3.3 BEYOND STEEPEST DESCENT

MAKING „GOOD" ITERATIVE METHOD

MIN $f(x)$, CONTINUOUS $\nabla f(x)$

START x_0

GENERATE SEQUENCE $\{x_k\}$ AS

$$x_{k+1} = x_k + t_k p_k \qquad t_k > 0$$

SUCH THAT

1) $f(x_{k+1}) < f(x_k)$ IF $\nabla f(x_k) \neq 0$

$\Leftrightarrow$ DESENT METHOD

2) $p_k^T \nabla f(x_k) < 0$ $\Leftarrow$ MOVING IN DES. DIRECTION

$\varphi_k(t) = f(x_k + t p_k)$

$\varphi_k'(0) = \nabla f(x_k)^T p_k < 0$

SO FOR t SMALL

$$f(x_k + t p_k) = \varphi_k(t) < \varphi(0) = f(x_k)$$

[TOO SMALL t IS BAD – SLOW CONV. $\Rightarrow$ 3)]

3) $\exists\; 0 < \beta < 1$ S.T. $p_k^T \nabla f(x_{k+1}) > \beta\, p_k^T \nabla f(x_k)$

t_k NOT TOO SMALL (BY 2 + 3)

$$p_k^T \nabla f(x_k + t_k p_k) > \beta \underbrace{p_k^T \nabla f(x_k)}_{<0 \text{ BY } 2)} > p_k^T \nabla(f_k)$$

$$p_k^T \nabla f(x_k + t_k p_k) - p_k^T \nabla f(x_k) > (\beta - 1) p_k^T \nabla f(x_k) > 0$$

IF $t_k \to 0$ THEN $= 0$ AND RIGHTHAND SIDE C. > 0

$\Rightarrow$ NO ARBITRARILY SMALL t_k

BUT WHAT IF t_k TOO LARGE (MAYBE NOT BEST DECREASE

GOING TOO FAR.

$\Rightarrow \exists\ 0 < \alpha < \beta < 1$ SUCH THAT

$$f(x_{k+1}) \le f(x_k) + \alpha t_k p_k^T \nabla f(x_k)$$

REWRITE

$$\frac{f(x_k) - f(x_{k+1})}{t_k} \ge \alpha \left[-p_k^T \nabla f(x_k) \right]$$

↳ RELATIVE DECREASE $\ge$ FRACTION OF POSSIBLE DECREASE.

DIFF VIEW: $M = \dfrac{-\alpha p_k^T \nabla f(x_k)}{\|p_k\|}$

$$\frac{f(x_k) - f(x_{k+1})}{\| x_k - x_{k+1} \|} = \frac{f(x_k) - f(x_{k+1})}{\|p_k\| \cdot t_k} \ge M$$

$$f(x_k) - f(x_{k+1}) \ge M \| x_k - x_{k+1} \| \quad \Rightarrow$$

THEOREM 3.3.1 WOLFE

LET $f(x)$ HAVE CONTINUOUS FIRST PARTIAL DERIVATIVES. AND BE BOUNDED FROM BELOW. LET $0 < \alpha < \beta < 1$,

IF $p_k, x_k \in \mathbb{R}^n$ S.T.

$$p_k^T \cdot \nabla f(x_k) < 0$$

THEN $\exists\ 0 \leq a_k < b_k$ SUCH THAT

- $4_)$ HOLDS $\forall t_k \in (0, b_k)$
- $3_)$ HOLDS $\forall t_k \in (a_k, b_k)$

HOW TO FIND t? DIFFICULT

- TRY: $t = 1$

WHILE $4_)$ VIOLATED:

$$f(x_k + t p_k) > f(x_k) + \alpha t p_k^T \nabla f(x_k)$$

REPLACE t BY $t \cdot \beta$ WHERE $\beta < 1$

[HOW TO FIND β? NUMERICAL METHODS]

CHOICE OF p_k:

- WANT $p_k^T \nabla f(x_k) < 0$ (*)

STEEPEST DESCENT: ... WITH $Q = I$

$$p_k = -\nabla f(x_k)$$

Q POSITIVE DEFINITE:

FOR $p_k = -Q \nabla f(x_k)$ (*) HOLDS

$$\{-\nabla f(x_k)^T Q \nabla f(x_k) < 0\}$$

RECALL NEWTON'S METHOD

$$p_k = -[Hf(x)]^{-1} \nabla f(x_k)$$

WORKS FOR $Hf(x)$ POS DEF.

