

最適化数学 講義ノート (担当: 関口 良行)

1 凸関数

二次関数 $f(x) = x^2$ は最小解を求めるのが楽であった. これは二次関数だけではなく, 同様の形をした凸関数にも当てはまる. 最適化では凸性は大事な概念である.

定義 (凸関数). $0 < \lambda < 1$ を満たす任意の λ と任意の $u, v \in \mathbb{R}^n$ に対して,

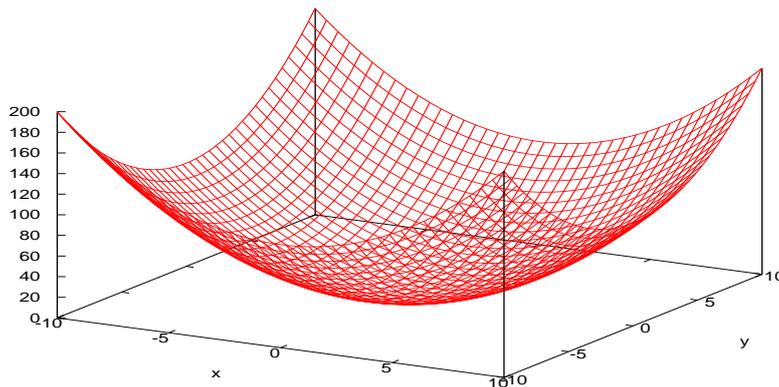
$$f((1-\lambda)u + \lambda v) \leq (1-\lambda)f(u) + \lambda f(v)$$

が成り立つとき, f を**凸関数**と呼ぶ. 等号が成り立つのが $u = v$ のときに限るとき f を**狭義凸関数**と呼ぶ.

注意. 凸関数は, 高校数学では関数が下に凸であるという表現で習った.

例 1. 例えば, $x^2, e^x, |x|$ は凸関数である. 一方, $x^{-1}, -\log x$ は $x > 0$ の部分にだけ制限すれば凸関数になる. このように定義域をある部分集合 D に制限すれば f が凸になるとき, f は D 上で凸であると言う.

\mathbb{R}^2 の場合に凸関数がどのような形をしているか考えてみよう. 下に凸関数 $x^2 + y^2$ のグラフを示す.



定義中, $(1-\lambda)u + \lambda v$ は u, v を結ぶ線分の内分点を表している. 関数が凸であるというのは, 点 $(x, f(x))$ と $(y, f(y))$ を結んだ線分が常に f のグラフの上にあることを表す.

$(u, f(u))$ における f のグラフに対する接平面は,

$$f(u) + f_x(u)(x_2 - x_1) + f_y(u)(y_2 - y_1)$$

となる. 凸関数のグラフの形より, 以下が成り立つことがわかる:

連続微分可能な関数 $f: R^2 \rightarrow R$ が凸ならば, 任意の $u = (x_1, y_1), v = (x_2, y_2) \in R^2$ に対して,

$$f(v) \geq f(u) + f_x(u)(x_2 - x_1) + f_y(u)(y_2 - y_1)$$

が成り立つ.

これは一般に n 変数関数に対しても成り立つ. 次の記号を導入する.

定義. 連続微分可能な関数 $f: \mathbb{R}^2 \rightarrow \mathbb{R}$ と $u = (x_1, \dots, x_n)$ に対して,

$$\nabla f(u) = \left(\frac{\partial f}{\partial x_1}(u), \dots, \frac{\partial f}{\partial x_n}(u) \right)$$

を f の u における**勾配ベクトル**と呼ぶ.

命題 1. $f: \mathbb{R}^n \rightarrow \mathbb{R}$ が凸ならば, 任意の u, v に対して,

$$f(v) \geq f(u) + \nabla f(u)(v - u)$$

が成り立つ. ここで, $\nabla f(u)(v - u)$ はベクトル $\nabla f(u)$ と $(v - u)$ の内積を表す.

