CHAPTER 3 - UNCONSTRAINED OPTIMIZATION ITERATIVE METHODS

IDEA: START AT X_0 AND BUILD A SEQUENCE X_K S.T. $\lim_{K \to \infty} X_K = X^*$, $X^* = \min f$.

- WANT FAST CONVERGENCE OF lim
- EASY TO COMPUTE X_i FROM X_{i-1}
- WORKING FOR GENERAL f

PLAN: NEWTON'S METHOD

: STEEPEST DESCENT METHOD

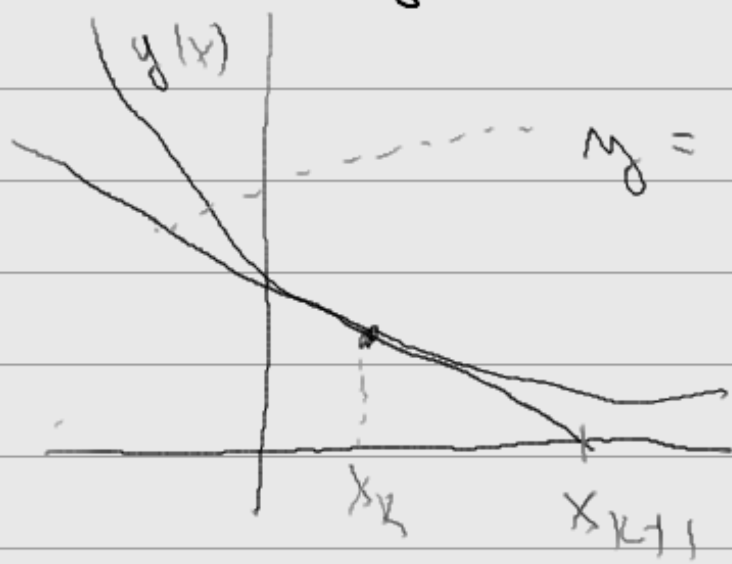
: IMPROVEMENTS

NEWTON'S METHOD

$f: \mathbb{R}^n \to \mathbb{R}$, FOR x^* = MIN : $\nabla f(x^*) = 0$

GOAL: FOR $g: \mathbb{R}^n \to \mathbb{R}$ FIND x ST $g(x) = 0$

IDEA FOR $g: \mathbb{R} \to \mathbb{R}$



$y = g'(x_k)(x - x_k) + g(x_k)$

SOLVE x FOR $y = 0$

$$x_{(k+1)} = x_k - \frac{g(x_k)}{g'(x_k)}$$

WORKS IF

- x_0 CLOSE TO OPT - GOOD INITIAL GUESS
- $g(x)$ NOT TOO WOBBLY

GENERALIZATION FOR $g: \mathbb{R}^n \to \mathbb{R}^n$

g_i - i^{th} COMPONENT OF g

WANT 0 FOR EACH COMPONENT

$$g_1(x_k) + \nabla g_1(x_k)^T (x - x_k) = 0$$

$$\vdots$$

$$g_n(x_k) + \nabla g_n(x_k)^T (x - x_k) = 0$$

OR JUST

$$g(x_k) + \nabla g(x_k)(x - x_k) = 0$$

JACOBIAN $ij = \dfrac{\partial g_i}{\partial x_j}(x)$

SUMMARY:

$g: \mathbb{R}^n \to \mathbb{R}^n$ DIFFERENTIABLE, $x_0 \in \mathbb{R}^n$

THEN NEWTON'S METHOD SEQUENCE $\{x_k\}$

IS: (A) $[\nabla g(x_k)](x_{k+1} - x_k) = -g(x_k)$

OR $\qquad k \geq 0$

(B) $x_{k+1} = x_k - [\nabla g(x_k)]^{-1} g(x_k)$

NOTES:

- IF NOT UNIQUE SOLUTION .. SEQ. UNDEFINED
- CONVERGENCE TO $g(x) = 0$ NOT GUARANTEED
- (B) NICE TO SEE x_{k+1}, (A) ACTUALLY SOLVED

NO NEED FOR INVERT

EX

$$x^2 + y^2 + z^2 = 3 \quad \leftarrow \text{SPHERE}$$
$$x^2 + y^2 - z = 1 \quad \leftarrow \text{PARABOLOID}$$
$$x + y + z = 3 \quad \leftarrow \text{PLANE}$$

– UNIQUE SOLUTION $x = y = z = 1$

FOR NEWTON:
$$g(x,y,z) = \begin{pmatrix} x^2+y^2+z^2-3 \\ x^2+y^2-z-1 \\ x+y+z-3 \end{pmatrix} = 0 \text{ WANT}$$

