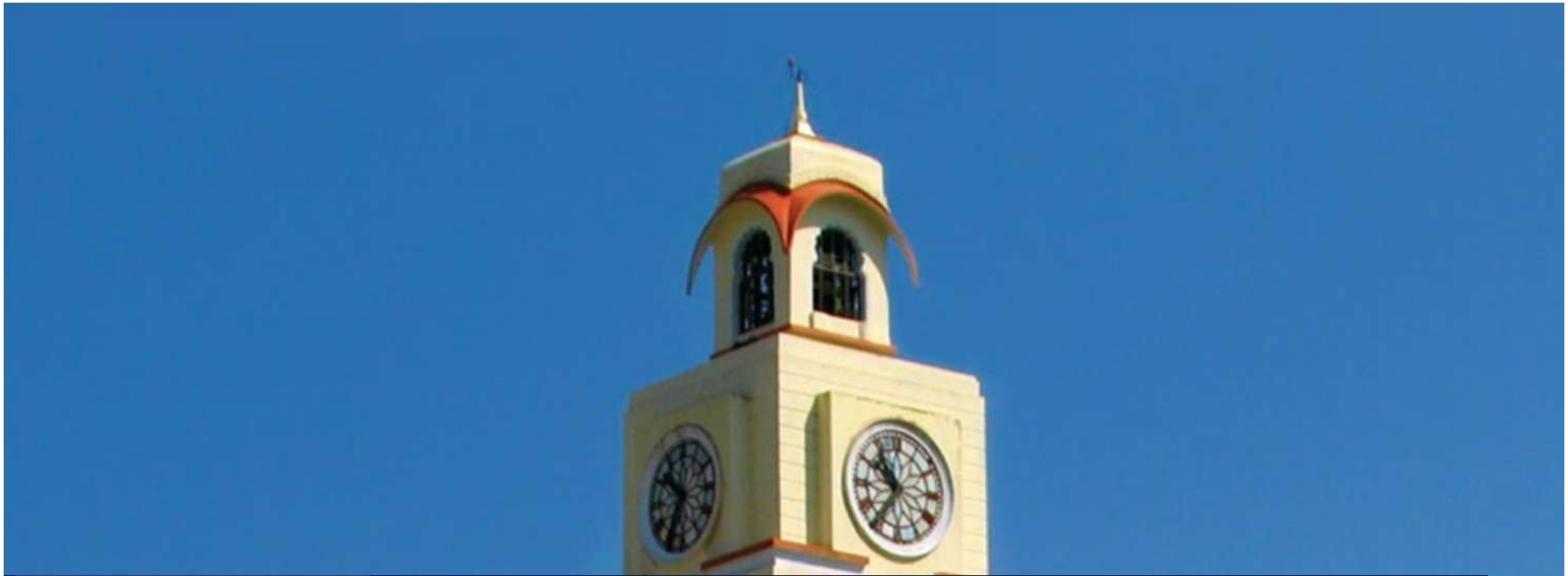
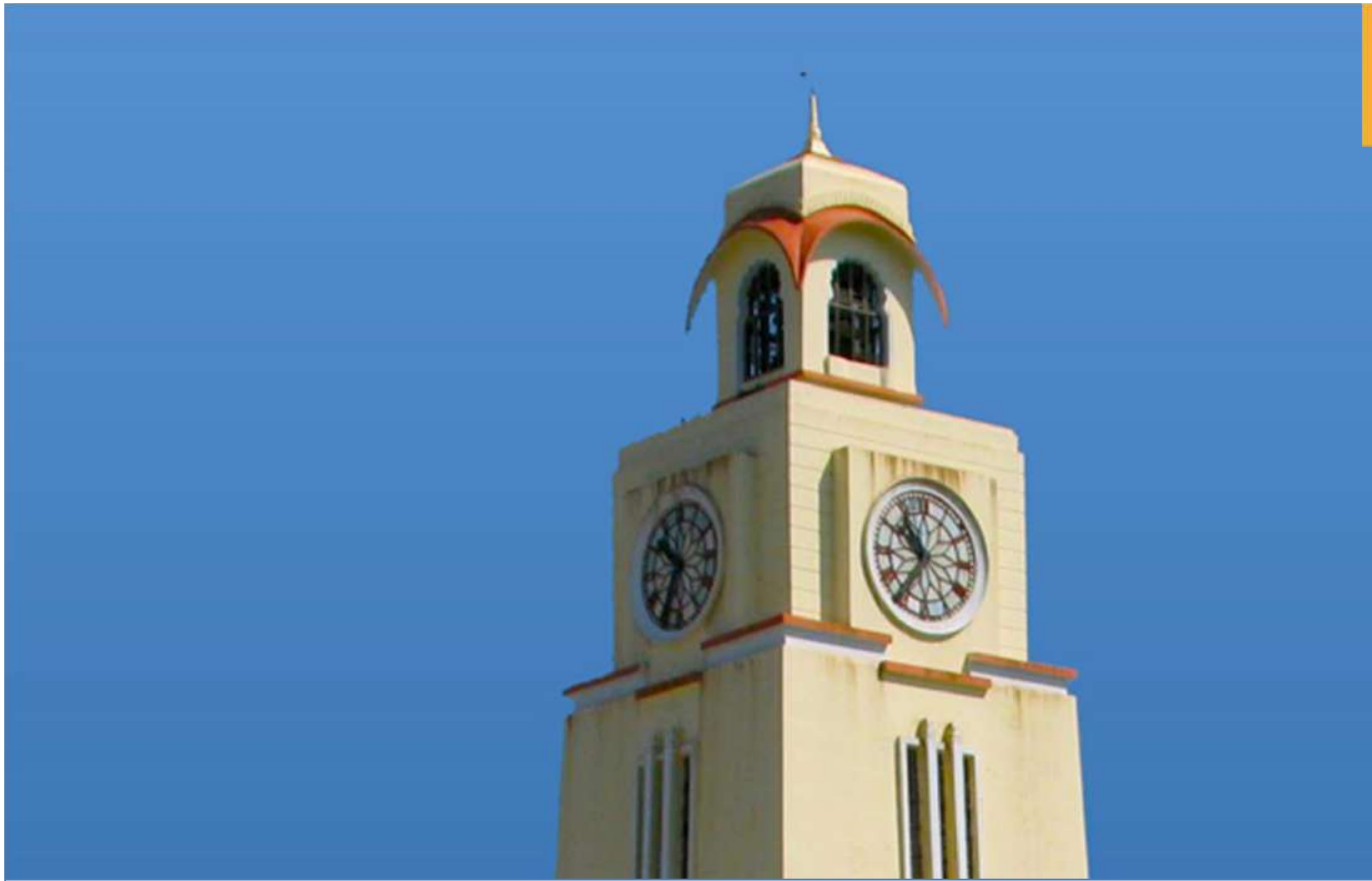