Proof. $u, v \in \mathbb{R}^n, 0 < \lambda < 1$ とする. 関数 f のテーラー展開より,

$$f((1 - \lambda)u + \lambda v) = f(u + \lambda(v - u)) = f(u) + \nabla f(u)(\lambda(v - u)) + o(\lambda\|u - v\|)$$

となる. 凸関数の定義より, $f((1 - \lambda)u + \lambda v) \leq (1 - \lambda)f(u) + \lambda f(v)$ なので,

$$\begin{aligned} (1 - \lambda)f(u) + \lambda f(v) &\geq f(u) + \lambda \nabla f(u)(v - u) + o(\lambda\|u - v\|) \\ \lambda f(v) &\geq \lambda f(u) + \lambda \nabla f(u)(v - u) + o(\lambda\|u - v\|) \\ f(v) &\geq f(u) + \nabla f(u)(v - u) + \frac{o(\lambda\|u - v\|)}{\lambda} \end{aligned}$$

となる. $\lambda \rightarrow 0$ とすると, $\frac{o(\lambda\|u - v\|)}{\lambda} \rightarrow 0$ となるので, 命題の不等式は示された.

□

1.1 凸関数の性質

1 変数関数 h に対しては, 以下のような性質が成り立つ.

h が凸関数

\Leftrightarrow すべての $t \in R$ で, $h'(t)$ が非減少 ($t_1 < t_2 \Rightarrow h'(t_1) \leq h'(t_2)$)

\Leftrightarrow すべての $t \in R$ で, $h''(t) \geq 0$

関数 $J(x) = x^2$ は実際に $J'(x) = 2x$, $J''(x) = 2$ なので, 上記の性質を満たすことがわかる.

このような性質は多変数の場合でも成り立つ. 一階微分は多変数の場合, 勾配ベクトルが対応するが, 二階微分はヘッセ行列が対応する.

定義. 連続 2 階微分可能な関数 $f: \mathbb{R}^n \rightarrow \mathbb{R}$ と $u \in \mathbb{R}^n$ に対して,

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

を**ヘッセ行列**と呼ぶ.

2 変数関数 $f: \mathbb{R}^2 \rightarrow \mathbb{R}$ の場合は,

$$\nabla^2 f(u) = \begin{bmatrix} f_{xx}(u) & f_{xy}(u) \\ f_{yx}(u) & f_{yy}(u) \end{bmatrix}$$

となる.

ここで, 行列の言葉を用意する.

定義. A を実数成分を持つ対称行列とする.

- すべての $v \in \mathbb{R}^n$ に対して, $v^T A v \geq 0$ が成り立つとき, A を**半正定値行列**と呼ぶ. また, 逆の不等号が成り立つとき, A を**半負定値行列**と呼ぶ.
- $v \neq 0$ を満たすすべての $v \in \mathbb{R}^2$ に対して, $v^T A v > 0$ が成り立つとき, A を**正定値行列**と呼ぶ. 逆の不等号が成り立つとき, A を**負定値行列**と呼ぶ.
- $v \in \mathbb{R}^2$ によって $v^T A v$ の符号が正にも負にもなるとき, A を**不定値行列**と呼ぶ.

例 2. $A = \begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix}$ のとする. 任意の $v = (v_1, v_2) \in \mathbb{R}^2$ に対して, $v^T A v = 2v_1^2 + v_2^2$ となるので A は正定値である. 一方, $B = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$ とすると, $v^T B v = v_1^2 + 2v_1 v_2 - v_2^2$ となるので, 不定値である.

命題 2. 連続微分可能な関数 $f: \mathbb{R}^n \rightarrow \mathbb{R}$ に対して以下が成り立つ;

- f が凸関数 \iff 任意の $u \in \mathbb{R}^n$ に対して, ヘッセ行列 $\nabla^2 f(u)$ が半正定値である.
- f が狭義凸関数 \iff 任意の $u \in \mathbb{R}^n$ に対して, ヘッセ行列 $\nabla^2 f(u)$ が正定値である.

