APM462
Lecture Notes

Yuchen Wang

September 19, 2019

Contents

1 Lecture 1 - September 6th

functions $\mathbb{R} \to \mathbb{R}$

Mean Value Theorem in 1 Dimension $g \in C^1$ on $\mathbb{R}$

$$\frac{g(x+h) - g(x)}{h} = g'(x + \theta h)$$

where $\theta \in (0, 1)$
Or equivalently,

$$g(x+h) = g(x) + hg'(x + \theta h)$$

1st Order Taylor Approximation $g \in C^1$ on $\mathbb{R}$

$$g(x+h) = g(x) + hg'(x) + o(h)$$

where $o(h)$ is "little o" of h, the error term.
Say a function $f(h) = o(h)$, this means $\lim_{h \to 0} \frac{f(h)}{h} = 0$
For example, for $f(h) = h^2$, we can say $f(h) = o(h)$,
since $\lim_{h \to 0} \frac{f(h)}{h} = \lim_{h \to 0} \frac{h^2}{h} = \lim_{h \to 0} h = 0$
proof: (Use MVT):
WTS : $g(x+h) - g(x) - hg'(x) = o(h)$

$$\begin{aligned} \lim_{h \to 0} \frac{[g(x+h) - g(x)] - hg'(x)}{h} &= \lim_{h \to 0} \frac{[hg'(x + \theta h)] - hg'(x)}{h} \\ &= \lim_{h \to 0} g'(x + \theta h) - g'(x) \\ &= \lim_{h \to 0} g'(x) - g'(x) \\ &= 0 \end{aligned}$$

■

2nd Order Mean Value Theorem $g \in C^2$ on $\mathbb{R}$

$$g(x+h) = g(x) + hg'(x) + \frac{h^2}{2} g'(x + \theta h)$$

for some $\theta \in (0, 1)$

proof:
WTS: $g(x+h)-g(x)-hg'(x)-\frac{h^2}{2}g''(x)=o(h^2)$

$$\begin{aligned}\lim_{h\to 0}\frac{g(x+h)-g(x)-hg'(x)-\frac{h^2}{2}g''(x)}{h^2} &= \lim_{h\to 0}\frac{[\frac{h^2}{2}g'(x+\theta h)]-\frac{h^2}{2}g''(x)}{h^2}\\ &= \lim_{h\to 0}\frac{1}{2}(g''(x+\theta h)-g''(x))\\ &= \lim_{h\to 0}\frac{1}{2}(g''(x)-g''(x))\\ &= 0\end{aligned}$$

■

2 Lecture 2 - September 9th

multivariate functions: $\mathbb{R}^n \to \mathbb{R}$

Recall: Definition of gradient Gradient of $f:\mathbb{R}^n\to\mathbb{R}$ at $x\in\mathbb{R}^n$ (denoted $\nabla f(x)$) if exists is a vector characterized by the property:

$$\lim_{\mathbf{v}\to\mathbf{0}}\frac{f(\mathbf{x}+\mathbf{v})-f(\mathbf{x})-\nabla f(\mathbf{x})\cdot\mathbf{v}}{||\mathbf{v}||}=0$$

In Cartesian coordinates, $\nabla f(\mathbf{x})=(\frac{\partial f}{\partial \mathbf{x}_1}(\mathbf{x}),\ldots,\frac{\partial f}{\partial \mathbf{x}_n}(\mathbf{x}))$

Mean Value Theorem in n dimension $f\in C^1$ on $\mathbb{R}^n$, then for any $\mathbf{x},\mathbf{v}\in\mathbb{R}^n$,

$$f(\mathbf{x}+\mathbf{v})=f(\mathbf{x})+\nabla f(\mathbf{x}+\theta\mathbf{v})\cdot\mathbf{v}$$

for some $\theta\in(0,1)$

proof: Reduce to 1-dimension case
$g(t):=f(\mathbf{x}+t\mathbf{v}), t\in\mathbb{R}$