BITS Pilani
Pilani Campus

Mathematical Foundations

MFDS Team



BITS Pilani
Pilani Campus

Mathematical Foundations

Webinar 3

Agenda

- Maxima and Minima
- Gradient of a vector-valued function
- Gradient descent (Steepest Descent)
- Gradient Descent Example
- Lagrange’s Method
- Previous Year Problems based on these topics.

1. Maxima and Minima of a function of three variables using the Hessian.

Consider a function f (x, y, z) of three variables.

Step 1: Critical Points

To find critical points, we need to solve the system of equations given by the gradient of f and is given by $\nabla f = \begin{bmatrix} \dfrac{\partial f}{\partial x} & \dfrac{\partial f}{\partial y} & \dfrac{\partial f}{\partial z} \end{bmatrix}^T = \begin{bmatrix} 0 & 0 & 0 \end{bmatrix}^T$

The solutions to this system represent critical points

Step 2: Hessian Matrix

The Hessian matrix, denoted as H, is a square matrix of second-order partial derivatives of f and is given by

$$H = \begin{bmatrix} \dfrac{\partial^2 f}{\partial x^2} & \dfrac{\partial^2 f}{\partial x \partial y} & \dfrac{\partial^2 f}{\partial x \partial z} \\ \dfrac{\partial^2 f}{\partial y \partial x} & \dfrac{\partial^2 f}{\partial y^2} & \dfrac{\partial^2 f}{\partial y \partial z} \\ \dfrac{\partial^2 f}{\partial z \partial x} & \dfrac{\partial^2 f}{\partial z \partial y} & \dfrac{\partial^2 f}{\partial z^2} \end{bmatrix}$$

Maxima and Minima Continued…

Step 3: Second derivative Test

The determinant of the Hessian matrix, det(H), is used to determine the nature of critical points:

$f''(x_0) > 0$

Let x_0 be a critical point of f.

i) If the Hessian matrix H is positive definite (i.e. all eigenvalues are strictly positive) then $\delta^T H(x_0)\delta > 0$. So f (x) > f ($x_0$) and hence x_0 is a point of local minima.

ii) If the Hessian matrix H is negative definite (i.e. all eigenvalues are strictly negative) then $\delta^T H(x_0)\delta < 0$. So f (x) < f ($x_0$) and hence x_0 is a point of local maxima.

iii) If the Hessian matrix H has both positive and negative eigenvalues then x_0 is a saddle point.

iv) If the determinant of Hessian matrix H is zero then the test is inconclusive.

$x = x_0$

This method helps classify critical points as maxima, minima, or saddle points in functions of three variables

Maxima and Minima Continued…

Example: 1

Discuss the maximum and minimum value of the function $f(x,y,z)=x^2+y^2+z^2+x-2z-xy$.

$\mathbb{R}^3 \rightarrow \mathbb{R}$

Solution: $f(x,y,z) = x^2 + y^2 + z^2 + x - 2z - xy.$

$$\nabla_{x,y,z} f(x,y,z) = [f_x \quad f_y \quad f_z] = [2x+1-y \quad 2y-x \quad 2z-2]$$

$$f_x = 2x+1-y \;,\; f_y = 2y-x, \quad f_z = 2z-2$$

To find critical points, we need to solve the system of equations given by the gradient of f is equal to zero. i.e

$$\nabla f = \begin{bmatrix} \frac{\partial f}{\partial x} & \frac{\partial f}{\partial y} & \frac{\partial f}{\partial z} \end{bmatrix}^T = \begin{bmatrix} 0 & 0 & 0 \end{bmatrix}^T$$

$$2x+1-y=0 \quad \Rightarrow y = 2x+1 \quad (1)$$

$$2y-x=0 \quad \Rightarrow x = 2y \quad (2)$$

$$2z-2=0 \quad \Rightarrow z = 1$$

Substituting (2) in (1) gives y= -1/3, x=-2/3 and z=1

Maxima and Minima Continued…

The only critical point is (-2/3,-1/3,1) where the function will be either minima or maxima.

