Lecture 12

12.1 Hessian Matrix

We start by defining Hessian matrix, which is a natural extension of the concept of the second derivative in higher dimensions, allowing us to analyze the rate of change and curvature of a function in multiple directions simultaneously.

Definition 12.1.1 ▶ Hessian

Suppose $f : \mathcal{O}_n \to \mathbb{R}$ is C^2 at $a \in \mathcal{O}_n$. The **Hessian of** f **at** a is defined as the matrix,

$$H_f(a) = \left(\frac{\partial^2 f}{\partial x_i \partial x_j} (a) \right)_{n \times n}$$

In explicit notation, it has the following form,

$$H_f = \begin{pmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f}{\partial x_2 \partial x_1} & \frac{\partial^2 f}{\partial x_2^2} & \cdots & \frac{\partial^2 f}{\partial x_2 \partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1} & \frac{\partial^2 f}{\partial x_n \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_n^2} \end{pmatrix}$$

It is important to note that for any function f that is twice continuously differentiable ($f \in C^2$), its Hessian matrix H_f is symmetric, meaning that $H_f = {H_f}^t$.

Example 12.1.1

Let, $f : \mathbb{R}^2 \to \mathbb{R}$ be a function defined by $f(x, y) = \sin^2 x + x^2 y + y^2$. Then,

$$Df = \begin{pmatrix} \sin 2x + 2xy & x^2 + 2y \end{pmatrix} : \mathbb{R}^2 \to \mathbb{R} \quad \text{is linear.}$$

The gradient is given by,

$$\boldsymbol{\nabla} f = \left\langle \sin 2x + 2xy, x^2 + 2y \right\rangle \in \mathbb{R}^2$$

And the Hessian matrix H_f is,

$$H_f = \begin{pmatrix} f_{xx} & f_{xy} \\ f_{xy} & f_{yy} \end{pmatrix} = \begin{pmatrix} 2(\cos 2x + y) & 2x \\ 2x & 2 \end{pmatrix} \quad [\because f \in C^2]$$

Now let's introduce some notation. Given $A = (a_{ij})_{n \times n} \in M_n(\mathbb{R})$ and $x \in \mathbb{R}^n$ we denote $Q_A(x)$ by,

$$Q_A(x) = x^t A x = \langle Ax, x \rangle_{\mathbb{R}^n}$$

$$= \begin{pmatrix} x_1 & x_2 & \cdots & x_n \end{pmatrix} \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{31} & a_{32} & \cdots & a_{3n} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$$

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

Definition 12.1.2 ▶ Quadratic Form

A function $f : \mathbb{R}^n \to \mathbb{R}$ is called a **Quadratic Form** if it can be expressed as $f(x) = Q_A(x)$ for all x and for some symmetric $A \in M_n(\mathbb{R})$

It is important to note that a Quadratic Form represents a homogeneous polynomial of degree 2. For instance, in the case of a bivariate polynomial $p(x, y) = a_{11}x^2 + a_{22}y^2 + a_{12}xy$, the matrix

$$A = \begin{pmatrix} a_{11} & \frac{1}{2}a_{12} \\ \frac{1}{2}a_{12} & a_{22} \end{pmatrix}$$

corresponds to the quadratic form $p(x, y) = Q_A(x, y)$, capturing the essential quadratic behavior of p.

12.2 Positive Definite, Negative Definite, Semi Definite Matrices

Definition 12.2.1 ▶ Positive Definite, Negative Definite, Semi Definite

- A symmetric matrix $A \in M_n(\mathbb{R})$ is called **Positive Definite** if

$$\langle Ax, x \rangle > 0 \ \forall \ x \in \mathbb{R}^n \setminus \{0\}$$

- A symmetric matrix $A \in M_n(\mathbb{R})$ is called **Negative Definite** if

$$\langle Ax, x \rangle < 0 \ \forall \ x \in \mathbb{R}^n \setminus \{0\}$$

- A symmetric matrix $A \in M_n(\mathbb{R})$ is called **Semi Definite** if

$$\langle Ax, x \rangle \geq 0 \ \forall \ x \in \mathbb{R}^n \setminus \{0\}$$

Example 12.2.1

1. I_n is positive definite because for any vector $x \in \mathbb{R}^n \setminus \{0\}$ the inner product

$$\langle I_n x, x \rangle = \|x\|^2 > 0$$

   is strictly positive.

2. For any matrix $A \in M_n(\mathbb{R})$, if there exists a matrix $B \in M_n(\mathbb{R})$ such that $A = B^t B$, we can examine the inner product $\langle Ax, x \rangle$ for any $x \in \mathbb{R}^n \setminus \{0\}$.

   Let's compute this inner product

$$\begin{aligned} \langle Ax, x \rangle &= \langle B^t Bx, x \rangle \\ &= x^t B^t Bx \\ &= (Bx)^t (Bx) \\ &= \|Bx\|^2 \quad \text{for all } x \in \mathbb{R}^n \setminus \{0\}. \end{aligned}$$

Therefore, we conclude that $\langle Ax, x\rangle \geq 0$ for all $x \in \mathbb{R}^n \setminus \{0\}$.

