最適化 復習

Kosuke Toda

目次

第1章

まえがき

様々な問題に対して，最も効率的になるように意思決定をする手法としてオペレーションズ・リサーチ (OR) というものがある [1]．これは，現実の問題を数理モデルに置換し，問題を解決する．最適化理論 (optimization theory) は，OR の基礎理論の 1 つであり，理論だけでなく，現実の問題を解くための手法を提供してきた [2]．本資料では，特に連続最適化に焦点を当てる．

第 2 章

数学的準備

2.1　諸定義

n 次元実数空間 $\mathbb{R}^n$ を考える．

- ε-近傍
  - $B(x, \varepsilon) = \{\boldsymbol{y} \in \mathbb{R}^n \mid \|\boldsymbol{y} - \boldsymbol{x}\| < \varepsilon\}$
  - ただし，$\|\cdot\|$ はユークリッドノルムであり，$\|\boldsymbol{x}\| = (\boldsymbol{x}^{\mathrm{T}}\boldsymbol{x})^{1/2}$ である．
- $X \subseteq \mathbb{R}^n$ が開集合 (open set) である
  - $\forall \boldsymbol{x} \in X$, $\exists \varepsilon > 0$; $B(\boldsymbol{x}, \varepsilon) \subseteq X$
- $X \subseteq \mathbb{R}^n$ が閉集合 (closed set) である
  - X の補集合 X^{c} が開集合である．
- $\boldsymbol{x}$ が $X \subseteq \mathbb{R}^n$ の内点 (interior point) である
  - $\boldsymbol{x} \in X \subseteq \mathbb{R}^n$ に対して，$\exists \varepsilon > 0$; $B(\boldsymbol{x}, \varepsilon) \subseteq X$ が成立する．
- X の内部 (interior)，$\mathrm{Int}(X)$
  - $X \subseteq \mathbb{R}^n$ の内点の集合．
- $\boldsymbol{x}$ が $X \subseteq \mathbb{R}^n$ の触点 (contact point) である
  - $\boldsymbol{x} \in X \subseteq \mathbb{R}^n$ に対して，$\forall \varepsilon > 0$; $B(\boldsymbol{x}, \varepsilon) \cap X \neq \emptyset$ が成立する．
- X の閉包 (closure)，$\mathrm{Cl}(X)$
  - $X \subseteq \mathbb{R}^n$ の触点の集合．
  - X を含む最小の閉集合．
- X の境界 (boundary)，$\mathrm{Bd}(X)$
  - $\mathrm{Bd}(X) = \mathrm{Cl}(X) \setminus \mathrm{Int}(X)$.
- $\boldsymbol{x}$ は X の集積点である
  - 任意の $\varepsilon > 0$ に対して，$B(\boldsymbol{x}, \varepsilon) \cap X$ が $\boldsymbol{x}$ と異なる要素を含む．
- 孤立点 (isolated point)
  - X の集積点でない X の触点

- X が有界である (bounded)
  - $\exists \varepsilon > 0$; $X \subseteq B(\mathbf{0}, \varepsilon)$.
- 収束
  - 点列 $\{\boldsymbol{x}^i\}$, $i = 1, 2, \ldots$ を考える．$\forall \varepsilon > 0$, $\exists I_\varepsilon$; $\|\boldsymbol{x}^i - \boldsymbol{x}\| < \varepsilon$, $i \geq I_\varepsilon$ となる点 $\boldsymbol{x}$ が存在するとき，$\boldsymbol{x}$ は点列 $\{\boldsymbol{x}^i\}$ の極限 (limit) といい，点列 $\{\boldsymbol{x}^i\}$ は $\boldsymbol{x}$ に収束する (converge) という．
- 点列 $\{\boldsymbol{x}^i\}$ の集積点
  - $\{\boldsymbol{x}^i\}$ の部分点列 $\{\boldsymbol{x}^{l_i}\}$ が点 $\boldsymbol{x}$ に収束するとき，点 $\boldsymbol{x}$ を点列 $\{\boldsymbol{x}^i\}$ の集積点という．

有界閉集合 $X \subseteq \mathbb{R}^n$ について，$\{\boldsymbol{x}^i\} \subseteq X$ なる無限点列は少なくとも 1 つの集積点をもつ．

2.2 関数

- 連続性
  - 関数 $\boldsymbol{f} : X \to \mathbb{R}^m$ を考える ($X \subseteq R^n$).

$$\forall \epsilon > 0, \exists \delta > 0;\ \|\boldsymbol{x} - \boldsymbol{x}\| < \delta \Rightarrow \|\boldsymbol{f}(\boldsymbol{x}) - \boldsymbol{f}(\boldsymbol{x}^0)\| < \varepsilon$$

