凸関数とその最適性条件

関数 $f:\mathbb{R}^n \to \mathbb{R}$ が凸関数であるとは，任意の $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$ に対して不等式

$$\lambda f(\boldsymbol{x}) + (1-\lambda) f(\boldsymbol{y}) \geq f(\lambda \boldsymbol{x} + (1-\lambda)\boldsymbol{y}), \quad \forall \lambda \in [0,1]$$

が成り立つことである．また，f が狭義凸関数であるとは，任意の $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$ に対して不等式

$$\lambda f(\boldsymbol{x}) + (1-\lambda) f(\boldsymbol{y}) > f(\lambda \boldsymbol{x} + (1-\lambda)\boldsymbol{y}), \quad \forall \lambda \in (0,1)$$

が成り立つことである．

命題 1. 微分可能な関数 $f:\mathbb{R}^n \to \mathbb{R}$ が凸関数であるための必要十分条件は，任意の $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$ に対して不等式

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

が成り立つことである．

証明．まず，f が凸関数であることを仮定する．凸関数の定義より，任意の $\lambda \in (0,1)$ に対して不等式

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

が成り立つ．この両辺を λ で割って，$\lambda \to +0$ とすると

$$\begin{aligned} f(\boldsymbol{y}) - f(\boldsymbol{x}) &\geq \frac{f(\boldsymbol{x} + \lambda(\boldsymbol{y} - \boldsymbol{x})) - f(\boldsymbol{x})}{\lambda} \\ &\to \nabla f(\boldsymbol{x})^\top (\boldsymbol{y} - \boldsymbol{x}) \quad (\lambda \to +0) \end{aligned}$$

が得られる [*1]．

次に，(1) を仮定する．任意の $\boldsymbol{x}_1, \boldsymbol{x}_2 \in \mathbb{R}^n$ と任意の $\lambda \in (0,1)$ に対して $\boldsymbol{x}$ を

$$\boldsymbol{x} = \lambda \boldsymbol{x}_1 + (1-\lambda)\boldsymbol{x}_2 \tag{2}$$

で定義する．このとき，(1) より

$$f(\boldsymbol{x}_i) \geq f(\boldsymbol{x}) + \nabla f(\boldsymbol{x})^\top (\boldsymbol{x}_i - \boldsymbol{x}), \quad i = 1, 2 \tag{3}$$

が成り立つ．(3) で $i=1$ とおいた式を λ 倍し，$i=2$ とおいた式を $(1-\lambda)$ 倍して辺々加えると

$$\lambda f(\boldsymbol{x}_1) + (1-\lambda) f(\boldsymbol{x}_2) \geq f(\boldsymbol{x}) + \nabla f(\boldsymbol{x})^\top [\lambda \boldsymbol{x}_1 + (1-\lambda)\boldsymbol{x}_2 - \boldsymbol{x}] = f(\boldsymbol{x}) \tag{4}$$

が得られる．(4) の右辺に (2) を代入すると

$$\lambda f(\boldsymbol{x}_1) + (1-\lambda) f(\boldsymbol{x}_2) \geq f(\lambda \boldsymbol{x}_1 + (1-\lambda)\boldsymbol{x}_2)$$

が得られるので，f は凸関数である．　□

[*1] $f(\boldsymbol{x} + \lambda(\boldsymbol{y} - \boldsymbol{x})) = f(\boldsymbol{x}) + \lambda \nabla f(\boldsymbol{x})^\top (\boldsymbol{y} - \boldsymbol{x}) + \mathrm{O}(\lambda^2)$ より $\dfrac{f(\boldsymbol{x} + \lambda(\boldsymbol{y} - \boldsymbol{x})) - f(\boldsymbol{x})}{\lambda} = \nabla f(\boldsymbol{x})^\top (\boldsymbol{y} - \boldsymbol{x}) + \mathrm{O}(\lambda)$ が成り立つ．

命題 2 (凸関数の最小化の最適性条件)**.** $f:\mathbb{R}^n\to\mathbb{R}$ が微分可能な凸関数ならば，任意の停留解 $\boldsymbol{x}^*$ は f の大域的最小解である．

証明. $\nabla f(\boldsymbol{x}^*)=\boldsymbol{0}$ のとき，命題 1 より，任意の $\boldsymbol{x}\in\mathbb{R}^n$ に対して不等式

$$f(\boldsymbol{x})\geq f(\boldsymbol{x}^*)+\nabla f(\boldsymbol{x}^*)^\top(\boldsymbol{x}-\boldsymbol{x}^*)=f(\boldsymbol{x}^*)$$

が成り立つ． □

命題 3. 2 回連続微分可能な（つまり，C^2 級の）関数 $f:\mathbb{R}^n\to\mathbb{R}$ が凸関数であるための必要十分条件は，任意の点 $\boldsymbol{x}\in\mathbb{R}^n$ において $\nabla^2 f(\boldsymbol{x})$ が半正定値であることである．

証明. まず，f が凸関数であることを仮定する．任意の点 $\boldsymbol{x}\in\mathbb{R}^n$ に対し，方向を表すベクトル $\boldsymbol{d}\in\mathbb{R}^n$ と実数 $\epsilon>0$ を用いて点 $\boldsymbol{x}+\epsilon\boldsymbol{d}$ を定める．テイラー展開と命題 1 より

$$\begin{aligned}f(\boldsymbol{x}+\epsilon\boldsymbol{d})&=f(\boldsymbol{x})+\epsilon\nabla f(\boldsymbol{x})^\top\boldsymbol{d}+\epsilon^2\frac{1}{2}\boldsymbol{d}^\top\nabla^2 f(\boldsymbol{x})\,\boldsymbol{d}+\mathrm{O}(\epsilon^2)\\&\geq f(\boldsymbol{x})+\epsilon\nabla f(\boldsymbol{x})^\top\boldsymbol{d}\end{aligned}$$

が得られる．従って

$$\epsilon^2\frac{1}{2}\boldsymbol{d}^\top\nabla^2 f(\boldsymbol{x})\,\boldsymbol{d}+\mathrm{O}(\epsilon^2)\geq 0 \tag{5}$$

が成り立つ．(5) の両辺を ϵ^2 で割り $\epsilon\to+0$ とすると

$$\boldsymbol{d}^\top\nabla^2 f(\boldsymbol{x})\,\boldsymbol{d}\geq 0 \tag{6}$$

が得られる．任意の方向 $\boldsymbol{d}$ に対して (6) が成り立つので，$\nabla^2 f(\boldsymbol{x})\succeq 0$ である．

次に，$\nabla^2 f(\boldsymbol{x})\succeq 0$ $(\forall\boldsymbol{x}\in\mathbb{R}^n)$ を仮定する．テイラーの定理より，任意の $\boldsymbol{x}$ および $\boldsymbol{d}$ に対して

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

を満たす $\theta\in(0,1)$ が存在する．$\nabla^2 f(\boldsymbol{x}+\theta\boldsymbol{d})\succeq 0$ より $\boldsymbol{d}^\top\nabla^2 f(\boldsymbol{x}+\theta\boldsymbol{d})\,\boldsymbol{d}\geq 0$ であることを用い，$\boldsymbol{y}=\boldsymbol{x}+\boldsymbol{d}$ とおくと

$$f(\boldsymbol{y})\geq f(\boldsymbol{x})+\nabla f(\boldsymbol{x})^\top(\boldsymbol{y}-\boldsymbol{x})$$

が得られる．つまり，命題 1 の条件 (1) が成り立つ． □

（以上）