Welcome to Ma221! (Apr 21)

Krylov Subspace Methods

GMRES and CG for $Ax=b$

Krylov Subspace $\mathcal{K}_k = \text{span}\{b, Ab, A^2b, \ldots, A^{k-1}b\}$

orthogonal basis $Q_k = [q_1, \ldots, q_k]$ known

$Q_u = [q_{k+1}, \ldots, q_n]$

Goal: find "best" approximation x_k to $A^{-1}b$ in $\mathcal{K}_k$: $x_k \in \mathcal{K}_k$

(1) Choose x_k to minimize $\|x_k - x\|_2$ $\quad x = A^{-1}b$

but don't have enough information

All we have is $H_k = Q_k^T A Q_k$ (T_k if $A=A^T$)

(2) Choose x_k to minimize residual

$\|r_k\|_2 \quad r_k = b - Ax_k$

2 Algorithms:

A general: use GMRES

(generalized minimum residual)

$A=A^T$ use MINRES

(3) Choose x_k so $r_k \perp \mathcal{K}_k$ $\quad r_k^T Q_k = 0$

"orthogonal residual property"

or a Galerkin condition

$A = A^T \Rightarrow$ use SYMMLQ

A general $\Rightarrow$ variant of GMRES

(4) A s.p.d. $\Rightarrow$ define norm

$$\|r\|_{A^{-1}} = (r^T A^{-1} r)^{1/2}$$

"best" solution minimizes

$$\begin{aligned}\|r_k\|_{A^{-1}}^2 &= r_k^T A^{-1} r_k \\ &= (b - Ax_k)^T A^{-1} (b - Ax_k) \\ &= (Ax - Ax_k)^T A^{-1} (Ax - Ax_k) \\ &= (x - x_k)^T A^T \cancel{A^{-1} A} (x - x_k) \\ &= (x - x_k)^T A (x - x_k) \\ &= \|x - x_k\|_A^2\end{aligned}$$

alg is Conjugate Gradient (CG)

Thm: A s.p.d. $\Rightarrow$ defs (3) and (4) of "best" are equivalent. CG cost: one matrix vector multiply + few dot products + saxpys per iteration, only need to store 3 vectors

GMRES: fewest assumptions on A

CG: (nearly) most " " "

see fig 6.8 in text for more options

also see link to Templates for Solution to Linear Systems on webpage

GMRES: choose x_k to minimize

$$\|r_k\|_2 = \|b - A x_k\|_2$$

$$x_k = Q_k y_k \in \mathcal{K}_k$$

$$\|r_k\|_2 = \|b - A Q_k y_k\|_2$$

$$= \Big\| b - A \underbrace{[Q_k, Q_o]}_{\substack{Q \\ \text{orthogonal}}} \begin{bmatrix} y_k \\ 0 \end{bmatrix} \Big\|_2$$

$$= \| Q^T (\text{''}) \|_2$$

$$= \Big\| Q^T b - \underbrace{Q^T A Q}_{H} \begin{bmatrix} y_k \\ 0 \end{bmatrix} \Big\|_2$$

$$= \Big\| \|b\|_2 e_1 - H \begin{bmatrix} y_k \\ 0 \end{bmatrix} \Big\|_2$$

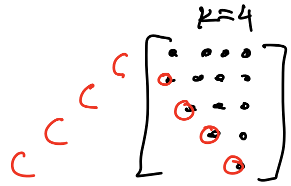
$$= \Bigg\| \|b\| e_1 - \begin{array}{cc} \text{known} & \text{unknown} \\ \left[\begin{array}{c|c} H_k & H_{ok} \\ \hline H_{ku} & H_{ou} \end{array}\right] & \end{array} \begin{bmatrix} y_k \\ 0 \end{bmatrix} \Bigg\|_2$$

$$H_{ku} = \begin{bmatrix} 0 & \boxed{H_{k+1,k}} \end{bmatrix}$$

$$= \Bigg\| \|b\|_2 e_1 - \begin{array}{cc} & k \\ k+1 & \begin{bmatrix} H_k \\ 0 \cdots 0 \; H_{k+1,k} \end{bmatrix} \end{array} y_k \Bigg\|_2$$

= $k+1$ by k least squares problem

Cheap to solve using Givens Rotations



cost $= O(k^2)$

or $O(k)$ per iteration

Conjugate Gradient Algorithm (CG)

Lemma: CG "best" in two senses:

(3) and (4) are equivalent

(3): choose x_k so $r_k \perp Q_k$, $r_k^T Q_k = 0$

(4) choose x_k to minimize $\|r_k\|_{A^{-1}}^2 = r_k^T A^{-1} r_k$

Both solved by

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

where $T_k =$ tridiagonal matrix

computed by Lanczos: $T_k = Q_k^T A Q_k$

also $r_{k+1} = \pm \|r_{k+1}\|_2 \cdot q_{k+1}$

Intuition for (*)

