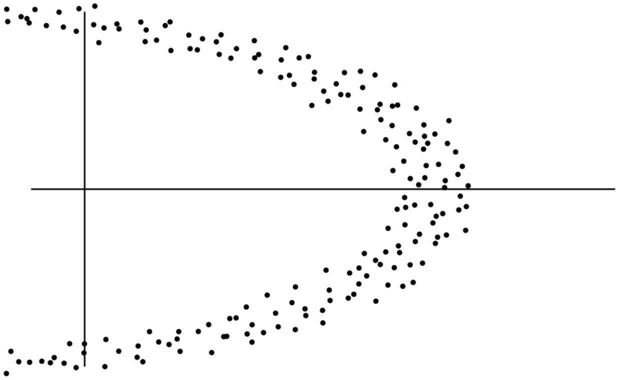


- Example 2:

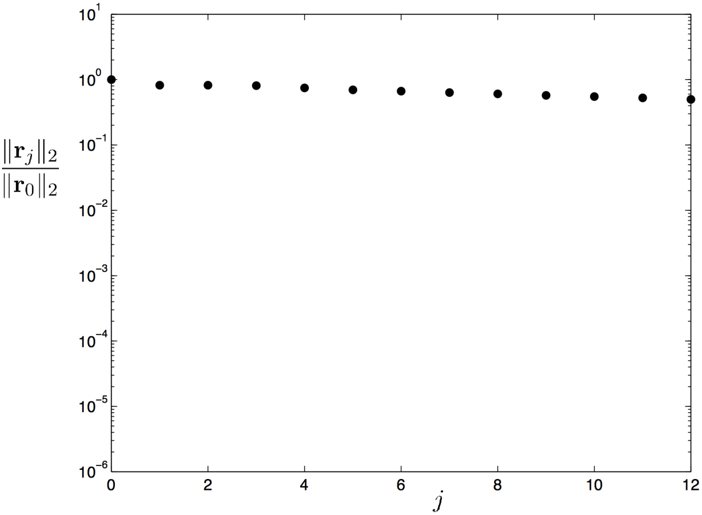
```
m = 200; B = 2*eye(m)+0.5*randn(m)/sqrt(m);
```

$\mathbf{A} = \mathbf{B} + \mathbf{D}$, $\quad$ $\mathbf{D}$ is the diagonal matrix with complex entries

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



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 $\mathcal{M} = \begin{bmatrix} \mathbf{A} & \mathbf{0} \\ \mathbf{0} & \mathbf{C}\mathbf{A}^{-1}\mathbf{B}^* \end{bmatrix}$, 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 $-\mathbf{C}\mathbf{A}^{-1}\mathbf{B}^*$ 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.