n 変数実数値関数の 2 回微分

ヘッセ行列

$\boldsymbol{x} = (x_1, ..., x_n)$ における f のヘッセ行列 $\nabla^2 f(\boldsymbol{x})$ を以下のように定義する.

$$\nabla^2 f(\boldsymbol{x}) = \begin{bmatrix} \dfrac{\partial^2 f(\boldsymbol{x})}{\partial x_1^2} & \dfrac{\partial^2 f(\boldsymbol{x})}{\partial x_1 \partial x_2} & \cdots & \dfrac{\partial^2 f(\boldsymbol{x})}{\partial x_1 \partial x_n} \\ \dfrac{\partial^2 f(\boldsymbol{x})}{\partial x_2 \partial x_1} & \dfrac{\partial^2 f(\boldsymbol{x})}{\partial x_2^2} & \cdots & \dfrac{\partial^2 f(\boldsymbol{x})}{\partial x_2 \partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \dfrac{\partial^2 f(\boldsymbol{x})}{\partial x_n \partial x_1} & \dfrac{\partial^2 f(\boldsymbol{x})}{\partial x_n \partial x_2} & \cdots & \dfrac{\partial^2 f(\boldsymbol{x})}{\partial x_n^2} \end{bmatrix}$$

n 変数実数値関数の 2 回微分

2 回微分可能の定義

$\displaystyle\lim_{\lambda\to 0}\frac{o(\lambda)}{\lambda}=0$ を満たすある o が存在して,

$$f(\boldsymbol{x}+\boldsymbol{d})=f(\boldsymbol{x})+\nabla f(\boldsymbol{x})^\top(\boldsymbol{d})+\frac{1}{2}(\boldsymbol{d})^\top\nabla^2 f(\boldsymbol{x})\boldsymbol{d}+o(\|\boldsymbol{d}\|^2)$$

が成り立つとき, x において f は 2 回微分可能であるという.
更に, $\nabla^2 f(x)$ が x について連続ならば, f は連続的 2 回微分可能であるという.

テイラーの定理

定理 4.6(テイラーの定理)

関数 $f : \mathbb{R}^n \to \mathbb{R}$ が開凸集合 $X \subseteq \mathbb{R}$ 上で 2 回連続微分可能とする.
このとき,

$$\forall \boldsymbol{x}, \boldsymbol{y} \in X, \tau \in (0, 1);$$

$$f(\boldsymbol{x}) - f(\boldsymbol{y}) = \nabla f(\boldsymbol{y})^\top (\boldsymbol{x} - \boldsymbol{y}) + \frac{1}{2}(\boldsymbol{x} - \boldsymbol{y})^\top \nabla^2 f((1-\tau)\boldsymbol{x} + \tau \boldsymbol{y})(\boldsymbol{x} - \boldsymbol{y})$$

定理 4.7

$X \subseteq \mathbb{R}^n$ が開凸集合で, $f : X \to \mathbb{R}$ が X 上で
連続的 2 回微分可能であるとする. このとき,

- f が X 上で凸関数 $\iff \forall \boldsymbol{x} \in X, \nabla^2 f(\boldsymbol{x}) \in \mathbb{S}^n_+$
- $\forall \boldsymbol{x} \in X, \nabla^2 f(\boldsymbol{x}) \in \mathbb{S}^n_{++} \Rightarrow f$ は X 上で狭義凸関数

定理 4.7 の証明 (1)

(1 つめの主張について)

(⇒)

f が凸関数であるとする。X が開集合であることから、X の任意の点のある近傍は X に含まれる。

したがっていま、任意の $d \in \mathbb{R}^n$, $\boldsymbol{x} \in X$ と、十分小さい $\lambda > 0$ に対し、$\boldsymbol{x} + \lambda d \in X$ となる。

よってテイラーの定理より、点 $\boldsymbol{x} + \lambda d, \boldsymbol{x}$ に対し、ある $\tau \in (0, 1)$ が存在して

$$
\begin{aligned}
f(\boldsymbol{x} + \lambda \boldsymbol{d}) &= f(\boldsymbol{x}) + \lambda \nabla f(\boldsymbol{x})^\top \boldsymbol{d} + \frac{1}{2} \lambda \boldsymbol{d}^\top \nabla^2 f((1 - \tau)(\boldsymbol{x} + \lambda \boldsymbol{d}) + \tau \boldsymbol{x}) \lambda \boldsymbol{d} \\
&= f(\boldsymbol{x}) + \lambda \nabla f(\boldsymbol{x})^\top \boldsymbol{d} + \frac{\lambda^2}{2} \boldsymbol{d}^\top \nabla^2 f(\boldsymbol{x} + (1 - \tau) \lambda \boldsymbol{d}) \boldsymbol{d}
\end{aligned} \tag{1}
$$