$$\begin{aligned}
g'(t) &= \frac{d}{dt} f(\mathbf{x} + t\mathbf{v}) \\
&= \sum_{i=1}^{n} \frac{\partial f}{\partial \mathbf{x}_i}(\mathbf{x} + t\mathbf{v}) \cdot \frac{d(\mathbf{x} + t\mathbf{v})_i}{dt} && \text{(by Chain Rule)} \\
&= \sum \frac{\partial f}{\partial \mathbf{x}_i}(\mathbf{x} + t\mathbf{v}) \cdot \frac{d(\mathbf{x}_i + t\mathbf{v}_i)}{dt} \\
&= \sum \frac{\partial f}{\partial \mathbf{x}_i}(\mathbf{x} + t\mathbf{v}) \cdot \mathbf{v}_i \\
&= \nabla f(\mathbf{x} + t\mathbf{v}) \cdot \mathbf{v} && (*)
\end{aligned}$$

$g \in C^1$ on $\mathbb{R}$
Using MVT in $\mathbb{R}$:

$$\begin{aligned}
f(\mathbf{x} + \mathbf{v}) &= g(1) \\
&= g(0 + 1) \\
&= g(0) + 1g'(0 + \theta 1) && (\theta \in (0, 1)) \\
&= g(0) + g'(\theta) \\
&= f(\mathbf{x}) + \nabla f(\mathbf{x} + \theta\mathbf{v}) \cdot \mathbf{v} && \text{(by (*))}
\end{aligned}$$

$\blacksquare$

1st Order Taylor Approximation in $\mathbb{R}^n$ $\quad f \in C^1$ on $\mathbb{R}^n$

$$f(\mathbf{x} + \mathbf{v}) = f(\mathbf{x}) + \nabla f(\mathbf{x}) \cdot \mathbf{v} + o(||\mathbf{v}||)$$

<u>*proof:*</u>

$$\begin{aligned}
\lim_{||\mathbf{v}|| \to 0} \frac{[f(\mathbf{x} + \mathbf{v}) - f(\mathbf{x})] - \nabla f(\mathbf{x}) \cdot \mathbf{v}}{||\mathbf{v}||} &= \lim_{||\mathbf{v}|| \to 0} \frac{[\nabla f(\mathbf{x} + \theta\mathbf{v}) \cdot \mathbf{v}] - \nabla f(\mathbf{x}) \cdot \mathbf{v}}{||\mathbf{v}||} \\
&= \lim_{||\mathbf{v}|| \to 0} [\nabla f(\mathbf{x} + \theta\mathbf{v}) - \nabla f(\mathbf{x})] \cdot \frac{\mathbf{v}}{||\mathbf{v}||} \\
&= 0 \quad (\tfrac{\mathbf{v}}{||\mathbf{v}||} \text{ is a unit vector, remains 1})
\end{aligned}$$

$\blacksquare$

2nd Order Mean Value Theorem in $\mathbb{R}^n$ $\quad f \in C^2$ on $\mathbb{R}^n$

$$f(\mathbf{x} - \mathbf{v}) = f(\mathbf{x}) + \nabla f(\mathbf{x}) \cdot \mathbf{v} + \frac{1}{2}\mathbf{v}^T \nabla^2 f(\mathbf{x} + \theta\mathbf{v}) \cdot \mathbf{v}$$

Remarks In this course, ∇^2 means Hessian, not Laplacian.

$$\nabla^2 f(\mathbf{x}) = \left(\frac{\partial^2 f}{\partial \mathbf{x}_i \partial \mathbf{x}_j}\right)_{1\le i,j\le n}(\mathbf{x}) = \begin{pmatrix} \frac{\partial f}{\partial_1^2} & \frac{\partial f}{\partial_1\partial_2} & \cdots \\ \frac{\partial f}{\partial_2\partial_1} & \cdots & \\ \vdots & & \end{pmatrix}$$

The Hessian matrix is symmetric. This is sometimes called Clairaut's Theorem.
note: $\mathbf{v}^T\nabla^2 f(\mathbf{x})\mathbf{v} = \sum_{1\le i,j\le n}\frac{\partial^2 f}{\partial \mathbf{x}_i\partial \mathbf{x}_j}f(\mathbf{x})\mathbf{v}_i\mathbf{v}_j$

