Assignment 1

Saturday, 28 August 2021 8:16 PM

Name: Adarsh Shah
Sr. No.:
04-03-06-10-51-21-1-19473

1. (5 points) Consider the polynomial

$$p(x,y,z) = x^4y^2 + x^2y^4 + z^6 - 3x^2y^2z^2.$$

Show that $f^* = \inf_{x,y,z} p(x,y,z) = 0$.

Arithmatic mean $= \dfrac{x^4y^2 + x^2y^4 + z^6}{3}$

Geometric mean $= (x^4y^2 \cdot x^2y^4 \cdot z^6)^{1/3}$

$= x^2y^2z^2$

We know, A.M. $\geq$ G.M.

$\therefore \dfrac{x^4y^2 + x^2y^4 + z^6}{3} \geq x^2y^2z^2$

$\therefore x^4y^2 + x^2y^4 + z^6 \geq 3x^2y^2z^2$

$\therefore x^4y^2 + x^2y^4 + z^6 - 3x^2y^2z^2 \geq 0$

$\therefore \inf p(x,y,z) \geq 0$

2. (15 points) Suppose $f \in C_L^1$, where $L > 0$; that is,

$$\|\nabla f(x) - \nabla f(y)\|_2 \le L\|x-y\|_2.$$

If $f \in C_L^1$, show that the functions

$$g(x) = \frac{L}{2}x^Tx - f(x) \quad \text{and} \quad h(x) = \frac{L}{2}x^Tx + f(x)$$

are convex. Then, show that

$$-\frac{L}{2}\|y-x\|^2 \le f(y) - f(x) - \langle \nabla f(x), y-x\rangle \le \frac{L}{2}\|y-x\|^2.$$

We know,

$f(\alpha x + (1-\alpha)y) \le \alpha f(x) + (1-\alpha) f(y)$

$g(x) = \frac{L}{2}x^Tx - f(x)$ $\qquad$ $h(x) = \frac{L}{2}x^Tx - f(x)$

Since $f \in C_L^1$

$f(y) = f(x) + \nabla f(x)^T \cdot (y-x) + O(\|y-x\|)$

$f(x) = f(y) + \nabla f(y)^T \cdot (x-y) + O(\|y-x\|)$

$\therefore\ f(y) - f(x) = f(x) - f(y) + (\nabla f(x) - \nabla f(y))^T \cdot (y-x)$

$\therefore\ f(y) - f(x) = \frac{1}{2}\langle \nabla f(x) - \nabla f(y), y-x\rangle$

$\therefore\ f(y) - f(x) = -\frac{1}{2}\langle \nabla f(y) - \nabla f(x), y-x\rangle$

$$\boxed{\therefore\ f(y) - f(x) - \langle \nabla f(y) - \nabla f(x), y-x\rangle = \frac{1}{2}\langle \nabla f(y) - \nabla f(x), y-x\rangle \quad (1)}$$

Now,

$\|\nabla f(y) - \nabla f(x)\|_2 \le L\|y-x\|_2$

$\therefore\ \langle \nabla f(y) - \nabla f(x), \nabla f(y) - \nabla f(x)\rangle \le L^2 \langle y-x, y-x\rangle$

$\therefore\ \langle \nabla f(y) - \nabla f(x), y-x\rangle^2 \le L^2 \cdot \langle y-x, y-x\rangle^2$

$\therefore\ -L\|y-x\|^2 \le \langle \nabla f(y) - \nabla f(x), y-x\rangle \le L\|y-x\|^2$

From (1), we get

$$\boxed{-\frac{L}{2}\|y-x\|^2 \le f(y) - f(x) - \langle \nabla f(y) - \nabla f(x), y-x\rangle \le \frac{L}{2}\|y-x\|^2}$$

$g(x) = \frac{L}{2} x^T x - f(x)$

$\nabla g(x) = L I x - \nabla f(x)$

$I \in \mathbb{R}^{n \times n}$ (Identity matrix)

$H_g(x) = L \cdot I \quad \therefore g \in C^2$

if $g(y) \geq g(x) + \nabla g(x)^T \cdot (y-x) \quad \forall x\, y \in \mathbb{R}^n$

$\rightarrow g$ is Convex

$g(y) = g(x) + \nabla g(x)^T \cdot (y-x) + \frac{1}{2} (y-x)^T \cdot H_g(x) \cdot (y-x)$

$$\therefore g(y) - g(x) - \nabla g(x)^T (y-x) = \frac{1}{2} (y-x)^T \cdot L \cdot I \cdot (y-x)$$

$$= \frac{L}{2} \|y-x\|_2^2$$

$$> 0$$

$\therefore$ g is Convex function.