Moreover, if $\langle Ax, x\rangle = 0$, then $Bx = 0$, which implies that x is in the kernel of B. Conversely, if A is positive definite, there is no non-zero vector x such that $\langle Ax, x\rangle = 0$. This implies that the columns of B are linearly independent.

In summary, when A can be written as $A = B^tB$ for some matrix B, we can conclude that A is positive semi-definite (and the converse also holds).

3. Consider the matrix

$$A = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$$

We can compute the quadratic form associated with A as $Q_A = x_1^2 - x_2^2$.

By examining this expression, we observe that it is the difference between the squares of two variables. This indicates that the sign of Q_A can change depending on the values of x_1 and x_2. Consequently, the matrix A is considered **indefinite**.

4. Consider the matrix

$$\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$$

To determine the definiteness of A, we can compute the quadratic form associated with A as $Q_A(x) = x_1^2 \geq 0$ for all $x = \begin{pmatrix} x_1 & x_2 \end{pmatrix}$. This indicates that the matrix A is **positive semi-definite**.

Consider a positive definite matrix A. For any non-zero vector h, we have,

$$\langle Ah, h\rangle = \|Ah\|\|h\| \cos\theta > 0$$

where θ is the angle between vectors Ah and h. Since the cosine of any angle θ in the interval $[0, \pi/2)$ is positive, we can conclude that,

$$\cos\theta > 0 \implies \boxed{0 \leq \theta < \frac{\pi}{2}}$$

Thus, for any positive definite matrix A, the angle θ between Ah and h satisfies $0 \leq \theta < \frac{\pi}{2}$.

It is worth noting that classifying positive definite matrices becomes more challenging for higher dimensions ($n > 2$). However, for 2×2 matrices, we can easily determine it from the next theorem.

Theorem 12.2.1

Let $A = \begin{pmatrix} a & b \\ b & c \end{pmatrix} \in M_2(\mathbb{R})$ be symmetric. Then,

(i) A is **Positive Definite** $\iff a > 0$ and $ac - b^2 > 0$

(ii) A is **Negative Definite** $\iff a < 0$ and $ac - b^2 > 0$

(iii) A is **Indefinite** $\iff ac - b^2 < 0$

Proof. We have $\langle Ah, h\rangle = h^tAh$ for any vector h. Now, consider a non-zero vector $\mathbf{x} = (x_1, x_2) \in \mathbb{R}^2\setminus\{(0,0)\}$ with $x_2 \neq 0$. Without loss of generality, we can scale $\mathbf{x}$ as $\mathbf{x} = (x, 1)$ for some $x \in \mathbb{R}$. Then, we have,

$$\langle A\mathbf{x}, \mathbf{x}\rangle = ax^2 + 2bx + c > 0 \;\forall\, x \in \mathbb{R}$$

If $x_2 = 0$, we can choose $\mathbf{x} = \begin{pmatrix} 1 & 0 \end{pmatrix}$ (after scaling). Then, $\langle A\mathbf{x}, \mathbf{x}\rangle = a$. Therefore, we can summarize the conditions as follows,

$$A \text{ is Positive Definite}$$

$$
\begin{aligned}
&\iff a > 0 \text{ and } ax^2 + bx + c > 0 \ \forall\ x \in \mathbb{R} \\
&\iff a > 0 \text{ and } (2b)^2 - 4ac < 0 \\
&\iff a > 0 \text{ and } ac - b^2 > 0
\end{aligned}
$$

Similarly, we can derive the conditions for negative definite and indefinite matrices as,

$$
\begin{aligned}
&A \text{ is Negative Definite} \\
\iff &a < 0 \text{ and } ax^2 + bx + c < 0 \ \forall\ x \in \mathbb{R} \\
\iff &a < 0 \text{ and } (2b)^2 - 4ac < 0 \\
\iff &a < 0 \text{ and } ac - b^2 > 0
\end{aligned}
$$

And finally for indefinite ones,

$$
\begin{aligned}
&A \text{ is Indefinite} \\
\iff &ax^2 + bx + c < 0 \text{ for some } x \in \mathbb{R} \\
&\quad \text{and } ax^2 + bx + c > 0 \text{ for some } x \in \mathbb{R} \\
\iff &(2b)^2 - 4ac > 0 \\
\iff &ac - b^2 < 0
\end{aligned}
$$

□

Lemma 12.2.1

Let, $a \in \mathcal{O}_n,\ A(x) = \begin{pmatrix} a_1(x) & a_2(x) \\ a_2(x) & a_3(x) \end{pmatrix}$. Suppose, A is continuous at a (i.e., a_i's are continuous at a). Then, A is Positive Definite at a would imply that A is Positive Definite in a neighborhood of a.

Proof. $A(a)$ is Positive Definite, i.e., $a_1(a) > 0$ and $a_1(a)a_3(a) - {a_2}^2(a) > 0$. As $a_1(x)$ and $a_1(x)a_3(x) - a_2^2(x)$ are polynomial of continuous functions, we can find an $\epsilon > 0$ such that both are positive in $B_\epsilon(a)$. □