3.2 METHOD OF STEEPEST DESCENT BY CAUCHY 1847

PROPOSITION

$f : \mathbb{R}^n \to \mathbb{R}$. FOR $x \in \mathbb{R}^n$ $v = -\nabla f(x)$ POINTS IN THE DIRECTION OF MOST RAPID DECREASE

3.2.1 METHOD OF STEEPEST DESCENT

$\{x_k\}$ IS THE SEQUENCE FOR STEEPEST DESCENT WITH INITIAL POINT x_0 IF

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

WHERE t_k IS MINIMUM OF

$$\varphi_k(t) = f(x_k - t \nabla f(x_k)), \quad t \geq 0$$

("GO IN THE DIRECTION $\nabla f(x_k)$ AS FAR AS POSSIBLE")

EXAMPLE: $f(x, y) = 4x^2 - 4xy + 2y^2$

WITH $x_0 = (2, 3)$

$$\nabla f(x, y) = (8x - 4y, -4x + 4y)$$

$$\varphi_0(t) = f(2 - 4t, 3 - 4t)$$

- USE $\varphi_0'(t)$ TO FIND MINIMUM

$$\varphi_0'(t) = -16(2-4t)$$

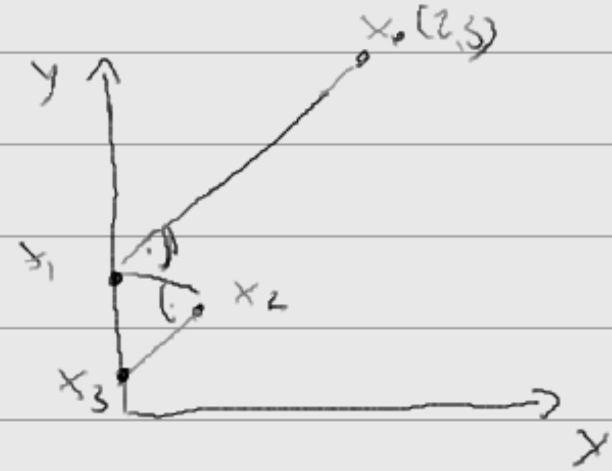
$\rightarrow$ CRITICAL POINT $t = \frac{1}{2}$, $\varphi_0''(\frac{1}{2}) = 64 > 0 \Rightarrow$ GLOBAL MINIMIZER

$$\Rightarrow x_1 = x_0 - \frac{1}{2} \nabla f(x_0) = (0, 1)$$

$$x_2 = \left(\frac{4}{10}, \frac{6}{10}\right)$$

$$x_3 = \left(0, \frac{3}{10}\right)$$

PLOT:



MOVES WITH RIGHT ANGLES

THEOREM 3.2.3

IF $\{x_k\}$ IS THE STEEPEST DESCENT SEQUENCE THEN $(x_{k+1} - x_k)$ IS ORTHOGONAL TO $(x_{k+2} - x_{k+1})$

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

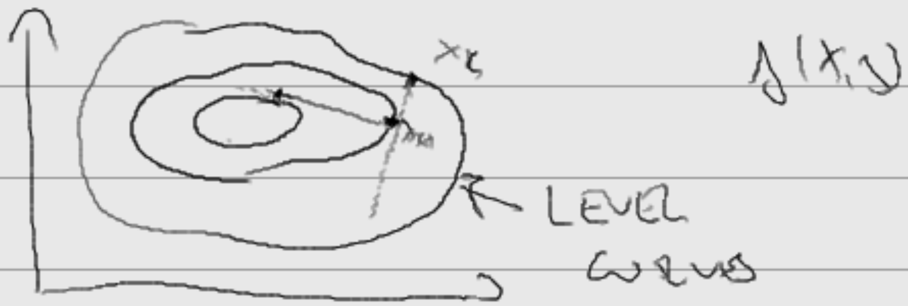
$$(x_{k+1} - x_k)^T (x_{k+2} - x_{k+1}) = t_{k+1} t_k \nabla f(x_k)^T \nabla f(x_{k+1}) \overset{?}{=} 0$$

$$\varphi_k(t) = f(x_k - t \nabla f(x^k))$$

$$0 = \varphi_k'(t_k) = -\nabla f(x^k - t_k \nabla f(x_k))^T \nabla f(x_k) =$$

$$= -\nabla f(x_{k+1})^T \nabla f(x_k) \qquad \square$$

(IMG)