The Hessian matrix $H = \begin{bmatrix} \frac{\partial^2 f}{\partial x^2} & \frac{\partial^2 f}{\partial x \partial y} & \frac{\partial^2 f}{\partial x \partial z} \\ \frac{\partial^2 f}{\partial y \partial x} & \frac{\partial^2 f}{\partial y^2} & \frac{\partial^2 f}{\partial y \partial z} \\ \frac{\partial^2 f}{\partial z \partial x} & \frac{\partial^2 f}{\partial z \partial y} & \frac{\partial^2 f}{\partial z^2} \end{bmatrix} = \begin{bmatrix} 2 & -1 & 0 \\ -1 & 2 & 0 \\ 0 & 0 & 2 \end{bmatrix}$

C_3

Further, The determinant of Hessian H is , $\det(H) = \begin{vmatrix} 2 & -1 & 0 \\ -1 & 2 & 0 \\ 0 & 0 & 2 \end{vmatrix} = 6 > 0$

R_3

The determinants of all the leading principal minors of H are given by .

$$M_1 = 2 > 0; \quad M_2 = \begin{vmatrix} 2 & -1 \\ -1 & 2 \end{vmatrix} = 3 > 0; \quad M_2 = \begin{vmatrix} 2 & -1 & 0 \\ -1 & 2 & 0 \\ 0 & 0 & 2 \end{vmatrix} = 6 > 0$$

If det(H) > 0 and all the leading principal minors of H are positive, then the critical point is a local minimum. Thus we have a minimum at (-2/3,-1/3,1) and the minimum value is -4/3.

$f\left(-\frac{2}{3}, -\frac{1}{3}, 1\right) = -\frac{4}{3}$

2.a Gradient of a vector-valued function with respect to input vector

For a function $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$ and a vector valued function $x = [x_{1,} x_{2,} \dots . x_n]^T \in \mathbb{R}^n$, the corresponding vector of function is given as

$$f(x) = \begin{bmatrix} f_1(x) \\ \vdots \\ f_m(x) \end{bmatrix} \in \mathbb{R}^m$$

$f(x) \in \mathbb{R}^m$

$f = \begin{pmatrix} f_1 \\ f_m \end{pmatrix}$

The gradient of $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$ with respect to x $\in \mathbb{R}^n$ is given by

$$\frac{\mathrm{d}\boldsymbol{f}(\boldsymbol{x})}{\mathrm{d}\boldsymbol{x}} = \begin{bmatrix} \boxed{\frac{\partial \boldsymbol{f}(\boldsymbol{x})}{\partial x_1}} & \cdots & \boxed{\frac{\partial \boldsymbol{f}(\boldsymbol{x})}{\partial x_n}} \end{bmatrix}$$

$$= \begin{bmatrix} \frac{\partial f_1(\boldsymbol{x})}{\partial x_1} & \cdots & \frac{\partial f_1(\boldsymbol{x})}{\partial x_n} \\ \vdots & & \vdots \\ \frac{\partial f_m(\boldsymbol{x})}{\partial x_1} & \cdots & \frac{\partial f_m(\boldsymbol{x})}{\partial x_n} \end{bmatrix} \in \mathbb{R}^{m \times n}.$$

$\mathbb{R}^n \rightarrow \mathbb{R}^m$

$\in \mathbb{R}^{m \times n}$

a) Gradient of a vector with respect to input vector

Example 2:

Let $f = [\ 3x_1^2 x_2 \quad 2x_1 + x_2^8\ ]^T$, then find the Gradient of f.

$f: \mathbb{R}^2 \to \mathbb{R}^2$

Solution: $\frac{df}{dx} = \begin{bmatrix} \frac{\partial f_1(x)}{\partial x_1} & \frac{\partial f_1(x)}{\partial x_2} \\ \frac{\partial f_2(x)}{\partial x_1} & \frac{\partial f_2(x)}{\partial x_2} \end{bmatrix}$

$= \begin{bmatrix} 6x_1x_2 & 3x_1^2 \\ 2 & 8x_2^7 \end{bmatrix}$

2×2

b) Gradient of a scalar with respect to a matrix

Let $f : \mathbb{R}^{MXN} \rightarrow \mathbb{R}$ and a matrix $A \in \mathbb{R}^{MXN}$, then the gradient of f is $\begin{bmatrix} \frac{\partial f_1(x)}{\partial x_1} & \frac{\partial f_1(x)}{\partial x_2} & \frac{\partial f_1(x)}{\partial x_3} \\ \frac{\partial f_2(x)}{\partial x_1} & \frac{\partial f_2(x)}{\partial x_2} & \frac{\partial f_2(x)}{\partial x_3} \\ \frac{\partial f_3(x)}{\partial x_1} & \frac{\partial f_3(x)}{\partial x_2} & \frac{\partial f_3(x)}{\partial x_3} \end{bmatrix}$

Example:3

Let f = $2\,x + 4\,y + 3\,z + 4\,w$ and the matrix $A = \begin{bmatrix} x & y \\ z & w \end{bmatrix}$ the find the gradient of f.

$+2m+2n$

Solution: The gradient of f is

$$\frac{df}{dA} = \begin{bmatrix} \frac{\partial f}{\partial x} & \frac{\partial f}{\partial y} \\ \frac{\partial f}{\partial z} & \frac{\partial f}{\partial w} \end{bmatrix} = \begin{bmatrix} 2 & 4 \\ 3 & 4 \end{bmatrix}$$

2×2

$\frac{\partial f}{\partial A}$ $A = [\quad]$

2×2 2×3

3. Proof one of the gradient identities involving trace

Example: 4

Consider the function $f(\mathrm{x}) = \mathrm{Tr}(\mathrm{A\,x})$, where $\mathrm{A} \in \mathrm{R}^{2\times 2}$ and $\mathrm{x} \in \mathrm{R}^{2\times 2}$. Determine the gradient of 'f' with respect to x, denoted as $\nabla_X f(\mathrm{x})$.