2nd Order Taylor Approximation in $\mathbb{R}^n$ $f \in C^2$ on $\mathbb{R}^n$

$$f(\mathbf{x}+\mathbf{v}) = f(\mathbf{x}) + \nabla f(\mathbf{x})\cdot\mathbf{v} + \frac{1}{2}\mathbf{v}^T\nabla^2 f(\mathbf{x})\mathbf{v} + o(||\mathbf{v}||^2)$$

proof:

$$\begin{aligned}\lim_{||\mathbf{v}||\to 0}\frac{[f(\mathbf{x}+\mathbf{v})-f(\mathbf{x})]-\nabla f(\mathbf{x})\cdot\mathbf{v}-\frac{1}{2}\mathbf{v}^T\nabla^2 f(\mathbf{x})\mathbf{v}}{||\mathbf{v}||^2} &= \lim_{||\mathbf{v}||\to 0}\frac{[\frac{1}{2}\mathbf{v}^T\nabla^2 f(\mathbf{x}+\theta\mathbf{v})\cdot\mathbf{v}]-\frac{1}{2}\mathbf{v}^T\nabla^2 f(\mathbf{x})\cdot\mathbf{v}}{||\mathbf{v}||^2} \\ &\quad \text{(By 2nd MVT)} \\ &= \lim_{||\mathbf{v}||\to 0}\frac{1}{2}\left(\frac{\mathbf{v}}{||\mathbf{v}||}\right)^T[\nabla^2 f(\mathbf{x}+\theta\mathbf{v})-\nabla^2 f(\mathbf{x})]\left(\frac{\mathbf{v}}{||\mathbf{v}||}\right) \\ &= 0\end{aligned}$$

■

Geometric Meaning of Gradient $f:\mathbb{R}^n\to\mathbb{R}$
Rate of change of f at $\mathbf{x}$ in direction $\mathbf{v}$ $(||\mathbf{v}||=1) = \frac{d}{dt}|_{t=0}f(\mathbf{x}+t\mathbf{v})$

$$\begin{aligned}\frac{d}{dt}|_{t=0}f(\mathbf{x}+t\mathbf{v}) &= \nabla f(\mathbf{x}+t\mathbf{v})\cdot\mathbf{v}|_{t=0} \\ &= \nabla f(\mathbf{x})\cdot\mathbf{v} \\ &= |\nabla f(\mathbf{x})||\mathbf{v}|\cos\theta \\ &= |\nabla f(\mathbf{x})|\cos\theta \qquad (||\mathbf{v}||=1)\end{aligned}$$

maximized at $\theta = 0$
So $\nabla f(\mathbf{x})$ points in the direction of steepest ascent.

Implicit Function Theorem $f:\mathbb{R}^{n+1}\to\mathbb{R}\in C^1$
Fix $(\mathbf{a},b)\in\mathbb{R}^n\times\mathbb{R}$ s.t. $f(\mathbf{a},b)=0$.
If $\nabla f(\mathbf{a},b)\neq 0$, then $\{(\mathbf{x},y)\in(\mathbb{R}^n\times\mathbb{R})|f(\mathbf{x},\mathbf{y})=0\}$ is locally (near $(\mathbf{a},b)$) the graph of a function.

Level Sets of f c-level set of $f := \{\mathbf{x} \in \mathbb{R}^n | f(\mathbf{x}) = c\}$

Fact gradient $\nabla f(\mathbf{x}_0) \perp$ level curve (through $\mathbf{x}_0$)

Definition of Convex Set $\Omega \subseteq \mathbb{R}^n$ is a convex set if $\mathbf{x}_1, \mathbf{x}_2 \in \Omega \Rightarrow s\mathbf{x}_1 + (1-s)\mathbf{x}_2 \in \Omega$ where $s \in [0,1]$

Definition of Convex Function A function f : convex $\Omega \subseteq \mathbb{R}^n$ is convex if

$$f(s\mathbf{x}_1 + (1-s)\mathbf{x}_2) \leq sf(\mathbf{x}_1) + (1-s)f(\mathbf{x}_2)$$

for all $\mathbf{x}_1, \mathbf{x}_2 \in \Omega$ and all $s \in [0,1]$

Remarks Second line above (or equal to) the graph

Definition of Concave Function A function f is concave if $-f$ is convex.