    が成立するとき，$\boldsymbol{f}$ は点 $\boldsymbol{x}^0$ で連続 (continuous) であるという．
  - 任意の $\boldsymbol{x} \in X$ で連続となるとき，関数 $\boldsymbol{f}$ は X で連続という．
- 実数値関数
  - 値域が実数集合の関数 $f : X \to \mathbb{R}$.
- 実数値関数のクラス
  - $f : X \to \mathbb{R}(X \subseteq \mathbb{R}^n$ は開集合) を考える．
  - f が連続であるとき，X 上で C^0 級と呼ばれ，$f \in C^0$ と記す．
  - $f \in C^0$ で，$\partial f(\boldsymbol{x})/\partial x_i$, $i = 1, 2, \ldots, n$ が存在し，連続であれば，f は X 上で C^1 級と呼ばれ，$f \in C^1$ と表す．
  - $f \in C^1$ で，$\partial^2 f(\boldsymbol{x})/\partial x_i \partial x_j$, $i, j = 1, 2, \ldots, n$ が存在し，連続であれば，f は X 上で C^2 級と呼ばれ，$f \in C^2$ と表す．
  - 以後 C^k 級も同様に定義される．
- 勾配ベクトル $\nabla f(\boldsymbol{x})$

$$\nabla f(\boldsymbol{x}) = \left(\frac{\partial f(\boldsymbol{x})}{\partial x_1}, \frac{\partial f(\boldsymbol{x})}{\partial x_2}, \cdots, \frac{\partial f(\boldsymbol{x})}{\partial x_n} \right)$$

- ヘッセ行列 $H(\boldsymbol{x}) = \nabla^2 f(\boldsymbol{x})$

$$\nabla^2 f(\boldsymbol{x}) = \begin{pmatrix} \dfrac{\partial^2 f(\boldsymbol{x})}{\partial {x_1}^2} & \dfrac{\partial^2 f(\boldsymbol{x})}{\partial x_1 \partial x_2} & \cdots & \dfrac{\partial^2 f(\boldsymbol{x})}{\partial x_1 \partial x_n} \\ \dfrac{\partial^2 f(\boldsymbol{x})}{\partial x_2 \partial x_1} & \dfrac{\partial^2 f(\boldsymbol{x})}{\partial {x_2}^2} & \cdots & \dfrac{\partial^2 f(\boldsymbol{x})}{\partial x_2 \partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \dfrac{\partial^2 f(\boldsymbol{x})}{\partial x_n \partial x_1} & \dfrac{\partial^2 f(\boldsymbol{x})}{\partial x_n \partial x_2} & \cdots & \dfrac{\partial^2 f(\boldsymbol{x})}{\partial {x_n}^2} \end{pmatrix}$$

■Weierstrass の定理　有界閉集合 $X \subseteq \mathbb{R}^n$ 上の連続な実数値関数 $f(\boldsymbol{x})$ は X 内の点で最大値，最小値をとる．

■平均値の定理　$f : X \to \mathbb{R}\,(X \subseteq \mathbb{R}^n)$, $f \in C^1$, $\boldsymbol{x}^1, \boldsymbol{x}^2 \in X$ に対して，

$$f(\boldsymbol{x}^1) = f(\boldsymbol{x}^2) + \nabla f(\theta \boldsymbol{x}^1 + (1-\theta)\boldsymbol{x}^2)^{\mathrm{T}}(\boldsymbol{x}^1 - \boldsymbol{x}^2)$$

を満たす $\theta \in (0,1)$ が存在する．

(注) 関数 $f : I \to \mathbb{R}\,(I \subseteq \mathbb{R})$, $f \in C^1$, $x_1, x_2 \in I$ に対しては，

$$f(x_1) = f(x_2) + f'(\theta x_1 + (1-\theta)x_2)(x_1 - x_2)$$

を満たす $\theta \in (0,1)$ が存在する．これは，点 $x_1, x_2 \in I$ を結ぶ線分内に，$f(x_1), f(x_2)$ を結ぶ直線と同じ傾きとなる接線を持つ点 $c = \theta x_1 + (1-\theta)x_2$, $\theta \in (0,1)$ が存在することに相当する．

