Welcome to Ma 221! Lecture 26, Fall 24

Conjugate Gradients (CG)

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

orthogonal $Q_k = [q_1, \ldots, q_k]$ orthogonal basis of $\mathcal{K}_k$

Goal: find "best" solution to $Ax = b$
for $x \in \mathcal{K}_k(A,b)$

(3) Choose x_k so $r_k = Ax_k - b \perp \mathcal{K}_k$

$r_k^T Q_k = 0$

"orthogonal residual"

(4) A s.p.d: "best solution"

minimizes $\|r_k\|_{A^{-1}}^2 = r_k^T A^{-1} r_k$

$= \|x_k - x\|_A^2$

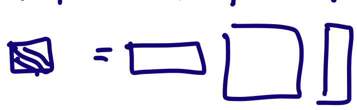
$= (x_k - x)^T A (x_k - x)$

Lemma: CG "best" because it satisfies
(3) and (4)

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

where T_k = tridiagonal matrix
from Lanczos:

$T_k = Q_k^T A Q_k$



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 exactly

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

Proof of Lemma: drop subscript k:

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

$$r = b - Ax, \quad T = Q^T A Q$$

$$\begin{aligned} Q^T r &= Q^T (b - Ax) \\ &= Q^T b - Q^T A x \\ &= e_1 \|b\|_2 - Q^T A (Q T^{-1} e_1 \|b\|_2) \\ &= e_1 \|b\|_2 - \underbrace{\underbrace{(Q^T A Q)}_{T} T^{-1}}_{I} e_1 \|b\|_2 \\ &= 0 \Rightarrow \text{prop 3) holds} \end{aligned}$$

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

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

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

$\geq \|r\|^2_{A^{-1}}$ which is minimum if middle $= 0$

$$\text{middle} = -(AQz)^T A^{-1} r - r^T A^{-1}(AQz)$$
$$= -2(AQz)^T A^{-1} r$$
$$= -2 z^T Q^T \underbrace{A A^{-1}}_{I} r$$
$$= -2 z^T \underbrace{Q^T r}_{0} \implies \text{prop 4) holds}$$

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

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

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

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

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 3 most recent vectors in memory

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

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

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

(2) p_k called conjugate (A conjugate): orthogonal w.r.t. inner product defined by A: $p_k^T A p_j = 0$ if $j \neq k$

Can also derive CG starting from (1) and (2), showing they satisfy (*)

T_k = s.p.d. and tridiagonal $\Rightarrow$ use Cholesky

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

$$\begin{bmatrix} \diagdown\!\!\diagdown & 0 \\ 0 & \end{bmatrix}$$

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

$$= \begin{bmatrix} 1 & & 0 \\ \ddots & \ddots & \\ 0 & & 1 \end{bmatrix} \cdot \begin{bmatrix} \diagdown & 0 \\ 0 & \diagdown \end{bmatrix} \begin{bmatrix} 1 & \ddots & 0 \\ & \ddots & \\ 0 & & 1 \end{bmatrix}$$

L_k has unit diagonal,

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

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

$$= Q_k \left(L_k D_k L_k^T\right)^{-1} e_1 \|b\|_2$$

$$= Q_k \left(L_k^{-T} D_k^{-1} L_k^{-1}\right) e_1 \|b\|_2$$

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

$$= P_k' \cdot y_k$$

where $P_k' = [p_1', p_2', \ldots, p_k']$, $y_k = \begin{bmatrix} y_{k-1} \\ s_k \end{bmatrix}$

eventual conjugate gradients p_k are scalar multiples of p_k'

Proof of 2): p_k' are A-conjugate, i.e. $P_k'^T A P_k'$ diagonal:

$$= (Q_k L_k^{-T})^T A (Q_k L_k^{-T})$$

$$= L_k^{-1} (Q_k^T A Q_k) L_k^{-T}$$

$$= L_k^{-1} \cdot (T_k) \cdot L_k^{-T}$$

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

$$= D_k$$

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

need $P_k' = [P_{k-1}', p_k']$ $y_k = \begin{bmatrix} s_1 \\ \vdots \\ s_k \end{bmatrix} = \begin{bmatrix} y_{k-1} \\ s_k \end{bmatrix}$