3 Lecture 3 - September 13rd

Basic Properties of convex functions Let $\Omega \subseteq \mathbb{R}^n$ be a convex set.

1. f_1, f_2 are convex functions on $\Omega \Rightarrow f_1 + f_2$ is a convex function on Ω.
2. f is a convex function, $a \geq 0 \Rightarrow af$ is a convex function.
3. f is a convex on $\Omega \Rightarrow$ The sublevel sets of f, $SL_c := \{\mathbf{x} \in \mathbb{R}^n | f(\mathbf{x}) \leq c\}$ is convex.

proof of (3):
Let $x_1, x_2 \in SL_C$, so that $f(x_1) \leq c$ and $f(x_2) \leq c$.
WTS: $sx_1 + (1-s)x_2 \in SL_c$ for any $s \in [0,1]$

$$\begin{aligned} f(sx_1 + (1-s)x_2) &\leq sf(x_1) + (1-s)f(x_2) && (f \text{ is convex}) \\ &\leq sc + (1-s)c \\ &= c \end{aligned}$$

$$\Rightarrow sx_1 + (1-s)x_2 \in SL_c$$

■

Example of a convex function Let $f : \mathbb{R} \to \mathbb{R}$, $f(x) = |x|$
Let $x_1, x_2 \in \mathbb{R}, s \in [0,1]$
Then

$$\begin{aligned} f(sx_1 + (1-s)x_2) &= |sx_1 + (1-s)x_2| \\ &\leq |sx_1| + |(1-s)x_2| \qquad \text{(by Triangle Inequality)} \\ &= s|x_1| + (1-s)|x_2| \\ &= sf(x_1) + (1-s)f(x_2) \end{aligned}$$

Then f is a convex function.

Theorem - Characterization of C^1 convex functions Let f : convex subset of $\mathbb{R}^n$ $\Omega \to \mathbb{R}$ be a C^1 function.
Then,
f is convex $\iff$ $f(y) \geq f(x) + \nabla f(x) \cdot (y - x)$ for all $x, y \in \Omega$

Remarks Tangent line below the graph.

<u>*proof:*</u>
($\Rightarrow$)
f is convex, then by definition,

$$\begin{aligned} f(s\mathbf{x}_1 + (1-s)\mathbf{x}_2) &\leq sf(\mathbf{x}_1) + (1-s)f(\mathbf{x}_2) \\ f(s\mathbf{x}_1 + (1-s)\mathbf{x}_2) - f(\mathbf{x}_2) &\leq s(f(\mathbf{x}_1) - f(\mathbf{x}_2)) \\ \frac{f(s\mathbf{x}_1 + (1-s)\mathbf{x}_2) - f(\mathbf{x}_2)}{s} &\leq f(\mathbf{x}_1) - f(\mathbf{x}_2) \\ \lim_{s \to 0} \frac{f(\mathbf{x}_2 + s(\mathbf{x}_1 - \mathbf{x}_2)) - f(\mathbf{x}_2)}{s} &\leq f(\mathbf{x}_1) - f(\mathbf{x}_2) \\ \nabla f(\mathbf{x}_2) \cdot (\mathbf{x}_1 - \mathbf{x}_2) &\leq f(\mathbf{x}_1) - f(\mathbf{x}_2) \\ &\text{(since } \tfrac{d}{ds}|_{s=0} f(\mathbf{x}_2 + s(\mathbf{x}_1 - \mathbf{x}_2)) = \nabla f(\mathbf{x}_2) \cdot (\mathbf{x}_1 - \mathbf{x}_2)) \\ f(\mathbf{x}_2) + \nabla f(\mathbf{x}_2) \cdot (\mathbf{x}_1 - \mathbf{x}_2) &\leq f(\mathbf{x}_1) \\ f(\mathbf{x}) + \nabla f(\mathbf{x}) \cdot (\mathbf{y} - \mathbf{x}) &\leq f(\mathbf{y}) \end{aligned}$$

where $0 \leq s \leq 1$
($\Leftarrow$)
Fix $\mathbf{x}_0, \mathbf{x}_1 \in \Omega$ and $s \in (0,1)$
Let $x = s\mathbf{x}_0 + (1-s)\mathbf{x}_1$