MAKES THE METHOD SLOW -> STEPS MAY BE TOO LARGE

THEOREM 3.2.5

IF $\{X_k\}$ IS THE STEEPEST DESCENT SEQUENCE AND $\nabla f(x_k) \neq 0$ THEN

$$f(x_k) > f(x_{k+1})$$

NOTE

ANY METHOD WHERE $\{X_k\}$ SATISFY $f(x_k) > f(x_{k+1})$ IS CALLED DESCENT METHOD

- CONVERGENCE:

THEOREM 3.2.6

LET $f(x)$ BE COERCIVE WITH CONTINUOUS FIRST PARTIAL DERIVATIVES. FOR ANY x_0 SOME SUBSEQUENCE OF $\{X_k\}$ CONVERGES. LIMIT OF ANY SUBSEQUENCE IS A CRITICAL POINT.

COROLLARY 3.2.7

IF f STRICTLY CONVEX, COERCIVE AND CONTINUOUS FIRST PARTIAL DERIVATIVES, THEN STEEPEST DESCENT CONVERGES TO GLOBAL MINIMIZER OF f.

→ CONVERGENCE SPEED

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

WHERE Q POSITIVE WITH EIGENVALUES

$\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n > 0$

MIN FROM GRADIENT $Qx^* = b$

LET $E(x) = \frac{1}{2}(x - x^*)^T Q (x - x^*)$

$\therefore \quad E(x) = f(x) + \underbrace{\frac{1}{2} x^{*T} Q x^*}_{\text{CONSTANT}}$ ← JUST MULTIPLY OUT

SAME GRADIENT

$g(x) = Qx - b$

⇒ BY METHOD

$x_{k+1} = x_k - \alpha_k g_k$ | BY MINIMIZING $f(x_k - \alpha g_k)$

WHERE

$g_k = Qx_k - b, \quad \alpha_k = \frac{g_k^T g_k}{g_k^T Q g_k}$

LEMMA:

$$E(X_{k+1}) = \left(1 - \frac{(g_k^T g_k)^2}{(g_k^T Q g_k)(g_k^T Q^{-1} g_k)}\right) E(X_k)$$

PROOF

BY COMPUTING $\dfrac{E(X_k) - E(X_{k+1})}{E(X_k)} = \ldots$

KANTOROVICH INEQUALITY

FOR POSITIVE DEFINITE Q WITH EIGENVALUES $\lambda_1 \geq \lambda_2 \geq \lambda_3 \geq \ldots \geq \lambda_n$ AND FOR ANY X

$$\frac{(x^T x)^2}{(x^T Q x)(x^T Q^{-1} x)} \geq \frac{4\lambda_1 \lambda_n}{(\lambda_1 + \lambda_n)^2}$$

BY COMBINING:

THEOREM

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

IF $R = \lambda_1 / \lambda_n$

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

-> IF R BIG - SLOW CONVERGENCE. REALLY HAPPENS. 1 BAD EIGENVALUE MAY DESTROY THE METHOD :-(

-> PROBLEM BECAUSE TOO LONG STEPS;

SIMPLIFICATION - USE ALWAYS JUST ONE COORDINATE WHERE DESCENT POSSIBLE

- POSSIBLY MANY STEPS, BUT EASIER

DRAWBACKS:

- $\varphi_k(t) = f(x_k - t \nabla f(x_k)),\ t \geq 0$ MAY BE DIFFICULT TO MINIMIZE
- CONVERGING MIGHT BE SLOW

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 $\nabla f(x_k) \neq 0$

⇔ DESENT METHOD

2) $p_k^T \nabla f(x_k) < 0$ ← 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. ⇒ 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 RIGHT HAND SIDE > 0

$\Rightarrow$ NO ARBITRARILY SMALL t_k

BUT WHAT IF t_k TOO LARGE (MAYBE NOT BEST DECREASE

GOING TOO FAR.

↳ $\exists\ 0 < \alpha < \beta < 1$ 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: $M = \frac{-\alpha p_k^T \nabla f(x_k)}{\|p_k\|}$

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

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