1/11/2024

- $f : \mathbb{R}^n \to \mathbb{R}$

$\frac{\partial f}{\partial x_i}$: is the partial derivative of f with respect to x_i

$$\frac{\partial f}{\partial \underset{\sim}{x}} = \nabla f \equiv \left[\frac{\partial f}{\partial x_1} \quad \frac{\partial f}{\partial x_2} \quad \cdots \quad \frac{\partial f}{\partial x_n}\right]^T$$, called the gradient of f.

if f has a local maximum or minimum at x_0

then $\nabla f \big|_{\underset{\sim}{x} = \underset{\sim}{x_0}} = [0 \quad 0 \quad \cdots \quad 0]^T$

the second order partial derivatives form a matrix

$$\frac{\partial^2 f}{\partial \underset{\sim}{x}\, \partial \underset{\sim}{x}^T} = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1 \partial x_1} & \frac{\partial^2 f}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f}{\partial x_2 \partial x_1} & \frac{\partial^2 f}{\partial x_2 \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_2 \partial x_n} \\ \vdots & \vdots & & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1} & \frac{\partial^2 f}{\partial x_n \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_n \partial x_n} \end{bmatrix}$$

we call the above matrix to be the Hessian matrix of f.

For a point x_0, if f has strict local minimum (or maximum) is equivalent to $\frac{\partial^2 f}{\partial \underset{\sim}{x}\, \partial x^T}\Big|_{x = x_0}$ is a positive definite (or negative definite) matrix.

- if A is a symmetric matrix then A is a pd matrix iff $A = B^T B$

a matrix $A_{n\times n}$ is a positive definite matrix if for some non-singular matrix B

$$\left[\begin{array}{cc} \frac{\partial^2 f}{\partial x_1 \partial x_1} & \frac{\partial^2 f}{\partial x_1 \partial x_2} \\ \frac{\partial^2 f}{\partial x_2 \partial x_1} & \frac{\partial^2 f}{\partial x_2 \partial x_2} \end{array}\right]_{x=x_0}$$

this is a pd matrix if

$$\left.\frac{\partial^2 f}{\partial x_1 \partial x_1}\right|_{x=x_0} > 0 \quad \text{and}$$

$$\left|\left[\begin{array}{cc} \frac{\partial^2 f}{\partial x_1 \partial x_1} & \frac{\partial^2 f}{\partial x_1 \partial x_2} \\ \frac{\partial^2 f}{\partial x_2 \partial x_1} & \frac{\partial^2 f}{\partial x_2 \partial x_2} \end{array}\right]\right| > 0$$

$x^T A x > 0 \; \forall x \neq 0$ and $= 0$ if $x = 0$

Suppose $A = B^T B$ for some B

$$x^T A x = x^T B^T B x$$
$$= (Bx)^T (Bx)$$
$$= y^T y$$
$$= \sum_{i=1}^{n} y_i^2$$

As long as $N(B) = \{0\}$, we have $x^T A x > 0$ for all non-null vector x.

If B is a square matrix, B must be non-singular

- $f(x) = a^T x = \sum_{i=1}^{n} a_i x_i$

$$\frac{\partial f}{\partial x_1} = a_1$$

$$\frac{\partial f}{\partial x_i} = a_i \quad \forall i$$

$$\frac{\partial f}{\partial \underline{x}} = [a_1 \;\; a_2 \;\; \cdots \;\; a_n]^T = a^T$$

$$\frac{\partial^2 f}{\partial \underset{\sim}{x} \, \partial \underset{\sim}{x}^T} = O_{n \times n}$$

- $f(x) = x^T A x$ where $A_{n\times n}$ matrix

$$= \sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij} x_i x_j$$

$$\frac{\partial f}{\partial x_1} = 2a_{11} x_1 + (a_{12} + a_{21}) x_2 + \cdots + (a_{1n} + a_{n1}) x_n$$

Similarly, $\frac{\partial f}{\partial x_i} = (a_{i1} + a_{1i}) x_1 + \cdots + 2 a_{ii} x_i$

$(a \quad a \quad) x$

$+ \cdots \quad (a_{n1} + a_{1n}) x_n$

$$\frac{\partial f}{\partial \underset{\sim}{x}} = \begin{bmatrix} 2a_{11}x_1 + (a_{12}+a_{21})x_2 + \cdots & (a_{1n}+a_{n1})x_n \\ \vdots & \\ (a_{1n}+a_{n1})x_1 + \cdots & + 2a_{nn}x_n \end{bmatrix}$$

$$= (A + A^T)\underset{\sim}{x}$$

$$\frac{\partial^2 f}{\partial x_2 \partial x_1} = (a_{12}+a_{21}), \quad \frac{\partial^2 f}{\partial x_1 \partial x_1} = 2a_{11}$$

$$\frac{\partial^2 f}{\partial x \partial x^T} = \begin{bmatrix} 2a_{11} & (a_{12}+a_{21}) & \cdots & (a_{n1}+a_{1n}) \\ \vdots & & & \\ (a_{n1}+a_{1n}) & \cdots & & 2a_{nn} \end{bmatrix}$$

$$= (A + A^T)$$

$f(X) = \text{trace}(AX) \qquad f: \mathbb{R}^{m \times n} \to \mathbb{R}$

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

$$\frac{\partial f}{\partial X} = \begin{bmatrix} \frac{\partial f}{\partial x_{11}} & \frac{\partial f}{\partial x_{12}} & \cdots & \frac{\partial f}{\partial x_{mn}} \end{bmatrix}^T$$

$$\frac{\partial^2 f}{\partial X \partial X^T} = \begin{bmatrix} \frac{\partial^2 f}{\partial x_{11} \partial x_{11}} & \cdots & \frac{\partial^2 f}{\partial x_{11} \partial x_{mn}} \\ \vdots & & \\ \frac{\partial^2 f}{\partial x_{mn} \partial x_{11}} & \cdots & \frac{\partial^2 f}{\partial x_{mn} \partial x_{mn}} \end{bmatrix}_{mn \times mn}$$

$f(x) \quad \text{trace}(AX) \qquad A \quad X$

$f(X) = \text{trace}(AX) \qquad {}_{n\times n} \ {}_{n\times n}$

$$\frac{\partial f}{\partial x_{11}} = \frac{\partial}{\partial x_{11}} tr(AX) = tr\left(A \frac{\partial X}{\partial x_{11}}\right)$$

$$\frac{\partial X}{\partial x_{11}} = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ \vdots & & & \\ 0 & \cdots & & 0 \end{bmatrix} = e_1 e_1^T$$

$$\frac{\partial X}{\partial x_{11}} = e_1^{(n)} e_1^{(n)T}, \quad \frac{\partial X}{\partial x_{ij}} = e_i^{(n)} e_j^{(n)T}$$

$$\frac{\partial f}{\partial x_{11}} = tr\left(A \frac{\partial X}{\partial x_{11}}\right) = tr\left(A e_1 e_1^T\right)$$

$$= tr\left(e_1^T A e_1\right)$$

$$= a_{11}$$

$$\frac{\partial f}{\partial x_{ij}} = tr\left(A \frac{\partial X}{\partial x_{ij}}\right) = tr\left(A e_i e_j^T\right)$$

$$= tr\left(e_j^T A e_i\right)$$

$$= a_{ij}$$

$$\frac{\partial f}{\partial X} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ \vdots & & & \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix} = A$$