Nonlinear Optimization

Derek Li

Contents

1 Review

1.1 One-Variable Calculus

Theorem 1.1 (Mean Value Theorem)**.** Let $g \in C^1$ on $\mathbb{R}$. We have

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

for some $\theta \in (0,1)$ and $\frac{g(x+h)-g(x)}{h}$ is the slope of secant line between $(x, g(x))$ and $(x+h, g(x+h))$. Or we can write $g(x+h) = g(x) + hg'(x+\theta h)$.

Theorem 1.2 (First Order Taylor Approximation)**.** Let $g \in C^1$ on $\mathbb{R}$. We have

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

where $o(h)$ is the error and we say a function $f(h) = o(h)$ to mean

$$\lim_{h \to 0} \frac{f(h)}{h} = 0.$$

Proof. Want to show $g(x+h) - g(x) - hg'(x) = o(h)$.

We have

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

□

Theorem 1.3 (Second Order Mean Value Theorem)**.** Let $g \in C^2$ on $\mathbb{R}$. We have

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

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

Theorem 1.4 (Second Order Taylor Approximation)**.** Let $g \in C^2$ on $\mathbb{R}$. We have

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

Proof. W.T.S. $g(x+h) - g(x) - hg'(x) - \frac{h^2}{2}g''(x) = o(h^2)$.

We have

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

□

1.2 Multi-variable Calculus

Definition 1.1 (Gradient)**.** Gradient of $f : \mathbb{R}^n \to \mathbb{R}$ at $\mathbf{x} \in \mathbb{R}^n, \nabla f(\mathbf{x})$, if exists is a vector characterized by the property

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

and $\nabla f(\mathbf{x}) = \left(\frac{\partial f}{\partial x_1}(\mathbf{x}), \cdots, \frac{\partial f}{\partial x_n}(\mathbf{x})\right)$.

The instantaneous rate of change of f at $\mathbf{x}$ in direction $\mathbf{v}$ (suppose w.l.o.g. $\|\mathbf{v}\| = 1$) is

$$\begin{aligned} \left.\frac{\mathrm{d}}{\mathrm{d}t}\right|_{t=0} f(\mathbf{x}+t\mathbf{v}) &= \nabla f(\mathbf{x}+t\mathbf{v})\cdot \mathbf{v}|_{t=0} \\ &= \nabla f(\mathbf{x})\cdot \mathbf{v} \\ &= |\nabla f(\mathbf{x})||\mathbf{v}|\cos\theta \\ &= |\nabla f(\mathbf{x})|\cos\theta, \end{aligned}$$

where θ is the angle between $\nabla f(\mathbf{x})$ and $\mathbf{v}$. Obviously, the instantaneous rate maximizes when $\theta = 0$. Therefore, when it is not equal to zero, $\nabla f(\mathbf{x})$ points in the direction of steepest ascent.

Theorem 1.5 (Mean Value Theorem in $\mathbb{R}^n$)**.** Let $f \in C^1$ on $\mathbb{R}^n$, then for any $\mathbf{x}, \mathbf{v} \in \mathbb{R}^n$, we have

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

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

Proof. Consider $g(t) = f(\mathbf{x} + t\mathbf{v})$, where $t \in \mathbb{R}$ and $g \in C^1$ on $\mathbb{R}$.

By Mean Value Theorem in $\mathbb{R}$, we have

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

for some $\theta \in (0, 1)$. □

Note.

$$g'(t) = \frac{\mathrm{d}}{\mathrm{d}t} f(\mathbf{x} + t\mathbf{v}) = \nabla f(\mathbf{x} + t\mathbf{v}) \cdot \mathbf{v}.$$

Theorem 1.6 (First Order Taylor Approximation in $\mathbb{R}^n$)**.** Let $f \in C^1$ on $\mathbb{R}^n$. We have

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

Proof. We have

$$\begin{aligned} \lim_{\|\mathbf{v}\| \to 0} \frac{f(\mathbf{x} + \mathbf{v}) - f(\mathbf{x}) - \nabla f(\mathbf{x}) \cdot \mathbf{v}}{\|\mathbf{v}\|} &= \lim_{\|\mathbf{v}\| \to 0} \frac{\nabla f(\mathbf{x} + \theta\mathbf{v}) \cdot \mathbf{v} - \nabla f(\mathbf{x}) \cdot \mathbf{v}}{\|\mathbf{v}\|} \\ &= \lim_{\|\mathbf{v}\| \to 0} [\nabla f(\mathbf{x} + \theta\mathbf{v}) - \nabla f(\mathbf{v})] \cdot \frac{\mathbf{v}}{\|\mathbf{v}\|} = 0. \end{aligned}$$

□

Theorem 1.7 (Second Order Mean Value Theorem in $\mathbb{R}^n$)**.** Let $f \in C^2$ on $\mathbb{R}^n$. We have

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

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

Note 1. Hessian matrix

$$\nabla^2 f(\mathbf{x}) = \left(\frac{\partial^2 f}{\partial x_i \partial x_j}(\mathbf{x}) \right)_{1 \leqslant i,j \leqslant n}$$

is a symmetric matrix because of Clairaut's Theorem.

