9.2 DESCENT PROPERTIES OF CD-METHOD

→ MINIMIZATION AS SUBSPACES TAKEN.

THEOREM: [EXPANDING SUBSPACE THEOREM]

LET $\{d_i\}_{i=0}^{n-1}$ BE Q-ORTHOGONAL NONZERO VECTORS FOR ANY K. THE SEQUENCE

$$x_{k+1} = x_k + \alpha_k d_k$$

$$\alpha_k = \frac{-g_k^T d_k}{d_k^T Q d_k} \qquad g_k = Q x_k - b$$

HAS PROPERTIES THAT x_k MINIMIZES

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

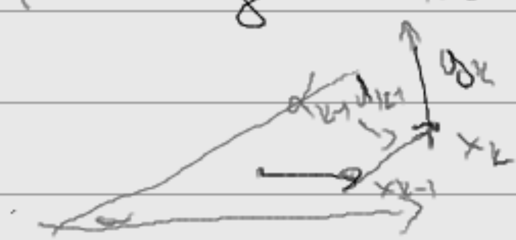
- LINE $x_{k-1} + \alpha d_{k-1}$; $\alpha \in \mathbb{R}$
- AFFINE SUBSPACE $x_0 + \mathcal{L}(d_0 \ldots d_k) = B_k$

PROOF:

SHOW ONLY IT CONTAINS THE LINE :-)

ENOUGH TO SHOW: $\nabla f(x_k) \perp$ TO B_k

(SINCE f STRICTLY CONVEX) g_k



BY INDUCTION ON K

9.3 CONJUGATE GRADIENT METHOD

- d_k NOT GIVEN IN ADVANCE BUT COMPUTED ON THE FLY - FROM $g_k = \nabla f(x_k)$ AS THEY ARE $\perp$ TO $\mathcal{B}_k$
- ALSO GETS A FLAVOR OF STEEPEST DESCENT

ALGORITHM:

$x_0 \in \mathbb{R}^n$, $d_0 = -g_0 = b - Qx_0$ AND

$$x_{k+1} = x_k + \alpha_k d_k$$

$$\alpha_k = \frac{-g_k^T d_k}{d_k^T Q d_k}$$

$$d_{k+1} = -g_{k+1} + \beta_k d_k$$

$$\beta_k = \frac{g_{k+1}^T Q d_k}{d_k^T Q d_k}$$

HOW TO SET β_k?:

$$d_k^T Q d_{k+1} = -d_k^T Q g_{k+1} + \beta_k d_k^T Q d_k = 0$$

$$d_i^T Q d_{k+1} = -d_i^T Q g_{k+1} + \underbrace{\beta_k d_i^T Q d_k}_{=0} =$$

$= ?$ ( ) OF FOLL. PAGE

CONJUGATE GRADIENT THEOREM

IF ALGORITHM DOES NOT STOP AT x_k, THEN

a) $\mathcal{L}(g_0, g_1, \dots g_k) = \mathcal{L}(g_0, Qg_0, \dots Q^k g_0)$

b) $\mathcal{L}(d_0, d_1 \dots d_k) = \mathcal{L}(g_0, Qg_0, \dots Q^k g_0)$

c) $d_k^T Q d_i = 0$

d) $\alpha_k = g_k^T g_k / d_k^T Q d_k$

e) $\beta_k = g_{k+1}^T g_{k+1} / g_k^T g_k$

PROOF:

INDUCTION FOR a, b, c AT THE SAME TIME:

$k \to k+1$; $g_{k+1} = Qx_{k+1} - b = Qx_k + \alpha_k Q d_k - b =$

$= g_k + \alpha_k Q d_k \Rightarrow g_{k+1} \in \mathcal{L}(g_0, \dots Q^{k+1} g_0)$

AND $g_{k+1} \notin \mathcal{L}(d_0, d_1 \dots d_k)$ AS IS ORTHOGONAL TO ALL

(UNLESS $g_{k+1} = 0$) $\Rightarrow$ a)

$d_{k+1} = -g_{k+1} + \beta d_k$ – PROVES b)

FOR c) : (SEE PREVIOUS PAGE)

$-d_i^T Q g_{k+1} = -g_{k+1}^T Q d_i$ AND

d) $Qd_i \in \mathcal{L}(d_0 \dots d_{i+1})$ WHICH ARE $\perp$ TO g_{k+1}

0 $d_{k+1} = -g_{k+1} + \beta_k d_k \quad / \cdot -g_{k+1}$ 0

d_k — $-g_{k+1}^T d_{k+1} = g_{k+1}^T g_{k+1} - \beta_k g_{k+1}^T d_k$

$K \to K+1$: $g_{k+1} \perp B_k$

$$g_{k+1} = Qx_{k+1} - b = Qx_k + \alpha_k Q d_k - b = $$
$$= g_k + \alpha_k Q d_k$$

SO BY α

$$d_k^T g_{k+1} = d_k^T g_k + \alpha_k d_k^T Q d_k = 0$$
$$d_i g_{k+1} = \underbrace{d_i^T g_k}_{\text{IND.} = 0} + d_k \underbrace{d_i^T Q d_k}_{= 0 \text{ Q-ORT.}} = 0$$

$\square$

COROLLARY: $g_k^T d_i = 0 \quad \forall i < k$

[THEOREM'S NAME - MAKING BIGGER SUBSPACE OVER WHICH f IS MINIMIZED]

RECALL:

$$E(x) = \frac{1}{2}(x - x^*)^T Q (x - x^*) = \frac{1}{2}x^T Q x + \overset{b^T x}{\overset{\parallel}{x^{*T} Q x}} +$$
$$+ \frac{1}{2} x^* Q^T x^* = f(x) + \underbrace{\frac{1}{2} x^* Q^T x^*}_{\text{CONSTANT}}$$

