LECTURE 4

SECOND ORDER OPTIMALITY CONDITIONS

Consider

$$f(x) = x_1^2 + 6x_1x_2 + x_2^2 + 4x_1 + 4x_2$$

$$= \frac{1}{2} x^T \underbrace{\begin{bmatrix} 2 & 6 \\ 6 & 2 \end{bmatrix}}_{A} x + \underbrace{\begin{bmatrix} 4 & 4 \end{bmatrix}}_{b^T} x$$

Note

$$\nabla f(x) = \begin{bmatrix} 2x_1 + 6x_2 + 4 \\ 6x_1 + 2x_2 + 4 \end{bmatrix}$$

$$= \underbrace{\begin{bmatrix} 2 & 6 \\ 6 & 2 \end{bmatrix}}_{A} x + \underbrace{\begin{bmatrix} 4 \\ 4 \end{bmatrix}}_{b}$$

and

$$\nabla^2 f(x) = \underbrace{\begin{bmatrix} 2 & 6 \\ 6 & 2 \end{bmatrix}}_{A}.$$

There is only one point x_* s.t.

$$\nabla f(x_*) = 0$$

$$\Longrightarrow \begin{bmatrix} 2 & 6 \\ 6 & 2 \end{bmatrix} \begin{bmatrix} (x_*)_1 \\ (x_*)_2 \end{bmatrix} = \begin{bmatrix} -4 \\ -4 \end{bmatrix} \Longrightarrow x_* = \begin{bmatrix} -1/2 \\ -1/2 \end{bmatrix}$$

* x_* may or may not be a local minimizer.
* All other points are not local minimizers.

Whether $x_* = \begin{bmatrix} -1/2 \\ -1/2 \end{bmatrix}$ is a local minimizer or not depends on the second derivatives, Specifically the Hessian matrix $\nabla^2 f(x) = \begin{bmatrix} 2 & 6 \\ 6 & 2 \end{bmatrix}$.

<u>DEFN</u> (Stationary Point)
Let $f: \mathbb{R}^n \to \mathbb{R}$ be smooth. A point $x_* \in \mathbb{R}^n$ is called a <u>stationary point</u> if $\nabla f(x_*) = 0$.

<u>THM</u> (Taylor's Thm with Second Order Remainder)
Let $f: \mathbb{R}^n \to \mathbb{R}$ be a twice continuously differentiable function. Given $x, p \in \mathbb{R}^n$. There exists a $t \in (0,1)$ such that

$$f(x+p) = f(x) + \nabla f(x)^T p + \frac{1}{2} p^T \nabla^2 f(x+tp) p$$

Suppose $x_* \in \mathbb{R}^n$ is a stationary point. For all $p \in \mathbb{R}^n$ it follows from Taylor's thm that

$$f(x_* + p) = f(x_*) + \underbrace{\nabla f(x_*)^T}_{0} p + \frac{1}{2} p^T \nabla^2 f(x_* + tp) p$$

$$\Longrightarrow \quad f(x_* + p) - f(x_*) = \frac{1}{2} p^T \nabla^2 f(x_* + tp) p$$

for some $t \in (0, 1)$.

Whether

$$f(x_* + p) \geq f(x_*) \text{ for all small } p \in \mathbb{R}^n$$

depends on the condition

$$p^T \nabla^2 f(x_*) p \geq 0 \text{ for all } p \in \mathbb{R}^n$$

<u>SECOND ORDER SUFFICIENT CONDITIONS</u>

$$p^T \nabla^2 f(x_*) p > 0 \text{ for all } p \in \mathbb{R}^n$$

$$\Longrightarrow$$

$$p^T \nabla^2 f(x_*) p > 0 \text{ for all small } p \in \mathbb{R}^n$$

$$\Longrightarrow \quad \left(\text{By CONTINUITY OF SECOND DERIVATIVES}\right)$$

$$p^T \nabla^2 f(x_* + tp) p > 0 \text{ for all small } p \in \mathbb{R}^n \text{ and all } t \in (0,1)$$

$$\Longrightarrow \quad \left(\text{By Taylor's thm with second order remainder}\right)$$

$$f(x_* + p) > f(x_*) \text{ for all small non-zero } p \in \mathbb{R}^n$$

$$\Longrightarrow$$

x_* is a (strict) local minimizer

DEFN (Positive Definiteness)

A symmetric matrix $A \in \mathbb{R}^{n \times n}$ is called

(i) positive definite if $p^T A p > 0$ for all nonzero $p \in \mathbb{R}^n$;

(ii) positive semidefinite if $p^T A p \geq 0$ for all $p \in \mathbb{R}^n$.

NOTATION

$A \succ 0$ — A is positive definite

$A \succeq 0$ — A is positive semidefinite

e.g.

