Lecture 23
GMRES and Other Krylov Subspace Methods

Songting Luo

Department of Mathematics
Iowa State University

MATH 562 Numerical Analysis II

Outline

1. GMRES and Other Krylov Subspace Methods

Outline

Krylov Subspace Algorithms

- CG only works for SPD matrices. It minimizes $\|\mathbf{x} - \mathbf{x}_*\|_{\mathbf{A}}$ and $\phi(\mathbf{x}) = \frac{1}{2}\mathbf{x}^T\mathbf{A}\mathbf{x} - \mathbf{x}^T\mathbf{b}$
- There have been many proposed extensions to nonsymmetric matrices, GMRES, BiCG, etc.
- GMRES (Generalized Minimal RESiduals) is one of most well known
- The basic idea of GMRES is to find $\mathbf{x}_n \in \mathcal{K}_n$ that minimizes $\|\mathbf{r}_n\| = \|\mathbf{b} - \mathbf{A}\mathbf{x}_n\|$
- This can be viewed as a least squares problem: Find a vector c s.t. $\|\mathbf{A}\mathbf{K}_n\mathbf{c} - \mathbf{b}\|$ is minimized, where $\mathbf{K}_n$ is the $m \times n$ Krylov matrix composed of basis vectors of $\mathcal{K}_n$, and

$$\mathbf{A}\mathbf{K}_n = [\mathbf{A}\mathbf{b}, \mathbf{A}^2\mathbf{b}, \ldots, \mathbf{A}^n\mathbf{b}]$$

- Orthogonal basis is often used, produced by Arnoldi iteration

Review: Arnoldi Iteration

- Let $\mathbf{Q}_n = [\mathbf{q}_1|\mathbf{q}_2|\cdots|\mathbf{q}_n]$ be $m \times n$ matrix with first n columns of $\mathbf{Q}$ and $\tilde{\mathbf{H}}_n$ be $(n+1) \times n$ upper-left section of $\mathbf{H}$.
- Start by picking a random $\mathbf{q}_1$ and then determine $\mathbf{q}_2$ and $\tilde{H}_1$
- The nth columns of $\mathbf{A}\mathbf{Q}_n = \mathbf{Q}_{n+1}\tilde{\mathbf{H}}_n$ can be written as

$$\mathbf{A}\mathbf{q}_n = h_{1n}\mathbf{q}_1 + \cdots + h_{nn}\mathbf{q}_n + h_{n+1,n}\mathbf{q}_{n+1}$$

Algorithm: Arnoldi Iteration

given random nonzero $\mathbf{b}$, let $\mathbf{q}_1 = \mathbf{b}/\|\mathbf{b}\|$
for $n = 1$ to $1, 2, 3, \ldots$
    $\mathbf{v} = \mathbf{A}\mathbf{q}_n$
    for $j = 1$ to n
        $h_{jn} = \mathbf{q}_j^* \mathbf{v}$
        $\mathbf{v} = \mathbf{v} - h_{jn}\mathbf{q}_j$
    $h_{n+1,n} = \|\mathbf{v}\|$
    $\mathbf{q}_{n+1} = \mathbf{v}/h_{n+1,n}$

Minimal Residual with Orthogonal Basis

- Let $\mathbf{Q}_n$ be Krylov matrix whose columns $\mathbf{q}_1, \mathbf{q}_2, \ldots$ span the successive Krylov subspaces
- Instead of find $\mathbf{x}_n = \mathbf{K}_n \mathbf{c}$, find $\mathbf{x}_n = \mathbf{Q}_n \mathbf{y}$ which minimizes $\|\mathbf{A}\mathbf{Q}_n \mathbf{y} - \mathbf{b}\|$
- For Arnoldi iteration, we showed that $\mathbf{A}\mathbf{Q}_n = \mathbf{Q}_{n+1} \tilde{\mathbf{H}}_n$, so

$$\|\mathbf{Q}_{n+1} \tilde{\mathbf{H}}_n \mathbf{y} - \mathbf{b}\| = \text{ minimum}$$

- Left multiplication by $\mathbf{Q}_{n+1}^*$ does not change the norm, so

$$\|\tilde{\mathbf{H}}_n \mathbf{y} - \mathbf{Q}_{n+1}^* \mathbf{b}\| = \text{ minimum}$$

- Finally, by construction, $\mathbf{Q}_{n+1}^* \mathbf{b} = \|\mathbf{b}\| \mathbf{e}_1$, so

$$\|\tilde{\mathbf{H}}_n \mathbf{y} - \|\mathbf{b}\| \mathbf{e}_1\| = \text{ minimum}$$

The GMRES Algorithm

Algorithm: GMRES

$\mathbf{q}_1 = \mathbf{b}/\|\mathbf{b}\|$
for $n = 1$ to $1, 2, 3, \ldots$
  Step n of Arnoldi iteration
  Find $\mathbf{y}$ to minimize $\|\tilde{\mathbf{H}}_n\mathbf{y} - \|\mathbf{b}\|\mathbf{e}_1\| = \|\mathbf{r}_n\|$
  $\mathbf{x}_n = \mathbf{Q}_n\mathbf{y}$

