AMS526: Numerical Analysis I
(Numerical Linear Algebra)
Lecture 23: Minimal Residual and GMRES (NLA§35, MC§11.4)

Xiangmin Jiao

SUNY Stony Brook
Outline

1. Minimal Residual for Symmetric Systems (MC§11.4.1)
2. Generalization to Nonsymmetric Systems
3. Convergence of GMRES
In Lanczos iteration for \( A \in \mathbb{R}^{m \times m} \), starting from \( q_1 = b/\|b\| \)

\[
AQ_k = Q_{k+1} \tilde{T}_k, \tag{1}
\]

where \( \tilde{T}_k \) is \((k + 1) \times k\); \( Q_k \) is composed of orthonormal basis of \( \mathcal{K}_k \)

Let \( x_k = Q_k y_k \). Then,

\[
r_k = b - Ax_k = b - AQ_k y_k = b - Q_{k+1} \tilde{T}_k y_k
\]

Let \( Q_k^T r_k = Q_k^T \left( b - Q_{k+1} \tilde{T}_k y_k \right) = 0 \) (i.e., \( r_k \perp \mathcal{K}_k \)), we obtain

\[
Q_k^T Q_{k+1} \tilde{T}_k y_k = Q_k^T b
\]

where \( Q_k^T Q_{k+1} \tilde{T}_k = T_k \) and \( Q_k^T b = \beta e_1 \in \mathbb{R}^k \) with \( \beta = \|b\| \)

Hence,

\[
T_k y_k = \beta e_1 \tag{2}
\]

where \( T_k = Q_k^T AQ_k \) is tridiagonal, and is SPD if \( A \) is SPD

It takes \( \mathcal{O}(1) \) flops to update Cholesky factorization of \( T_k \) and then \( \mathcal{O}(k) \) flops to solve (2)
How to Generalize CG to Symmetric Indefinite Systems?

- If $A$ is not positive definite, Cholesky factorization of $T$ breaks down.
- What if we use LU or QR for $T_k$ instead?
  - SYMMLQ uses QR factorization of $T_k$ to solve tridiagonal system.
  - SYMMLQ enforces $r_k \perp K_k(A, b)$ like CG, but it does not minimize any specific norm in $e_k$, since $\| \cdot \|_A$ is no longer a proper norm.
- Alternative approach is MINRES, which minimize 2-norm of residual at each step.
- Both SYMMLQ and MINRES were developed by C. C. Paige and M. A. Saunders (1975).
Minimal Residual Method (MINRES)

- Similar to CG, consider Lanczos iteration of $A$ with $q_1 = b/\|b\|$
  \[
  AQ_k = Q_{k+1} \tilde{T}_k
  \]

Let $x_k = Q_k y_k$. Then,
\[
r_k = b - Ax_k = b - AQ_k y_k = b - Q_{k+1} \tilde{T}_k y_k
\]

- Since $q_1 = b/\|b\|$, $b \in \text{range}(Q_{k+1})$. Therefore,
  \[
  \|r_k\| = \|b - Q_{k+1} \tilde{T}_k y_k\| = \|Q_{k+1}^T (b - Q_{k+1} \tilde{T}_k y_k)\| = \|\beta e_1 - \tilde{T}_k y_k\|
  \]
  where $Q_{k+1}^T b = \beta e_1 \in \mathbb{R}^{k+1}$ with $\beta = \|b\|$.

- At each step, we solve $(k + 1) \times k$ least-squares problem
  \[
  \tilde{T}_k y_k \approx \beta e_1
  \]
  to minimize $\|r_k\| = \|\tilde{T}_k y_k - \beta e_1\|$, where $\tilde{T}_k$ is tridiagonal.

- It takes $O(1)$ flops to update QR of $\tilde{T}_k$ using Given’s rotation and then $O(k)$ flops to solve (3).
Comparison of CG, MINRES and SYMMLQ

<table>
<thead>
<tr>
<th></th>
<th>CG</th>
<th>SYMMLQ</th>
<th>MINRES</th>
</tr>
</thead>
<tbody>
<tr>
<td>objective</td>
<td>( \min |e_k|_A )</td>
<td>( r_k \perp \mathcal{K}_k(A, b) )</td>
<td>( \min |r_k|_2 )</td>
</tr>
<tr>
<td>internal solver</td>
<td>( T = LDL^T )</td>
<td>( T = LQ )</td>
<td>( \tilde{T} = QR )</td>
</tr>
<tr>
<td>monotonicity</td>
<td>( |e_k|_A )</td>
<td>-</td>
<td>( |r_k|_2 )</td>
</tr>
<tr>
<td>flops per iteration</td>
<td>8( m )</td>
<td></td>
<td>9( m )</td>
</tr>
</tbody>
</table>

In MINRES, residual decreases monotonically.

