最小 2 乗法と統計的回帰分析

数学概論 A (2023 年 6 月 21 日)

今野 良彦
大阪公立大学

概 要

19 世紀の初頭に Carl Friedrich Gauss と Adrien-Marie Legendre によって独立に発見された最小 2 乗法は現代においても重要な役割を果たす. 線型代数の言葉を使い, この手法の理解を目指す. さらに統計的回帰分析のおける最小 2 乗法とその発展形の現代的な役割を簡単に解説する.

目 次

[1]Least absolute shrinkage and selection operator.

0 記号一覧

記号	説明
$\mathbb{N}$	自然数全体の集合.
$\mathbb{R}$	実数全体の集合.
$\mathbb{R}^n$	有限 n 個の実数空間の直積集合.
$\mathsf{Mat}(n, m; \mathbb{R})$	n 行 m 列の実行列全体のなす集合. $\mathsf{Mat}(n, n; \mathbb{R})$ を $\mathsf{Mat}(n; \mathbb{R})$ と書く.
$\boldsymbol{I}_n$	n 次の単位行列.
$\mathbf{0}_n$	$\mathbb{R}^n$ の零ベクトル.
$\boldsymbol{A}^\top$	m 行 n 列の行列 $\boldsymbol{A}$ の転置行列.
$\boldsymbol{A}^{-1}$	正方行列 $\boldsymbol{A}$ の逆行列.
$\boldsymbol{A}^\dagger$	m 行 n 列の行列 $\boldsymbol{A}$ の Moore-Penrose の一般化逆行列. これは n 行 m 列の行列.
$\mathsf{range}(\boldsymbol{A})$	行列 $\boldsymbol{A}$ が定めた線型変換の値域 (像).
$\mathsf{ker}(\boldsymbol{A})$	行列 $\boldsymbol{A}$ が定めた線型変換の核.
$\mathsf{rank}(\boldsymbol{A})$	行列 $\boldsymbol{A}$ の階数 (ランク).
$\mathsf{span}(\boldsymbol{A})$	行列 $\boldsymbol{A}$ の列で張られるベクトル空間.
$\det$	行列式.
$\mathrm{tr}\,(\boldsymbol{A})$	正方行列 $\boldsymbol{A}$ のトレース. すなわち, 対角成分の和.
$*$	真の母数を表す記号. 数学的な意味はない.
$\sim$	確率変数を表す記号. $\widetilde{x}$ と書いたら, これは確率変数であることを示す. ただし, この規則は補遺では適用しない.
$\lvert \cdot \rvert_2$	Euclid ノルム. すなわち, ベクトルの成分の 2 乗の和.
$\lvert \cdot \rvert_1$	絶対値ノルム. すなわち, ベクトルの成分の絶対値の和.
$\lvert \cdot \rvert_0$	零でない成分の個数で与えられるノルム.
$\lVert \boldsymbol{A} \rVert$	行列 $\boldsymbol{A}$ が表す線型変換の作用素ノルム. $\boldsymbol{A}$ が正方行列の場合には, $\lVert \boldsymbol{A} \rVert$ は行列 $\boldsymbol{A}$ の固有値の絶対値の最大値.
$\mathbb{P}$	確率.
$\mathsf{N}(0, \sigma^2)$	平均が 0 で分散が σ^2 の正規分布.
$\mathrm{argmin}_x f(x)$	関数 $f(x)$ の最小値を与える点すべてからなる集合.

文字の使い方:

- スカラーは通常の字体の小文字で表す.

- ベクトルは太文字の小文字で表し, 縦ベクトルとする.
- 行列は太文字の大文字で表す.

1 最小 2 乗法の起源と考え方

統計学において, 最小 2 乗法の最初の遭遇は 2 変数間の関係を表す散布図 (scatter diagram) に直線を引くときであろう. 直線で変数間の関係を代表するもの (回帰直線と呼ばれる) をみつけたいという欲望を実現するための 1 つの方法と位置づけることができる. Legendre は 1805 年に最小 2 乗法の計算方法を明示した[2]. 子午線データに最小 2 乗法を適用して, 「メートルの長さ」を決定したのである.

$n \geq 3\,(n \in \mathbb{N})$ とし, データ $(x_1,\, y_1),\, (x_2,\, y_2),\, \ldots,\, (x_n,\, y_n)$ を観測したとする. ただし, 話を簡単にするために $x_1,\, x_2,\, \ldots,\, x_n$ はすべて異なる値と仮定する. 直線は異なる 2 点で 1 つ定まるので, 単純にデータの 2 点を結ぶことにより直線をひくと $\dfrac{n(n-1)}{2}$ 個の直線ができる. ある意味で直線をひくためにはデータが過剰である.

いま, データの関係を代表する直線の式を

$$y = \beta_1 + \beta_2 x \tag{1}$$

と書く. すると n 点 $\{(x_i,\, y_i)\}_{i=1}^{n}$ を通る直線はない. すなわち, 連立方程式

$$\begin{cases} \beta_1 + x_1\beta_2 = y_1 \\ \beta_1 + x_2\beta_2 = y_2 \\ \quad\vdots \\ \beta_1 + x_2\beta_2 = y_n \end{cases}$$

をみたす解 $(\beta_1,\, \beta_2)$ は存在しない. そのような直線は存在しないが, β_1 と β_2 をデータからうまい具合に定めたいわけである. そのために $x_i\,(i = 1,\, 2,\, \ldots,\, n)$ と直線 $y = \beta_1 + \beta_2 x$ によって予想された値を $\widehat{y}_i := \beta_1 + \beta_2 x_i$ と書く. 実際のデータの値 y_i と x_i と直線 (1) による y_i の予想値 $\widehat{y}_i$ の差 $e_i := y_i - \widehat{y}_i$ を**残差**ということにする. すると問題は残差の全体 (これはデータに対する直線の当てはまりの悪さを表す量と考えることができる) をなるべく小さくする直線をさがせたよい. たとえば

$$\mathrm{minimize}_{\beta_1,\,\beta_2} \sum_{i=1}^{n} |e_i| \tag{2}$$

$$\mathrm{minimize}_{\beta_1,\,\beta_2} \sum_{i=1}^{n} e_i^2 \tag{3}$$

の最小化問題を解けばよい. (2) で与えられる最小化問題の解を最小絶対値乗解とよび, (3) で与えられる最小化問題の解を最小 2 乗解とよぶ. こ

[2] たたし, 最小 2 乗法の発明の priority については統計学史においては見解の一致にいまだいないようである.