-> MINIMIZES DISTANCE TO x^* IN EVERY STEP - x^* OPT. SOL OF $E(x)$ AND HENCE ALSO OF $f(x)$

e) NOTE $g_{k+1}^T g_k = 0$ SINCE $g_k \in \mathcal{G}(d_0 \dots d_k)$

$$g_{k+1} = g_k + \alpha_k Q d_k \qquad / g_{k+1}^T$$

$$g_{k+1}^T g_{k+1} = \overbrace{g_{k+1}^T g_k}^{=0} + \alpha_k \, g_{k+1}^T Q d_k$$

$$g_{k+1}^T Q d_k = \frac{1}{\alpha_k} \, g_{k+1}^T g_{k+1}$$

$$\beta_k = \frac{g_{k+1}^T Q d_k}{d_k^T Q d_k} = \frac{d_k^T Q d_k}{g_k^T g_k} \cdot \frac{g_{k+1}^T g_{k+1}}{d_k^T Q d_k}$$

D

d, e MAY BE MORE CONVENIENT TO COMPUTE
[β_k DOES NOT USE Q ... NO HESSIAN → LATER]

9.4 C-G AND EIGENVALUES

THEOREM h_j:

$$X_{k+1} = X_0 + P_k(Q) \, g_0$$

WHERE P_k IS A POLYNOMIAL OF DEG. K

$$E_{k+1}(X_{k+1}) = \frac{1}{2} (X_{k+1} - X^*)^T Q (X_{k+1} - X^*)^T = \dots$$

$$\left(\begin{array}{l} X_{k+1} - X^* = X_0 - X^* + P_k(Q) \, g_0 = \\ = X_0 - X^* + P_k(Q) Q (X_0 - X^*) = \end{array} \right. \quad \left| \begin{array}{l} Q x_0 - b \\ b = Q x^* \end{array} \right.$$

$$= [I + Q P_k(Q)] (X_0 - X^*)$$

$$= \frac{1}{2} (X_0 - X^*)^T Q [I + Q P_k(Q)]^2 (X_0 - X^*)$$

REWRITE PROBLEM:

$E(X_{k+1})$ MINIMIZED OVER ALL POLYNOMIALS

$$X_{k+1} = X_0 + \gamma_0 g_0 + \ldots \gamma_k Q^k g_0$$

THEOREM:

X_{k+1} FROM C-G METHOD SATISFIES

$$E(X_{k+1}) = \min_{P_k} \tfrac{1}{2}(X_0 - X^*)^T Q [I + QP_k(Q)]^2 (X_0 - X^*)$$

OVER ALL POLYNOMIALS OF DEGREE k.

MEDITATION:

$e_1 \ldots e_n$. ORTHONORMAL EIGENVECTORS OF Q WITH EIGENVALUES $\lambda_1 \ldots \lambda_n$

THEN $X_0 - X^* = \delta_1 e_1 + \delta_2 e_2 + \ldots + \delta_n e_n$

SO

$$E(X_0) = \tfrac{1}{2}(X_0 - X^*)^T Q (X_0 - X^*) = \tfrac{1}{2}\sum \lambda_i \delta_i^2 \Rightarrow$$

$\forall P_k$ FOR $E(X_{k+1}) \leq \frac{1}{2}\sum (1 + \lambda_i P_k(\lambda_i))^2 \lambda_i \delta_i^2$

$$\leq \max_{\lambda_i} (1 + \lambda_i P_k(\lambda_i))^2 E(X_0)$$

NOTE: LOCALLY BETTER THAN STEEPEST DESCENT

AS IS S.T. $x_{k+1} = x_k - \alpha_k g_k \Rightarrow$ IT

IS A POINT IN THE SUBSPACE LIKE x_{k+1} FROM

CG, BUT CG OPTIMAL.

- EIGENVALUES USED IN THE FOLLOWING:

9.5 PARTIAL CONJUGATE GRADIENT METHOD

- RESTART EVERY $m+1$ STEPS

$m = 0 \Rightarrow$ STEEPEST DESCENT

$m = n-1 \Rightarrow$ CONJUGATE GRADIENT METHOD

CONSIDER

$$(P) \begin{cases} \min & \frac{1}{2} x^T Q x - b^T x \\ \text{s.t.} & c^T x = 0 \end{cases}$$

USING PENALTY METHOD (COURANT-BELTRAMI)

$$\min \frac{1}{2} x^T Q x - b^T x + \frac{1}{2}\mu (c^T x)^2$$

$$\Rightarrow \min \frac{1}{2} x^T (Q + \mu c c^T) x$$

SOLVING FOR $\mu \to \infty$

EIGENVALUES OF $Q \in [a, A]$

EIGENVALUES OF $Q + \mu c c^T \in [a, A]$ WHERE

RECALL STEEPEST DESCENT METHOD:

CONVERGES AS

$$E(X_{k+1}) \leq \left(\frac{\lambda_1 - \lambda_n}{\lambda_1 + \lambda_n}\right) E(X_k)$$

DISASTER IN HUGE μ

THEOREM (PARTIAL CONJUGATE GRADIENT)

LET Q HAS $n-m$ EIGENVALUES IN $[a, b]$,

$a > 0$ AND REMAINING ARE $> b$

IF X_{k+1} FROM X_k BY $m+1$ STEPS

(RESTART VERSION) THEN

$$E(X_{k+1}) \leq \left(\frac{b-a}{b+a}\right)^2 E(X_k)$$

HELPS AVOID TROUBLE WITH PENALTY

METHOD