Dec 2, 2020

Last time: $f(x) = \frac{1}{2}x^TAx - x^Tb$ $\quad \nabla f(x) = Ax - b$

GD: $x_{k+1} = x_k - \alpha_k g_k \quad g_k = \nabla f(x_k)$ $\qquad \left(1 - \frac{1}{\kappa_2(A)}\right)^{k/2}$

CG: $x_{k+1} = \operatorname{argmin} f(x)$ s.t. $x_k \in \mathcal{K}_k$ $\qquad \left(1 - \frac{1}{\sqrt{\kappa_2(A)}}\right)^{k/2}$

How is CG a gradient method? residual $\Leftrightarrow$ $-$gradient

CG + Lanczos

$q_0 = 0, \ \beta_0 = \|b\|_2, \ q_1 = b/\beta_0, \ x_0 = 0, \ w_0 = 0$

for $k = 1, 2, \ldots$

- $\alpha_k = q_k^T A q_k$
- search direction: $w_k = q_k - l_{k-1} w_{k-1}$
- CG step: $x_k = x_{k-1} + \eta_k w_k$
- $\beta_k q_{k+1} = Aq_k - \alpha_k q_k - \beta_{k-1} q_{k-1}$

$Q_k^T A Q_k = T_k$

$T_k = L_k D_k L_k^T$

$W_k L_k^T = Q_k$

Claim 1: $g_k = -r_k \propto q_{k+1}$

Proof: $g_k = Ax_k - b$

$x_k, b \in K_k(A, b)$

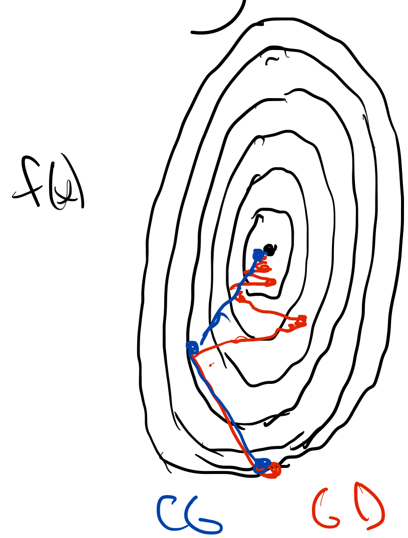
$g_k \in K_{k+1}(A, b)$

$-g_k = r_k \perp K_k(A, b)$ (HW 6)

$g_k \propto q_{k+1}$

Corollary 2:

residuals/gradients orthog



Claim 3: w_k are "A-orthogonal"

- $w_i^T A w_j = 0 \quad i \neq j$

Proof: $W_k^T A W_k$

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

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

$= L_k^{-1} L_k D_k L_k^T L_k^{-T}$

$= D_k$

$$w_k = g_k - l_{k-1} w_{k-1}$$

(1) $$= c\, g_{k-1} - l_{k-1} w_{k-1}$$

Search direction is

$$p_k \propto w_k$$

$$p_k = g_{k-1} + \tau_{k-1} p_{k-1}$$

$$\underbrace{\cancel{p_{k-1}^T A p_k}}_{0 \text{ (claim 3)}} = p_{k-1}^T A g_{k-1} + p_{k-1}^T \tau_{k-1} A p_{k-1}$$

$$\Rightarrow \tau_{k-1} = -p_{k-1}^T A g_{k-1} / p_{k-1}^T A p_{k-1} \quad \text{(E4)}$$

$$p_k^T A p_k = p_k^T A g_{k-1} + \underbrace{\cancel{p_k^T \tau_{k-1} A p_{k-1}}}_{0 \text{ (C3)}}$$

$$p_k^T A g_{k-1} = p_k^T A p_k \quad \text{(E5)}$$

$$A x_k - b = A x_{k-1} - A \mu_k p_k - b$$

$$g_k = g_{k-1} - \mu_k A p_k$$

$$g_{k-1}^T \qquad g_{k-1}^T(\qquad)$$

(C2) $$0 = g_{k-1}^T g_{k-1} - \mu_k g_{k-1}^T A p_k$$

$$\mu_k = \frac{g_{k-1}^T g_{k-1}}{p_k^T A p_k} \quad \text{(E5)}$$

$$g_{k-1} = g_{k-2} - \mu_{k-1} A p_{k-1}$$

$$g_{k-1}^T \qquad g_{k-1}^T(\qquad)$$

$$\Downarrow$$

$$g_{k-1}^T g_{k-1} = -\mu_{k-1} g_{k-1}^T A p_{k-1} \quad \text{(E6)}$$

$$g_{k-2}^T \qquad g_{k-2}^T(\qquad)$$

$$\Downarrow$$

$$g_{k-2}^T g_{k-2} = \mu_{k-1} g_{k-2}^T A p_{k-1} \quad \text{(E7)}$$

$$\tau_{k-1} \overset{(E4)}{=} -\frac{p_{k-1}^T A g_{k-1}}{p_{k-1}^T A p_{k-1}} \underset{(E5)}{\overset{(E6)}{=}} \frac{-g_{k-1}^T g_{k-1} / \mu_k}{p_{k-1}^T A g_{k-2}} \underset{(E7)}{=} -\frac{g_{k-1}^T g_{k-1}}{g_{k-2}^T g_{k-2}}$$

CG (gradient view)

$x_0 = 0 = p_0 \quad g_0 = b \quad g_{-1} = \infty$

for $k = 1, 2, \ldots$

$\tau_{k-1} = g_{k-1}^T g_{k-1} / g_{k-2}^T g_{k-2}$ decrease in residual

$p_k = g_{k-1} + \tau_{k-1} p_{k-1}$ search direction

$\mu_k = g_{k-1}^T g_{k-1} / p_k^T A p_k$ step length

$x_k = x_{k-1} - \mu_k p_k$ CG step

$g_k = g_{k-1} - \mu_k A p_k$ update gradient

Theorem: $\mathrm{span}\{x_1, \ldots, x_k\} = \mathrm{span}\{p_1, \ldots, p_k\} = \mathrm{span}\{g_0, \ldots, g_{k-1}\} = K_k(A, b)$

$k=1$: $g_0 = b \quad p_1 = b, \quad x_1 = 0 - \mu_k p_1 = \frac{b^T b}{b^T A b} b \quad K_1(A, b) = \mathrm{span}\{b\}$

$g_k = g_{k-1} - \mu_k A p_k \qquad g_{k-1}, p_k \in K_k \quad \checkmark$

$p_{k+1} = g_k + \tau_k p_k \quad \checkmark$

$x_{k+1} = x_k - \mu_k p_{k+1} \qquad x_k \in K_k \quad \checkmark$

"linear" CG

$$f(x) = \frac{1}{2}x^T A x - x^T b$$

$$g(x) = Ax - b$$

$$p_k = g_{k-1} + \frac{g_{k-1}^T g_{k-1}}{g_{k-2}^T g_{k-2}} p_{k-1}$$

$$\mu_k = \underset{\mu}{\operatorname{argmin}} f(x_k - \mu_k p_k)$$

nonlinear CG

$f(x)$ convex

$g(x) = \nabla f(x)$

same

Line search