れらは凸関数であり, その性質から解が存在することはわかる[3].

最小 2 乗解を求めるために

$$h(\beta_1,\, \beta_2) = \sum_{i=1}^{n} \bigl\{ y_i - (\beta_1 + \beta_2 x_i) \bigr\}^2$$

とおく. これを $\beta_1,\, \beta_2$ に関して偏微分して整理すると

$$\left\{ \begin{array}{rl} n\beta_1 + \bigl(\sum_{i=1}^n x_i\bigr)\beta_2 & = \sum_{i=1}^n y_i \\ \bigl(\sum_{i=1}^n x_i\bigr)\beta_1 + \bigl(\sum_{i=1}^n x_i^2\bigr)\beta_2 & = \sum_{i=1}^n x_i y_i \end{array} \right. \tag{4}$$

を得る. これを解けばよい. この最適問題は狭義凸最適問題なので, 最適解は一意的に存在するので, この必要条件 (4) を解けば解が与えられる.

2 連立 1 次方程式と過剰決定問題

前節の議論を線型代数の言葉を用いて考えてみる. $n \geq d\,(d,\, n \in \mathbb{N})$ とする. d 個の未知数 $\beta_1,\, \beta_2,\, \ldots,\, \beta_d$ に関する n 個の (実係数の)1 次方程式の系

$$\left\{ \begin{array}{c} x_{11}\beta_1 + x_{12}\beta_2 + \cdots + x_{1d}\beta_d = y_1 \\ x_{21}\beta_1 + x_{22}\beta_2 + \cdots + x_{2d}\beta_d = y_2 \\ \vdots \\ x_{n1}\beta_1 + x_{n2}\beta_2 + \cdots + x_{nd}\beta_d = y_n \end{array} \right.$$

を考える. 係数の行列とベクトルを

$$\boldsymbol{X} := \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1d} \\ x_{21} & x_{22} & \cdots & x_{2d} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{nd} \end{pmatrix}; \quad \boldsymbol{\beta} := \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_d \end{pmatrix} \in \mathbb{R}^d; \quad \boldsymbol{y} := \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} \in \mathbb{R}^n$$

と記せば, 連立 1 次方程式系は

$$\boldsymbol{X}\boldsymbol{\beta} = \boldsymbol{y} \tag{5}$$

と書き直せる.

定理 1. 方程式 (5) が解を持つための必要十分条件は

$$\mathsf{rank}(\boldsymbol{X}) = \mathsf{rank}([\boldsymbol{X} : \boldsymbol{y}])$$

[3]この問題は狭義凸関数の最適問題である. 節 C の議論から最小解は一意的に存在することがわかる.

である．すなわち

$$\begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1d} \\ x_{21} & x_{22} & \cdots & x_{2d} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{nd} \end{pmatrix} \quad \text{と} \quad \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1d} & y_1 \\ x_{21} & x_{22} & \cdots & x_{2d} & y_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{nd} & y_n \end{pmatrix}$$

が同じランクをもつことである．

Proof. 行列 $\boldsymbol{X}$ の各列を $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_d$ と記すと

$$\boldsymbol{X}\boldsymbol{\beta} = \boldsymbol{y} \quad \Longleftrightarrow \quad \beta_1 \boldsymbol{x}_1 + \beta_2 \boldsymbol{x}_2 + \cdots + \beta_d \boldsymbol{x}_d = \boldsymbol{y}$$

となる．ここで

$$\begin{aligned} \mathsf{span}(\boldsymbol{X}) &:= \left\{ x_1 \boldsymbol{x}_1 + x_2 \boldsymbol{x}_2 + \cdots + x_d \boldsymbol{x}_d \in \mathbb{R}^n;\ x_1,\ x_2,\ \ldots,\ x_d \in \mathbb{R} \right\} \\ \mathsf{span}([\boldsymbol{X} : \boldsymbol{y}]) &:= \left\{ x_1 \boldsymbol{x}_1 + x_2 \boldsymbol{x}_2 + \cdots + x_d \boldsymbol{x}_d + x_{d+1} \boldsymbol{y} \in \mathbb{R}^n;\ x_1,\ x_2,\ \ldots,\ x_{d+1} \in \mathbb{R} \right\} \end{aligned}$$

とおくと

$$\begin{aligned} \mathsf{rank}(\boldsymbol{X}) &= \mathsf{dim}\big(\mathsf{span}(\boldsymbol{X})\big), \\ \mathsf{rank}([\boldsymbol{X} : \boldsymbol{y}]) &= \mathsf{dim}\big(\mathsf{span}(]\boldsymbol{X} : \boldsymbol{y}])\big) \end{aligned}$$

となる．だから

$$\mathsf{rank}(\boldsymbol{X}) = \mathsf{rank}([\boldsymbol{X} : \boldsymbol{y}]) \quad \Longleftrightarrow \quad \mathsf{span}(\boldsymbol{X}) = \mathsf{span}([\boldsymbol{X} : \boldsymbol{y}])$$

がわかる．よって

$$\mathsf{rank}(\boldsymbol{X}) = \mathsf{rank}([\boldsymbol{X} : \boldsymbol{y}]) \quad \Longleftrightarrow \quad \boldsymbol{y} \in \mathsf{span}(\boldsymbol{X})$$

がわかる． □

系 2. 方程式 (5) が一意的な解をもつための必要十分条件は

$$\mathsf{rank}(\boldsymbol{X}) = d \quad \text{かつ} \quad \mathsf{rank}(\boldsymbol{X}) = \mathsf{rank}([\boldsymbol{X} : \boldsymbol{y}])$$

である．

Proof. $\boldsymbol{\beta}_0$ を方程式 $\boldsymbol{X}\boldsymbol{\beta} = \boldsymbol{y}$ の一つの解とする．斉次(せいじ) 方程式 $\boldsymbol{X}\boldsymbol{\beta} = \mathbf{0}_n$ の解を $\widetilde{\boldsymbol{\beta}}$ を用いて $\boldsymbol{\beta}_0 + \widetilde{\boldsymbol{\beta}}$ と書ける．ただし $\mathbf{0}_n$ は $\mathbb{R}^n$ の零ベクトルである．したがって解が一意であるための必要十分条件は

$$\mathsf{ker}(\boldsymbol{X}) := \left\{ \boldsymbol{\beta} \in \mathbb{R}^d;\ \boldsymbol{X}\boldsymbol{\beta} = \mathbf{0}_n \right\} = \{\mathbf{0}_d\} \quad \Longleftrightarrow \quad \mathsf{rank}(\boldsymbol{X}) = d.$$

