Welcome back to Ma 221! Lecture 40, Dec 1

Finish CG, some Preconditioning

Derived CG from (*) $x_k = Q_k T_k^{-1} e_1 \|b\|_2$

Need recurrences for

x_k = solution

r_k = residual

p_k = conjugate gradient

(R_r') $r_k = r_{k-1} - A \cdot p_k \cdot \nu_k$

(R_x') $x_k = x_{k-1} + p_k \nu_k$

(R_p') $p_k = r_{k-1} + \mu_k p_{k-1}$

Need formulas for ν_k, μ_k

(see book for details)

eg: take (R_r') multiply by r_{k-1}^T

$\Rightarrow$ $r_{k-1}^T r_k = 0 = r_{k-1}^T r_{k-1} - r_{k-1}^T A p_k \cdot \nu_k$

$\Rightarrow$ $\nu_k = \dfrac{r_{k-1}^T r_{k-1}}{r_{k-1}^T A p_k}$ (see book)

Final Formulas:

$$\nu_k = \frac{r_{k-1}^T r_{k-1}}{p_k^T A p_k}$$

$$\mu_k = \frac{r_{k-1}^T r_{k-1}}{r_{k-2}^T r_{k-2}}$$

Altogether: CG for $Ax=b$

$k=0$, $x_0=0$, $r_0=b$, $p_1=b$

repeat

$k = k+1$

$z = A \cdot p_k$

$\nu_k = (r_{k-1}^T r_{k-1}) / (p_k^T \cdot z)$

$x_k = x_{k-1} + \nu_k \cdot p_k$

$r_k = r_{k-1} - \nu_k \cdot z$

$\mu_k = \dfrac{r_k^T r_k}{r_{k-1}^T r_{k-1}}$

$p_{k+1} = r_k + \mu_{k+1} \cdot p_k$

until $\|r_k\|_2$ small enough

Convergence:

Thm $\dfrac{\|r_k\|_{A^{-1}}}{\|r_0\|_{A^{-1}}} \leq \dfrac{2}{1 + \dfrac{2k}{\sqrt{\kappa(A)} - 1}}$

when $\kappa(A) \gg 1$ need

$O(\sqrt{\kappa(A)})$ steps to converge

proof uses Chebyshev polynomials.

For d-dimensional Poisson on n^d mesh

$\kappa(A) \sim O(h^2) \Rightarrow$ CG take $O(n)$ steps to converge

Preconditioning to accelerate convergence

How to change $Ax=b$ cheaply to converge faster:

Simplest $M^{-1}Ax = M^{-1}b$ where $M \approx A$

(1) M^{-1} cheaper to multiply by than A

(2) $M^{-1}A$ better conditioned than A

Idea straightforward for GMRES, just multiply by A and M^{-1}, CG trickier

A s.p.d. $\Rightarrow$ choose M s.p.d., but

$M^{-1}A$ not s.p.d.

M s.p.d. $\Rightarrow$ $M = Q\Lambda Q^T \Rightarrow M^{1/2} = Q\Lambda^{1/2}Q^T$

Imagine applying CG to

(*) $\underbrace{(M^{-1/2}AM^{-1/2})}_{\text{s.p.d.}}(M^{1/2}x) = M^{-1/2}b$ (Same as $Ax=b$)

$M^{-1}A$ and $M^{-1/2}AM^{-1/2} = M^{\frac{1}{2}}(M^{-1}A)M^{-\frac{1}{2}}$ are similar

$\Rightarrow$ same $\frac{\lambda_{max}}{\lambda_{min}}$

Preconditioned CG (equivalent to CG on (*))

$k=0,\ x(0)=0,\ r(0)=b,\ p(1)=M^{-1}b,\ y(0)=M^{-1}r(0)$

repeat

$k = k+1$

$z = A \cdot p(k)$

$\nu(k) = (y(k-1)^T \cdot r(k-1)) / (p(k)^T \cdot z)$

$x(k) = x(k-1) + \nu(k) \cdot p(k)$

$r(k) = r(k-1) - \nu(k) \cdot z$

$y(k) = M^{-1} \cdot r(k)$

$\mu(k+1) = (y(k)^T \cdot r(k)) / (y(k-1)^T \cdot r(k-1))$

$p(k+1) = y(k) + \mu(k+1) \cdot p(k)$

until $\|r(k)\|_2$ small enough

Thm 6.9: equivalent to (*)

HW Q 6.14 (see errata on p 317)

How to pick a good M?