また、f は凸関数なので、勾配不等式より、

$$
f(\boldsymbol{x} + \lambda \boldsymbol{d}) \geq f(\boldsymbol{x}) + \lambda \nabla f(\boldsymbol{x})^\top \boldsymbol{d} \tag{2}
$$

定理 4.7 の証明 (1)

(2)-(1) より、

$$0 \geq -\frac{\lambda^2}{2}\boldsymbol{d}^\top \nabla^2 f(\boldsymbol{x} + (1-\tau)\lambda \boldsymbol{d})\boldsymbol{d}$$
$$\iff \boldsymbol{d}^\top \nabla^2 f(\boldsymbol{x} + (1-\tau)\lambda \boldsymbol{d})\boldsymbol{d} \geq 0$$

λ はいくらでも小さく取れるから、$\lambda \to 0$ として、

$$\lim_{\lambda \to 0} \boldsymbol{d}^\top \nabla^2 f(\boldsymbol{x} + (1-\tau)\lambda \boldsymbol{d})\boldsymbol{d} = \boldsymbol{d}^\top \nabla^2 f(\boldsymbol{x})\boldsymbol{d} \geq 0$$

定理 4.7 の証明 (1)

(⇐) ヘッセ行列 $\nabla^2 f(x)$ が任意の $x \in X$ に対して半正定値であるとする。
テイラーの定理より、
ある $\tau \in (0,1)$ が存在して、任意の $\boldsymbol{x}, \boldsymbol{y} \in X$ に対して、

$$\begin{aligned} f(\boldsymbol{y}) &= f(\boldsymbol{x}) + \nabla f(\boldsymbol{x})^\top (\boldsymbol{y} - \boldsymbol{x}) + \frac{1}{2}(\boldsymbol{y} - \boldsymbol{x})^\top \nabla^2 f(\boldsymbol{x} + \tau(\boldsymbol{y} - \boldsymbol{x}))(\boldsymbol{y} - \boldsymbol{x}) \\ &\geq f(\boldsymbol{x}) + \nabla f(\boldsymbol{x})^\top (\boldsymbol{y} - \boldsymbol{x}) \end{aligned} \tag{3}$$

($\nabla^2 f(\boldsymbol{x} + \tau(\boldsymbol{y} - \boldsymbol{x}))$ の半正定値性を用いた)
これは勾配不等式にほかならない。ゆえに f は凸関数である。 □

定理 4.7 の証明 (2)

(2 つめの主張について)

$\forall \boldsymbol{x} \in X, \nabla^2 f(x) \in \mathbb{S}^n_{++}$ とする。(3) より、

$$\begin{aligned} &\forall \boldsymbol{x}, \boldsymbol{y} \in X; \\ &\qquad f(\boldsymbol{y}) > f(\boldsymbol{x}) + \nabla f(\boldsymbol{x})^\top (\boldsymbol{y} - \boldsymbol{x}) \\ &\qquad f(\boldsymbol{x}) > f(\boldsymbol{y}) + \nabla f(\boldsymbol{y})^\top (\boldsymbol{x} - \boldsymbol{y}) \end{aligned}$$

これと X が凸であることから、

$$f(\boldsymbol{y}) > f((1-\lambda)\boldsymbol{x} + \lambda \boldsymbol{y}) + \nabla f((1-\lambda)\boldsymbol{x} + \lambda \boldsymbol{y})^\top (1-\lambda)(\boldsymbol{y} - \boldsymbol{x}) \tag{4}$$

$$f(\boldsymbol{x}) > f((1-\lambda)\boldsymbol{x} + \lambda \boldsymbol{y}) + \nabla f((1-\lambda)\boldsymbol{x} + \lambda \boldsymbol{y})^\top \lambda(\boldsymbol{x} - \boldsymbol{y}) \tag{5}$$

(4)×λ+(5)×$(1-\lambda)$:

$$(1-\lambda) f(\boldsymbol{x}) + \lambda f(\boldsymbol{y}) > f((1-\lambda)\boldsymbol{x} + \lambda \boldsymbol{y})$$

これより f は X 上で狭義凸。 □