□

3　最小 2 乗問題

過剰決定下では, 方程式 $\boldsymbol{X}\boldsymbol{\beta} = \boldsymbol{y}$ の解は存在しない. その代わりに近似解を以下のように考えよう. ベクトル $\boldsymbol{\beta}$ に対する**残差ベクトル** $\boldsymbol{e}$ を

$$\boldsymbol{e} := \boldsymbol{X}\boldsymbol{\beta} - \boldsymbol{y} \tag{6}$$

と定める. 近似解として残差ベクトル $\boldsymbol{e}$ ができるだけ小さくなるものを採用する.

いま, ベクトル $\boldsymbol{e} = (e_1, e_2, \ldots, e_d)^\top$ に対して

$$|\boldsymbol{e}|_1 := \sum_{i=1}^{d} |e_i|, \quad |\boldsymbol{e}|_2 := \sqrt{\sum_{i=1}^{d} e_i^2} \tag{7}$$

とおく. このとき, 最小 2 乗解 $\widehat{\boldsymbol{\beta}}^{\mathrm{ls}}$ と最小絶対値解 $\widehat{\boldsymbol{\beta}}^{\mathrm{ab}}$ をそれぞれ

$$\begin{aligned} \left|\boldsymbol{X}\widehat{\boldsymbol{\beta}}^{\mathrm{Ls}} - \boldsymbol{y}\right|_2 &\le \left|\boldsymbol{X}\boldsymbol{\beta} - \boldsymbol{y}\right|_2 \qquad (\forall \boldsymbol{\beta} \in \mathbb{R}^d), \\ \left|\boldsymbol{X}\widehat{\boldsymbol{\beta}}^{\mathrm{ab}} - \boldsymbol{y}\right|_1 &\le \left|\boldsymbol{X}\boldsymbol{\beta} - \boldsymbol{y}\right|_1 \qquad (\forall \boldsymbol{\beta} \in \mathbb{R}^d) \end{aligned}$$

をみたすもので定めることにする. これらを方程式 5) の近似解と考える.

注意 3. 凸関数の性質から

$$\left\{\widehat{\boldsymbol{\beta}}^{\mathrm{ab}} \in \mathbb{R}^d;\, |\boldsymbol{X}\widehat{\boldsymbol{\beta}}^{\mathrm{ab}} - \boldsymbol{y}|_1 \le |\boldsymbol{X}\boldsymbol{\beta} - \boldsymbol{y}|_1 \, (\forall \boldsymbol{\beta} \in \mathbb{R}^d)\right\} \ne \varnothing$$

となることが知られている.

注意 4. 最小 2 乗解が広く用いられるのは, 解析的に扱いやすいことが一つの理由であろう.

例 5.

$$\boldsymbol{X} := \begin{pmatrix} 2 & 0 \\ -1 & 1 \\ 0 & 2 \end{pmatrix}, \quad \boldsymbol{y} := \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix}$$

とする. すなわち

$$\begin{cases} 2\beta_1 = 1 \\ -\beta_1 + \beta_2 = 0 \\ 2\beta_2 = -1 \end{cases}$$

を考える. すると定理 1 から直ちに解が存在しないことがわかる. 最小 2 乗解は, $\boldsymbol{\beta}$ に関する最小化問題

$$\text{minimize}\left\{(2\beta_1 - 1)^2 + (-\beta_1 + \beta_2)^2 + (2\beta_2 + 1)^2\right\} \tag{8}$$

の解として与えられる．すると

$$\widehat{\boldsymbol{\beta}}^{\mathrm{ls}} = \frac{1}{3}\begin{pmatrix} 1 \\ -1 \end{pmatrix} \tag{9}$$

がわかる．さらに

$$\widehat{\boldsymbol{e}} := \boldsymbol{X}\widehat{\boldsymbol{\beta}}^{\mathrm{ls}} - \boldsymbol{y} = \frac{1}{3}\begin{pmatrix} -1 \\ -2 \\ 1 \end{pmatrix}, \quad |\boldsymbol{X}\widehat{\boldsymbol{\beta}}^{\mathrm{ls}} - \boldsymbol{y}|_2^2 = \frac{2}{3}$$

を得る．一方, 連立 1 次方程式

$$\begin{cases} 2\beta_1 = 1 \\ 2\beta_2 = -1 \end{cases}$$

を考える．上の連立 1 次方程式の解 $\boldsymbol{\beta}_0$ は

$$\boldsymbol{\beta}_0 = \frac{1}{2}\begin{pmatrix} 1 \\ -1 \end{pmatrix}$$

で与えられ

$$|\boldsymbol{X}\boldsymbol{\beta}_0 - \boldsymbol{y}|_2^2 = 1$$

となる．よって, $\boldsymbol{X}\widehat{\boldsymbol{\beta}}^{\mathrm{ls}}$ より $\boldsymbol{X}\boldsymbol{\beta}_0$ は $\boldsymbol{y}$ からより離れていることがわかる．すなわち

$$\boldsymbol{X}\widehat{\boldsymbol{\beta}}^{\mathrm{ls}} = \frac{1}{3}\begin{pmatrix} 2 \\ -1 \\ 0 \end{pmatrix} - \frac{1}{3}\begin{pmatrix} 0 \\ 1 \\ 2 \end{pmatrix} = \frac{2}{3}\begin{pmatrix} 1 \\ -1 \\ -1 \end{pmatrix}$$

が $\boldsymbol{y}$ に最も近いことがわかる．これは $\boldsymbol{y}$ の $\boldsymbol{X}$ の列で張られる部分空間への直交射影になっていることがわかる． □

3.1 最小 2 乗解の求め方

いま

$$f(\boldsymbol{\beta}) = |\boldsymbol{X}\boldsymbol{\beta} - \boldsymbol{y}|_2^2, \quad \nabla := \left(\frac{\partial}{\partial\beta_1}, \frac{\partial}{\partial\beta_2}, \ldots, \frac{\partial}{\partial\beta_d}\right)^{\top},$$
$$\boldsymbol{\beta} = (\beta_1, \beta_2, \ldots, \beta_d)^{\top}$$

