ESE 415

Lecture 4

9/9/2024

Second-Order conditions for Optimality

Useful results:

<u>Theorem (Extended Law of the Mean)</u> $D \subseteq \mathbb{R}^n$ open set, $f \in C^2(D, \mathbb{R})$ and $x^* \in D$. Then for all $x \in \mathbb{R}^n$

$$f(x) = f(x^*) + (\nabla f)(x^*)^T (x - x^*) + \frac{1}{2}(x - x^*)^T (Hf)(z)(x - x^*)$$

for some point $z \in \underbrace{[x, x^*] := \{ tx + (1-t)x^* : t \in [0,1]\}}_{\text{straight line in } \mathbb{R}^n \text{ connecting } x \text{ and } x^*.}$

<u>Other reformulations</u> of the above theorem: (Taylor)

$$f(x) = f(x^*) + (\nabla f)(x^*)^T (x - x^*) + o(\|x - x^*\|)$$

or

$$f(x) = f(x^*) + (\nabla f)(x^*)^T (x - x^*) + \frac{1}{2}(x - x^*)^T (Hf)(x^*)(x - x^*) + o(\|x - x^*\|^2$$

<u>Recall o-notation:</u>

Let $g : \mathbb{R} \to \mathbb{R}$, then

"$g(x) = o(x)$" $\overset{\text{def.}}{\Longleftrightarrow}$ $\lim\limits_{x \to 0} \left| \frac{g(x)}{x} \right| = 0.$ (asymptotically, g goes to zero faster than x)

(as $x \longrightarrow 0$)

different from Computer Science usage ($N \to \infty$)

Theorem (2nd-order NECESSARY condition)

Let x^* be a local minimizer of $f \in C^2(\mathbb{R}^n, \mathbb{R})$, then

$$(\nabla f)(x^*) = 0 \quad \underline{\text{and}} \quad (Hf)(x^*) \succcurlyeq 0.$$

<u>Proof</u>: $\nearrow$ HW2, Problem 5

Theorem (2nd-order SUFFICIENT conditions)

Let $f \in C^2(\mathbb{R}^n, \mathbb{R})$ and $(\nabla f)(x^*) = 0$. The point $x^* \in \mathbb{R}^n$ is a

1. global minimizer of f if $\quad \forall_{x \in \mathbb{R}^n} \ (Hf)(x) \succcurlyeq 0$.
2. strict global min. of f if $\quad \forall x \in \mathbb{R}^n \ (Hf)(x) \succ 0$.
3. strict local min. of f if $\quad (Hf)(x^*) \succ 0$.

<u>Proof</u>: 1. Since $(Hf)(x) \succcurlyeq 0$ for all $x \in \mathbb{R}^n$, we will also have $(Hf)(z) \succcurlyeq 0$ in the Extended Law of the Mean:

$$f(x) = f(x^*) + \cancel{\underbrace{(\nabla f)(x^*)}_{=0}{}^T (x-x^*)} + \underbrace{\frac{1}{2}(x-x^*)^T \underbrace{(Hf)(z)}_{\succcurlyeq 0}(x-x^*)}_{\geqslant 0}$$

$$\Rightarrow \forall_{x \in \mathbb{R}^n} \ f(x) \geq f(x^*).$$

2. Analogously.

3. Taylor expansion with $(\nabla f)(x^*) = 0$:

$$f(x) = f(x^*) + \frac{1}{2}(x-x^*)^T (Hf)(x^*)(x-x^*) + o(\|x-x^*\|^2).$$

From Problem 6 in Homework 2: $x^T A x \geq \lambda_{min}\|x\|^2$

symmetric and positive semidefinite

$$\Rightarrow f(x) = f(x^*) + \underbrace{\frac{\lambda_{min}}{2}\|x-x^*\|^2}_{\geq 0 \ \checkmark} + \underbrace{o(\|x-x^*\|^2)}$$

could be negative and ruin our plan of showing $f(x) \geq f(x^*)$

BUT:

$$\frac{o(\|x-x^*\|^2)}{\|x-x^*\|^2} \longrightarrow 0 \quad \text{as} \quad x \to x^* \quad (\|x-x^*\| \to 0)$$

meaning: $\exists_{\delta>0} \ \forall x \in B_\delta(x^*) \quad |o(\|x-x^*\|^2)| \leq \|x-x^*\|^2 \times \underbrace{\frac{\lambda_{min}}{4}}$

constant we introduced in order to bound negative impact of

"worst-case analysis":

$$o(\|x-x^*\|^2) \geq -\frac{\lambda_{min}}{4}\|x-x^*\|^2$$

$$\Rightarrow f(x) \geq f(x^*) + \frac{\lambda_{min}}{2}\|x-x^*\|^2 - \frac{\lambda_{min}}{4}\|x-x^*\|^2$$

$= +\frac{\lambda_{min}}{4}\|x-x^*\|^2 \geq 0$ ("by design")

$$\Rightarrow f(x) \geq f(x^*) \quad \text{for all } x \in B_\delta(x^*).$$

$\blacksquare$

Recap of tests for positive (semi) definiteness:

Theorem: Let $A \in \mathbb{R}^{n\times n}$ be a symmetric matrix.

- $A \succ 0 \iff x^T A x > 0 \ \forall x \neq 0 \iff \lambda_i > 0 \ \forall i = 1, \dots, n$
- $A \succeq 0 \iff x^T A x \geq 0 \ \forall x \in \mathbb{R}^n \iff \lambda_i \geq 0 \ \forall i = 1, \dots, n$
- A is indefinite $\iff \exists i,j \quad \lambda_i > 0$ and $\lambda_j < 0$

Next: Sylvester criterion based on "principal minors":

Recall: $\Delta_k \dots$ kth principal minor of a symmetric $A \in \mathbb{R}^{n\times n}$

$$A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & & & \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{pmatrix}, \quad \Delta_1 = a_{11}, \quad \Delta_2 = \det\begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}, \dots, \Delta_n = \det A$$

Theorem (Sylvester criterion): $A \in \mathbb{R}^{n\times n}$ symmetric, $\Delta_k \dots$ kth principal minor of A. Then

- $A \succ 0 \iff \Delta_k > 0 \quad \forall k = 1, \dots, n$
- $A \prec 0 \iff (-1)^k \Delta_k > 0 \quad \forall k = 1, \dots, n$

Remark:
- $\Delta_k \geq 0 \ \forall k = 1, \dots, n$ does not imply $A \succeq 0$.
- However, $\Delta_1 > 0, \Delta_2 > 0, \dots, \Delta_{n-1} > 0, \Delta_n = 0$ actually does imply $A \succeq 0$.