→ Multiplying $Q_k^T b = e_1 \|b\|_2$ projects b onto $\mathcal{K}_k$

→ Multiplying by T_k^{-1} solves projected problem

→ Multiplying by Q_k maps projection back to $\mathbb{R}^n$

Proof: drop subscript k:

$Q = Q_k, \quad T = T_k, \quad x = Q T^{-1} e_1 \|b\|_2$

$r = b - Ax \qquad T = Q^T A Q \quad (\text{spd})$

$$
\begin{aligned}
Q^T r &= Q^T (b - Ax) \\
&= Q^T b - Q^T A x \\
&= e_1 \|b\|_2 - \underbrace{Q^T A}(Q T^{-1} e_1 \|b\|_2) \\
&= e_1 \|b\|_2 - \underbrace{T T^{-1}}_{I} e_1 \|b\|_2 \\
&= 0
\end{aligned}
$$

show that x minimizes $\|r\|_{A^{-1}}^2$

$x' = x + Qz, \quad r' = b - Ax' = r - AQz$

$$
\begin{aligned}
\|r'\|_{A^{-1}}^2 &= r'^T A^{-1} r' \\
&= (r - AQz)^T A^{-1} (r - AQz) \\
&= r^T A^{-1} r + \underline{\quad ? \quad} + (AQz)^T A^{-1} (AQz) \\
&= \|r\|_{A^{-1}}^2 + \underbrace{\quad ? \quad}_{\text{if} = 0} + \|AQz\|_{A^{-1}}^2
\end{aligned}
$$

then $\|r'\|_A^2 \geq \|r\|_{A^{-1}}^2 = \text{minimum}$

$$
\begin{aligned}
&-2 (AQz)^T A^{-1} r \\
&= -2 z^T Q^T \underbrace{A A^{-1}}_{= I} r \\
&= -2 z^T \underbrace{Q^T r}_{= 0} = 0
\end{aligned}
$$

$$r_k = b - A \underbrace{x_k}_{\in \mathcal{K}_k} \in \mathcal{K}_{k+1}$$

r_k in span of $\mathcal{K}_{k+1}$ but not $\mathcal{K}_k$

Q_{k+1} but not Q_k

$\Rightarrow r_k$ must be multiple of q_{k+1} because $r_k^T Q_k = 0$

$\Rightarrow \quad r_k = \pm \|r_k\|_2 \cdot q_{k+1}$

Derive CG starting 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

only keep most recent vectors in memory

(1) p_k called gradient because each step of CG moves x_k in direction p_k

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

until x_k minimizes $\|r_k\|_{A^{-1}}$ over all choices of ν

(2) p_k called conjugate (A-conjugate): othogonal w.r.t inner product of A:

$$p_k^T A p_j = 0 \quad \text{if } k \neq j$$

Can also derive CG by starting with properties (1) and (2), showing they satisfy (*)

T_k = spd + tridiagonal ⇒ use Cholesky

$T_k = L_k' L_k'^T$, L_k' lower bidiagonal

$$T_k = L_k' L_k'^T = L_k D_k L_k^T$$

L_k unit diagonal, D_k diagonal

$$D_k(i,i) = \left(L_k'(i,i)\right)^2$$

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

$$= Q_k (L_k D_k L_k^T)^{-1} e_1 \|b\|$$

$$= \left[Q_k L_k^{-T}\right] \cdot \left[D_k^{-1} L_k^{-1} e_1 \|b\|_2\right]$$

$$= P_k' \cdot y_k$$

$$P_k' = [p_1', p_2', \dots, p_k']$$

eventual conjugate gradients p_k are scalar multiples of p_k'

Can prove property (2)

Lemma: p_k' are A-conjugate

$$P_k'^T A P_k' \text{ diagonal}$$

Proof: $P_k'^T A P_k' = (Q_k L_k^{-T})^T A (Q_k L_k^{-T})$

$$= L_k^{-1} \underbrace{Q_k^T A Q_k} L_k^{-T}$$

$$= L_k^{-1} \quad T_k \quad L_k^{-T}$$

$$= \underbrace{L_k^{-1} (L_k}_{I} D_k \underbrace{L_k^T) L_k^{-T}}_{I}$$

$$= D_k$$

Need to derive recurrences for columns p_k' of P_k' and components of y_k

Need $P_k' = [P_{k-1}', p_k']$

and $y_k = \begin{bmatrix} s_1 \\ \vdots \\ s_k \end{bmatrix} = \begin{bmatrix} y_{k-1} \\ s_k \end{bmatrix}$