IF $Hf(x)$ NOT POSITIVE DEF. TAKE

$$Q = Hf(x) + \mu_k I \quad (MU)$$

FIND μ_k:

$$x^T(Hf(x) + \mu_k I)x = x^T Hf(x) x + \|x\|^2 \mu_k$$

WANT ≥ 0. HOW SMALL?

$$\bar{\mu}_k = \max_{\|x\|=1} |x^T Hf(x) x|$$

... EX $\bar{\mu}_k = |\lambda_1|$ LARGEST EIGEN VALUE

$$x^T(Hf(x) + \mu_k I)x = \|x\|^2 \cdot \left[\frac{x^T}{\|x\|} Hf(x) \frac{x}{\|x\|} + \mu_k\right] \geq$$

$$\geq \|x^2\|(-\bar{\mu}_k + \mu_k) > 0$$

$\Rightarrow$ $Hf(x_k) + \mu_k I$ IS POSITIVE DEFINITE

MIX OF NEWTON &
& STEEPEST DESCENT

ALGORITHM:

INPUT x_k

1) COMPUTE μ_k S.T.

$Hf(x_k) + \mu_k I$ IS POSITIVE DEFINITE

2) SOLVE p_k:

$$(Hf(x_k) + \mu_k I)\, p_k = -\nabla f(x_k)$$

3) COMPUTE t_k TO SATISFY WOLFE

4) $x_{k+1} = x_k + t_k p_k$

APPROXIMATE $f(x)$ BY

NEWTON
HAS $Hf(x)$

$$Q_k(x) = f(x_k) + \nabla f(x_k)^T (x - x_k) + \frac{1}{2}(x - x_k)^T A_k (x - x_k)$$

WHERE $A_k = Hf(x_k) + \mu_k I$

IN PRACTICE:

$$p_k = -A_k^{-1} \nabla f(x_k)$$

MAY BE TOO LARGE (NUMERICAL ISSUES)

IDEA:

KEEP THE STEP LENGTH $l_k = \| t_k p^{(k)} \|$

AND COMPUTE X_{k+1} AS:

$$X_{k+1} - X_k = (A_k + \lambda_k I)^{-1} (\nabla f(x^{(k)}))$$

WHERE $\| X_{k+1} - X_k \| = l_k$

IS λ_k RUINING THE SOLUTION ??

THEOREM 5.5.1

LET f HAS CONTINUOUS SECOND PARTIAL DERIVATIVES, $X_k \in \mathbb{R}^n$ AND

$$Q_k(X) = f(X_k) + \nabla f(X_k)^T (X - X_k) + \frac{1}{2} (X - X_k)^T A_k (X - X_k)$$

THERE EXISTS $\lambda_k \geq 0$ SUCH THAT MINIMIZER X_{k+1} OF $Q_k(X)$ SUBJECT TO

$\| X - X_k \| \leq l_k$ IS A SOLUTION TO

$$(A_k + \lambda_k I)(X - X_k) = -\nabla f(X_k)$$

IF $\| -A_k^{-1} (\nabla f(X_k)) \| \leq l_k$, PICK $\lambda_k = 0$, OTHERWISE $\lambda_k > 0$ ST $\| X_{k+1} - X_k \| = l_k$

PROOF APPLY KKT

$$(3)\begin{cases} \text{MIN } Q_k(x) \\ \text{S.T. } \|x - x_k\|^2 - l_k^2 \le 0 \end{cases}$$

$$L(x, \mu_k) = f(x_k) + \nabla f(x_k)^T (x - x_k) + \tfrac{1}{2}(x - x_k)^T A_k (x - x_k) +$$
$$+ \mu_k (\|x - x_k\|^2 - l_k^2)$$

WHERE μ_k IS KKT MULTIPLIER (λ^* FROM KKT)

x_{k+1} $\therefore L(x^*, \mu_k) \le L(x, \mu_k)$ OR GRADIENT CONDITION -

$$0 = \nabla f(x_k) + A_k (x_{k+1} - x_k) + 2\mu_k (x_{k+1} - x_k)$$

SO

$$(A_k + 2\mu_k)(x^{k+1} - x^k) = -\nabla f(x_k)$$

LET $\lambda_k = 2\mu_k$ ∎

λ_k CAN BE CHOSEN TO PRESERVE OPT. SOLUTION.

→ GOOD APPROXIMATION TO $f(x)$

HOPEFULLY

l_k .. PLACE WE TRUST THE APPROXIMATION IS GOOD.