AMS526: Numerical Analysis I (Numerical Linear Algebra)

Lecture 25: Preconditioning; Other Krylov Subspace Methods

Xiangmin Jiao

SUNY Stony Brook

December 11, 2008

Outline

1. Preconditioning

2. GMRES and Other Krylov Subspace Methods

Preconditioning

- Motivation: Convergence of iterative methods heavily depends on the condition number of the equation
- Main idea of preconditioning is to introduce a nonsingular matrix $\boldsymbol{M}$ such that $\boldsymbol{M}^{-1}\boldsymbol{A}$ has a smaller condition number. Thereafter, solve

  $$\boldsymbol{M}^{-1}\boldsymbol{A}\boldsymbol{x} = \boldsymbol{M}^{-1}\boldsymbol{b},$$

  which has the same solution as $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$
- Criteria of $\boldsymbol{M}$
  - Good approximation of $\boldsymbol{A}$
  - Ease of inversion
- However, if $\boldsymbol{A}$ is symmetric, $\boldsymbol{M}^{-1}\boldsymbol{A}$ breaks symmetry. How to resolve this problem?

Preconditioning of CG

- When preconditioning a symmetric matrix, use SPD matrix $\boldsymbol{M}$
- Let its Cholesky factorization be $\boldsymbol{M} = \boldsymbol{R}\boldsymbol{R}^*$. Solve

$$\boldsymbol{R}^{-1}\boldsymbol{A}\boldsymbol{R}^{-T}(\boldsymbol{R}^T\boldsymbol{x}) = \boldsymbol{R}^{-1}\boldsymbol{b}$$

- In practice, algorithm can be organized so that only $\boldsymbol{M}^{-1}$ (instead of $\boldsymbol{R}^{-1}$) appears

Algorithm: Preconditioned Conjugate Gradient Method

```
$\boldsymbol{x}_0 = 0$, $\boldsymbol{r}_0 = \boldsymbol{b}$, $\boldsymbol{p}_0 = \boldsymbol{M}^{-1}\boldsymbol{r}_0$, $\boldsymbol{z}_0 = \boldsymbol{p}_0$
for $n = 1$ to $1, 2, 3, \ldots$
        $\alpha_n = (\boldsymbol{r}_{n-1}^T \boldsymbol{z}_{n-1})/(\boldsymbol{p}_{n-1}^T \boldsymbol{A}\boldsymbol{p}_{n-1})$        step length
        $\boldsymbol{x}_n = \boldsymbol{x}_{n-1} + \alpha_n \boldsymbol{p}_{n-1}$        approximate solution
        $\boldsymbol{r}_n = \boldsymbol{r}_{n-1} - \alpha_n \boldsymbol{A}\boldsymbol{p}_{n-1}$        residual
        $\boldsymbol{z}_n = \boldsymbol{M}^{-1}\boldsymbol{r}_n$        preconditioning
        $\beta_n = (\boldsymbol{r}_n^T \boldsymbol{z}_n)/(\boldsymbol{r}_{n-1}^T \boldsymbol{z}_{n-1})$        improvement this step
        $\boldsymbol{p}_n = \boldsymbol{r}_n + \beta_n \boldsymbol{p}_{n-1}$        search direction
```

Commonly Used Preconditioners

- **Jacobi preconditioning**: $M = \text{diag}(A)$. Very simple and cheap, might improve certain problems but usually insufficient
- **Block-Jacobi preconditioning**: Let M be composed of block-diagonal instead of diagonal.
- **Classical iterative methods**: Precondition by applying one step of Jacobi, Gauss-Seidel, SOR, or SSOR
- **Incomplete factorizations**: Perform Gaussian elimination but ignore fill
- **Multigrid (coarse-grid approximations)**: For a PDE discretized on a grid, a preconditioner can be formed by transferring the solution to a coarser grid, solving a smaller problem, then transferring back. This is the most efficient approach if applicable