$f : \mathbb{R}^{2\times 2} \to \mathbb{R}$

Solution:

$$A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \text{ and } \mathrm{x} = \begin{pmatrix} x_{11} & x_{12} \\ x_{21} & x_{22} \end{pmatrix}$$

$$A\mathrm{x} = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}\begin{pmatrix} x_{11} & x_{12} \\ x_{21} & x_{22} \end{pmatrix}$$

$$A\mathrm{x} = \begin{pmatrix} a_{11}x_{11} + a_{12}x_{21} & a_{11}x_{12} + a_{12}x_{22} \\ a_{21}x_{11} + a_{22}x_{21} & a_{21}x_{12} + a_{22}x_{22} \end{pmatrix}$$

trace

$$f(\mathrm{x}) = Tr(\mathrm{A\,x})$$

$$f(\mathrm{x}) = (a_{11}x_{11} + a_{12}x_{21} + a_{21}x_{12} + a_{22}x_{22})$$

$\mathrm{Tr}\ A^T$

$$\nabla_X f(\mathrm{x}) = \begin{pmatrix} \frac{\partial f(\mathrm{x})}{\partial x_{11}} & \frac{\partial f(\mathrm{x})}{\partial x_{12}} \\ \frac{\partial f(\mathrm{x})}{\partial x_{21}} & \frac{\partial f(\mathrm{x})}{\partial x_{22}} \end{pmatrix} = \begin{pmatrix} a_{11} & a_{21} \\ a_{12} & a_{22} \end{pmatrix} = A^T$$

4. Gradient descent (Steepest Descent)

Definition: Gradient Descent is an optimization algorithm used to minimize a function by iteratively moving in the direction of steepest descent.

Purpose: The primary goal of Gradient Descent is to find the local minimum of a function. It's widely used in machine learning for optimizing models and finding optimal parameters.

How it Works:

- At each iteration, the algorithm computes the gradient of the objective function with respect to the parameters.
- It then updates the parameters in the opposite direction of the gradient to reduce the value of the function.
- This process continues until convergence is achieved or a stopping criterion is met.

Algorithm

Data : $x_0 \in \mathbb{R}^n$

Step 0: Set $i = 0$

Step 1: If $\nabla f(x_i) = 0$ then stop

else, compute search direction $h_i = -\nabla f(x_i)$.

Step 2: Compute the size $\lambda_i \in \arg\min_{\lambda \geq 0} f(x_i + \lambda \cdot h_i)$

Step 3: Set $x_{i+1} = x_i + \lambda_i h_i$ go to step 1.

5. Derivation of Step size (τ)

- Suppose we want to minimize the **quadratic function:** $f(x) = \frac{1}{2}x^T Q x - b^T x$
- Let the iterative formula for minimum point be $x_{k+1} = x_k - \tau_k \nabla f(x_k)$

Where τ_k be the step size

- Let $g(\tau) = f(x_{k+1}) = f\left(x_k - \tau_k \nabla f(x_k)\right)$

$$g(\tau) = \frac{1}{2}\left[x_k - \tau_k \nabla f(x_k)\right]^T Q\left[x_k - \tau_k \nabla f(x_k)\right] - b^T\left[x_k - \tau_k \nabla f(x_k)\right]$$

$$g(\tau) = a\tau^2 + d\tau + c$$

$$a = \frac{1}{2}\left[\nabla f(x_k)\right]^T Q \nabla f(x_k)$$

$$d = \left[b^T - x^T Q\right]\nabla f(x_k) = -\left[\nabla f(x_k)\right]^T \nabla f(x_k)$$

Continued…

Here $g(\tau)$ is quadratic and concave function

- Condition for extrema of $\frac{dg}{d\tau}=0$ is

$$\frac{dg}{d\tau}=2a\tau+d=0 \Rightarrow \tau=-\frac{d}{2a}$$

$$\tau_k=\frac{\left[\nabla f\left(x_k\right)\right]^T \nabla f\left(x_k\right)}{\left[\nabla f\left(x_k\right)\right]^T Q \nabla f\left(x_k\right)}$$

Therefore

$$\tau=\frac{S^T S}{S^T Q S}$$

$S = \nabla f(x_k)$

6. Step size for Quadratic function

Let $f(x,y)=0.5ax^2+0.5by^2+cxy$

$$S=\nabla f=\begin{bmatrix}\frac{\partial f}{\partial x}\\ \frac{\partial f}{\partial y}\end{bmatrix}=\begin{bmatrix}ax+cy\\ by+cx\end{bmatrix}$$

$$H=\begin{bmatrix}f_{xx} & f_{xy}\\ f_{yx} & f_{yy}\end{bmatrix}=\begin{bmatrix}a & c\\ c & b\end{bmatrix}$$

Hence the step size of the steepest descent algorithm is given by

$$\tau=\frac{S^T S}{S^T HS}=\frac{\begin{bmatrix}ax+cy\\ by+cx\end{bmatrix}^T\begin{bmatrix}ax+cy\\ by+cx\end{bmatrix}}{\begin{bmatrix}ax+cy\\ by+cx\end{bmatrix}^T\begin{bmatrix}a & c\\ c & b\end{bmatrix}\begin{bmatrix}ax+cy\\ by+cx\end{bmatrix}}$$

$$\tau=\frac{(ax+cy)^2+(by+cx)^2}{\left[(a^2+c^2)x+(ac+bc)y\right][ax+cy]+\left[(ac+bc)x+(b^2+c^2)y\right][by+cx]}$$

7. Problems

Q1. Find the minimum of f(x,y) = $3x^2+y^2$ by

a) Computing the gradient of f and τ

b) Solve with initial values x_0 = 1 and y_0 = 3

Solution:

$$f(x,y) = 3x^2 + y^2 \text{ with } (x_0, y_o) = (1,3)$$

$$f(1,3) = 3.1^2 + 3^2 = 12$$

Step1: We find $\nabla f = \begin{bmatrix} \frac{\partial f}{\partial x} \\ \frac{\partial f}{\partial y} \end{bmatrix}$

$$s = \nabla f = \begin{bmatrix} 6x \\ 2y \end{bmatrix} \text{ and } H = \begin{bmatrix} 6 & 0 \\ 0 & 2 \end{bmatrix}$$

$-S_i, \nabla f_i = S$

Q1 Continued…

Step2: Compute $\tau_i = \dfrac{s^T s}{s^T H s}$ is the step size.

