Amir Hossein Eyvazkhani

1747696

Ex 3

1

(a) $\frac{df}{dx_1} = 2x_1+2 = 0 \Rightarrow x_1 = -1$

$\frac{df}{dx_2} = 6x_2 + 0.5 \Rightarrow x_2 = \frac{-1}{12}$

$\Rightarrow x^* = (-1, \frac{-1}{12}) \Rightarrow$

$p^* = (-1)^2 + 3(\frac{-1}{12})^2 + 2(-1) + 0.5(\frac{-1}{12})$

$= -1.02$

$\frac{df}{d^2x_1} = 2 > 0$

$\frac{df}{d^2x_2} = 6 > 0$

$\rightarrow$ it's a minimum

(b) $\nabla f = \begin{pmatrix} 2x_1+2 \\ 6x_2+0.5 \end{pmatrix}$

$x_0 = (3,-1)$, $f_0 = 17.5$

$\nabla f = \begin{pmatrix} 8 \\ -5.5 \end{pmatrix} \Rightarrow x_1 = x_0 - \alpha\nabla f = (3,-1) - (0.2)(8,-5.5) = (1.4, 0.1)$

$x_1 = (1.4, 0.1)$, $f_1 = 4.48$

$\nabla f = \begin{pmatrix} 4.8 \\ 1.1 \end{pmatrix} \Rightarrow x_2 = x_1 - \alpha\nabla f = (1.4, 0.1) - (0.2)(4.8, 1.1) = (0.44, -0.12)$

$x_2 = (0.44, -0.12)$, $f_2 = 1.05$

$\nabla f = \begin{pmatrix} 2.88 \\ -0.22 \end{pmatrix} \Rightarrow x_3 = x_2 - \alpha\nabla f = (0.44, -0.12) - (0.2)(2.88, -0.22)$
$= (-0.136, -0.076)$

$x_3 = (-0.136, -0.076)$, $f_3 = -0.27$

yes as the values clearly demonstrate, the algorithm is minimizing it.

(c) $\nabla f = \begin{pmatrix} 2x_1+2 \\ 6x_2+0.5 \end{pmatrix}$

$x_0 = (3,-1)$, $f_0 = 17.5$

$\nabla f = (8,-5.5) \Rightarrow x_1 = x_0 - u\nabla f = (-1, 1.75)$

$x_1 = (-1, 1.75)$, $f_1 = 9.06$

$\nabla f = (0, 11) \Rightarrow x_2 = x_1 - u\nabla f = (-1, -3.75)$

$x_2 = (-1, -3.75)$, $f_2 = 39.31$

$\nabla f = (0, -22) \Rightarrow x_3 = x_2 - u\nabla f = (-1, 7.25)$

$x_3 = (-1, 7.25)$, $f_3 = 160.3$

$\Rightarrow$ It's not minimizing. The u (step size) is not suitable. (too large).

(d) The function is: $L_\theta = \frac{1}{N} \|y - x^T\theta\|_2^2$

$$= \frac{1}{N} \sum (y - x^T\theta)^2$$

$$\nabla L = \frac{dL}{d\theta} = \frac{2}{N}(y - x^T\theta)(-x^T) = 0$$

$$\Rightarrow (y - x^T\theta)(-x^T) = 0 \Rightarrow -x^T y + x^T x \theta = 0$$

$$\Rightarrow \theta = (x^T x)^{-1}(x^T y)$$

2

a) $\nabla f = \begin{pmatrix} 2x_1 \\ 2x_2 \end{pmatrix}$

$$f(x+u\Delta x) > f(x) + \alpha u \nabla f(x)^T \Delta x \xrightarrow{\Delta x = -\nabla f} f(x-u\nabla f) > f(x) - \alpha u \|\nabla f\|_2^2$$

$$\Rightarrow f((x_1,x_2) - u(2x_1,2x_2)) > f(x_1,x_2) - \alpha u (4x_1^2 + 4x_2^2)$$

$$\Rightarrow (1-2u)^2 x_1^2 + (1-2u)^2 x_2^2 > (1-4\alpha u) x_1^2 + (1-4\alpha u) x_2^2$$

$$\Rightarrow (1-2u)^2 \cancel{(x_1^2+x_2^2)} > (1-4\alpha u) \cancel{(x_1^2+x_2^2)}$$

$$\Rightarrow 1 - 4u + 4u^2 > 1 - 4\alpha u \Rightarrow 4u(u-1+\alpha) > 0$$

$$\Rightarrow \mu - 1 + \alpha > 0$$

b) $x = (0.5, 1) \Rightarrow \Delta x = -\nabla f = (1, 2)$

$\mu = 10$, $\alpha = 0.5$, $\beta = 0.1$

while ($\mu - 0.5 > 0$) → derived from previous part

$\mu = \mu (0.1)$

iter 0 : $\mu = 10$

iter 1 : 1 → at 2 iterations the algorithm stops and the final μ is 0.1

iter 2: $\mu = 0.1$,

iter 3 X