$$\begin{cases} f(\mathbf{x}_0) & \geq f(\mathbf{x}) + \nabla f(\mathbf{x}) \cdot (\mathbf{x}_0 - \mathbf{x}) \\ & = f(\mathbf{x}) + \nabla f(\mathbf{x}) \cdot (1-s)(\mathbf{x}_0 - \mathbf{x}_1) \\ f(\mathbf{x}_1) & \geq f(\mathbf{x}) + \nabla f(\mathbf{x}) \cdot (\mathbf{x}_1 - \mathbf{x}) \\ & = f(\mathbf{x}) + \nabla f(\mathbf{x}) \cdot s(\mathbf{x}_1 - \mathbf{x}_0) \end{cases}$$

$$\begin{cases} sf(x_0) & \geq sf(\mathbf{x}) + \nabla f(\mathbf{x}) \cdot (1-s) \cdot s(\mathbf{x}_0 - \mathbf{x}_1) \\ (1-s)f(\mathbf{x}_1) & \geq (1-s)f(\mathbf{x}) + \nabla f(\mathbf{x}) \cdot (1-s) \cdot s(\mathbf{x}_1 - \mathbf{x}_0) \end{cases}$$

Then

$$sf(\mathbf{x}_0) + (1-s)f(\mathbf{x}_1) \geq f(x) + 0$$

Then f is convex. ■

4 Lecture 4 - September 16th

C^1 criterion for convexity

$$f : \Omega \to \mathbb{R} \text{ is convex } \iff f(y) \geq f(x) + \nabla f(x) \cdot (y - x)$$

for all $x, y \in \Omega$

Theorem: C^2 criterion for convexity Let $f \in C^2$ on $\Omega \subseteq \mathbb{R}^n$ (here we assume $\Omega \subseteq \mathbb{R}^n$ is a convex set containing an interior point)
Then

$$f \text{ is convex on } \Omega \iff \nabla^2 f(x) \geq 0$$

for all $x \in \Omega$

Remark 1 Let A be an $n \times n$ matrix.
"$A \geq 0$" means A is positive semi-definite:

$$v^T A v \geq 0$$

for all $v \in \mathbb{R}^n$

Remark 2 In $\mathbb{R}$,

$$f \text{ is convex } \iff f'(x) \geq 0$$

for all $x \in \Omega$
("concave up" in first year calculus)

proof for Theorem:
Recall 2nd order MVT:

$$f(y) = f(x) + \nabla f(x) \cdot (y - x) + \frac{1}{2}(y - x)^T \nabla^2 f(x + s(y - x)) \cdot (y - x)$$

for some $s \in [0, 1]$
$(\Leftarrow)$
Since $\nabla^2 f(x) \geq 0$, then

$$\frac{1}{2}(y - x)^T \nabla^2 f(x + s(y - x)) \cdot (y - x) \geq 0$$

Then

$$f(y) \geq f(x) + \nabla f(x) \cdot (y - x)$$

for all $x, y \in \Omega$.
Then by C^1 criterion, f is convex.
$(\Rightarrow)$
Assume f is convex on Ω.
Suppose for contradiction that $\nabla^2 f(x)$ is not positive semi-definite at some $x \in \Omega$.
Then $\exists v \neq 0$ s.t. $v^T \nabla^2 f(x) v < 0$ v could be arbitrarily small and > 0
Let $y = x + v$, then

$$(y - x)^T \nabla^2 f(x + s(y - x)) \cdot (y - x) < 0$$

for all $s \in [0, 1]$
Then by MVT,

$$f(y) < f(x) + \nabla f(x) \cdot (y - x)$$

for some $x, y \in \Omega$, and this contradicts the C^1 criterion. ■

4.1 Minimization and Maximization of Convex Functions

Theorem f : convex $\Omega \subseteq \mathbb{R}^n \to \mathbb{R}$ is a convex function.
Suppose $\Gamma := \{x \in \Omega | f(x) = \min_{\Omega} f(x)\} \neq \emptyset$
(i.e. minimizer exists)
Then Γ is a convex set, and any local minimum of f is a global minimum of f.
proof:

Let $m = \min_{\Omega} f(x)$.