$\tau_i = \dfrac{S^T S}{S^T H S}$

$$\tau_i = \frac{\begin{bmatrix} 6x_i & 2y_i \end{bmatrix}\begin{bmatrix} 6x_i \\ 2y_i \end{bmatrix}}{\begin{bmatrix} 6x_i & 2y_i \end{bmatrix}\begin{bmatrix} 6 & 0 \\ 0 & 2 \end{bmatrix}\begin{bmatrix} 6x_i \\ 2y_i \end{bmatrix}}$$

$$\tau_i = \frac{36x_i^2 + 4y_i^2}{216x_i^2 + 8y_i^2}$$

Step:3 Iterate the minimum point

$x_{i+1} = x_i - \tau_i \nabla f_i$

$$x_{i+1} = x_i - \tau_i \nabla f_i$$

$$\begin{bmatrix} x_{i+1} \\ y_{i+1} \end{bmatrix} = \begin{bmatrix} x_i \\ y_i \end{bmatrix} - \left(\frac{36x_i^2 + 4y_i^2}{216x_i^2 + 8y_i^2} \right) \begin{bmatrix} 6x_i \\ 2y_i \end{bmatrix}$$

Q1 Continued…

1st Iteration

i=0 $$\begin{bmatrix} x_1 \\ y_1 \end{bmatrix} = \begin{bmatrix} x_0 \\ y_0 \end{bmatrix} - \left(\frac{36x_0^2 + 4y_0^2}{216x_0^2 + 8y_0^2} \right) \begin{bmatrix} 6x_0 \\ 2y_0 \end{bmatrix}$$

$$\begin{bmatrix} x_1 \\ y_1 \end{bmatrix} = \begin{bmatrix} 1 \\ 3 \end{bmatrix} - \left(\frac{36.1^2 + 4.3^2}{216.1^2 + 8.3^2} \right) \begin{bmatrix} 6.1 \\ 2.3 \end{bmatrix}$$

$$\begin{bmatrix} x_1 \\ y_1 \end{bmatrix} = \begin{bmatrix} 1 \\ 3 \end{bmatrix} - \left(\frac{72}{288} \right) \begin{bmatrix} 6 \\ 6 \end{bmatrix}$$

$$\begin{bmatrix} x_1 \\ y_1 \end{bmatrix} = \begin{bmatrix} -0.5 \\ 1.5 \end{bmatrix}$$

$$f(-0.5, 1.5) = 3(-0.5)^2 + (1.5)^2 = 3$$

2nd Iteration

i=1 $$\begin{bmatrix} x_2 \\ y_2 \end{bmatrix} = \begin{bmatrix} x_1 \\ y_1 \end{bmatrix} - \left(\frac{36x_1^2 + 4y_1^2}{216x_1^2 + 8y_1^2} \right) \begin{bmatrix} 6x_1 \\ 2y_1 \end{bmatrix}$$

$$\begin{bmatrix} x_2 \\ y_2 \end{bmatrix} = \begin{bmatrix} -0.5 \\ 1.5 \end{bmatrix} - \left(\frac{36.(-0.5)^2 + 4.(1.5)^2}{216.(-0.5)^2 + 8.(1.5)^2} \right) \begin{bmatrix} 6.(-0.5) \\ 2.(1.5) \end{bmatrix}$$

$$\begin{bmatrix} x_2 \\ y_2 \end{bmatrix} = \begin{bmatrix} -0.5 \\ 1.5 \end{bmatrix} - \left(\frac{18}{72} \right) \begin{bmatrix} -3 \\ 3 \end{bmatrix}$$

$$\begin{bmatrix} x_2 \\ y_2 \end{bmatrix} = \begin{bmatrix} 0.25 \\ 0.75 \end{bmatrix}$$

$$f(0.25, 0.75) = 3(0.25)^2 + (0.75)^2 = 0.75$$

Q1 Continued…

3rd Iteration

i=2 $$\begin{bmatrix} x_3 \\ y_3 \end{bmatrix} = \begin{bmatrix} x_2 \\ y_2 \end{bmatrix} - \left(\frac{36x_2^2 + 4y_2^2}{216x_2^2 + 8y_2^2} \right) \begin{bmatrix} 6x_2 \\ 2y_2 \end{bmatrix}$$

$$\begin{bmatrix} x_3 \\ y_3 \end{bmatrix} = \begin{bmatrix} 0.25 \\ 0.75 \end{bmatrix} - \left(\frac{36.(0.25)^2 + 4.(0.75)^2}{216.(0.25)^2 + 8.(0.75)^2} \right) \begin{bmatrix} 6.(0.25) \\ 2.(0.75) \end{bmatrix}$$

$$\begin{bmatrix} x_3 \\ y_3 \end{bmatrix} = \begin{bmatrix} 0.25 \\ 0.75 \end{bmatrix} - \left(\frac{4.5}{18} \right) \begin{bmatrix} 1.5 \\ 4.5 \end{bmatrix}$$

$$\begin{bmatrix} x_3 \\ y_3 \end{bmatrix} = \begin{bmatrix} -0.125 \\ -0.375 \end{bmatrix}$$

