Welcome back to Ma221! Lecture 39, Nov 29

Krylov Subspace Methods: GMRES and CG

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

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

$Q_u = [q_{k+1} \cdots q_n]$ unknown

GMRES: choose x_k to minimize

$\|r_k\|_2 = \|b - Ax_k\|_2 \quad x_k = Q_k y_k \in \mathcal{K}_k$

$\|r_k\|_2 = \|b - AQ_k y_k\|_2$

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

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

$= \| Q^T b - \underset{\substack{\| \\ H}}{Q^T A Q} \begin{bmatrix} y_k \\ 0 \end{bmatrix} \|_2$

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

$= \| \|b\|_2 \cdot e_1 - \begin{array}{c} k \\ \left[\begin{array}{c|c} H_k & H_{uk} \\ \hline H_{ku} & H_u \end{array}\right] \end{array} \begin{array}{c} k \\ \\ \end{array} \begin{bmatrix} y_k \\ 0 \end{bmatrix} \|_2$

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

$= \| \|b\|_2 \cdot e_1 - \begin{bmatrix} H_k \\ 0 \cdots 0 \; H_{k+1,k} \end{bmatrix} [y_k] \|_2$

$\uparrow$

$k+1$ by k least squares problem

cheap to solve using Given rotations

$$\begin{array}{c} k=4 \\ \begin{bmatrix} \bullet & \bullet & \bullet & \bullet \\ \bullet & \bullet & \bullet & \bullet \\ & \bullet & \bullet & \bullet \\ & & \bullet & \bullet \\ & & & \bullet \end{bmatrix} \end{array} \qquad \text{cost} = O(k^2) \text{ or } k \text{ per iteration}$$

Conjugate Gradients (CG)

Lemma: CG "best" in two ways (equivalent)

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

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

Both solved by

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

T_k = tridiagonal from Lanczos $\quad T_k = Q_k^T A Q_k$

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

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: $Q = Q_k$, $T = T_k$

$$x = QT^{-1}e_1 \|b\|_2$$

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

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

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

$$\begin{aligned}
\|r'\|_{A^{-1}}^2 &= r'^T A^{-1} r' \\
&= (r - AQz)^T A^{-1} (r - AQz) \\
&= r^T A^{-1} r - 2(AQz)^T A^{-1} r + (AQz)^T A^{-1} (AQz) \\
&= \|r\|_{A^{-1}}^2 - 2(z^T Q^T A^T) A^{-1} r + \|AQz\|_{A^{-1}}^2 \\
&= \|r\|_{A^{-1}}^2 - 2 z^T Q^T r + \|AQz\|_{A^{-1}}^2 \\
&= \|r\|_{A^{-1}}^2 + \|AQz\|_{A^{-1}}^2 \\
&\geq \|r\|_{A^{-1}}^2 \qquad \text{QED}
\end{aligned}$$

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

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

in span of Q_{k+1} but not in span of Q_k

$\Rightarrow r_k$ multiple of q_{k+1}

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

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

(2) p_k called conjugate (A-conjugate) because p_k orthogonal w.r.t A:

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

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

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

$= L_k \underbrace{D_k}_{\text{diagonal}} L_k^T \qquad L_k(i,i) = 1$

$$= \begin{bmatrix} 1 & & 0 \\ \vdots & \ddots & \\ 0 & \cdots & 1 \end{bmatrix} \begin{bmatrix} \ddots & & \\ & \ddots & \\ & & \ddots \end{bmatrix} \begin{bmatrix} 1 & \cdots & 0 \\ & \ddots & \vdots \\ 0 & & 1 \end{bmatrix}$$

L_k unit diagonal, D_k diagonal

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

$$(*) \quad x_k = Q_k T_k^{-1} e_1 \|b\|_2$$
$$= Q_k (L_k D_k L_k^T)^{-1} e_1 \|b\|_2$$
$$= [Q_k L_k^{-T}] \cdot [D_k^{-1} L_k^{-1} e_1 \|b\|_2]$$
$$= \quad P_k' \quad \cdot \quad y_k$$

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

eventual conjugate gradients p_k are scalar multiples of p_k'

Prove property (2):

Lemma: p_k' are A-conjugate, or $P_k'^T A P_k'$ diagonal

proof:
$$P_k'^T A P_k' = [Q_k L_k^{-T}]^T A [Q_k L_k^{-T}]$$
$$= L_k^{-1} \underbrace{Q_k^T A Q_k}_{} L_k^{-T}$$
$$= L_k^{-1} \quad T_k \quad L_k^{-T}$$
$$= \underbrace{L_k^{-1} (L_k}_{} D_k \underbrace{L_k^T) L_k^{-T}}_{}$$
$$= D_k$$

Need 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

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

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

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

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 to bottom,
L_{k-1} and D_{k-1} are leading k-1 by k-1 submatrices of L_k and D_k

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

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

$$y_k = D_k^{-1} L_k^{-1} e_1 \|b\|_2$$

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

$$= \left[\frac{D_{k-1}^{-1} L_{k-1}^{-1} e_1 \|b\|}{s_k}\right] = \left[\frac{y_{k-1}}{s_k}\right]$$

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

$$= [Q_{k-1} L_{k-1}^{-T}, p_k'] = [P_{k-1}', p_k']$$

To get recurrence for p_k':

equate last columns of $Q_k = P_k' L_k^T$

$$L_k^T = \left[\begin{array}{c|c} 1 \; \ddots \; 1 & \begin{matrix} 0 \\ \vdots \\ 0 \\ x \end{matrix} \\ \hline 0 & 1 \end{array}\right] \leftarrow l_{k-1}$$

$$q_k = p_k' + p_{k-1}' \cdot l_{k-1}$$

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

Need recurrence for r_k: use (R_x)

(R_r) $r_k = b - A x_k$

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

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

All recurrences:

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

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

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

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

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

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

Need recurrences for ν_k, μ_k