とおくと最小 2 乗解 $\widehat{\boldsymbol{\beta}}^{\mathrm{ls}}$ は方程式

$$\nabla f(\boldsymbol{\beta}) := \left(\frac{\partial f}{\partial \beta_1}(\boldsymbol{\beta}),\, \frac{\partial f}{\partial \beta_2}(\boldsymbol{\beta}),\, \ldots,\, \frac{\partial f}{\partial \beta_d}(\boldsymbol{\beta})\right)^{\top} = \boldsymbol{0}_d \tag{10}$$

をみたさなければならない. すると (10) の左辺は

$$\nabla f(\boldsymbol{\beta}) = 2\boldsymbol{X}^{\top}(\boldsymbol{X}\boldsymbol{\beta} - \boldsymbol{y}) \tag{11}$$

と書けることがわかる. 実際, 以下の議論からわかる. $\mathbb{R}^d$ のベクトル $\boldsymbol{x}$ の第 k 成分を $(\boldsymbol{x})_k(x_k$ とも記す), 行列 $\boldsymbol{X}$ の第 (i, j) 成分を $(\boldsymbol{X})_{ij}$ (x_{ij} とも記す) と記すことにする. すると $k = 1, 2, \ldots, d$ に対して

$$\begin{aligned}\big(\nabla f(\boldsymbol{\beta})\big)_k &= \sum_{i=1}^{n} 2\left(\sum_{j=1}^{d} x_{ij}\beta_j - y_i\right)x_{ik} = \sum_{i=1}^{n} 2(\boldsymbol{X}^{\top})_{ki}(\boldsymbol{X}\boldsymbol{\beta} - \boldsymbol{y})_i \\ &= \big(2\boldsymbol{X}^{\top}(\boldsymbol{X}\boldsymbol{\beta} - \boldsymbol{y})\big)_k\end{aligned}$$

からわかる.

(10) と (11) から

$$\boldsymbol{X}^{\top}\boldsymbol{X}\widehat{\boldsymbol{\beta}}^{\mathrm{ls}} = \boldsymbol{X}^{\top}\boldsymbol{y} \tag{12}$$

を得る. この方程式を**正規方程式**という.

さらに $\boldsymbol{X}$ の列ベクトルが 1 次独立のとき, 正方行列 $\boldsymbol{X}^{\top}\boldsymbol{X}$ は可逆となり, (12) から

$$\widehat{\boldsymbol{\beta}}^{\mathrm{ls}} = (\boldsymbol{X}^{\top}\boldsymbol{X})^{-1}\boldsymbol{X}^{\top}\boldsymbol{y}$$

となることがわかる. さらに d 行 n 列の行列 $(\boldsymbol{X}^{\top}\boldsymbol{X})^{-1}\boldsymbol{X}^{\top}$ を $\boldsymbol{X}^{\dagger}$ と記し, n 行 d 列の行列 $\boldsymbol{X}$ の Moore-Penrose **の一般化逆行列**とよぶ. すなわち

$$\widehat{\boldsymbol{\beta}}^{\mathrm{ls}} = \boldsymbol{X}^{\dagger}\boldsymbol{y}$$

である.

注意 6. 一般には $\boldsymbol{X}^{\dagger}\boldsymbol{X} \neq \boldsymbol{I}_d$, $\boldsymbol{X}\boldsymbol{X}^{\dagger} \neq \boldsymbol{I}_n$ である. ただし $\boldsymbol{I}_d$ は d 次の単位行列である. Moore-Penrose の一般化逆行列の定義については, 補遺の節 B を参照のこと. □

4　最小 2 乗法によるデータのあてはめ

2 次元のデータ $(y,\, x)$ が

$$\widetilde{y} = \sin x + \widetilde{\epsilon} \tag{13}$$

というモデルで表現されているとする．変数 $\widetilde{y}$ を**応答変数**といい，x を**説明変数** (ランダムではない変数) とよぶことにする．$\widetilde{\epsilon}$ は誤差 (確率変数) で観測できない量である．通常，(13) の左辺の一項目の関数は未知なので，同一条件のもとでの繰り返しの観測されたデータに基づき一項目の関数を何らかの形で回復 (推定) することを考える．

たとえば，$\widetilde{\epsilon} \sim N(0,\, 0.04)$(平均が 0 で分散が 0.04 の正規分布) に従うと仮定して，疑似的にデータを作成した結果が以下である．

x	0	1	2	3	4	5	6
y	-0.0343	1.0081	0.8326	0.04047	-0.7585	-0.9285	-0.2110
x	7	8	9	10			
y	0.6626	0.08492	0.2761	-0.6962			

すなわち

$$(x_1,\, y_1) = (0,\, -0.0343),\ (x_2,\, y_2) = (1,\, 1.0081),\ \ldots,\ (x_{11},\, y_{11}) = (10,\, -0.6962)$$

である．

まず，多項式曲線

$$f(x) = \beta_1 + \beta_2 x + \beta_3 x^2 + \cdots + \beta_{11} x^{10}$$

をデータにあてはめてみよう．ここで $\beta_1,\, \beta_2,\, \ldots,\, \beta_{11}$ は係数である．データから連立方程式

$$\begin{cases} \beta_1 + \beta_2 x_1 + \beta_3 x_1^2 + \cdots + \beta_{11} x_1^{10} = y_1 \\ \beta_1 + \beta_2 x_2 + \beta_3 x_2^2 + \cdots + \beta_{11} x_2^{10} = y_2 \\ \quad\vdots \\ \beta_1 + \beta_2 x_{11} + \beta_3 x_{11}^2 + \cdots + \beta_{11} x_{11}^{10} = y_{11} \end{cases}$$