Outline

1. Preconditioning

2. GMRES and Other Krylov Subspace Methods

Minimizing Residual

- CG only works for SPD matrices.
- 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 is the following
    - Find $\boldsymbol{x}_n \in \mathcal{K}_n$ that minimizes $\|\boldsymbol{r}_n\| = \|\boldsymbol{b} - \boldsymbol{A}\boldsymbol{x}_n\|$
    - This can be viewed as a least squares problem: Find a vector $\boldsymbol{c}$ s.t.$\|\boldsymbol{A}\boldsymbol{K}_n\boldsymbol{c} - \boldsymbol{b}\|$ is minimized, where $\boldsymbol{K}_n$ is the $m \times n$ Krylov matrix composed of basis vectors of $\mathcal{K}_n$
    - Orthogonal basis is often used, produced by Arnoldi iteration

Review: Arnoldi Iteration

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

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

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

Minimal Residual with Orthogonal Basis

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

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

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

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

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

$$\left\|\tilde{\boldsymbol{H}}_n \boldsymbol{y} - \|\boldsymbol{b}\| \boldsymbol{e}_1\right\| = \text{minimum.}$$

The GMRES Algorithm

```
Algorithm: GMRES
  $q_1 = b/\|b\|$
  for $n = 1$ to $1, 2, 3, \ldots$
      Step $n$ of Arnoldi iteration
      Find $y$ to minimize $\left\|\tilde{H}_n y - \|b\| e_1\right\| = \|r_n\|$
      $x_n = Q_n y$
```

- The residual $\|r_n\|$ does not need to be computed explicitly from x_n
- Least squares problem has Hessenberg structure, solve with QR factorization of $\tilde{H}_n$ (computed by updating the factorization of $\tilde{H}_{n-1}$)
- Memory and cost grow with n. Restart the algorithm by clearing accumulated data (might stagnate the method)

Convergence of GMRES

- GMRES converges monotonically and it converges after at most m steps
- The residual $\boldsymbol{r}_n = p_n(\boldsymbol{A})\boldsymbol{b}$, where $p_n \in P_n$ is a degree n polynomial with $p(0) = 1$, so GMRES also finds a minimizing polynomial such that $\|p_n(\boldsymbol{A})\boldsymbol{b}\| = \text{minimum}$
- Based on a polynomial analysis, diagonalizable $\boldsymbol{A} = \boldsymbol{V}\boldsymbol{\Lambda}\boldsymbol{V}^{-1}$ converges as
$$\frac{\|\boldsymbol{r}_n\|}{\|\boldsymbol{b}\|} \leq \kappa(V) \inf_{p_n \in P_n} \sup_{\lambda_i \in \Lambda(\boldsymbol{A})} |p_n(\lambda_i)|$$
- In other words, if $\boldsymbol{A}$ 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 $\Lambda(\boldsymbol{A})$ decreases quickly with n, then GMRES converges quickly

Other Krylov Subspace Methods

- CG on the Normal Equations (CGN)
  - Solve $\boldsymbol{A}^T\boldsymbol{A}\boldsymbol{x} = \boldsymbol{A}^T\boldsymbol{b}$ using Conjugate Gradients
  - Poor convergence due to squared condition number (i.e., $\kappa(\boldsymbol{A}^T\boldsymbol{A}) = \kappa(\boldsymbol{A})^2$).
- BiConjugate Gradients (BiCG)
  - Makes residuals orthogonal to another Krylov subspace, based on $\boldsymbol{A}^*$
  - Memory requirements only constant number of vectors
  - Convergence sometimes comparable to GMRES, but unpredictable
- Conjugate Gradients Squared (CGS)
  - Avoids multiplication by $\boldsymbol{A}$, sometimes twice as fast convergence
- Quasi-Minimal Residuals (QMR) and Stabilized BiCG (Bi-CGSTAB)
  - Variants of BiCG with more regular convergence