3 Taylor Series

3.1 Taylor Examples

- Real-vauled function $f : \mathbb{R} \to \mathbb{R}$,

$$f(x) = f\left(x^{(0)}\right) + \frac{x - x^{(0)}}{1!} f'\left(x^{(0)}\right) + \frac{\left(x - x^{(0)}\right)^2}{2!} f''\left(x^{(0)}\right) + R_3$$

.

- The linear approximation of f about the point $\boldsymbol{x}_0$, $f : \mathbb{R}^n \to \mathbb{R}$,

$$f(\boldsymbol{x}) = f\left(\boldsymbol{x}^{(0)}\right) + \begin{bmatrix} \quad Df\left(\boldsymbol{x}^{(0)}\right) \quad \end{bmatrix} \begin{bmatrix} \\ \boldsymbol{x} - \boldsymbol{x}^{(0)} \\ \\ \end{bmatrix} + R_2.$$

- The quadratic approximation of f about the point $\boldsymbol{x}_0$, $f : \mathbb{R}^n \to \mathbb{R}$,

$$\begin{aligned} f(\boldsymbol{x}) = f\left(\boldsymbol{x}^{(0)}\right) &+ \begin{bmatrix} \quad Df\left(\boldsymbol{x}^{(0)}\right) \quad \end{bmatrix} \begin{bmatrix} \\ \boldsymbol{x} - \boldsymbol{x}^{(0)} \\ \\ \end{bmatrix} \\ &+ \frac{1}{2!} \begin{bmatrix} \quad \left(\boldsymbol{x} - \boldsymbol{x}^{(0)}\right)^\top \quad \end{bmatrix} \begin{bmatrix} \\ D^2 f\left(\boldsymbol{x}^{(0)}\right) \\ \\ \end{bmatrix} \begin{bmatrix} \\ \quad \boldsymbol{x} - \boldsymbol{x}^{(0)} \quad \\ \\ \end{bmatrix} + R_3. \end{aligned}$$

- For *quadratic approximation*, if we assume that $f \in \mathcal{C}^3$, term R_3 can be written as

$$\begin{aligned} f(\boldsymbol{x}) = f\left(\boldsymbol{x}_0\right) &+ \frac{1}{1!} Df\left(\boldsymbol{x}_0\right)\left(\boldsymbol{x} - \boldsymbol{x}_0\right) \\ &+ \frac{1}{2!}\left(\boldsymbol{x} - \boldsymbol{x}_0\right)^\top D^2 f\left(\boldsymbol{x}_0\right)\left(\boldsymbol{x} - \boldsymbol{x}_0\right) + O\left(\left\|\boldsymbol{x} - \boldsymbol{x}_0\right\|^3\right). \end{aligned}$$

3.2 Directional derivative

- The directional derivative is the rate of increase of f at $\boldsymbol{x}$ in the direction of $\boldsymbol{d}$,

$$\frac{\partial f}{\partial \boldsymbol{d}} \triangleq \lim_{\alpha \to 0} \frac{f(\boldsymbol{x} + \alpha \boldsymbol{d}) - f(\boldsymbol{x})}{\alpha} = \frac{d}{d\alpha} f(\boldsymbol{x} + \alpha \boldsymbol{d})\bigg|_{\alpha = 0} = \boldsymbol{d}^\top \nabla f(\boldsymbol{x}),$$

where

$$f(\boldsymbol{x} + \alpha \boldsymbol{d}) = f(\boldsymbol{x}) + \alpha \boldsymbol{d}^\top \nabla f(\boldsymbol{x}) + R_2.$$

Short derivation:
Given *linear approximation* of f about the point $\boldsymbol{x}_0$, $f : \mathbb{R}^n \to \mathbb{R}$,

$$f(\boldsymbol{x}) = f\left(\boldsymbol{x}^{(0)}\right) + \left[\quad Df\left(\boldsymbol{x}^{(0)}\right) \quad\right] \left[\boldsymbol{x} - \boldsymbol{x}^{(0)}\right] + R_2.$$

Thus,

$$f(\boldsymbol{x} + \alpha\boldsymbol{d}) = f\left(\boldsymbol{x}\right) + \left[\quad Df\left(\boldsymbol{x}\right) \quad\right] \left[\boldsymbol{x} + \alpha\boldsymbol{d} - \boldsymbol{x}\right] + R_2.$$

Substitute back,

$$\begin{aligned}\frac{\partial f}{\partial \boldsymbol{d}} \triangleq \lim_{\alpha\to 0} \frac{f(\boldsymbol{x}+\alpha\boldsymbol{d}) - f(\boldsymbol{x})}{\alpha} &= \lim_{\alpha\to 0} \frac{f(\boldsymbol{x}) + \left[\quad Df\left(\boldsymbol{x}\right) \quad\right]\left[\alpha\boldsymbol{d}\right] - f\left(\boldsymbol{x}\right)}{\alpha} \\ &= \left[\quad Df\left(\boldsymbol{x}\right) \quad\right]\left[\ \boldsymbol{d}\ \right] \\ &= \boldsymbol{d}^\top \nabla f(\boldsymbol{x}).\end{aligned}$$