$A_1 = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}$ is positive definite

$p^T A_1 p = 2p_1^2 + 2p_2^2 > 0$ for all $p = \begin{bmatrix} p_1 \\ p_2 \end{bmatrix} \neq 0$

$A_2 = \begin{bmatrix} 2 & 0 \\ 0 & 0 \end{bmatrix}$ is positive semidefinite

$p^T A_2 p = 2p_1^2 \geq 0$ for all $p = \begin{bmatrix} p_1 \\ p_2 \end{bmatrix}$

($p^T A_2 p = 0$ if $p = \begin{bmatrix} 0 \\ c \end{bmatrix}$)

THM (Second Order Sufficient Conditions)

Suppose $x_* \in \mathbb{R}^n$ and $f : \mathbb{R}^n \to \mathbb{R}$ is twice continuously differentiable.

$$\nabla f(x_*) = 0 \text{ and } \nabla^2 f(x_*) \succ 0$$

$$\Longrightarrow$$

x_* is a (strict) local minimizer

EXAMPLE

$$f(x) = x_1^2 + x_2^2 + x_1 + x_2$$

$$= \frac{1}{2} x^T \underbrace{\begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}}_{A} x + \underbrace{\begin{bmatrix} 1 & 1 \end{bmatrix}}_{b^T} x$$

Gradient

$$\nabla f(x) = \begin{bmatrix} 2x_1 + 1 \\ 2x_2 + 1 \end{bmatrix}$$

$$= \underbrace{\begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}}_{A} x + \underbrace{\begin{bmatrix} 1 \\ 1 \end{bmatrix}}_{b}$$

Hessian

$$\nabla^2 f(x) = \underbrace{\begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}}_{A}$$

① $x_* = \begin{bmatrix} -1/2 \\ -1/2 \end{bmatrix}$ is a stationary point such that $\nabla f(x_*) = 0$.

② $\nabla^2 f(x_*) = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix} \succ 0$

By the second order sufficient conditions $x_* = \begin{bmatrix} -1/2 \\ -1/2 \end{bmatrix}$ is a local minimizer.

SECOND ORDER NECESSARY CONDITIONS

$$p^T \nabla^2 f(x_*) p \not\geq 0 \quad \text{for some } p \in \mathbb{R}^n$$

$$\Longrightarrow$$

$$p^T \nabla^2 f(x_*) p < 0 \quad \text{for some small } p \in \mathbb{R}^n$$

$$\Longrightarrow \left(\begin{array}{l}\text{By CONTINUITY}\\ \text{OF SECOND DERIVATIVES}\end{array}\right)$$

$$p^T \nabla^2 f(x_* + tp) p < 0 \quad \text{for some small } p \in \mathbb{R}^n \text{ and all } t \in (0,1)$$

$$\Longrightarrow \left(\begin{array}{l}\text{By Taylor's Thm}\\ \text{with Second Order Remainder}\end{array}\right)$$

$$f(x_* + p) < f(x_*) \quad \text{for some } p \in \mathbb{R}^n$$

$$\left(\begin{array}{l}p \text{ can be chosen}\\ \text{arbitrarily small}\end{array}\right)$$

$$\Longrightarrow$$

x_* is not a local minimizer

<u>THM</u> (Second Order Necessary Conditions)

Suppose $f : \mathbb{R}^n \to \mathbb{R}$ is twice continuously differentiable.

$$x_* \in \mathbb{R}^n \text{ is a local minimizer}$$

$$\Longrightarrow$$

$$\nabla f(x_*) = 0 \quad \text{and} \quad \nabla^2 f(x_*) \succeq 0$$

EXAMPLE

Consider again

$$f(x) = x_1^2 + 6x_1x_2 + x_2^2 + 4x_1 + 4x_2$$

$$= \frac{1}{2} x^T \underbrace{\begin{bmatrix} 2 & 6 \\ 6 & 2 \end{bmatrix}}_{A} x + \underbrace{[4 \quad 4]}_{b^T} x$$

Gradient

$$\nabla f(x) = \begin{bmatrix} 2x_1 + 6x_2 + 4 \\ 6x_1 + 2x_2 + 4 \end{bmatrix}$$

$$= \underbrace{\begin{bmatrix} 2 & 6 \\ 6 & 2 \end{bmatrix}}_{A} x + \underbrace{\begin{bmatrix} 4 \\ 4 \end{bmatrix}}_{b}$$

Hessian

$$\nabla^2 f(x) = \begin{bmatrix} 2 & 6 \\ 6 & 2 \end{bmatrix}$$

① $x_* = \begin{bmatrix} -1/2 \\ -1/2 \end{bmatrix}$ is a stationary point such that $\nabla f(x_*) = 0$.

② $\nabla^2 f(x_*) = \begin{bmatrix} 2 & 6 \\ 6 & 2 \end{bmatrix} \not\succeq 0$

since $[1 \quad -1] \begin{bmatrix} 2 & 6 \\ 6 & 2 \end{bmatrix} \begin{bmatrix} 1 \\ -1 \end{bmatrix} = -8 > 0$

By second order necessary conditions f has no local minimizer. (In particular $x_* = \begin{bmatrix} -1/2 \\ -1/2 \end{bmatrix}$ is not a local minimizer.)