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

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

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

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

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

D

d, e MAY BE MORE CONVENIENT TO COMPUTE
[β_k DOES NOT USE Q ... NO HESSIAN $\leadsto$ 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

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

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

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

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

$$E(X_{k+1}) = \min_{P_k} \tfrac{1}{2}(x_0 - x^*)^T Q [I + Q P_k(Q)]^2 (x_0 - x^*)$$

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 $E(X_{k+1}) \le \tfrac{1}{2}\sum (1 + \lambda_i P_k(\lambda_i))^2 \lambda_i \delta_i^2$

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

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

9.6 EXTENSIONS TO NONQUADRATIC PROBLEMS

MIN $f(x)$

$g_k \longleftrightarrow \nabla f(x_k)$

$Q \longleftrightarrow Hf(x_k)$

1) PRECISE FOR QUADRATIC f

LIKE NEWTON'S METHOD – APPROX f

BY QUADRATIC FUNCTION

2) DOES NOT TERMINATE AFTER $n-1$ STEPS :(

a) JUST CONTINUE

b) RESTART ... PREF.

RES

$x_0, \quad g_0 = \nabla f(x_0), \quad d_0 = -g_0$

$n-1 \times$

$$x_{k+1} = x_k + \alpha_k d_k \qquad \alpha_k = \frac{-g_k^T d_k}{d_k^T Hf(x) d_k}$$

$$g_{k+1} = \nabla f(x_{k+1})$$

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

$$\beta_k = \frac{g_{k+1} Hf(x_k) d_k}{d_k^T Hf(x_k) d_k}$$

+ NO NEED TO FIND t_k

− $Hf(x_k)$ NEEDED,

TWEAKS: α_k & β_k TO AVOID $Hf(x_k)$

- FIND α_k AS $\min f(x_k + \alpha d_k)$
  (1 DIMENSIONAL PROBLEM)
- β_k

  a) FLETCHER-REEVES

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

  WE HAD BEFORE

  b) POLAK-RIBIERE

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

  → BETTER IN EXPERIMENTS.

(a, b, IDENTICAL TO ORIGINAL FOR QUADRATIC CASE)

CONVERGENCE

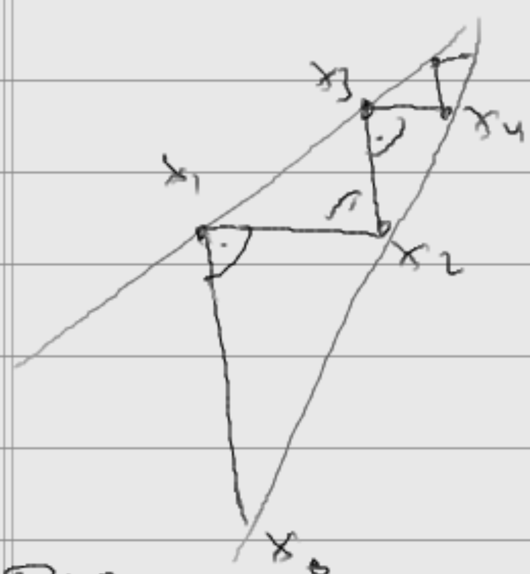
SIMILAR TO STEEPEST DESCENT, HOPEFULLY BETTER $|x_{k+1} - x^*|^2 \leq c|x_k - x^*|^2$

9.7. PARALLEL TANGENTS (PARTAN)

STEEPEST DESCENT TWEAK (? FROM ANAL. OF TANGENTS OF LEVEL CURVES OF QUADRATIC FUNCTIONS)

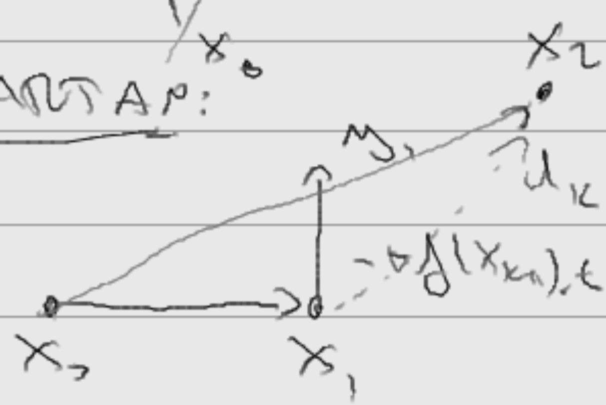
$$x_{k+1} = x_k - t_k \nabla f(x_k)$$

WHERE t_k IS MIN OF $g(t) = f(x_k - t\nabla f(x_k))$



↖ STEEPEST DESCENT

PARTAN:



THEOREM PARTAN IS EQUIVALENT TO CONJUGATE GRADIENTS FOR QUADRATIC f

- CONVERGENCE LIKE STEEPEST DESCENT (HOPE BETTER)
- TWO LINE SEARCHES (MIN $g(t)$ AND ONCE MORE FOR X_{K+1} (BUT NEED NOT BE PRECISE AS IN CONJ GRAD METHOD)