を得る．ここで $d = 2,\, 3,\, \ldots,\, 11$ に対して

$$\boldsymbol{X}_d := \begin{pmatrix} 1 & x_1 & x_1^2 & \ldots & x_1^{d-1} \\ 1 & x_2 & x_2^2 & \ldots & x_2^{d-1} \\ \vdots & \vdots & \ddots & & \vdots \\ 1 & x_{11} & x_{11}^2 & \ldots & x_{11}^{d-1} \end{pmatrix} \in \mathsf{Mat}(11,\, 11;\, \mathbb{R}), \quad \boldsymbol{y} := \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_{11} \end{pmatrix} \in \mathbb{R}^{11},$$

$$\boldsymbol{\beta}_d := \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_d \\ \mathbf{0}_{11-d} \end{pmatrix} \in \mathbb{R}^{11}$$

とおき, $\boldsymbol{X} := \boldsymbol{X}_{11}$ と書く．

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta}_{11}$$

と書ける．行列 $\boldsymbol{X}$ は Vandermonde 行列と呼ばれるもので,

$$\det \boldsymbol{X} = \prod_{1 \leq i < j \leq 11} (x_i - x_j) \tag{14}$$

となることが知られている．よって $x_i \neq x_j \, (i \neq j)$ のとき $\boldsymbol{X}$ は正則となるから

$$\boldsymbol{\beta}_{11} = \left(\boldsymbol{X}^\top \boldsymbol{X}\right)^{-1} \boldsymbol{X}^\top \boldsymbol{y} = \boldsymbol{X}^\dagger \boldsymbol{y}$$

を得る．計算結果は

$$\widehat{\boldsymbol{\beta}}_{11}^{\mathrm{ls}} = \begin{pmatrix} -0.0343 \\ 16.2400 \\ -38,0984 \\ 37.8369 \\ -20.2842 \\ 6.5035 \\ -1.3100 \\ \underline{0.1677} \\ \underline{-0.0133} \\ \underline{0.0006} \\ \underline{-0.0000} \end{pmatrix}$$

となる．7 次以上の係数は非常に小さくなっている．

$\beta_8 = \beta_9 = \beta_{10} = \beta_{11} = 0$ となるものを考える．すなわち

$$\boldsymbol{\beta}_6 = (\beta_0,\, \beta_1,\, \ldots,\, \beta_6,\, \mathbf{0}_4^\top)^\top \in \mathbb{R}^{11}$$

である．すると定理 1 から

$$
\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta}_6 \iff \begin{cases} \beta_1 + \beta_2 x_1 + \beta_3 x_1^2 + \cdots + \beta_7 x_1^6 = y_1 \\ \beta_1 + \beta_2 x_2 + \beta_3 x_2^2 + \cdots + \beta_7 x_2^6 = y_2 \\ \vdots \\ \beta_1 + \beta_2 x_{11} + \beta_3 x_{11}^2 + \cdots + \beta_7 x_{11}^6 = y_{11} \end{cases}
$$

をみたす $\boldsymbol{\beta}_6$ は存在しないので，解の近似値を最小 2 乗法で求めよう．$x_i \neq x_j \, (i \neq j)$ のとき，$\boldsymbol{X}_6$ の列は 1 次独立である．よって，前節の議論から

$$
\widehat{\boldsymbol{\beta}}_6^{\mathrm{ls}} = \begin{pmatrix} \boldsymbol{X}_6^{\dagger}\boldsymbol{y} \\ \boldsymbol{0}_4 \end{pmatrix}, \qquad \boldsymbol{X}_6^{\dagger} = \left(\boldsymbol{X}_6^{\top}\boldsymbol{X}_6\right)^{-1}\boldsymbol{X}_6^{\top}
$$

となる．計算結果は

$$
\widehat{\boldsymbol{\beta}}_6^{\mathrm{ls}} = \begin{pmatrix} -0.0260 \\ 1.0636 \\ 0.3067 \\ 0.1426 \\ -0.0146 \\ 0.0005 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}
$$

となる．

注意 7. $\widehat{\boldsymbol{\beta}}_{11}$ で定まる 10 次の多項式は $\widehat{\boldsymbol{\beta}}_6$ で定まるものと比較すると大きく振動していることがわかる．この 10 次の多項式はデータの**過学習** (over-fitting) を起こしている．まさに「過ぎたるは猶及(なおおよ)ばざるが如し」で中庸が大切ということだ．しかし，データからどの程度で中庸かを判定するのはやさしいことではない．□

5 ridge 解と lasso [4]解

$$
\boldsymbol{y} = \begin{pmatrix} 1 \\ 2 \end{pmatrix}, \quad \boldsymbol{x}_1 = \begin{pmatrix} 1 \\ 2 \end{pmatrix}, \quad \boldsymbol{x}_2 = \begin{pmatrix} c \\ 2 \end{pmatrix}, \quad \boldsymbol{X} = (\boldsymbol{x}_1, \boldsymbol{x}_2), \quad c \in \mathbb{R}
$$

[4]Least absolute shrinkage and selection operator.

とする．方程式

$$\boldsymbol{X}\boldsymbol{\beta} = \boldsymbol{y} \tag{15}$$

を考える．ただし $\boldsymbol{\beta} = (\beta_1, \beta_2)^\top$ である．定理 1 から $c \neq 1$ のとき解は存在しないことがわかる．$c \neq 1$ のとき，最小 2 乗解を求める[5]と

$$\widehat{\boldsymbol{\beta}}^{\mathrm{ls}} = \left(\boldsymbol{X}^\top \boldsymbol{X}\right)^{-1} \boldsymbol{X}^\top \boldsymbol{y} = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \tag{16}$$

となる．さらに $c = 1$ のとき，最小 2 乗解 $\widehat{\boldsymbol{\beta}}^{\mathrm{ls}}$ は方程式 (15) の 1 つの解となっている．$c = 1$ のときの斉次方程式

$$\boldsymbol{X}\boldsymbol{\beta} = \mathbf{0} \quad \Longrightarrow \quad \begin{pmatrix} 1 & 1 \\ 2 & 2 \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

の解は

$$\boldsymbol{\beta} = b \begin{pmatrix} -1 \\ 1 \end{pmatrix}, \qquad (b \in \mathbb{R})$$

と表現できる．よって，$c = 1$ のとき，方程式 (15) の解は

$$\boldsymbol{\beta} = \widehat{\boldsymbol{\beta}}^{\mathrm{ls}} + b \begin{pmatrix} -1 \\ 1 \end{pmatrix} =: \widehat{\boldsymbol{\beta}}^{\mathrm{ls}}[b] \qquad (b \in \mathbb{R})$$

