2 凸関数

2.1 凸関数の性質

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 2.1 (凸関数)**.** $0 < \lambda < 1$ を満たすすべての λ とすべての $u, v \in \mathbb{R}^n$ 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[1)] . 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.

1) 高校数学では下に凸な関数と呼ぶが, 大学では単に凸関数と呼ぶ.

例 2.1. 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,

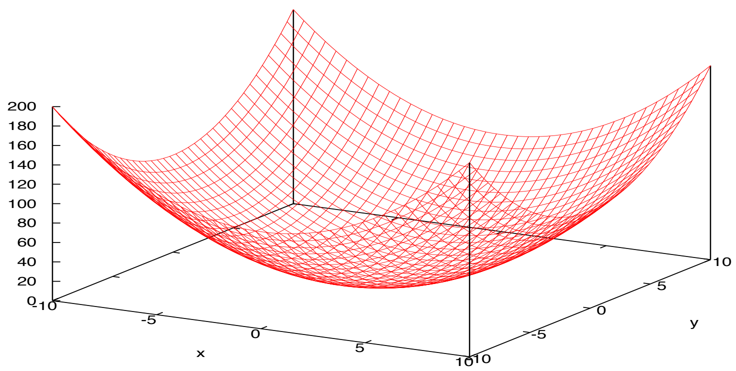
$$f(u) + f_x(u)(x_2 - x_1) + f_y(u)(y_2 - y_1)$$

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 $f\colon R^2 \to R$ が凸ならば, すべての $u = (x_1, y_1), v = (x_2, y_2) \in R^2$ 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 2.2. 連続微分可能な関数 $f\colon \mathbb{R}^2 \to \mathbb{R}$ と $u = (x_1, \ldots, x_n)$ に対して,

$$\nabla f(u) = (f_x(u), f_y(u))$$

を f の u における勾配ベクトルと呼ぶ. 一般に, n 変数関数 f に関しては勾配ベクトルは, 以下のように定義される;

$$\nabla f(u) = (f_{x_1}(u), f_{x_2}(u), \ldots, f_{x_n}(u)).$$

例 2.2. $f(x,y) = x^2 + 3y^2$ のとき, (a,b) における勾配ベクトルは, $f_x = 2x$, $f_y = 6y$ より,

$$\nabla f(a,b) = (f_x(a,b), f_y(a,b)) = (2a, 6b)$$

となる.

命題 2.1. $f\colon \mathbb{R}^n \to \mathbb{R}$ が凸ならば, すべての u, v に対して,

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

が成り立つ. ここで, $f(u) \cdot (v - u)$ はベクトル $f(u)$ 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,

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

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,

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

$$\lambda f(v) \geq \lambda f(u) + \lambda \nabla f(u) \cdot (u - v) + o(\lambda \|u - v\|)$$
$$f(v) \geq f(u) + \nabla f(u) \cdot (u - v) + \frac{o(\lambda \|u - v\|)}{\lambda}$$

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

例 2.3. 例 *2.2* の関数 $f(x,y) = x^2 + 3y^2$ は凸関数である *(*後で *2.4* において示す*)*. 命題 2.1 より, $u = (x_1, y_1), v = (x_2, y_2)$ とすると,

$$\begin{aligned} x_2^2 + 3y_2^2 = f(x_2, y_2) \\ &\geq f(x_1, y_1) + \nabla f(x_1, y_1) \cdot (x_2 - x_1, y_2 - y_1) \\ &= x_1^2 + 3y_1^2 + (2x_1, 6y_1) \cdot (x_2 - x_1, y_2 - y_1) \end{aligned}$$

なので, すべての x_1, y_1, x_2, y_2 に対して, 不等式

$$x_2^2 + 3y_2^2 \geq x_1^2 + 3y_1^2 + 2x_1(x_2 - x_1) + 6y_1(y_2 - y_1)$$

が成り立つ.

2.1.1 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 $f(x) = x^2$ は凸関数だが, 実際に $f'(x) = 2x$, $f''(x) = 2$ 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. 多変数の場合, 一階微分は勾配ベクトルが対応し, 二階微分は以下のヘッセ行列が対応する;

定義 2.3. 2 変数関数 $f\colon \mathbb{R}^2 \to \mathbb{R}$ と $u \in \mathbb{R}^n$ に対して,

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

をヘッセ行列と呼ぶ.

例えば, $f(x,y) = x^3 + 2xy + 3y^2$ とすると, 勾配ベクトルは $\nabla f(x,y) = (3x^2 + 2y, 2x + 6y)$ で, ヘッセ行列は

$$\nabla^2 f(x,y) = \begin{bmatrix} 6x & 2 \\ 2 & 6 \end{bmatrix}$$

となる．一般の n 変数関数に対しては，ヘッセ行列は

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

と定義される．ヘッセ行列を用いると以下のような凸性の判定方法がある．

定理 2.1. *連続微分可能な関数 $f\colon \mathbb{R}^n \to \mathbb{R}$ に対して以下が成り立つ;*

- f が凸関数 $\Leftrightarrow$ すべての $u \in \mathbb{R}^n$ で，

$$v^T \nabla^2 f(u) v \geq 0 \ (\text{すべての } v \in \mathbb{R}^n)$$

  が成り立つ[2)]．

- f が狭義凸関数 $\Longleftarrow$ すべての $u \in \mathbb{R}^n$ で，

$$v^T \nabla^2 f(u) v > 0 \ (\text{すべての } v \in \mathbb{R}^n)$$

  が成り立つ．

2) 一階微分を用いた場合は，f が凸関数 $\Longleftrightarrow$ すべての $u, v \in \mathbb{R}^n$ に対して，$\langle \nabla f(u) - \nabla f(v), u - v \rangle \geq 0$.