■Taylor の定理　$f : X \to \mathbb{R}\,(X \subseteq \mathbb{R}^n)$, $f \in C^2$, $\boldsymbol{x}^1, \boldsymbol{x}^2 \in X$ に対して，

$$f(\boldsymbol{x}^1) = f(\boldsymbol{x}^2) + \nabla f(\boldsymbol{x}^2)^{\mathrm{T}}(\boldsymbol{x}^1 - \boldsymbol{x}^2)^{\mathrm{T}} + \frac{1}{2}(\boldsymbol{x}^1 - \boldsymbol{x}^2)^{\mathrm{T}}\nabla^2 f(\boldsymbol{x}^2)(\boldsymbol{x}^1 - \boldsymbol{x}^2) + o(\|\boldsymbol{x}^1 - \boldsymbol{x}^2\|^2)$$

が成立する．ただし，o は $\lim_{t\to 0} o(t)/t = 0$ なる関数である．同様に，$f \in C^1$ に対して，

$$f(\boldsymbol{x}^1) = f(\boldsymbol{x}^2) + \nabla f(\boldsymbol{x}^2)^{\mathrm{T}}(\boldsymbol{x}^1 - \boldsymbol{x}^2)^{\mathrm{T}} + o(\|\boldsymbol{x}^1 - \boldsymbol{x}^2\|^1)$$

が成立する．

(注) 関数 $f : I \to \mathbb{R}\,(I \subseteq \mathbb{R})$, $f \in C^1$, $x_1, x_2 \in I$ に対しては，

$$f(x_1) = f(x_2) + f'(x_2)(x_1 - x_2) + \frac{f''(x_2)}{2}(x_1 - x_2)^2 + o((x_1 - x_2)^2)$$
$$f(x_1) = f(x_2) + f'(x_2)(x_1 - x_2) + o(x_1 - x_2)$$

が成立する．o を用いずに表すと，

$$f(x_1) = f(x_2) + f'(x_2)(x_1 - x_2) + \frac{f''(x_2)}{2}(x_1 - x_2)^2 + \cdots + \frac{f^{(n-1)}(x_2)}{(n-1)!}(x_1 - x_2)^{n-1} + R_n$$
$$R_n \coloneqq \frac{f^{(n)}(\theta x_1 + (1-\theta)x_2)}{n!}(x_1 - x_2)^n$$

を満たす $\theta \in (0,1)$ が存在する．$n = 1$ のときは平均値の定理．つまり，Taylor の定理は平均値の定理の一般化と考えることができる．

■陰関数の定理　$h_i : \mathbb{R}^n \to \mathbb{R}$, $\boldsymbol{x}^0 = (x_1^0, x_1^0, \ldots, x_n^0)^{\mathrm{T}} \in \mathbb{R}^n$ の近傍で，$h_i \in C^p\,(p \geq 1)$

$$h_i(\boldsymbol{x}^0) = 0,\ i = 1, 2, \ldots, m$$

が成立するとする．このとき，ヤコビ行列 (Jacobian matrix)

$$J(\boldsymbol{x}^0) = \begin{pmatrix} \dfrac{\partial h_1(\boldsymbol{x}^0)}{\partial x_1} & \cdots & \dfrac{\partial h_1(\boldsymbol{x}^0)}{\partial x_m} \\ \vdots & \vdots & \vdots \\ \dfrac{\partial h_m(\boldsymbol{x}^0)}{\partial x_1} & \cdots & \dfrac{\partial h_m(\boldsymbol{x}^0)}{\partial x_m} \end{pmatrix}$$

が正則ならば，ある $\varepsilon > 0$ に対して $\hat{\boldsymbol{x}}^0 = (x^0_{m+1}, \ldots, x^0_n)^{\mathrm{T}} \in \mathbb{R}^{n-m}$ の近傍 $U = B(\hat{\boldsymbol{x}}^0, \varepsilon)$ が存在し，$\hat{\boldsymbol{x}} \in U$ に対して，

1. $\phi_i \in C^p$, $i = 1, 2, \ldots, m$
2. $x_i = \phi(\hat{\boldsymbol{x}})$, $i = 1, 2, \ldots, m$
3. $h_i(\phi_1(\hat{\boldsymbol{x}}), \phi_2(\hat{\boldsymbol{x}}), \ldots, \phi_m(\hat{\boldsymbol{x}}), \hat{\boldsymbol{x}}) = 0, i = 1, 2, \ldots, m$