Note 2.

$$\mathbf{v}^T \nabla^2 f(\mathbf{x}) \cdot \mathbf{v} = \sum_{1 \leqslant i,j \leqslant n} \frac{\partial^2 f}{\partial x_i \partial x_j}(\mathbf{x}) v_i v_j.$$

Theorem 1.8 (Second Order Taylor Approximation in $\mathbb{R}^n$)**.** Let $f \in C^2$ on $\mathbb{R}^n$. We have

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

Proof. We have

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

□

Theorem 1.9 (Implicit Function Theorem)**.** Let $f : \mathbb{R}^{n+1} \to \mathbb{R}$ be a C^1 function. Fix $(\mathbf{a}, b) \in \mathbb{R}^n \times \mathbb{R}$ s.t. $f(\mathbf{a}, b) = 0$ and $\nabla f(\mathbf{a}, b) \neq 0$. We have $\{(\mathbf{x}, y) \in \mathbb{R}^n \times \mathbb{R} | f(\mathbf{x}, y) = 0\}$ is locally the graph of a function.

Definition 1.2 (Level Set)**.** $\{\mathbf{x} \in \mathbb{R}^n | f(\mathbf{x}) = c\}$ is called c-level set of f.

Theorem 1.10. Gradient $\nabla f(\mathbf{x}_0) \perp$ level curve through $\mathbf{x}_0$.

Definition 1.3 (Convex Set)**.** $\Omega \subseteq \mathbb{R}^n$ is a convex set if for all $\mathbf{x}_1, \mathbf{x}_2 \in \Omega$, we have line segment between $\mathbf{x}_1, \mathbf{x}_2 : s\mathbf{x}_1 + (1-s)\mathbf{x}_2 \in \Omega, s \in [0, 1]$.

Definition 1.4 (Convex Function)**.** A function $f : \Omega \subseteq \mathbb{R}^n \to \mathbb{R}$ is convex if

$$f(s\mathbf{x}_1 + (1-s)\mathbf{x}_2) \leqslant s f(\mathbf{x}_1) + (1-s) f(\mathbf{x}_2),$$

for all $\mathbf{x}_1, \mathbf{x}_2 \in \Omega$ and all $s \in [0, 1]$, where Ω is a convex set.

Example 1.1. $f : \mathbb{R}^n \to \mathbb{R}$ given by $f(\mathbf{x}) = \|\mathbf{x}\|$ is convex.

Proof. Let $\mathbf{x}_1, \mathbf{x}_2 \in \mathbb{R}^n$. We have

$$\begin{aligned}
f(s\mathbf{x}_1 + (1-s)\mathbf{x}_2) &= \|s\mathbf{x}_1 + (1-s)\mathbf{x}_2\| \leqslant \|s\mathbf{x}_1\| + \|(1-s)\mathbf{x}_2\| \\
&= s\|\mathbf{x}_1\| + (1-s)\|\mathbf{x}_2\| = s_1 f(\mathbf{x}_1) + (1-s) f(\mathbf{x}_2).
\end{aligned}$$

□

Definition 1.5 (Concave Function)**.** A function f concave if $-f$ is convex.

Note. The linear function is both convex and concave.

Theorem 1.11 (Basic Properties of Convex Function)**.** 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 convex function on Ω.

(2) f is convex functions and $a \geqslant 0 \Rightarrow af$ is a convex function.

(3) f is a convex function on $\Omega \Rightarrow \mathrm{SL}_c := \{\mathbf{x} \in \mathbb{R}^n | f(x) \leqslant c\}$, the sub-level sets are convex.

Proof of (3). W.T.S. for $\mathbf{x}_1, \mathbf{x}_2 \in \mathrm{SL}_c$, $s\mathbf{x}_1 + (1-s)\mathbf{x}_2 \in \mathrm{SL}_c$ for any $s \in [0, 1]$.