と表現できる．$c = 1$ のとき，$\widehat{\boldsymbol{\beta}}^{\mathrm{ls}}[b]$ は最小 2 乗解となる．一方，$c \neq 1$ のとき

$$\left|\widehat{\boldsymbol{\beta}}^{\mathrm{ls}}[b] - \widehat{\boldsymbol{\beta}}^{\mathrm{ls}}\right|_2^2 = \sqrt{2}|b|$$

となる．b は任意だったので，c が 1 に近いときには，データ $\boldsymbol{y}$ のすこしの変化で最小 2 乗解は大きく変化することを予想される．c が 1 に近いとき，行列 $\boldsymbol{X}$ は**線型共変性** をもつという．このような場合には，最小 2 乗解は不安定になる．データ $\boldsymbol{y}$ のすこし変化で最小 2 乗解は大きく変化する．

行列 $\boldsymbol{X}$ が線型共変性をもつときの不安定性を回避するために，つぎのような最適問題を考える．固定した $r \in (0, \infty)$ に対して

$$\boldsymbol{\beta} \mapsto |\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}|_2^2 + r|\boldsymbol{\beta}|_2^2$$

を $\boldsymbol{\beta}$ に関して最小化してみよう．最小値を与える点を **ridge 解**といい，$\widehat{\boldsymbol{\beta}}^{\mathrm{ridge}}[r]$ と記すことにする．上の式の左辺の第一項はデータの当てはまり

[5] 2×2 の正則行列の逆行列の計算をすることになる．

のよさを測る項で, 第二項は $\boldsymbol{\beta}$ に長さについての罰則と考えることができる. これは

$$(\boldsymbol{\beta}\text{ のデータへの適合度 }) + (\boldsymbol{\beta}\text{ への罰則})$$

と解釈できる.

$\widehat{\boldsymbol{\beta}}^{\mathrm{ridge}}[r]$ が ridge 解であるための必要十分条件は

$$\big(\boldsymbol{X}^\top\boldsymbol{X} + r\boldsymbol{I}_2\big)\widehat{\boldsymbol{\beta}}^{\mathrm{ridge}}[r] = \boldsymbol{X}^\top\boldsymbol{y}$$

であることがわかる. c の値に関わらず $\boldsymbol{X}^\top\boldsymbol{X} + r\boldsymbol{I}_2$ は正則であるので

$$\begin{aligned}\widehat{\boldsymbol{\beta}}^{\mathrm{ridge}}[r] &= \big(\boldsymbol{X}^\top\boldsymbol{X} + r\boldsymbol{I}_2\big)^{-1}\boldsymbol{X}^\top\boldsymbol{y}\\ &= \frac{1}{4(1-c)^2+(9+c^2)r+r^2}\begin{pmatrix} 4(1-c)^2+5r \\ (4+r)c\end{pmatrix}\end{aligned} \tag{17}$$

となる. $r>0$ を固定すると $\widehat{\boldsymbol{\beta}}^{\mathrm{ridge}}[r]$ は c に関して連続的に変化することがわかる. さらに

$$\lim_{r\to 0+}\widehat{\boldsymbol{\beta}}^{\mathrm{ridge}}[r] = \begin{cases} \dfrac{1}{2}\begin{pmatrix}1\\1\end{pmatrix} = \widehat{\boldsymbol{\beta}}^{\mathrm{ls}}[1/2] & (c=1), \\ \begin{pmatrix}1\\0\end{pmatrix} = \widehat{\boldsymbol{\beta}}^{\mathrm{ls}}[0] = \widehat{\boldsymbol{\beta}}^{\mathrm{ls}} & (c\neq 1)\end{cases}$$

となる. とくに $c=1$ のとき $\widehat{\boldsymbol{\beta}}^{\mathrm{ls}}[1/2]$ は方程式 (15) をみたすもののうち, 長さ $|\boldsymbol{\beta}|_2 = \sqrt{(1-b)^2+b^2}$ を最小にするものになっている.

注意 8. $c=1$ のとき, $\boldsymbol{X} = \begin{pmatrix}1 & 1\\ 2 & 2\end{pmatrix}$ は正則でない. 一般化逆行列 $\boldsymbol{X}^\dagger$ がみたすべき以下の関係式を解く.

① $\boldsymbol{X}\boldsymbol{X}^\dagger\boldsymbol{X} = \boldsymbol{X}$, ② $\boldsymbol{X}^\dagger\boldsymbol{X}\boldsymbol{X}^\dagger = \boldsymbol{X}^\dagger$, ③ $\big(\boldsymbol{X}\boldsymbol{X}^\dagger\big)^\top = \boldsymbol{X}\boldsymbol{X}^\dagger$,
④ $\big(\boldsymbol{X}^\dagger\boldsymbol{X}\big)^\top = \boldsymbol{X}^\dagger\boldsymbol{X}$.

煩わしい計算すると $\boldsymbol{X}^\dagger = \dfrac{1}{10}\begin{pmatrix}1 & 2\\ 1 & 2\end{pmatrix}$ を得る[6]. このことから $c=1$ のとき

$$\boldsymbol{X}^\dagger\boldsymbol{y} = \frac{1}{10}\begin{pmatrix}1 & 2\\ 1 & 2\end{pmatrix}\begin{pmatrix}1\\2\end{pmatrix} = \frac{1}{2}\begin{pmatrix}1\\1\end{pmatrix} = \widehat{\boldsymbol{\beta}}^{\mathrm{ls}}[1/2]$$

[6]① ～ ④ の関係式から $\boldsymbol{X}^\dagger$ をみつけるのは大変である. しかし, $\boldsymbol{X}$ と $\boldsymbol{X}^\dagger$ は ① ～ ④ の関係式をみたしていることは比較的簡単に確認できる. 一般論から ① ～ ④ の関係式をみたすものが唯一存在することが知られているので, $\boldsymbol{X}^\dagger = \dfrac{1}{10}\begin{pmatrix}1 & 2\\ 1 & 2\end{pmatrix}$ となることがわかる.

