連続最適化で使う数学

山下信雄

数理最適化と数学

- 高校の授業では，数学Ⅰ，Ⅲの一部として「数理最適化」を学んでいる.
- 数理最適化では，数学と計算機の力を使って，最適解を与える。
- 数理最適化を活用するためには，数学と計算機の基礎的な知識が必要.
- 本授業で必要とする数学：

  線形代数，微分

記号

- 列 $\{x^1, x^2, \cdots\} \subseteq R^n$ を $\{x^k\}$ で表す。

- ノルム：

$$\|x\| = \sqrt{x_1^2 + \cdots + x_n^2}$$

微分

関数 $f : R^n \to R$ の x_i の微分を偏微分という.

$$\frac{\partial f(x)}{\partial x_i} = \lim_{h \to 0} \frac{f(x + he_i) - f(x)}{h}$$

ただし, $e_i = (0, \cdots, \underset{i}{1}, \cdots, 0)^\top$ である

２回偏微分：

$$\frac{\partial^2 f(x)}{\partial x_i \partial x_j} = \frac{\partial}{\partial x_i}\left(\frac{\partial f(x)}{\partial x_j}\right)$$

多変数関数の微分（勾配）

次のベクトルを $f : R^n \to R$ の勾配という。

$$\nabla f(x) = \begin{pmatrix} \dfrac{\partial f(x)}{\partial x_1} \\ \dfrac{\partial f(x)}{\partial x_2} \\ \vdots \\ \dfrac{\partial f(x)}{\partial x_n} \end{pmatrix}$$

勾配の例

1次関数 $$f(x) = c^\top x = \sum_{i=1}^{n} c_i x_i$$

$$\nabla f(x) = c$$

２次関数 $$f(x) = x^\top Q x = \sum_{j=1}^{n} \sum_{i=1}^{n} Q_{ij} x_i x_j$$

$$\nabla f(x) = \left(Q + Q^\top\right) x$$

多変数関数の微分(ヘッセ行列)

次の行列を関数 f のヘッセ行列という。

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

ヘッセ行列の例

1次関数 $f(x) = c^\top x = \sum_{i=1}^{n} c_i x_i$

$$\nabla^2 f(x) = 0$$

２次関数 $f(x) = x^\top Q x = \sum_{j=1}^{n} \sum_{i=1}^{n} Q_{ij} x_i x_j$

$$\nabla^2 f(x) = Q + Q^\top$$

線形代数(半正定値行列)

半正定値行列：すべての $v \in R^n$ にたいして

$$v^\top A v \geq 0$$

が成り立つ行列 A

正定値行列：すべての0でない $\boldsymbol{v} \in \boldsymbol{R}^n$ にたいして

$$v^\top A v > 0$$

が成り立つ行列 A

(半)正定値行列の性質

- A が半正定値行列であれば， $A^\top$ も半正定値行列である.
- A が正定値行列であれば， A^{-1} も正定値行列である.
- 正定値対称行列の固有値はすべて正である.
- 半正定値対称行列の固有値はすべて非負の実数である.

テーラー展開(1)

p を $R \to R$ の関数とする.

１次のテーラー展開:

$$p(t) = p(0) + p'(0)t + o(t)$$

２次のテーラー展開:

$$p(t) = p(0) + p'(0)t + \frac{1}{2}p''(0)t^2 + o(t^2)$$

テーラー展開（2）

$f:R^n \to R,\ d \in R^n$ とし，$p(t) = f(x+td)$ とする．

このとき, 合成関数の微分を考えれば,

$$p(0) = f(x),\ p'(0) = \nabla f(x)^\top d,\ p''(0) = d^\top \nabla^2 f(x) d$$

を得る．

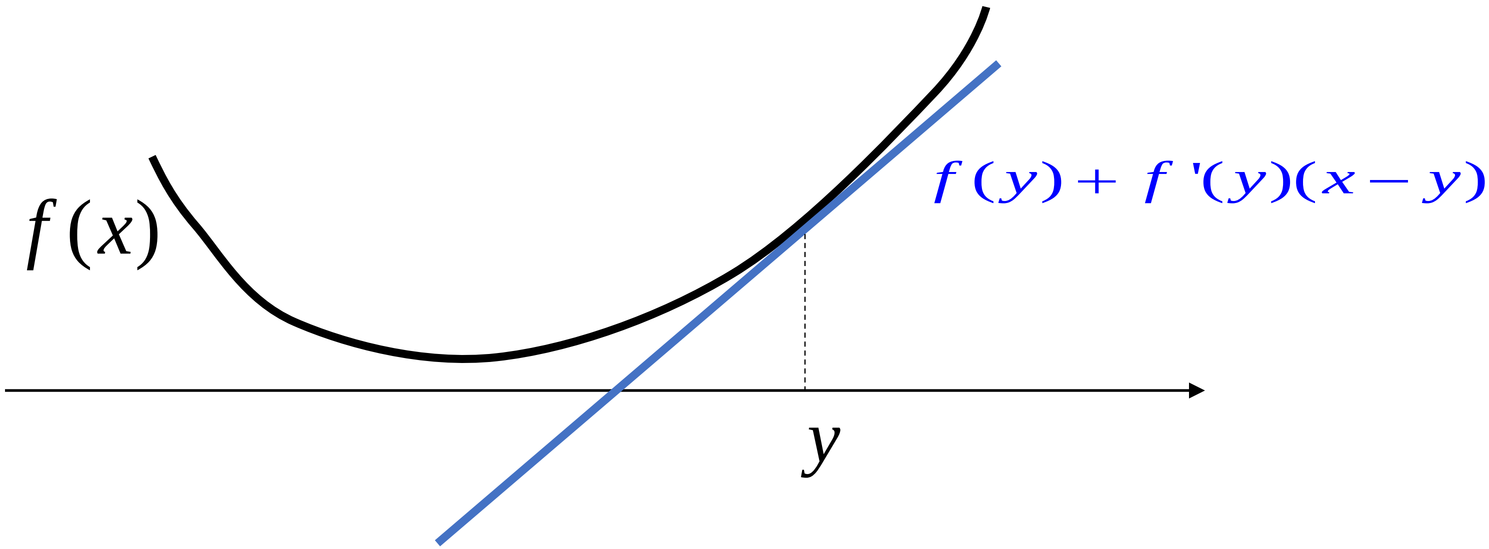
$$f(x+td) = f(x) + t\nabla f(x)^\top d + o(t)$$