Convergence of MINRES

- If $k_*$ is dimension of smallest invariant space that contains $r_0$, then MINRES terminates in $k_*$ steps in exact arithmetic.
- In particular, if $A$ has $n$ distinct eigenvalues, MINRES converges in $\leq n$ iterations.
- MINRES minimizes $\|p_k(A)r_0\|_2$ at $k$th step, with $r_0 = b$, where $p_k$ is degree-$k$ polynomial $p_k(x) = 1 + c_1x + c_2x^2 + \cdots + c_kx^k$.
- Analogous to CG,
  $$\frac{\|r_k\|_2}{\|r_0\|_2} \leq \inf_{p_k} \max_{\lambda} |p_k(\lambda)|$$

  - If $A$ is SPD, $\inf_{p_k} \max_{\lambda} |p_k(\lambda)| \leq 2 \left( \frac{\sqrt{\kappa} - 1}{\sqrt{\kappa} + 1} \right)^k \approx 2 \left( 1 - \frac{2}{\sqrt{\kappa}} \right)^k$ for large $\kappa$. Hence, MINRES also takes up to $\mathcal{O}(\sqrt{\kappa})$ iterations.
  - If $A$ is indefinite, MINRES may take up to $\mathcal{O}(\kappa)$ iterations in worst case (Benzi, Golub & Liesen (2005 *Acta Numerica*)))
Outline

1. Minimal Residual for Symmetric Systems (MC§11.4.1)
2. Generalization to Nonsymmetric Systems
3. Convergence of GMRES
Overview of Methods for Nonsymmetric Systems

- Various extensions of Krylov subspace methods have been proposed for nonsymmetric matrices (GMRES, BiCG, BiCGSTAB, QMR, etc.)
- Three different approaches
  - Generalization of minimal residual with Arnoldi iteration (GMRES, NLA§35, MC§11.4.3)
  - CG-style method with nonsymmetric Lanczos iteration (BiCG, NLA§39, MC§11.4.5)
  - Quasi minimal residual with nonsymmetric Lanczos iteration (QMR, NLA§39, MC§11.4.5)
- GMRES (Generalized Minimal RESidual) is one of most well known, developed by Saad and Schultz (1986)
Review: Arnoldi Iteration

• In Arnoldi iteration with $A \in \mathbb{R}^{m \times m}$ and $q_1 = b/\|b\|$, 

$$AQ_k = Q_{k+1} \tilde{H}_k,$$  \hspace{1cm} (4)

where $Q_k = [q_1 \mid q_2 \mid \cdots \mid q_k]$ is composed orthonormal basis of $\mathcal{K}_k(A, b)$, and $\tilde{H}_k$ is $(k + 1) \times k$ upper Hessenberg.

Algorithm: Arnoldi Iteration

given random nonzero $b$, let $q_1 = b/\|b\|$

for $k = 1, 2, 3, \ldots$

\[ v = Aq_k \]

for $j = 1$ to $k$

\[ h_{jk} = q_j^*v \]

\[ v = v - h_{jk}q_j \]

\[ h_{k+1,k} = \|v\| \]

\[ q_{k+1} = v/h_{k+1,k} \]

This is essentially modified Gram-Schmidt orthogonalization of $\mathcal{K}_k(A, b)$
Derivation of GMRES

- Similar to MINRES, let $x_k = Q_k y_k$. Then,

$$r_k = b - Ax_k = b - AQ_k y_k = b - Q_{k+1} \tilde{H}_k y_k$$

- Since $q_1 = b/\|b\|$, $b \in \text{range}(Q_{k+1})$. Therefore,

$$\|r_k\| = \|b - Q_{k+1} \tilde{H}_k y_k\| = \|Q_{k+1}^T \left(b - Q_{k+1} \tilde{H}_k y_k \right)\| = \|\beta e_1 - \tilde{H}_k y_k\|,$$

where $Q_{k+1}^T b = \beta e_1 \in \mathbb{R}^{k+1}$ with $\beta = \|b\|

- At each step, we solve $(k + 1) \times k$ least-squares problem

$$\tilde{H}_k y_k \approx \beta e_1 \quad (5)$$

to minimizes $\|r_k\| = \|\tilde{H}_k y_k - \beta e_1\|$
High-Level GMRES Algorithm

