Nonlinear Optimization

Derek Li

Contents

1 Review

1.1 One-Variable Calculus

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

1.1.2 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

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

$\square$

1.1.3 Second Order MVT

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

1.1.4 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

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

1.2.2 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:

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

1.2.3 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

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

□

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

1.2.5 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

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

□

1.2.6 Geometric Meaning of Gradient

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.