解説. $a, b \in \mathbb{R}^n$ に対して, $h(\lambda) = f((1-\lambda)a + \lambda b)$ とおく. すると 「 f が凸 \iff 任意の $a, b \in \mathbb{R}^n$ に対して h が凸」 が成り立つ. また, h は一変数関数なので, 「 h が凸 $\iff h''(\lambda) \geq 0$ 」 が成り立つ. ここで, $u = (1-\lambda)a + \lambda b$ と $v = b - a$ に対して, $h''(\lambda) = v^T \nabla^2 f(u) v$ となる. これより命題が導かれる. \square

1.2 正值性の調べ方

さて、関数のヘッセ行列が任意の点で半正定値であれば、その関数は凸になることが分かった。それでは、行列が半正定値であるとはどのように調べたらよいだろうか？

1.2.1 固有値を用いた判定法

それには行列の理論による以下のような定理が有用である。ある数 a が非負であるとは、 $a \geq 0$ であることを表し、ある数が正であるとは $a > 0$ であることを表す。

定理 3. A を実対称行列とする。以下の主張が成り立つ：

- (1). A が半正定値 (半負定値) $\Leftrightarrow A$ の固有値がすべて非負 (非正)
- (2). A が正定値 (負定値) $\Leftrightarrow A$ の固有値がすべて正 (負)
- (3). A が不定値 $\Leftrightarrow A$ が正と負の固有値を持つ

解説. A は実対称行列なので、ある直交行列 P ($P^T P = E$) が存在して、 $P^{-1} A P = \Lambda$ と対角化できる。ここで、 Λ は A の固有値を対角要素に持つ対角行列である。すると、 $P^{-1} = P^T$ なので、 $u = P^T v$ と変数変換すると、

$$v^T A v = (P u)^T A (P u) = u^T (P^T A P) u = u^T \Lambda u = \lambda_1 u_1^2 + \cdots + \lambda_n u_n^2$$

となる。ここで、 $u = (u_1, \dots, u_n)$ 、 λ_i は A の固有値である。これより上の主張が導かれる。 \square

例 3. $f(x, y) = 3x^2 - 2xy + 3y^2$ とする。 $\nabla f(x, y) = (6x + 2y, 2x + 6y)$ 、 $\nabla^2 f(x, y) = \begin{bmatrix} 6 & 2 \\ 2 & 6 \end{bmatrix}$ となる。ヘッセ行列は定数行列であり、この固有値は 4, 8 になる。よって、ヘッセ行列は正定値で、 f は狭義凸関数である。

1.2.2 行列式を用いた判定法

2×2 行列であれば、以下のように行列式を用いて判定できることがわかる。

実対称行列を $A = \begin{bmatrix} a & b \\ b & d \end{bmatrix}$ とし、 $a \neq 0$ とする。任意の $(x, y) \in \mathbb{R}^2$ に対して、

$$\begin{bmatrix} x & y \end{bmatrix} A \begin{bmatrix} x \\ y \end{bmatrix} = ax^2 + 2bxy + dy^2 = a \left(x + \frac{by}{a} \right)^2 + \frac{y^2}{a} (ad - b^2) = a \left(x + \frac{by}{a} \right)^2 + \frac{y^2}{a} \begin{vmatrix} a & b \\ b & d \end{vmatrix}$$

となる ($\begin{vmatrix} a & b \\ b & d \end{vmatrix}$ は A の行列式 $|A|$ を表す)。この式より、

命題 4.

(1). $a > 0, |A| > 0 \Rightarrow A$ は正定値

(2). $a < 0, |A| > 0 \Rightarrow A$ は負定値

(3). $|A| < 0 \Rightarrow A$ は不定値

解説. (1), (2) は命題の前の式から証明できる. (3) については, 線形代数で学んだように, 行列式の値は固有値の積と等しい (λ_1, λ_2 を固有値とすると, $|A| = \lambda_1 \lambda_2$). よって, 2×2 行列の場合, 行列式が負ということは, 二つある固有値の符号が異なるということである. \square

例 4. 例 3 のヘッセ行列 $\nabla^2 f(x, y) = \begin{bmatrix} 6 & 2 \\ 2 & 6 \end{bmatrix}$ の正值性を行列式を用いて調べる. $f_{xx}(x, y) = 6 > 0, |\nabla^2 f(x, y)| = 32 > 0$ なのでヘッセ行列は正定値である.