

クレジット:

UTokyo Online Education 数理手法Ⅲ 2018 寒野善博

ライセンス:

利用者は、本講義資料を、教育的な目的に限ってページ単位で利用することができます。特に記載のない限り、本講義資料はページ単位でクリエイティブ・コモンズ 表示-非営利-改変禁止 ライセンスの下に提供されています。

<http://creativecommons.org/licenses/by-nc-nd/4.0/>

本講義資料内には、東京大学が第三者より許諾を得て利用している画像等や、各種ライセンスによって提供されている画像等が含まれています。個々の画像等を本講義資料から切り離して利用することはできません。個々の画像等の利用については、それぞれの権利者の定めるところに従ってください。



凸関数とその最適性条件

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

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

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 $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ に対して不等式

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

が成り立つことである。

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

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

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

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

が成り立つ。この両辺を λ で割って、 $\lambda \rightarrow +0$ とすると

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

が得られる^{*1}。

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

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

で定義する。このとき、(1) より

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

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

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

が得られる。(4) の右辺に (2) を代入すると

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

が得られるので、 f は凸関数である。 □

^{*1} $f(\mathbf{x} + \lambda(\mathbf{y} - \mathbf{x})) = f(\mathbf{x}) + \lambda \nabla f(\mathbf{x})^\top (\mathbf{y} - \mathbf{x}) + O(\lambda^2)$ より $\frac{f(\mathbf{x} + \lambda(\mathbf{y} - \mathbf{x})) - f(\mathbf{x})}{\lambda} = \nabla f(\mathbf{x})^\top (\mathbf{y} - \mathbf{x}) + O(\lambda)$ が成り立つ。

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

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

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

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 $f: \mathbb{R}^n \rightarrow \mathbb{R}$ が凸関数であるための必要十分条件は, 任意の点 $\mathbf{x} \in \mathbb{R}^n$ において $\nabla^2 f(\mathbf{x})$ 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 $\mathbf{x} \in \mathbb{R}^n$ に対し, 方向を表すベクトル $\mathbf{d} \in \mathbb{R}^n$ と実数 $\epsilon > 0$ を用いて点 $\mathbf{x} + \epsilon \mathbf{d}$ を定める. テイラー展開と命題 1 より

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

が得られる. 従って

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

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

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

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

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

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

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

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

が得られる. つまり, 命題 1 の条件 (1) が成り立つ. □

(以上)