$$f(-0.125, -0.375) = 3(-0.125)^2 + (-0.375)^2 = 0.1875$$

4th Iteration

i=3 $$\begin{bmatrix} x_4 \\ y_4 \end{bmatrix} = \begin{bmatrix} x_3 \\ y_3 \end{bmatrix} - \left(\frac{36x_3^2 + 4y_3^2}{216x_3^2 + 8y_3^2} \right) \begin{bmatrix} 6x_3 \\ 2y_3 \end{bmatrix}$$

$$\begin{bmatrix} x_1 \\ y_1 \end{bmatrix} = \begin{bmatrix} -0.125 \\ 0.375 \end{bmatrix} - \left(\frac{36.(-0.125)^2 + 4.(0.375)^2}{216.(-0.125)^2 + 8.(0.375)^2} \right) \begin{bmatrix} 6.(-0.125) \\ 2.(0.375) \end{bmatrix}$$

$$\begin{bmatrix} x_4 \\ y_4 \end{bmatrix} = \begin{bmatrix} -0.125 \\ 0.375 \end{bmatrix} - \left(\frac{1}{4} \right) \begin{bmatrix} -0.75 \\ 0.75 \end{bmatrix}$$

$$\begin{bmatrix} x_4 \\ y_4 \end{bmatrix} = \begin{bmatrix} 0.0625 \\ 0.1875 \end{bmatrix}$$

$$f(0.0625, 0.1875) = 3(0.0625)^2 + (0.1875)^2 = 0.046875$$

At the end of 10th iteration, we get $\begin{bmatrix} x_{10} \\ y_{10} \end{bmatrix} = \begin{bmatrix} -0.001953 \\ 0.005859 \end{bmatrix}$ and the minimum value is

$$f(-0.001953, 0.005859) = 3(-0.001953)^2 + (0.005859)^2 = 0.000046$$

8. Lagrange's Method of Multipliers

Property: The gradient of a function is perpendicular to the contour lines, the tangents to the contour lines of *f* and *g* are parallel if and only if the gradients of *f* and *g* are parallel.

Thus we need points (x, y) where $g(x, y) = c$ and

$$\nabla_{x,y} f = \lambda \nabla_{x,y} g,$$

for some λ

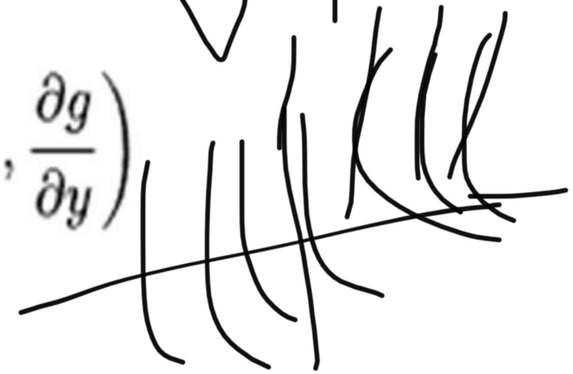
where

$$\nabla_{x,y} f = \left(\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y} \right), \qquad \nabla_{x,y} g = \left(\frac{\partial g}{\partial x}, \frac{\partial g}{\partial y} \right)$$

$f(x, y) = c$

$g(x, y) = c$

∇f



Example:

Q:

An asteroid is entering the atmosphere of moon. The shape of the asteroid is described by the equation $4x2+y2+4z2=16$. The temperature on the surface of the asteroid after one month was observed to be represented by equation $8x2+4yz-16z+600$. Is it possible to find the point on surface of asteroid with maximum temperature? If yes, find it?

$(2, 0, 0)$ 600
$+32$
632

$f(x,y,z) = 8x^2 + 4yz$

$4x^2+y^2+4z^2=16$

Solution:

$$\text{Let } f(x,y,z) = 8x^2 + 4yz - 16z + 600, \quad \text{———} \quad (1)$$

$$g(x,y,z) = 4x^2 + y^2 + 4z^2 - 16 \quad \text{———} \quad (2)$$

$$\nabla f = \lambda \nabla g \Rightarrow \frac{\partial f}{\partial x} i + \frac{\partial f}{\partial y} j + \frac{\partial f}{\partial z} k = \lambda \left[\frac{\partial g}{\partial x} i + \frac{\partial g}{\partial y} j + \frac{\partial g}{\partial z} k \right]$$

Solution Continued…

$$16x\,i + 4z\,j + (4y-16)\,k = \lambda[8x\,i + 2y\,j + 8z\,k]$$

$16x = \lambda 8x \rightarrow \lambda = 2;$

$4z = \lambda 2y \Rightarrow z = y;$

$4y - 16 = 16z$

$\Rightarrow y = \frac{-4}{3},\ z = \frac{-4}{3}$

$\nabla f = \lambda \nabla g$

substituting in equation (1), $4x^2 + \left(\frac{-4}{3}\right)^2 + 4\left(\frac{-4}{3}\right)^2 = 16,\ x = \pm\frac{4}{3}$

Hence at the point $\left(\pm\frac{4}{3}, \frac{-4}{3}, \frac{-4}{3}\right)$ on the surface of asteroid with maximum temperature is **1928/3=342.67**

$= 646.67$

9. Previous Year Problems

Q1:

(x, y)

Answer the following questions with justifications.

A signal tower is placed at $(0, 0)$. The cost to place a receiver at a point (x, y) is equal to the square of the Euclidean distance from that point to the point at which the signal tower is placed. The region of interest is $\{(x, y) | 7 - 2x + y \leq 0, 0 \leq x + y + 5\}$.

$\sqrt{x^2+y^2}$

i) Help the analyst to formulate the optimization problem to find the point at which the cost to place the receiver is minimum in the region of interest. (1.5Marks)

ii) Write down the Lagrangian for the constrained optimization problem so obtained in i) (0.5 Marks)

Previous Year Problems continued….