$$f(x+td) = f(x) + t\nabla f(x)^\top d + \frac{t^2}{2} d^\top \nabla^2 f(x) d + o(t^2)$$

凸関数の必要十分条件①
微分可能なとき

定理　f は微分可能であるとする．このとき f が凸関数であるための必要十分条件は次の不等式が成り立つことである．

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



定理の証明

（必要性）f を凸関数とする．さらに

$$p(\alpha) = f(\alpha x + (1-\alpha)y) = f(y + \alpha(x-y))$$

とする．このとき，$p(0) = f(y),\ p(1) = f(x)$ であり，

$$p(\alpha) = f(\alpha x + (1-\alpha)y) \le \alpha f(x) + (1-\alpha) f(y) = \alpha p(1) + (1-\alpha) p(0)$$

式を整理すると，

$$p(1) - p(0) \ge \frac{p(\alpha) - p(0)}{\alpha}$$

であり，$\alpha \to 0$ とすると，$p(1) - p(0) \ge p'(0)$

合成関数の微分から $p'(0) = \nabla f(y)^\top (x-y)$

よって，

$$f(x) \ge f(y) + \nabla f(y)^\top (x-y)$$

定理の証明

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

（十分性）任意の$\bar{x}, \bar{y} \in R^n, \alpha \in [0,1]$ に対して,

$$f(\alpha\bar{x} + (1-\alpha)\bar{y}) \leq \alpha f(\bar{x}) + (1-\alpha) f(\bar{y})$$

が成り立つことを示す.

$x = \bar{y},\ y = \alpha\bar{x} + (1-\alpha)\bar{y}$ を代入すると

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

$x = \bar{x},\ y = \alpha\bar{x} + (1-\alpha)\bar{y}$ を代入すると

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

始めの不等式を$1-\alpha$ 倍し，次の不等式をα倍して

辺々を足すと,

$$
\begin{aligned}
&\alpha f(\bar{x}) + (1-\alpha) f(\bar{y}) \\
&\quad \geq \qquad \alpha f(\alpha\bar{x} + (1-\alpha)\bar{y}) + \alpha(1-\alpha)\nabla f(\alpha\bar{x} + (1-\alpha)\bar{y})^\top(\bar{x} - \bar{y}) \\
&\qquad + (1-\alpha) f(\alpha\bar{x} + (1-\alpha)\bar{y}) + \alpha(1-\alpha)\nabla f(\alpha\bar{x} + (1-\alpha)\bar{y})^\top(\bar{y} - \bar{x}) \\
&\quad = f(\alpha\bar{x} + (1-\alpha)\bar{y})
\end{aligned}
$$

凸関数の必要十分条件②
2回微分可能なとき

定理　f は2回微分可能であるとする.
このとき f が凸関数であるための必要十分条件は
すべての $x \in R^n$ に対してヘッセ行列 $\nabla^2 f(x)$ が
半正定値行列になることである.

高校の数学Ⅲで,

$$f''(x) > 0 \ \forall x \in R \ \Rightarrow \quad \text{下に凸}$$

としていたものの一般化

証明

(必要性) 対偶で示す.

ある $x \in R^n$ で $\nabla^2 f(x)$ は半正定値行列でないとする.

このとき, $d^\top \nabla f(x) d < 0$ となる $d \in R^n$ が存在する.

「テーラー展開(2)」のスライドより

$$f(x+td) = f(x) + t\nabla f(x)^\top d + \frac{t^2}{2} d^\top \nabla^2 f(x) d + o(t^2)$$

よって, t が十分小さい正の数のとき

$$\frac{f(x+td) - f(x) - t\nabla f(x)^\top d}{t^2} = \frac{1}{2} d^\top \nabla^2 f(x) d + \frac{o(t^2)}{t^2} < 0$$

つまり

$$f(x+td)-f(x)-t\nabla f(x)^{\top}d<0$$

を得る．

いま， $y=x+td$ とすると，$td=y-x$

$$f(y)-f(x)-\nabla f(x)^{\top}(y-x)<0$$

つまり

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

よって， f は凸関数でない．

(十分性)

テーラーの定理より，任意の $x, y \in R^n$ に対して

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

が成り立つ$t \in [0,1]$ が存在する．

仮定より， $\nabla^2 f(x+t(y-x))$ は半正定値行列であるから，

$$\frac{1}{2}(y-x)^\top \nabla^2 f(x+t(y-x))(y-x) \geq 0$$

よって，

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

を得る．これは f が凸関数であることを示している．