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 $\Leftrightarrow$ すべての $a, b \in \mathbb{R}^n$ 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 $\Leftrightarrow$ $h''(\lambda) \geq 0$」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． □

例 2.4. $f(x,y) = x^2 + 3y^2$ とすると，勾配ベクトルは $\nabla f(x,y) = (2x, 6y)$ で，ヘッセ行列は

$$\nabla^2 f(x,y) = \begin{bmatrix} 2 & 0 \\ 0 & 6 \end{bmatrix}$$

となる．一般にヘッセ行列は成分に変数 x, y を持つが，この関数では変数を持たない行列となる．つまりすべての点で一つの決まった行列になっている．定理 *2.1* を用いて f が凸関数かどうか調べよう．ベクトル $v = (v_1, v_2)$ とする．いま，すべての点 (x,y) で，

$$v^T \nabla f(x,y) v = \begin{bmatrix} v_1 & v_2 \end{bmatrix} \begin{bmatrix} 2 & 0 \\ 0 & 6 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = 2v_1^2 + 6v_2^2$$

となる．また，すべての v_1, v_2 に対して，$2v_1^2 + 6v_2^2 \geq 0$ となるので，定理 *2.1* より f は凸関数である．

ここで, 行列の言葉を用意する.

定義 2.4. 行列 A とベクトル $v = (v_1, v_2, \ldots, v_n)$ に対して, $v_1, \ldots, v_n$ を変数に持ち, 値を実数にとる関数

$$g(v) = v^T A v$$

を 2 次形式と呼ぶ.

例 2.5. $A = \left[\begin{smallmatrix} 2 & 0 \\ 0 & 1 \end{smallmatrix}\right]$ のとすると,

$$v^T A v = \begin{bmatrix} v_1 & v_2 \end{bmatrix} \begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = 2v_1^2 + v_2^2$$

は 2 次形式である. 一方, $B = \left[\begin{smallmatrix} 1 & 1 \\ 1 & -1 \end{smallmatrix}\right]$ とすると,

$$v^T B v = \begin{bmatrix} v_1 & v_2 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = v_1^2 + 2v_1 v_2 - v_2^2$$

は 2 次形式である.

定義 2.5. 行列 A を実数の成分を持つ対称行列とする.

- すべての $v \in R^n$ 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 $v \in R^2$ 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.
- $v \in R^2$ によって $v^T A v$ の符号が正にも負にもなるとき, A を不定と言う.

例 2.6. 例 *2.5* の行列の正値性を調べてみよう. $A = \left[\begin{smallmatrix} 2 & 0 \\ 0 & 1 \end{smallmatrix}\right]$ に対して, 二次形式は $v^T A v = 2v_1^2 + v_2^2$ となるので A は正定値である. 一方, $B = \left[\begin{smallmatrix} 1 & 1 \\ 1 & -1 \end{smallmatrix}\right]$ に対して, $v^T B v = v_1^2 + 2v_1 v_2 - v_2^2$ となる. $v = (1, 0)$ とすると, $v^T B v = 1 > 0$, $v = (0, 1)$ とすると $v^T B v = -1 < 0$ なので, 不定である.

これらの言葉を用いると, 定理 *2.1* は以下のように書ける.

定理 2.2 (定理 2.1 の言い換え)**.** 連続微分可能な関数 $f \colon \mathbb{R}^n \to \mathbb{R}$ に対して以下が成り立つ;

- f が凸関数 $\Longleftrightarrow$ すべての $u \in \mathbb{R}^n$ 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 $\Longleftarrow$ すべての $u \in \mathbb{R}^n$ に対して, ヘッセ行列 $\nabla^2 f(u)$ が正定値である.

2.1.2 正値性の調べ方

さて, 関数のヘッセ行列がすべて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?

2.1.2.1 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.

定理 2.3. *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 0 以上 (0 以下)
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 $\Leftrightarrow$ 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 $(P^TP = E)$ が存在して, $P^{-1}AP = \Lambda$ 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, $u = P^Tv$ と変数変換すると,

$$v^TAv = (Pu)^TA(Pu) = u^T(P^TAP)u = u^T\Lambda u = \lambda_1 u_1^2 + \cdots + \lambda_n u_n^2$$

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

例 2.7. $f(x,y) = 3x^2 - 2xy + 3y^2$ とする. $\nabla f(x,y) = (6x+2y, 2x+6y)$, $\nabla^2 f(x,y) = \left[\begin{smallmatrix} 6 & 2 \\ 2 & 6 \end{smallmatrix}\right]$ 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.

2.1.2.2 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. すべての $(x,y) \in R^2$ に対して,

$$\begin{bmatrix} x & y \end{bmatrix} A \begin{bmatrix} x \\ y \end{bmatrix} = ax^2 + 2bxy + dy^2 = a\left(x + \frac{by}{a}\right)^2 + \frac{y^2}{a}(ad - b^2) = a\left(x + \frac{by}{a}\right)^2 + \frac{1}{a}\begin{vmatrix} a & b \\ b & d \end{vmatrix} y^2$$

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 2.2.

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

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. □

例 2.8. 例 *2.7* のヘッセ行列 $\nabla^2 f(x, y) = \left[\begin{smallmatrix} 6 & 2 \\ 2 & 6 \end{smallmatrix}\right]$ の正値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.