If true, get recurrence

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

$= P'_{k-1} \cdot y_{k-1} + p'_k \cdot s_k$

$= x_{k-1} + p'_k \cdot s_k$

also need recurrences for p'_k and s_k

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

Since Cholesky works "top-down", L_{k-1} and D_{k-1}
are also leading $k-1$ by $k-1$ submatrices
of L_k and D_k

$$T_k = L_k D_k L_k^T$$
$$= \left[\begin{array}{c|c} L_{k-1} & \begin{matrix}0\\0\\ \vdots \\0\end{matrix} \\ \hline 0 \cdots 0\, l_{k-1} & 1 \end{array}\right] \cdot \left[\begin{array}{c|c} D_{k-1} & 0 \\ \hline 0 & d_k \end{array}\right] \left[\begin{array}{c|c} L_{k-1} & 0 \\ \hline 0 \cdots 0\, l_{k-1} & 1 \end{array}\right]^T$$

$$\Rightarrow L_k^{-1} = \left[\begin{array}{c|c} L_{k-1}^{-1} & 0 \\ \hline \text{stuff} & 1 \end{array}\right]$$

$$y_k = D_k^{-1} L_k^{-1} e_1 \|b\|_2$$
$$= \left[\begin{array}{c|c} D_{k-1}^{-1} & 0 \\ \hline 0 & d_k^{-1} \end{array}\right] \left[\begin{array}{c|c} L_{k-1}^{-1} & 0 \\ \hline \text{stuff} & 1 \end{array}\right] e_1 \cdot \|b\|_2$$
$$= \begin{bmatrix} D_{k-1}^{-1} L_{k-1}^{-1} e_1 \|b\|_2 \\ s_k \end{bmatrix} = \begin{bmatrix} y_{k-1} \\ s_k \end{bmatrix}$$

$$P'_k = Q_k \cdot L_k^{-T} = [Q_{k-1}, q_k] \cdot \left[\begin{array}{c|c} L_{k-1}^{-T} & \text{stuff} \\ \hline 0 & 1 \end{array}\right]$$
$$= [Q_{k-1} L_{k-1}^{-T}, p'_k] = [P'_{k-1}, p'_k]$$

To get recurrence for p_k', write

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

equate last columns:

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

or

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

Also need recurrence for r_k: use (R_x)

$$(R_r) \quad r_k = b - A x_k$$
$$= \underbrace{b - A(x_{k-1}}_{} + p_k' s_k)$$
$$= r_{k-1} - A p_k' s_k$$

All recurrences:

$$(R_r) \quad r_k = r_{k-1} - A p_k' s_k$$
$$(R_x) \quad x_k = x_{k-1} + p_k' \cdot s_k$$
$$(R_p) \quad p_k' = q_k - \ell_{k-1} p_{k-1}'$$

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

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

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

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

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

just 3 vectors: p, x, r

Need formulas for ν_k and μ_k:

Take (Rr'), multiply by r_{k-1}^T, get

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

$$0 = \quad ''$$

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

More stable version:

$$(Rp')^T \cdot A p_k \Rightarrow$$

$$(p_k = r_{k-1} + \mu_k p_{k-1})^T A p_k$$

$$\Rightarrow p_k^T A p_k = r_{k-1}^T A p_k + \mu_k \cdot \underbrace{p_{k-1}^T A p_k}_{=0}$$

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

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

Formula for μ_k

$$(Rp')^T A p_{k-1} \Rightarrow$$

$$(p_k = r_{k-1} + \mu_k p_{k-1})^T A p_{k-1}$$

$$\Rightarrow \underbrace{p_k^T A p_{k-1}}_{=0} = r_{k-1}^T A p_{k-1} + \mu_k p_{k-1}^T A p_{k-1}$$

$$\Rightarrow \mu_k = \frac{-r_{k-1}^T A p_{k-1}}{p_{k-1}^T A p_{k-1}}$$

avoid extra dot product in numerator

$$(R r')^T r_k \Rightarrow$$

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

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

$$\Rightarrow \nu_k = \frac{-r_k^T r_k}{p_k^T A r_k}$$

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

$$\Rightarrow \frac{r_k^T r_k}{r_{k-1}^T r_{k+1}} = -\frac{p_k^T A r_k}{p_k^T A p_k} = \mu_{k+1}$$

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

Put it altogether to get CG for $Ax=b$

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

repeat

$k = k+1$

$z = A \cdot p_k$

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

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

(Rr') $r_k = r_{k-1} - \nu_k \cdot z$

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

$(R_{p'})$ $p_{k+1} = r_k + \nu_{k+1} \cdot p_k$

until $\|r_k\|_2$ small enough

cost: 2 dot products $r_k^T r_k$ and $p_k^T z$

3 saxpys

1 $A \cdot p_k$

Convergence:

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

if cond(A) large, need $O(\sqrt{\text{cond}(A)})$ steps to converge

For d-dimensional Poisson on mesh with n vertices in each direction, $\text{cond}(A) \sim O(n^2)$

$\Rightarrow$ CG takes $O(n)$ steps to converge