Q1i) Solution :

The formulation is

$$\min x^2 + y^2$$

subject to constraint $7 - 2x + y \le 0,$

$$x + y + 5 \ge 0$$

$\sqrt{x^2+y^2}$ or x^2+y^2

λ_1

λ_2

Q1ii) Solution:

The Lagrangian is given by

$$L(x, y, \lambda_1, \lambda_2) = x^2 + y^2 + \lambda_1(7 - 2x + y) + \lambda_2(-5 - x - y).$$

$\frac{\partial L}{\partial x} = 0$ | $\frac{\partial L}{\partial y} = 0$

innovate achieve lead

Previous Year Problems continued….

Q2:

Answer the following questions with justifications.

Consider the function $f(x) = ax^3 + bx^2 + cx + d$ where $a > 0$.

(a) Find the condition on a, b, c such that the function has two distinct critical points. Calculate the critical points in terms of a, b, c. Identify the nature of each critical point (i.e maxima or minima) [3 Marks]

(b) Define a zone of attraction around each local minimum to be the region around it such that if gradient descent starts at any point in the region, it would end up at the given local minimum. Find the zone of attraction for each local minimum, if any, of the critical points. Justify your answer mathematically. [3 Marks]

Previous Year Problems continued….

Q2)a) Solution :

To find the critical points we take $\frac{df}{dx} = 3ax^2 + 2bx + c = 0$, to obtain two roots $x_1 = \frac{-b-\sqrt{b^2-3ac}}{3a}$ and $x_2 = \frac{-b+\sqrt{b^2-3ac}}{3a}$.

In order for the the two roots to be real and distinct, the quantity under the square-root sign needs to be strictly positive, i.e $b^2 - 3ac > 0$.

We see that the second derivative $\frac{d^2 f}{dx^2} = 6ax + 2b$.

$\frac{\partial^2 f}{\partial a^2}(x_1) < 0$

The second derivative is negative for the critical point x_1 and positive for x_2 which means x_1 is a maxima and x_2 is a minima.

Previous Year Problems continued….

Q2)b) Solution :

As identified in part (a), there is a single local minimum $x_2 = \frac{-b+\sqrt{b^2-3ac}}{3a}$.

We can rewrite $\frac{df}{dx} = 3a(x + \frac{b}{3a})^2 + (c - \frac{b^2}{3a})$

We can rewrite $\frac{df}{dx} = 3a(x + \frac{b}{3a})^2 + (c - \frac{b^2}{3a})$. Solving for $\frac{df}{dx} < 0$ we see that $\frac{df}{dx} < 0$ when $x > \frac{-b-\sqrt{b^2-3ac}}{3a}$ and $x < \frac{-b+\sqrt{b^2-3ac}}{3a}$. When $x > \frac{-b+\sqrt{b^2-3ac}}{3a}$, $\frac{df}{dx} > 0$.

gradient descent will take any point on the left of the local minimum x_2 but greater than $\frac{-b-\sqrt{b^2-3ac}}{3a}$ to x_2 in a sufficiently large number of steps with a suitable step-size.

Similarly since the derivative is positive on the right of x_2, gradient descent will take any point on the right of x_2 to x_2 in a sufficiently large number of steps with a suitable step-size. Thus the zone of attraction for the local minimum $x_2 = \frac{-b+\sqrt{b^2-3ac}}{3a}$ is $\frac{-b-\sqrt{b^2-3ac}}{3a} < x \leq \infty$.

Previous Year Problems continued….

Q3:

(3 marks)

Consider a quadratic function $f(x, y) = x^2 + \beta y^2$ where $\beta \in \mathbb{R}$ is an unknown constant. Also assume that $\beta > 0$. Consider the problem of minimizing this function using gradient descent algorithm:

(i) Derive a closed form expression (involving β) for the optimal step-size α for the first iteration of gradient descent if the initial point is

$$\begin{bmatrix} x_0 \\ y_0 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

(ii) If it is given to you that the optimal step-size $\alpha = 0.5$, derive the value of constant β using the formula derived in (i).

(3 marks)

Previous Year Problems continued….

Q3)a) Solution:

The gradient of $f(x, y)$ at (x_0, y_0) is as follows

$$\nabla f(x_0, y_0) = \begin{bmatrix} 2x_0 \\ 2\beta y_0 \end{bmatrix} = \begin{bmatrix} 2 \\ 2\beta \end{bmatrix}$$

The optimal step size is obtained as solution of following 1 dimensional optimization problem

$$\text{argmin}_\alpha f\left(\begin{bmatrix} x_0 \\ y_0 \end{bmatrix} - \alpha \nabla f(x_0, y_0) \right)$$

$x_{k+1} = x_k - \alpha \nabla f_k$

after substituting $x_0 = 1$ and $y_0 = 1$, we get that, this is equivalent to solving

$$\text{argmin}_\alpha f(1 - 2\alpha, 1 - 2\alpha\beta)$$

In other words we need to find the minimum of

$$g(\alpha) = (1 - 2\alpha)^2 + \beta(1 - 2\alpha\beta)^2$$

To find minimum of $g(\alpha)$ with respect to α, we find $g'(\alpha)$ and equate it to zero

$$g'(\alpha) = 2\beta^3\alpha - \beta^2 + 2\alpha - 1 = 0$$

Hence the closed form expression for α is given by

$$\alpha = \frac{1 + \beta^2}{2 + 2\beta^3}$$

innovate achieve lead

Previous Year Problems continued….

Q3)b) Solution:

If it is given that $\alpha = 0.5$, substitutng in the repvious expression we get the equation

$$0.5 = \frac{1 + \beta^2}{2 + 2\beta^3}$$

Rearranging we get the expression $\beta^2(\beta - 1) = 0$. Hence potential value of β is 0 or 1. Its given in question that $\beta \neq 0$.
Hence final answer is $\beta = 1$.