- The residual $\|\mathbf{r}_n\|$ does not need to be computed explicitly from $\mathbf{x}_n$

The GMRES Algorithm

- Least squares problem has Hessenberg structure, solved with QR factorization of $\tilde{\mathbf{H}}_n$
- If QR factorization of $\tilde{\mathbf{H}}_n$ is constructed from scratch, then it costs $O(n^2)$ flops, due to Hessenberg structure
- However, QR factorization of $\tilde{\mathbf{H}}_n$ can be updated from that of $\tilde{\mathbf{H}}_{n-1}$, using Givens rotation within $O(n)$ work
- However, memory and cost grow with n.
- In practice, restart the algorithm by clearing accumulated data. This might stagnate the method

GMRES and Polynomial Approximation

- GMRES can be interpreted as finding polynomial $p_n \in \mathbf{P}_n$ for $n = 1, 2, 3, \ldots$ where

$$\mathbf{P}_n = \{\text{polynomial } p \text{ of degree } \leqslant n \text{ with } p(0) = 1\}$$

  such that $\|p_n(\mathbf{A})\mathbf{b}\|$ is minimized

- Note that $\mathbf{r} = \mathbf{b} - \mathbf{A}\mathbf{K}_n\mathbf{c}$, where

$$\mathbf{A}\mathcal{K}_n\mathbf{c} = (c_1\mathbf{A} + c_2\mathbf{A}^2 + \cdots + c_n\mathbf{A}^n)\mathbf{b}$$

  and $\mathbf{r} = (1 - c_1\mathbf{A} - c_2\mathbf{A}^2 - \cdots - c_n\mathbf{A}^n)\mathbf{b}$.

- In other words, $p_n(z) = 1 - z(c_1 + c_2 z + \cdots + c_n z^n)$
- Invariance of GMRES
  - Scale invariance: If we change $\mathbf{A} \to \sigma\mathbf{A}$ and $\mathbf{b} \to \sigma\mathbf{b}$, then $\mathbf{r}_n \to \sigma\mathbf{r}_n$
  - Invariance under unitary similarity transformations: If change $\mathbf{A} \to \mathbf{U}\mathbf{A}\mathbf{U}^*$ for some unitary matrix $\mathbf{U}$ and $\mathbf{b} \to \mathbf{U}\mathbf{b}$, then $\mathbf{r}_n \to \mathbf{U}^*\mathbf{r}_n$

Convergence of GMRES

- GMRES converges monotonically and it converges after at most m steps
- Based on a polynomial analysis, diagonalizable $\mathbf{A} = \mathbf{V}\boldsymbol{\Lambda}\mathbf{V}^{-1}$ converges as
$$\frac{\|\mathbf{r}_n\|}{\|\mathbf{b}\|} \leqslant \kappa(\mathbf{V}) \inf_{p_n \in \mathbf{P}_n} \sup_{\lambda_i \in \boldsymbol{\Lambda}(\mathbf{A})} |p_n(\lambda_i)|$$
- In other words, if $\mathbf{A}$ is not far from normal (i.e., eigenvectors are nearly orthogonal), and if properly normalized degree n polynomials can be found whose size on the spectrum $\boldsymbol{\Lambda}(\mathbf{A})$ decreases quickly with n, then GMRES converges quickly

Other Krylov Subspace Methods

- CG on the Normal Equations (CGN)
  - Solve $\mathbf{A}^*\mathbf{A}\mathbf{x} = \mathbf{A}^*\mathbf{b}$ using Conjugate Gradients
  - Poor convergence due to squared condition number (i.e., $\kappa(\mathbf{A}^*\mathbf{A}) = \kappa(\mathbf{A})^2$)
  - One advantage is that it applies least squares problems without modification
- BiConjugate Gradients (BCG/BiCG)
  - Makes residuals orthogonal to another Krylov subspace, based on $\mathbf{A}^*$
  - It can be implemented with three-term recurrences, so memory requirements is smaller
  - Convergence sometimes comparable to GMRES, but unpredictable
- Conjugate Gradients Squared (CGS)
  - Avoids multiplication by $\mathbf{A}^*$ in BCG, sometimes twice as fast convergence as BCG
- Quasi-Minimal Residuals (QMR) and Stabilized BiCG (Bi-CGSTAB)
  - Variants of BiCG with more regular convergence

Some References

- **Templates for the Solution of Linear Systems: Building Blocks for Iterative Methods** (by R. Barrett and M. Berry and T. F. Chan and J. Demmel and J. Donato and J. Dongarra and V. Eijkhout and R. Pozo and C. Romine and H. Van der Vorst)
- **Templates for the Solution of Algebraic Eigenvalue Problems: a Practical Guide** (by Zhaojun Bai, James Demmel, Jack Dongarra, Axel Ruhe, and Henk van der Vorst)