3.3 Taylor series

♠ $\mathcal{THEOREM}$5.8 Taylor's Theorem. Assume that a function $f : \mathbb{R} \to \mathbb{R}$ is m times continuously differentiable (i.e., $f \in \mathcal{C}^m$) on an interval $[a, b]$. Denote $h = b - a$. Then

$$f(b) = f(a) + \frac{h}{1!}f^{(1)}(a) + \frac{h^2}{2!}f^{(2)}(a) + \cdots + \frac{h^{m-1}}{(m-1)!}f^{(m-1)}(a) + R_m$$

where m-th reminder term can be written with the m-th derivative of f, $f^{(m)}$, and

$$R_m = \frac{h^m(1-\theta)^{m-1}}{(m-1)!}f^{(m)}(a+\theta h) = \frac{h^m}{m!}f^{(m)}\left(a + \theta' h\right)$$

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

♠ $\mathcal{THEOREM}$5.9 Mean value theorem If a function $\boldsymbol{f} : \mathbb{R}^n \to \mathbb{R}^m$ is differentiable on an open set $\Omega \subset \mathbb{R}^n$, then for any pair of points $\boldsymbol{x}, \boldsymbol{y} \in \Omega$, there exists a matrix $\boldsymbol{M}$ such that

$$\boldsymbol{f}(\boldsymbol{x}) - \boldsymbol{f}(\boldsymbol{y}) = \boldsymbol{M}(\boldsymbol{x} - \boldsymbol{y}).$$

3.4 Derivation details

For quadratic approximation, we can think about it in this way,

$$
\begin{aligned}
f(\boldsymbol{x}^{(0)} + \alpha \boldsymbol{d}) &= \phi(\alpha) \\
&= \phi(0) + \frac{\alpha}{1!}\phi^{'}(0) + \frac{\alpha^2}{2!}\phi^{''}(0) + R_3
\end{aligned}
$$

First order item,

$$
\frac{d\phi}{d\alpha} = \begin{bmatrix} Df\left(\boldsymbol{x}^{(0)} + \alpha \boldsymbol{d}\right) \end{bmatrix} \begin{bmatrix} \boldsymbol{d} \end{bmatrix}, \quad \phi^{'}(0) = \begin{bmatrix} Df\left(\boldsymbol{x}^{(0)}\right) \end{bmatrix} \begin{bmatrix} \boldsymbol{d} \end{bmatrix}.
$$

Second order term ,

$$
\begin{aligned}
\frac{d^2\phi}{d\alpha^2} &= \frac{d}{d\alpha}\left(\frac{d\phi}{d\alpha}\right) \\
&= \frac{d}{d\alpha}\left(\begin{bmatrix} Df\left(\boldsymbol{x}^{(0)} + \alpha \boldsymbol{d}\right) \end{bmatrix} \begin{bmatrix} \boldsymbol{d} \end{bmatrix}\right) \\
&= \frac{d}{d\alpha}\left(\begin{bmatrix} \boldsymbol{d}^\top \end{bmatrix} \begin{bmatrix} \nabla f\left(\boldsymbol{x}^{(0)} + \alpha \boldsymbol{d}\right) \end{bmatrix}\right) \\
&= \begin{bmatrix} \boldsymbol{d}^\top \end{bmatrix} \frac{d}{d\alpha}\left(\begin{bmatrix} \nabla f\left(\boldsymbol{x}^{(0)} + \alpha \boldsymbol{d}\right) \end{bmatrix}\right) \\
&= \begin{bmatrix} \boldsymbol{d}^\top \end{bmatrix} \begin{bmatrix} \vdots \\ \frac{d}{d\alpha}\left(\frac{\partial f}{\partial x_i}\left(\boldsymbol{x}^{(0)} + \alpha \boldsymbol{d}\right)\right) \\ \vdots \end{bmatrix} \\
&= \begin{bmatrix} \boldsymbol{d}^\top \end{bmatrix} \begin{bmatrix} \vdots & & \vdots \\ \frac{\partial^2 f}{\partial x_1 x_i}\left(\boldsymbol{x}^{(0)} + \alpha \boldsymbol{d}\right) & \dots & \frac{\partial^2 f}{\partial x_n x_i}\left(\boldsymbol{x}^{(0)} + \alpha \boldsymbol{d}\right) \\ \vdots & & \vdots \end{bmatrix} \begin{bmatrix} \boldsymbol{d} \end{bmatrix} \\
&= \begin{bmatrix} \boldsymbol{d}^\top \end{bmatrix} \begin{bmatrix} \mathbf{F}\left(\boldsymbol{x}^{(0)} + \alpha \boldsymbol{d}\right) \end{bmatrix} \begin{bmatrix} \boldsymbol{d} \end{bmatrix}
\end{aligned}
$$

In summary,

$$
\begin{aligned}
\phi(\alpha)|_{\alpha=1} &= f\left(\boldsymbol{x}\right) \\
&= f(\boldsymbol{x}^{(0)} + 1\boldsymbol{d}) \\
&= f(\boldsymbol{x}^{(0)}) \quad + \begin{bmatrix} Df(\boldsymbol{x}^{(0)}) \end{bmatrix} \begin{bmatrix} \boldsymbol{d} \end{bmatrix} \\
&\qquad + \tfrac{1}{2} \begin{bmatrix} \boldsymbol{d}^\top \end{bmatrix} \begin{bmatrix} \mathbf{F}\left(\boldsymbol{x}^{(0)} + 1\boldsymbol{d}\right) \end{bmatrix} \begin{bmatrix} \boldsymbol{d} \end{bmatrix} + R_3
\end{aligned}
$$

[Ref]: Edwin K.P. Chong, Stanislaw H. Żak, “PART I MATHEMATICAL REVIEW” in “An introduction to optimization”, 4th Edition, John Wiley and Sons, Inc. 2013.