Previous Year Problems continued….

Q4:

Answer the following questions with justifications.

Consider a point in 2D space $\mathcal{P} = (4, 2)$ and a line represented by the equation $f(x, y) = 0$. Using the method of Lagrange multipliers, derive the closest point on this line to the given point $\mathcal{P}$. You can assume that the closeness between two points is measured by square of euclidean distance. Derive the closest point to $\mathcal{P}$ when

i) $f(x, y) = x - 2y + 3$

ii) $f(x, y) = x + 2y + 5$

Also derive the distance to $\mathcal{P}$ in both cases. (4 marks)

$d^2(x.y) = x^2 + y^2$

Previous Year Problems continued….

Q4)a) Solution:

The line is $x-2y+3=0$. The optimization problem to be minimized to obtain the closest point is given by

$$\min\,(4-x)^2+(2-y)^2$$
$$\text{subject to } x-2y+3=0$$

(x, y) $(4, 2)$

$(x-4)^2+(y-2)^2$

Construct the lagrangian equation as

$$L(x,y,\lambda)=(4-x)^2+(2-y)^2+\lambda(x-2y+3)=0$$

Solve the equation $\nabla L(x,y,\lambda)=\mathbf{0}$

$2(x-4)+\lambda=0, 2(y-2)-2\lambda=0, x-2y+3=0$

Then , $x=\frac{17}{5}, y=\frac{16}{5}$

And distance $=\sqrt{(4-\frac{17}{5})^2+(2-\frac{16}{5})^2}=\sqrt{\frac{9}{5}}$

Previous Year Problems :

Q4)b) Solution:

The line is $x+2y+5=0$. The optimization problem to be minimized to obtain the closest point is given by

$$\min\ (4-x)^2 + (2-y)^2$$
$$\text{subject to } x + 2y + 5 = 0$$

Construct the lagrangian equation as

$$L(x,y,\lambda) = (4-x)^2 + (2-y)^2 + \lambda(x+2y+5) = 0$$

Solve the equation $\nabla L(x,y,\lambda) = \mathbf{0}$

$2(x-4)+\lambda = 0, 2(y-2)+2\lambda = 0, x+2y+5 = 0$

$\frac{\partial L}{\partial x} = 0$

$\frac{\partial L}{\partial y} = 0$ $\frac{\partial L}{\partial \lambda} = 0$

Then , $x = \frac{7}{5}, y = \frac{-16}{5}$

And distance $= \sqrt{(4-\frac{7}{5})^2 + (2-\frac{-16}{5})^2} = \sqrt{\frac{169}{5}}$

Previous Year Problems :

Q5:

Answer the following questions with justifications.

A data science intern arrived at a loss function given by

$$f(x) = 3x^4 - 20x^3 + 36x^2 + 10.$$

i) Help the intern to find the stationary points and classify and hence the global minima. [3 marks]

ii) Suggest the intern whether $x = 0.5$ or $x = 3.5$ is a better initial condition to find global minima using simple gradient descent method with reasons. [1 mark]

Previous Year Problems :

Q5)i) Solution:

Given $f(x) = 3x^4 - 20x^3 + 36x^2 + 10$. So, we have

$$\begin{aligned} f'(x) = &\quad 12x^3 - 60x^2 + 72x = 0 \\ \Rightarrow &\quad 12x(x-2)(x-3) = 0 \\ \Rightarrow &\quad x = 0, 2, 3 \end{aligned}$$

0, 3, lo

$$\begin{aligned} f''(x) = &\quad 36x^2 - 120x + 72 \\ \Rightarrow f''(0) = &\quad 72 > 0 \Rightarrow \text{ Pt of minima} \\ \Rightarrow f''(2) = &\quad -24 < 0 \Rightarrow \text{ Pt of maxima} \\ \Rightarrow f''(3) = &\quad 36 > 0 \Rightarrow \text{ Pt of minima} \end{aligned}$$

2

Now $f(0) = 10 < f(3) = 37$. Therefore 0 is a pt of global minima.

Q5)ii) Solution:

Clearly 0.5 is closer to 0 (global minima) and 3.5 is closer to 3 (local minima). So, 0 is a better initial condition.

Previous Year Problems :

Q6:

The set of interest for a data scientist was

$$M = \{(x, y) \in \mathbb{R}^2 | x^2 + y^2 \leq 5, x + 2y = 4, x \geq 0, y \geq 0\}.$$

i) Formulate a constrained optimization problem to find a point in M nearest to the origin $(0, 0)$. [1 mark]

ii) Write the Lagrangian function for the above problem. [1 mark]

(x, y)

$d(x, y) = \sqrt{x^2 + y^2}$

Previous Year Problems :

Q6)i) Solution:

Distance from origin to a point (x, y) is $(x^2+y^2)^{\frac{1}{2}}$. Therefore the problem is

$$\min(x^2 + y^2)^{\frac{1}{2}}$$

subject to constraints

$$x^2 + y^2 - 5 \leq 0,$$
$$x + 2y - 4 = 0,$$
$$-x \leq 0, -y \leq 0.$$

$z = (x^2+y^2$

$= (x, y$

$\alpha_1, \alpha_2, \alpha_3, \gamma$

Q6)ii) Solution:

The Lagrangian function is given by

$$L(x, y, \alpha_1, \alpha_2, \alpha_3, \gamma) = (x^2 + y^2)^{\frac{1}{2}} + \alpha_1(x^2 + y^2 - 5) - \alpha_2 x - \alpha_3 y + \gamma(x + 2y - 4)$$

THANK YOU