Algorithm: GMRES

\[ q_1 = \frac{b}{\| b \|} \]

**for** \( k = 1, 2, 3, \ldots \)

- Step \( k \) of Arnoldi iteration
  - Update QR factorization of \( \tilde{H}_k \)
  - Use QR to solve \( \tilde{H}_k y_k \approx \beta e_1 \)
  - \( x_k = Q_k y_k \)

- \( \tilde{H}_k \) is upper Hessenberg. It takes \( O(k) \) flops to update QR factorization of \( \tilde{H}_n \) using Given’s rotation and \( O(k^2) \) flops to solve (5)
- \( \| r_k \| = \| \tilde{H}_k y_k - \beta e_1 \| \), so no need to compute explicitly from \( x_k \)
- However, Arnoldi iteration takes \( O(mk) \) flops, besides matrix-vector multiplication
- It is also possible to orthogonalize \( K_k(A, b) \) using Householder instead of Gram-Schmidt
Outline

1. Minimal Residual for Symmetric Systems (MC§11.4.1)

2. Generalization to Nonsymmetric Systems

3. Convergence of GMRES
Convergence of GMRES for Diagonalizable Matrices

- Suppose $A$ is diagonalizable. If $k_*$ is dimension of smallest invariant space that contains $r_0$, then GMRES terminates in $k_*$ steps in exact arithmetic. Here, $S$ is invariant if $v \in S \Rightarrow Av \in S$
- $\|r_k\|_2$ decreases monotonically in GMRES
- GMRES minimizes $\|p_k(A)r_0\|_2$ at $k$th step, with $r_0 = b$, where $p_k$ is degree-$k$ polynomial $p_k(x) = 1 + c_1x + c_2x^2 + \cdots + c_kx^k$
- If $A = V\Lambda V^{-1}$, then $\|p_k(A)\| \leq \|V\|\|p_k(\Lambda)\|\|V^{-1}\| = \kappa(V)\|p_k(\Lambda)\|$
- Since $\|r_k\| = \|p_k(A)r_0\| \leq \|p_k(A)\|\|r_0\|$, we have
  \[
  \frac{\|r_k\|^2}{\|r_0\|^2} \leq \inf_{p_k} \|p_k(A)\| \leq \kappa(V)\inf_{p_k} \|p_k(\Lambda)\| = \kappa(V)\inf_{p_k} \sup_\lambda \|p_k(\lambda)\|
  \]
- If $A$ is not far from normal (i.e., $\kappa(V)$ is small), GMRES converges similarly to MINRES
  - If $A$ is positive definite, then it may take $\mathcal{O}(\sqrt{\kappa})$ steps
  - If $A$ is indefinite, worst case is $\sim \mathcal{O}(\kappa)$ steps (this bound is not sharp)
- For nonnormal matrices, factor $\kappa(V)$ may be very pessimistic
Convergence of GMRES for General Matrices

- If $A$ is not diagonalizable, if $k_*$ is dimension of smallest generalized invariant space that contains $r_0$, then GMRES terminates in $k_*$ steps in exact arithmetic. Here, $S$ is generalized invariant subspace if $v \in S \Rightarrow A^k v \in S$ for $k \geq 0$, is related to generalized eigenvectors.

- If $A = VJV^{-1}$, where $J$ is the Jordan matrix, then

\[
\| p_k(A) \| \leq \| V \| \| p_k(J) \| \| V^{-1} \| = \kappa(V) \| p_k(J) \| 
\]

and

\[
\frac{\| r_k \|_2}{\| r_0 \|_2} \leq \inf_{p_k} \| p_k(A) \| \leq \kappa(V) \inf_{p_k} \| p_k(J) \| 
\]

- Bound on $\| p_k(J) \|$ is more complicated to describe; see Tichý, Liesen, and Faber (2007)
Practical Considerations of GMRES

- In (full) GMRES, memory and computational cost both grow with $k$
- In practice, (full) GMRES is almost never used
- Instead, GMRES with restart, denoted by GMRES$(k)$, is used, which runs GMRES for $k$ steps, and then restart to solve residual equation
- Fundamental difficulties with restarted GMRES$(k)$
  - if $k$ is too large, GMRES$(k)$ is expensive in terms of cost per iteration
  - if $k$ is too small, GMRES$(k)$ may stagnate
- No practical guideline for optimal $k$ for any input matrix
- Practical applications reply on preconditioners for faster convergence