$h(x) = \frac{L}{2} x^T x + f(x)$

$\nabla h(x) = L \cdot I x + \nabla f(x)$

$H_h(x) = L \cdot I \quad \therefore h \in C^2$

To show $h(x)$ is convex,

it is sufficient to show

$h(y) \geq h(x) + \nabla h(x)^T \cdot (y-x)$

$\forall x, y \in \mathbb{R}^n$

$$h(y) = h(x) + \nabla h(x)^T \cdot (y-x) + \frac{1}{2}(y-x)^T \cdot H_h(x) \cdot (y-x)$$

$$\therefore h(y) - h(x) - \nabla h(x)^T \cdot (y-x) = \frac{1}{2}(y-x)^T H_h(x) \cdot (y-x)$$

$$= \frac{L}{2}(y-x)^T \cdot (y-x)$$

$$= \frac{L}{2}\|y-x\|^2$$

$$> 0$$

$$\therefore h(y) > h(x) + \nabla h(x)^T (y-x)$$

$\therefore$ h is convex

3. (10 points) You are each given 1000 pairs of data points (x_i, y_i), where $x_i \in \mathbb{R}^5$ and $y_i \in \mathbb{R}$. We know that the data is generated by the equation

$$y_i = w^T x_i + b.$$

Using the provided data, find w that minimizes the least squares error between y_i and $w^T x_i$. Furthermore, for the general case where $x \in \mathbb{R}^n$ and we are given m data points, what is the closed form solution to this problem? Is this solution unique?

Suppose the number of linearly independent data points is less than n - how would you solve this problem, and is the solution unique?

Let $Y = (y_1 .. y_m)^T \in \mathbb{R}^{m \times 1}$

$W = (w_1 ... w_n, b)^T \in \mathbb{R}^{n+1 \times 1}$

$X \in \mathbb{R}^{m \times (n+1)}$ where x_{ij} is a datapoint, $x_{i\,n+1} = 1$

$$\boxed{\therefore W = \underset{\substack{w_i \in \mathbb{R},\, i \leq n,\, i \in \mathbb{N} \\ b \in \mathbb{R}}}{\operatorname{argmin}} (Y - X \cdot W)^T \cdot (Y - X \cdot W)}$$

Evaluating, $f(w) = (Y - Xw)^T (Y - Xw)$

$= (Y^T - w^T X^T)(Y - Xw)$

$$= Y^T Y - Y^T X W - W^T X^T Y + W^T X^T X W$$

$$\therefore f(w) = \|Y\|_2^2 + W^T X^T X W - 2 \cdot X^T Y \qquad (\because U^T V = V^T U)$$

$$\therefore \nabla f(w) = 2 \cdot X^T X . w - 2 \cdot X^T y \quad \text{—— } 1$$

$$\therefore H_{f(x)} = 2 \cdot x^T \cdot x \cdot I \qquad (I \in \mathbb{R}^{n \times n})$$

$$= 2 \cdot x^T x \cdot I$$

$\therefore$ H is a convex function if $X^T X \succ 0$

$\therefore \nabla f(x) = 0 \rightarrow \quad \omega = (X^T X)^{-1} \cdot X^T Y$

and unique solution exists if $X^T X \succ 0$

If $m < n$, $X^T X$ is not Positive definite

$\therefore$ The unique solution does not exists.

SR NO : 19473

Least Mean Square Error : 1.3969886255800368

w0: 0.425

w1: 1.971

w2: -2.127

w3: 0.072

w4: 4.079

b: -0.138

4. (10 points) You are each given a grey box which takes $x \in \mathbb{R}^2$ and returns $f(x)$. You know that $f(x) = x^T A x + b^T x$, where $b = [1, 1]^T$ but A is unknown. You may use the grey box to evaluate **only** 1000 points. "Estimate" whether $f(x)$ has a global minimum or not by estimating a single real number and checking it's sign. State what that number is.

$$f(x) = x^T A x + b^T x \geq \alpha \|x\|^2 + b^T x$$

if $\alpha < 0 \rightarrow f(x)$ is unbounded below

$$\alpha \leq \frac{f(x) - b^T x}{|x|^2}$$

Doing iterations over a unit circle and estimating α.

alpha ~-0.262429. The global minimum does not exist as alpha is not positive at x = (-0.23575893550942717, -0.9718115683235417)

5. (10 points) You are each given a black box function which returns $f(x)$ and $\nabla f(x)$. It is not known if $f(x)$ is convex or coercive. Use the following iteration to try and find a minimum:

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

Question: Starting at $x_0 = [10, 10, 10]$, how many iterations will it take until you reach an ε-approximate point if (a) $\varepsilon = 0.01$, (b) $\varepsilon = 0.001$, and (c) $\varepsilon = 0.0001$?

SR no : 19473

e=0.01 k= 3

e=0.001 k= 3

e=0.0001 k= 3