Since $\mathbf{x}_1, \mathbf{x}_2 \in \mathrm{SL}_c$, we have $f(\mathbf{x}_1) \leqslant c, f(\mathbf{x}_2) \leqslant c$. Because f is a convex function, we have

$$f(s\mathbf{x}_1 + (1-s)\mathbf{x}_2) \leqslant sf(\mathbf{x}_1) + (1-s)f(\mathbf{x}_2) \leqslant sc + (1-s)c = c.$$

Thus, $s\mathbf{x}_1 + (1-s)\mathbf{x}_2 \in \mathrm{SL}_c$. □

Theorem 1.12 (Characterization of C^1 convex function)**.** Let $f : \Omega \to \mathbb{R}$ be a C^1 function and Ω is a convex subset of $\mathbb{R}^n$. Then f is convex on Ω iff

$$f(\mathbf{y}) \geqslant f(\mathbf{x}) + \nabla f(\mathbf{x}) \cdot (\mathbf{y} - \mathbf{x}),$$

for all $\mathbf{x}, \mathbf{y} \in \Omega$.

Proof. ($\Rightarrow$) Suppose f is convex. By definition,

$$\begin{aligned} & f(s\mathbf{x}_1 + (1-s)\mathbf{x}_2) \leqslant sf(\mathbf{x}_1) + (1-s)f(\mathbf{x}_2), 0 \ \leqslant s \leqslant 1. \\ \Rightarrow & \frac{f(s\mathbf{x}_1 + (1-s)\mathbf{x}_2) - f(\mathbf{x}_2)}{s} \leqslant f(\mathbf{x}_1) - f(\mathbf{x}_2), 0 < s \leqslant 1. \\ \Rightarrow & \lim_{s \to 0} \frac{f(\mathbf{x}_2 + s(\mathbf{x}_1 - \mathbf{x}_2)) - f(\mathbf{x}_2)}{s} \leqslant f(\mathbf{x}_1) - f(\mathbf{x}_2). \end{aligned}$$

Recall that $\partial_{\mathbf{x}_1 - \mathbf{x}_2} f(\mathbf{x}_2) := \lim_{s \to 0} \frac{f(\mathbf{x}_2 + s(\mathbf{x}_1 - \mathbf{x}_2)) - f(\mathbf{x}_2)}{s}$, i.e., the directional derivative of f at $\mathbf{x}_2$ in the direction $\mathbf{x}_1 - \mathbf{x}_2$.

Since f is C^1, we have

$$\nabla f(\mathbf{x}_2) \cdot (\mathbf{x}_1 - \mathbf{x}_2) \leqslant f(\mathbf{x}_1) - f(\mathbf{x}_2) \Rightarrow f(\mathbf{x}_1) \geqslant f(\mathbf{x}_2) + \nabla f(\mathbf{x}_2) \cdot (\mathbf{x}_1 - \mathbf{x}_2),$$

i.e.,

$$f(\mathbf{y}) \geqslant f(\mathbf{x}) + \nabla f(\mathbf{x}) \cdot (\mathbf{y} - \mathbf{x}).$$

($\Leftarrow$) Fix $\mathbf{x}_0, \mathbf{x}_1 \in \Omega, s \in (0,1)$. Let $\mathbf{x} = s\mathbf{x}_0 + (1-s)\mathbf{x}_1$. We have

$$\begin{aligned} f(\mathbf{x}_0) &\geqslant f(\mathbf{x}) + \nabla f(\mathbf{x}) \cdot (\mathbf{x}_0 - \mathbf{x}) = f(\mathbf{x}) + \nabla f(\mathbf{x}) \cdot (1-s)(\mathbf{x}_0 - \mathbf{x}_1), \\ f(\mathbf{x}_1) &\geqslant f(\mathbf{x}) + \nabla f(\mathbf{x}) \cdot (\mathbf{x}_1 - \mathbf{x}) = f(\mathbf{x}) + \nabla f(\mathbf{x}) \cdot s(\mathbf{x}_1 - \mathbf{x}_0). \end{aligned}$$

Therefore,

$$\begin{aligned} sf(\mathbf{x}_0) &\geqslant sf(\mathbf{x}) + \nabla f(\mathbf{x}) \cdot s(1-s)(\mathbf{x}_0 - \mathbf{x}_1), \\ (1-s)f(\mathbf{x}_1) &\geqslant (1-s)f(\mathbf{x}) + \nabla f(\mathbf{x}) \cdot s(1-s)(\mathbf{x}_1 - \mathbf{x}_0). \end{aligned}$$

Thus,

$$sf(\mathbf{x}_0) + (1-s)f(\mathbf{x}_1) \geqslant sf(\mathbf{x}) + (1-s)f(\mathbf{x}) = f(\mathbf{x}) = f(s\mathbf{x}_0 + (1-s)\mathbf{x}_1),$$

for all $\mathbf{x}_0, \mathbf{x}_1 \in \Omega$ and $s \in (0,1)$.

When $s = 0$ or $s = 1, sf(\mathbf{x}_0) + (1-s)f(\mathbf{x}_1) = f(s\mathbf{x}_0 + (1-s)\mathbf{x}_1)$.

In conclusion, f is convex on Ω. $\square$