となる陰関数 (implicit function)ϕ_i, $i = 1, 2, \ldots, m$ が存在する[*1]．

(注) 2 変数関数を考える．関数 $f : \Omega \to \mathbb{R}\,(\Omega \subseteq \mathbb{R}^2)$, $f \in C^p$, $p \geq 1$, $(x_1, x_2) \in \Omega$ とする．$f(x_1, x_2) = 0$ のとき，$x = x_1$ を含む開区間 I および I 上で定義された C^p 級の関数 ϕ がただ 1 つ存在し，$\phi(x_1) = x_2$ および

$$u(x) \coloneqq f(x, \phi(x)) = 0 \;\; (\forall x \in I)$$

が成立する．

2.3 凸集合

$X \subseteq \mathbb{R}^n$ とする．$\forall \boldsymbol{x}^1, \boldsymbol{x}^2 \in X$, $\forall \lambda \in [0, 1]$ に対して，$\lambda \boldsymbol{x}^1 + (1 - \lambda)\boldsymbol{x}^2 \in X$ が成立するとき，X は凸集合 (convex set) である[*2]．

補題 1　$X_1, X_2, \ldots \subseteq \mathbb{R}^n$ を凸集合とすると，$\displaystyle\bigcap_{i=1,2,\ldots} X_i$ も凸集合となる．

補題 2（分離超平面の存在）　$X \subseteq \mathbb{R}^n$ を空でない閉凸集合，$\boldsymbol{y} \notin X$ とする．このとき，

$$\boldsymbol{a}^{\mathrm{T}} \boldsymbol{x} \geq b > \boldsymbol{a}^{\mathrm{T}} \boldsymbol{y}, \; \forall x \in X$$

となる分離超平面 (separating hyperplane)$\{\boldsymbol{x} \in \mathbb{R}^n \,|\, \boldsymbol{a}^{\mathrm{T}} \boldsymbol{x} = b\}$ $(\boldsymbol{a} \neq \boldsymbol{0})$ が存在する．

補題 3　$X \subseteq \mathbb{R}^n$ を空でない凸集合，$\boldsymbol{y} \in \mathrm{Bd}(X)$ とする．このとき，

$$\boldsymbol{a}^{\mathrm{T}} \boldsymbol{x} \geq \boldsymbol{a}^{\mathrm{T}} \boldsymbol{y}, \forall \boldsymbol{x} \in X$$

なる $\boldsymbol{a} \neq \boldsymbol{0}$ が存在する．

[*1] 解析学の教科書では，陰関数が唯一存在すると記されている．
[*2] 2 点 $\boldsymbol{x}^1, \boldsymbol{x}^2$ を結ぶ線分上の任意の内分点は，$x_\lambda = \lambda \boldsymbol{x}^1 + (1 - \lambda)\boldsymbol{x}^2$, $\lambda \in [0, 1]$ と書ける．

第 3 章

付録

この付録には，補題や定理の証明を記載する．

証明 1（補題 1 の証明） 項が 2 つのとき，つまり，$X_1, X_2 \subseteq \mathbb{R}^n$ が凸集合ならば，$X_1 \cap X_2$ も凸集合である (*) を証明する．$\forall \boldsymbol{x}^1, \boldsymbol{x}^2 \in X_1 \cap X_2$ に対し，$\boldsymbol{x}^1$ と $\boldsymbol{x}^2$ を結ぶ線分は，$\boldsymbol{x}^1 \in X_1, \boldsymbol{x}^2 \in X_1$ より X_1 に含まれる ($\because X_1$ は凸集合)．同様に X_2 にも含まれる．よって，$\boldsymbol{x}^1$ と $\boldsymbol{x}^2$ を結ぶ線分は，$X_1 \cap X_2$ に含まれる．つまり，$X_1 \cap X_2$ は凸集合である．補題 1 は，(*) を繰り返し用いることで示される． □

補題 2 の証明には，定理 1，補題 4 を用いる．

定理 1（射影定理） X を Hilbert 空間[*1]とし，$L \subset X$ を閉部分空間とする．このとき，

$$\boldsymbol{u} \in X \, (\text{given}) \Rightarrow \exists! \boldsymbol{v} \in L \; s.t. \; (\boldsymbol{u} - \boldsymbol{v}, \boldsymbol{w}) = 0, \, \forall \boldsymbol{w} \in L$$

が成立する (射影が一意に存在する)．

補題 4 定理 1 の X が $X = \mathbb{R}^n$ であり，$L \subseteq \mathbb{R}^n$ が空でない閉凸集合とする．$\boldsymbol{u} \in \mathbb{R}^n$ の L への射影を $\boldsymbol{v}$ とする．このとき，$\forall \boldsymbol{w} \in L$ に対して，

$$(\boldsymbol{u} - \boldsymbol{v})^{\mathrm{T}} (\boldsymbol{w} - \boldsymbol{v}) \leq 0 \tag{3.1}$$