$$\Gamma = \{x \in \Omega | f(x) = m\} = \{x \in \Omega | f(x) \leq m\}$$

(sublevel set)
Then by Basic Properties of Convex Sets, Γ is convex.
Let x be a local minimum of f.
Suppose for contradiction that $\exists y$ s.t. $f(y) < f(x)$
(i.e. x is not a global minimum)

$$\begin{aligned} f(sy + (1-s)x) &\le sf(y) + (1-s)f(x) \\ &< sf(x) + (1-s)f(x) && (f(y) < f(x)) \\ &= f(x) \end{aligned}$$

for all $s \in (0,1)$
As s approaches 0, s approaches x.
Then we have $\lim_{s\to 0} f(sy + (1-s)x) = f(x) < f(x)$.
which is a contradiction. ■

5 Lecture 5 - September 18th

Theorem If $f : \Omega \subseteq \mathbb{R}^n \to \mathbb{R}$ is a convex function, and Ω is convex and compact, then

$$\max_{\Omega} f = \max_{\partial\Omega} f$$

Remarks Maximum value of f is attained (also) on the boundary of Ω
proof:
Since Ω is closed, $\partial\Omega \subseteq \Omega$, so $\max_{\Omega} f \ge \max_{\partial\Omega} f$.
Suppose $f(x_0) = \max_{\Omega} f$ for some $x_0 \notin \partial\Omega$. Let L be an arbitrary line through x_0.
By convexity and compactness of Ω, L meets $\partial\Omega$ at two points x_1, x_2.
Let $x_0 + sx_1 + (1-s)x_2$ for $s \in (0,1)$

$$\begin{aligned} f(x_0) = f(sx_1 + (1-s)x_2) &\le sf(x_1) + (1-s)f(x_2) && (f \text{ convex}) \\ &\le \max\{f(x_1), f(x_2)\} \\ &\le \max_{\partial\Omega} f \\ &= f(x_0) = \max_{\Omega} f \end{aligned}$$

This implies that

$$\max_{\Omega} f = \max_{\partial\Omega} f$$

as wanted. ■

Example

$$|ab| \leq \frac{1}{p}|a|^p + \frac{1}{q}|b|^q$$

where $p, q > 1$ s.t. $\frac{1}{p} + \frac{1}{q} = 1$.
Special cases:

1. $$p = q = 2, |ab| \leq \frac{|a|^2 + |b|^2}{2}$$

2. $$p = 3, q = \frac{3}{2}, |ab| \leq \frac{1}{3}|a|^3 + \frac{2}{3}|b|^{\frac{3}{2}}$$

proof:
Since function $f(x) = -\log(x)$ is convex, then

$$\begin{aligned}
(-\log)|ab| &= (-\log)|a| + (-\log)|b| \\
&= \frac{1}{p}(-\log)|a|^p + \frac{1}{q}(-\log)|b|^q \\
&\geq (-\log)(\frac{1}{p}|a|^p + \frac{1}{q}|b|^q) \\
(-\log)|ab| &\geq (-\log)(\frac{1}{p}|a|^p + \frac{1}{q}|b|^q) \\
\log|ab| &\leq \log(\frac{1}{p}|a|^p + \frac{1}{q}|b|^q) \\
|ab| &\leq \frac{1}{p}|a|^p + \frac{1}{q}|b|^q \qquad \text{(exponential function is increasing)}
\end{aligned}$$

5.1 Basics of Unconstrained Optimization

Extreme Value Theorem Suppose $f : \mathbb{R}^n \to \mathbb{R}$ is continues, and compact set $K \subseteq \mathbb{R}^n$ Then the problem

$$\min_{x \in K} f(x)$$

has a solution.

Recall

1. $$K \subseteq \mathbb{R}^n \text{ compact} \iff K \text{ closed and bounded}$$

2. If $h_1, \ldots, h_k$ and $g_1, \ldots, g_m$ are continuous functions on $\mathbb{R}^n$, then the set of all points $x \in \mathbb{R}^n$ s.t.

$$\begin{cases} h_i(x) = 0 & \text{for all } i \\ g_j(x) \leq 0 & \text{for all } j \end{cases}$$

is a closed set.

3. If such a set is also bounded