START WITH
$$X_0 = (1,0,1) \qquad \nabla g(x) = \begin{pmatrix} 2x & 2y & 2z \\ 2x & 2y & -1 \\ 1 & 1 & 1 \end{pmatrix}$$

USE [A] FOR STEP:

$$\underbrace{\begin{pmatrix} 2 & 0 & 1 \\ 2 & 0 & -1 \\ 1 & 1 & 1 \end{pmatrix}}_{\nabla g(x) =} \underbrace{\begin{pmatrix} x-1 \\ y \\ z-1 \end{pmatrix}}_{(x - x_0)} = \underbrace{\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}}_{-g(x_0)}$$

SOLUTION $X_1 = \left(\frac{3}{2}, \frac{1}{2}, 1\right)$ CONTINUE...

$X_2 = \left(\frac{5}{4}, \frac{3}{4}, 1\right)$, $X_3 = \left(\frac{9}{8}, \frac{7}{8}, 1\right)$ ... NICE

IF $x_0 = (0,0,0)$ SYSTEM – HAS NO SOLUTION

CONVERGENCE GUARANTEE: KANTOROVICH

- ALWAYS CONVERGES WITH ANY $t_0 \in \mathbb{R}$

$$g(t) = \frac{a}{2}t^2 - \frac{1}{b}t + \frac{c}{b} \text{ IF } abc \le \frac{1}{2}$$

TO $t^* = \frac{1}{ab}\left(1 - \sqrt{1 - 2abc}\right)$

GENERALIZATION:

IF X_0 AND a, b, c S.T. $abc \le \frac{1}{2}$

1, $\| (\nabla g(x_0))^{-1} \| \le b$

2, $\| (\nabla g(x_0))^{-1} g(x_0) \| \le c$

3, $\exists \delta > 0$ S.T. $\| \nabla g(x) - \nabla g(y) \| \le a \| x - y \|$

WHENEVER $\| x - x_0 \| < \delta$, $\| y - x_0 \| < \delta$

↳ USED TO JUSTIFY THAT REFINEMENT OF X AS INITIAL POINT WILL EVENTUALLY WORK. NOT TO FIND X_0

FOR MINIMIZATION OF f USE $g = \nabla f$ AND $\nabla g(x) = Hf(x)$

THEOREM 31.3 NEWTONS METHOD

$f: \mathbb{R}^n \to \mathbb{R}$, $x_0 \in \mathbb{R}^n$. NEWTONS SEQUENCE FOR MINIMIZING $f(x)$ IS:

A) $[Hf(x_k)](x_{k+1} - x_k) = -\nabla f(x_k)$

B) $x_{k+1} = x_k - (Hf(x_k))^{-1} \nabla f(x_k)$

DIFFERENT VIEW:

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

$\uparrow$

QUADRATIC FUNCTION BEST FIT FOR f AT x_k (SAME DERIVATIVES) - CONVEX f_k

- FIND MINIMIZER MEANS CRITICAL POINT y

$$\nabla f_k(y) = 0 = \nabla f(x_k) + Hf(x_k)(y - x_k)$$

$\Rightarrow [Hf(x_k)](y - x_k) = -\nabla f(x_k)$ WHICH IS A)

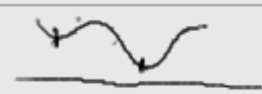
COROLLARY

FOR $f(x) = a + b^T x + \frac{1}{2} x^T A x$ IF A POS DEF, NEWTON METHOD WORKS IN 1 STEP.

- PERFECT FIT IS AGAIN f

NOT QUADRATIC $f(x)$ → CONVERGENCE NOT GUARANTEED EVEN IF UNIQUE MIN x^*, x_0 CLOSE TO x^*. (HW EX)

NOTE: NEWTON DOES NOT DISTINGUISH MIN, MAX OR SADDLE POINTS. MAY DIVERGE OR LOCAL ONES!



"IF $Hf(x)$" POSITIVE DEFINITE NEWTONS GO IN THE "RIGHT" DIRECTION

THEOREM 3.1.5

SEQUENCE $\{x_k\}$ FROM NEWTONS METHOD IF $Hf(x_k)$ IS POSITIVE DEFINITE AND $\nabla f(x_k) \neq 0$ THEN

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

IS A DESCENT DIRECTION:

$\exists \varepsilon > 0 \ \forall \ 0 < t < \varepsilon$

$$f(x_k + t p_k) < f(x_k)$$