となっていることがわかる．一方 $c \neq 1$ のとき $\boldsymbol{X} = \begin{pmatrix} 1 & c \\ 2 & 2 \end{pmatrix}$ は正則で，

$$\boldsymbol{X}^{\dagger} = \left(\boldsymbol{X}^{\top}\boldsymbol{X}\right)^{-1}\boldsymbol{X}^{\top} = \frac{1}{4-8c-4c^2}\begin{pmatrix} 4-4c & -2c+2c^2 \\ -4+4c & 2-2c \end{pmatrix}$$

となるので

$$\boldsymbol{X}^{\dagger}\boldsymbol{y} = \frac{1}{4-8c-4c^2}\begin{pmatrix} 4-4c & -2c+2c^2 \\ -4+4c & 2-2c \end{pmatrix}\begin{pmatrix} 1 \\ 2 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \widehat{\boldsymbol{\beta}}^{\mathrm{ls}}[0]$$

となる．以上から c の値に関わらず

$$\lim_{r \to 0+} \widehat{\boldsymbol{\beta}}^{\mathrm{ridge}}[r] = \boldsymbol{X}^{\dagger}\boldsymbol{y}$$

と書けることがわかる． □

多項式回帰の場合のように $\boldsymbol{\beta}$ の成分の多くが 0 であるような「疎らな」$\boldsymbol{\beta}$ に対しては固定した $r \in (0, \infty)$ に対して

$$\boldsymbol{\beta} \mapsto |\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}|_2^2 + 2r|\boldsymbol{\beta}|_1 \tag{18}$$

を $\boldsymbol{\beta}$ に関して最小化をするものを用いるとよいことが知られいる．すなわち，罰則項を成分の絶対値の和とする．これを **lasso 解**といい，$\widehat{\boldsymbol{\beta}}^{\mathrm{lasso}}[r]$ と記すことにする．第 4 節の多項式回帰の問題の $r = 1$ のときの lasso 解は

$$\widehat{\boldsymbol{\beta}}^{\mathrm{lasso}}[1] = \begin{pmatrix} 0.1448 \\ 0.2691 \\ 0.0674 \\ 0.0769 \\ -0.0334 \\ -0.0674 \\ 0.0386 \\ -0.0085 \\ 0.0001 \\ 0.0000 \end{pmatrix}$$

となる．

注意 9. ridge 解と lasso 解はともに新たな母数 r の選択が必要となる．「よい」r の選択を交差検証法などのデータの情報にのみですることができる．

注意 10. 凸関数の性質から最適問題 (18) の解は一意に存在することがわかる. しかし, 最適問題 (18) を陽に解くことはできない. この問題を繰り返し計算で解くためのアルゴリズムは統計解析環境 R [7] には実装されている.

注意 11. $\boldsymbol{\beta}$ の次元た大きいモデルを高次元統計的回帰モデルという. 次元の高い回帰ベクトルの場合, $\boldsymbol{\beta}$ が疎らな構造をもつことがよくある. このような状況では, lasso はよい「精度」をもつことが理論的にも調べられている. この点については次の節で簡単に説明する.

6 統計的回帰モデルと最小 2 乗法

統計的回帰モデルを導入する.

$$\widetilde{y}_i = \beta_1^* + \beta_2^* x_{i1} + \cdots + \beta_d^* x_{i,d-1} + \widetilde{\epsilon}_i \qquad (i = 1, 2, \ldots, n). \tag{19}$$

ここで $\beta_i^* \in \mathbb{R}$ は未知で, $\widehat{\epsilon}_1, \widetilde{\epsilon}_2, \ldots, \widetilde{\epsilon}_n$ は観測誤差 $\big($確率空間 $(\Omega, \mathcal{A}, \mathbb{P})$ 上で定義された確率変数列$\big)$ で, 独立同一に正規分布 $\mathsf{N}(0, \sigma^2)$ (平均が 0 で分散が σ^2 の正規分布を表す) に従うとする. また $\sigma > 0$ は既知とする.

観測データは確率変数 $\widetilde{y}_i$ の実現値と $(x_{i1}, x_{i2}, \ldots, x_{id})$ の組が n 個である. このモデルをベクトルと行列を用いて表現する.

$$\widetilde{\boldsymbol{y}} := \begin{pmatrix} \widetilde{y}_1 \\ \widetilde{y}_2 \\ \vdots \\ \widetilde{y}_n \end{pmatrix}, \qquad \underbrace{\boldsymbol{X}}_{n \text{ 行 } d \text{ 列}} := \begin{pmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1,d-1} \\ 1 & x_{21} & x_{22} & \cdots & x_{2,d-1} \\ \vdots & & & & \\ 1 & x_{n1} & x_{n2} & \cdots & x_{n,d-1} \end{pmatrix},$$

$$\boldsymbol{\beta}^* := \begin{pmatrix} \beta_1^* \\ \beta_2^* \\ \vdots \\ \beta_d^* \end{pmatrix}, \qquad \widetilde{\boldsymbol{\epsilon}} := \begin{pmatrix} \widetilde{\epsilon}_1 \\ \widetilde{\epsilon}_2 \\ \vdots \\ \widetilde{\epsilon}_n \end{pmatrix}$$

とおくと

$$\widetilde{\boldsymbol{y}} = \boldsymbol{X}\boldsymbol{\beta}^* + \widetilde{\boldsymbol{\epsilon}}$$

と書ける. 議論を簡単にするために

$$\mathsf{rank}\boldsymbol{X} = \min(d, n)$$

[7] オープンソース・フリーソフトウェアの統計解析向けのプログラミング言語及びその開発実行環境のこと. 詳しくは RjpWiki`http://www.okadajp.org/RWiki/` を参照せよ.

を仮定する．ただし $\min(d,\,n)$ は d と n の大きくない方を表す．

$r>0$ とし，最小 2 乗推定量 $\widehat{\boldsymbol{\beta}}^{\mathrm{ls}}$ と lasso $\widehat{\boldsymbol{\beta}}^{\mathrm{lasso}}[r]$ は

$$\widehat{\boldsymbol{\beta}}^{\mathrm{ls}} = \boldsymbol{X}^{\dagger}\widetilde{\boldsymbol{y}}, \tag{20}$$

$$\widehat{\boldsymbol{\beta}}^{\mathrm{lasso}}[r] \in \mathrm{argmin}_{\boldsymbol{\beta}\in\mathbb{R}^d}\left\{\left|\widetilde{\boldsymbol{y}}-\boldsymbol{X}\boldsymbol{\beta}\right|_2^2 + 2r\left|\boldsymbol{\beta}\right|_1\right\} \tag{21}$$