If true, get recurrence

(Rx) $x_k = P_k' y_k = [P_{k-1}', p_k'] \cdot \begin{bmatrix} y_{k-1} \\ s_k \end{bmatrix}$

$$= P_{k-1}' y_{k-1} + p_k' \cdot s_k$$

$$= x_{k-1} + p_k' s_k$$

also need recurrences for p_k' and s_k

Since Lanczos constructs T_k row by row
T_{k-1} is leading $k-1 \times k-1$ submatrix of T_k

Since Cholesky also row by row, L_{k-1} and D_{k-1} are leading $k-1 \times k-1$ submatrices of L_k and D_k

$$T_k = L_k D_k L_k^T$$
$$= \begin{bmatrix} L_{k-1} & \begin{matrix} 0 \\ 0 \\ 0 \end{matrix} \\ 0\,0 \cdots 0\, l_{k-1} & 1 \end{bmatrix} \begin{bmatrix} D_{k-1} & \\ & d_k \end{bmatrix} \begin{bmatrix} L_{k-1} & \begin{matrix} 0 \\ \vdots \\ 0 \end{matrix} \\ 0 \cdots 0\, l_{k-1} & 1 \end{bmatrix}^T$$

$$\Rightarrow L_k^{-1} = \begin{bmatrix} L_{k-1}^{-1} & \begin{matrix} 0 \\ \vdots \\ 0 \end{matrix} \\ \text{stuff} & 1 \end{bmatrix}$$

$$y_k = D_k^{-1} L_k^{-1} e_1 \|b\|_2$$
$$= \begin{bmatrix} D_{k-1}^{-1} & \\ & d_k^{-1} \end{bmatrix} \begin{bmatrix} L_{k-1}^{-1} & \begin{matrix} 0 \\ \vdots \\ 0 \end{matrix} \\ \text{stuff} & 1 \end{bmatrix} e_1 \cdot \|b\|_2$$
$$= \begin{bmatrix} D_{k-1}^{-1} L_{k-1}^{-1} e_1 \|b\|_2 \\ s_k \end{bmatrix} = \begin{bmatrix} y_{k-1} \\ s_k \end{bmatrix}$$

$$P_k' = Q_k \cdot L_k^{-T} = [Q_{k-1}, q_k] \begin{bmatrix} L_{k-1}^{-T} & \text{stuff} \\ 0 \cdots 0 & 1 \end{bmatrix}$$
$$= [Q_{k-1} L_{k-1}^{-T}, p_k'] = [P_{k-1}', p_k']$$

To get recurrence for p_k', write

$$Q_k = P_k' \cdot L_k^T \qquad L_k^T = \begin{bmatrix} 1 & & 0 \\ & \ddots & 0 \\ & & x \\ & & 1 \end{bmatrix}$$

equate last columns:

$$q_k = p_k' + p_{k-1}' \cdot L_k(k, k-1)$$

or

$$(R_p) \quad p_k' = q_k - p_{k-1}' \cdot l_{k-1}$$

Need recurrence for r_k: use (R_x)

$$(R_r) \quad r_k = b - A x_k$$

$$= \underbrace{b - A(x_{k-1}}_{} + p_k' s_k)$$

$$= r_{k-1} - A p_k' \cdot s_k$$

All recurrences:

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

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

$$(R_p) \quad p_k' = q_k - l_{k-1} \cdot p_{k-1}'$$

Substitute $q_k = r_{k-1} / \|r_{k-1}\|_2$

$$p_k = \|r_{k-1}\|_2 \cdot p_k'$$

$$(R_r') \quad r_k = r_{k-1} - A p_k (s_k / \|r_{k-1}\|_2)$$

$$= r_{k-1} - A \cdot p_k \cdot \nu_k$$

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

(Rp') $p_k = r_{k-1} - (\|r_{k-1}\|_2 \cdot \ell_{k-1} / \|r_{k-2}\|_2) \cdot p_{k-1}$

$= r_{k-1} + \mu_k \cdot p_{k-1}$

Need formulas for scalars ν_k, μ_k

see book for details

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

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

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

final formulas!

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

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

All together

CG for $Ax=b$

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

repeat

$k=k+1$

$z = Ap_k$

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

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

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

$$\beta_{k+1} = \frac{r_k^T \cdot r_k}{r_{k-1}^T \cdot r_{k-1}}$$

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

until $\|r_k\|_2$ small enough

Convergence:

Thm $$\frac{\|r_k\|_{A^{-1}}}{\|r_0\|_{A^{-1}}} \le \frac{2}{1 + \frac{2k}{\sqrt{\underbrace{\kappa(A)}_{\text{condition number}} - 1}}}$$

when $\kappa(A)$ large, need

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

For d-dimensional Poisson on mesh with n vertices in each direction: $\kappa(A) \approx O(n^2)$

so CG takes $O(n)$ steps to converge