Goals: (1) cheap to multiply by M^{-1}

(2) $\text{cond}(M^{-1}A) \ll \text{cond}(A)$

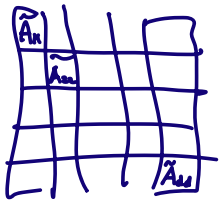
(1) If A s.p.d. $M = \text{diag}(A_{11}, A_{22}, \ldots, A_{nn})$

"Jacobi preconditioning"

$M^{-1}A = I - R_J$

$\rho(R_J) \ll 1 \Rightarrow M^{-1}A$ well conditioned

Thm (van der Sluis, 1969) if M diagonal then $M = \text{diag}(A_{11}, \ldots, A_{nn})$ within factor n of optimal for minimizing $\text{cond}(M^{-1/2} A M^{-1/2})$

(2) $M = \mathrm{diag}(\tilde{A}_{11}, \cdots, \tilde{A}_{dd})$ $A =$ 

more expensive than plain diagonal, but can be faster

"block Jacobi preconditioning"

each $\tilde{A}_{ii}$ could correspond to a different subset of physical simulation

Thm (D., 2023) above choice of M within factor of d of optimal over any block diagonal M, for minimizing $\mathrm{cond}(M^{-1/2} A M^{-1/2})$

(3) "Incomplete Cholesky"

compute partial Cholesky $LL^T \approx A$ by limiting fill-in in L

Easiest: no fill-in. $A_{ij} = 0 \Rightarrow L_{ij} = 0$

ILU(0) $\Rightarrow$ need triangular solve to

(can also do ILU(k)) Multiply by $M^{-1} = (L^T)^{-1} L^{-1}$

$\mathrm{cost} = O(\mathrm{nnz}(L)) = O(\mathrm{nnz}(A))$

same cost as Ax

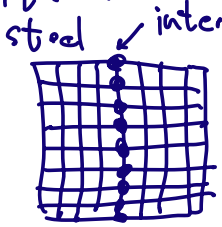
(4) A few steps of multigrid (perhaps just applied to $\tilde{A}_{ii}$, or to coarsest mesh

(5) Domain Decomposition

Ex: discretize FE problem on 2D mesh

Left is steel, right is concrete

Interface between them



$n \times n$ mesh

$$A = \begin{bmatrix} A_{ss} & 0 & A_{si} \\ 0 & A_{cc} & A_{ci} \\ A_{is} & A_{ic} & A_{ii} \end{bmatrix} \quad \begin{matrix} \frac{n(n-1)}{2} \\ \frac{n(n-1)}{2} \\ n \end{matrix}$$

$$A = \begin{bmatrix} I & 0 & 0 \\ 0 & I & 0 \\ A_{is}A_{ss}^{-1} & A_{ic}A_{cc}^{-1} & I \end{bmatrix} \begin{bmatrix} I & 0 & 0 \\ 0 & I & 0 \\ 0 & 0 & S \end{bmatrix} \begin{bmatrix} A_{ss} & 0 & A_{si} \\ 0 & A_{cc} & A_{ci} \\ 0 & 0 & I \end{bmatrix}$$

$$\text{Schur complement} = S = A_{ii} - A_{is}A_{ss}^{-1}A_{si} - A_{ic}A_{cc}^{-1}A_{ci}$$

$$A^{-1} = \begin{bmatrix} A_{ss}^{-1} & 0 & -A_{ss}^{-1}A_{si} \\ 0 & A_{cc}^{-1} & -A_{cc}^{-1}A_{ci} \\ 0 & 0 & I \end{bmatrix} \cdot \begin{bmatrix} I & & \\ & I & \\ & & S^{-1} \end{bmatrix} \cdot \begin{bmatrix} I & 0 & 0 \\ 0 & I & 0 \\ -A_{is}A_{ss}^{-1} & -A_{ic}A_{cc}^{-1} & I \end{bmatrix}$$

need to multiply by A^{-1} approximately

useful if good preconditioners for A_{ss}^{-1}, A_{cc}^{-1} available

Ex: multiply $(A_{ss}^{-1}A_{si})x = A_{ss}^{-1}(A_{si}x)$

$A_{si}x$: spmv

A_{ss}^{-1}: use preconditioned solver for A_{ss}

What about S^{-1}?

Can multiply by S:

$$Sx = (A_{ii}x) - A_{is}(A_{ss}^{-1}(A_{si}x)) - A_{ic}(A_{cc}^{-1}(A_{ci}x))$$

Given ability to multiply S, use CG to solve $Sx = b$, ie $S^{-1}b$

Often S much better conditioned, and smaller than A

Thm $\text{cond}(S) \leq \text{cond}(A)$ if A s.p.d.

Ex for Poisson $\kappa(S) = \sqrt{\kappa(A)}$

called "non overlapping domain decomposition"

can also do "overlapping domain decomposition"