が成立する．

証明 2（定理 1 の証明）

証明 3（補題 4 の証明） $\boldsymbol{u}$ の L への射影 $\boldsymbol{v}$ と任意の点 $\boldsymbol{w} \in L$ を結ぶ線分上の点 $\boldsymbol{x}_\lambda = (1 - \lambda)\boldsymbol{v} + \lambda \boldsymbol{w}, \, 0 < \lambda < 1$ を考える．L は凸より，$\boldsymbol{x}_\lambda$ も L に含まれる．$\boldsymbol{v}$ の定義より，

$$\|\boldsymbol{v} - \boldsymbol{u}\|^2 \leq \|((1 - \lambda)\boldsymbol{v} + \lambda \boldsymbol{w}) - \boldsymbol{u}\|^2 \tag{3.2}$$

[*1] 距離 (ノルム) を持つ集合をノルム空間という．内積を持つ線形空間を内積空間という．ノルム空間 X 内の任意のコーシー列が収束するとき，X は完備であるといい，完備性を持つノルム空間 X を Banach 空間という．また，内積空間 X 上の点 $\boldsymbol{u} \in X$ に対し，$\|\boldsymbol{u}\| = \sqrt{(\boldsymbol{u}, \boldsymbol{u})}$ を内積から誘導されるノルムと呼ぶ．内積から誘導されるノルム空間 X が Banach 空間であるとき，X を Hilbert 空間という．実数空間 $\mathbb{R}^n$ は完備性を持つ．

(3.2) 式を整理すると，

$$
\begin{aligned}
(\boldsymbol{v}-\boldsymbol{u}, \boldsymbol{v}-\boldsymbol{u}) &\leq \|(\boldsymbol{v}-\boldsymbol{u})+\lambda(\boldsymbol{w}-\boldsymbol{v})\|^2 \\
&\leq (\boldsymbol{v}-\boldsymbol{u}, \boldsymbol{v}-\boldsymbol{u})+2\lambda(\boldsymbol{v}-\boldsymbol{u})(\boldsymbol{w}-\boldsymbol{v})+\lambda^2(\boldsymbol{w}-\boldsymbol{v}, \boldsymbol{w}-\boldsymbol{v})
\end{aligned}
$$

いま，$\lambda \neq 0, \lambda > 0$ であり，$L \subseteq \mathbb{R}^n$ であるから，

$$2(\boldsymbol{u}-\boldsymbol{v})(\boldsymbol{w}-\boldsymbol{v}) \leq \lambda\|\boldsymbol{w}-\boldsymbol{v}\|^2$$

よって，$\lambda \to 0$ とすると，(3.1) 式を得る． □

証明 4（補題 2 の証明） $\boldsymbol{y}$ の X への射影を $\bar{\boldsymbol{y}}$ とする．$\boldsymbol{y} \notin X$ より，$\boldsymbol{y} \neq \bar{\boldsymbol{y}}$ である．よって，

$$\|\bar{\boldsymbol{y}}-\boldsymbol{y}\|^2 = (\bar{\boldsymbol{y}}-\boldsymbol{y})^{\mathrm{T}}(\bar{\boldsymbol{y}}-\boldsymbol{y}) > 0 \tag{3.3}$$

である．また，補題 4 より，

$$(\boldsymbol{y}-\bar{\boldsymbol{y}})^{\mathrm{T}}(\boldsymbol{x}-\bar{\boldsymbol{y}}) \leq 0,\ x \in X \tag{3.4}$$

が成立する．(3.3) 式，(3.4) 式より，

$$(\bar{\boldsymbol{y}}-\boldsymbol{y})^{\mathrm{T}}\boldsymbol{x} \geq (\bar{\boldsymbol{y}}-\boldsymbol{y})^{\mathrm{T}}\bar{\boldsymbol{y}} > (\bar{\boldsymbol{y}}-\boldsymbol{y})^{\mathrm{T}}\boldsymbol{y},\ x \in X$$

を得る．ここで，$\boldsymbol{a} = \bar{\boldsymbol{y}}-\boldsymbol{y}$, $b = (\bar{\boldsymbol{y}}-\boldsymbol{y})^{\mathrm{T}}\bar{\boldsymbol{y}}$ と置くと補題 2 が示される． □

参考文献

[1] オペレーションズ・リサーチとは. 公益社団法人日本オペレーションズ・リサーチ学会. https://www.orsj.or.jp/whatisor/whatisor.html. (参照 2020/3/23)

[2] 茨木俊秀. (2011). 最適化の数学. 共立出版株式会社.