で定義できる．ただし

$$|\boldsymbol{x}|_2 = \sqrt{\sum_{i=1}^{n} x_i^2}, \quad |\boldsymbol{x}|_1 = \sum_{i=1}^{n}|x_i| \qquad \big(\forall \boldsymbol{x} = (x_1,\,x_2,\,\ldots,\,x_n)^{\top} \in \mathbb{R}^n\big)$$

である．$f(x)$ を関数としたとき，$\mathrm{argmin}_x f(x)$ の意味は関数

$$x \mapsto f(x)$$

を最小にする点すべてのなす集合である．たとえば

$$\mathrm{argmin}_{0\le x\le 4\pi}\sin x = \left\{\frac{3}{2}\pi,\ \frac{7}{2}\pi\right\}$$

となる．

注意 12. 前節までの $\widetilde{\boldsymbol{y}}$ と $\widehat{\boldsymbol{\beta}}^{\mathrm{ls}}$ はデータの値から計算されたものである．したがって，ランダムではない．しかし，この節では，ランダムベクトル $\widetilde{\boldsymbol{y}}$ で表現される $\widehat{\boldsymbol{\beta}}^{\mathrm{ls}}$ と $\widehat{\boldsymbol{\beta}}^{\mathrm{lasso}}[r]$ ははランダムベクトルであることに注意せよ．さらに，(24) は確率ベクトル $\widetilde{\boldsymbol{y}}$ で表現されるので，ランダムな量である．したがって，$\omega\in\Omega$ ごとに定まる $\widetilde{\boldsymbol{y}}$ によって与えらる関数の最小化問題を解いたものが $\widehat{\boldsymbol{\beta}}^{\mathrm{lasso}}[r]$ であり，$\omega\in\Omega$ ごとに定まるランダムな量である．(22) の $\left|\boldsymbol{X}\widehat{\boldsymbol{\beta}}^{\mathrm{ls}} - \boldsymbol{X}\boldsymbol{\beta}^*\right|_2^2$ はランダムな量となるので，評価するために確率を計算したり，期待値を計算することが必要となる． □

統計的回帰モデル (19) のもとで以下が成立する．$\forall a>0$ に対して

$$\mathbb{P}\left(\frac{1}{n}\left|\boldsymbol{X}\widehat{\boldsymbol{\beta}}^{\mathrm{ls}} - \boldsymbol{X}\boldsymbol{\beta}^*\right|_2^2 \le C_1\sigma^2\frac{d}{n}\right) \ge 1-a \tag{22}$$

となる．ただし $\mathsf{Pr}(\,\cdot\,)$ は確率で，$C_1>0$ は $d,\,n,\,\sigma,\,\boldsymbol{\beta}^*,\,a$ に依存しない定数である．

$k\in\mathbb{N}$ は $k<\min(n,\,d)$ とする．以下の仮定をおく．

- $|\boldsymbol{\beta}^*|_0 \le k$ とする．ただし $|\boldsymbol{\beta}^*|_0$ はベクトル $\boldsymbol{\beta}^*$ の零でない成分の個数を表す．

- $$\left\|\boldsymbol{X}^\top \boldsymbol{X} - \boldsymbol{I}_d\right\| \le \frac{1}{32k} \tag{23}$$

  とする. ただし, 正方行列 $\boldsymbol{A}$ に対して, $\|\boldsymbol{A}\|$ は $\boldsymbol{A}$ の固有値の絶対値の最大のものを表す.

注意 13. 上の最初の仮定は回帰ベクトル $\boldsymbol{\beta}$ が疎であることを仮定している. 高次元の設定では, 疎な回帰ベクトルが自然に表れることが知られている. 2 番目の仮定は技術的な仮定である. $\boldsymbol{X}$ がよく計画された行列になっていることを意味する. しかし, 通常, k は観測者には未知であるので, (23) は観測者には確認できない条件となる. □

このとき, $\forall a > 0$ に対して

$$r = \sigma\sqrt{\frac{2\log(2d)}{n}} + \sigma\sqrt{\frac{2\log(1/a)}{n}}$$

とおくと

$$\mathbb{P}\left(\frac{1}{n}\left|\boldsymbol{X}\widehat{\boldsymbol{\beta}}^{\mathrm{lasso}}[r] - \boldsymbol{X}\boldsymbol{\beta}^*\right|_2^2 \le C_2 k\sigma^2 \frac{\log(2d/a)}{n}\right) \ge 1 - a \tag{24}$$

となる. ただし $C_2 > 0$ は d, n, k, σ, $\boldsymbol{\beta}^*$, a に依存しない定数である.

(22) と (24) と比較すると $\boldsymbol{\beta}^*$ が疎で d が大きいときに, $\widehat{\boldsymbol{\beta}}^{\mathrm{lasso}}[r]$ の精度は $\widehat{\boldsymbol{\beta}}^{\mathrm{ls}}$ のものよりよいことがみてとれる.

7 文献についての注意

節 1 は [7], [13] を参考にした. 節 2 は [11] からの借用である. 節 3 は [3] を参考にした. 節 4 は [12] からの借用である. 節 5 は [6] からの借用である. 節 6 は [8] を参考にした. 補遺 A は [1] から, 補遺 B は [10] からの借用である. 補遺 C は [2] と [14] からの借用である.

参考文献

[1] Bisgard, J. (2021). Analysis and Linear algebra: The Singular Value Decomposition and Applications. *Student Mathematical Library* **94**. American Mathematical Society.

[2] Borwein, J.M., Lewis, A.S. (2006). Convex Analysis and Nonlinear Optimization Theory and Examples, 2nd. edition. Spriner.

[3] Boyd, S., Vandenberghe, L. (2018). Introduction to Applied Linear algebra. Cambridge University Press.

[4] Gallier, J. (2001). Geometrical Methods and Applications. *Texts in Applied Mathematics* **38**. Springer.

[5] Kaipio, J., Somerslo, E. (2005). Statistical and Computational Inverse Problems. *Applied Mathematical Sciences* **60**. Springer.

[6] Lederer, J. (2022). Fundamentals of High-Dimensional Statistics with Exercies and R Labs. *Springer Text in Statistics*. Springer.

[7] Nyblom, J. (2023+). Note on Legendre 's Method of Least Squares. To appear in *Stattistical Science*.

[8] Rigollet, P., Hütter, J.-C. (2017). High Dimensionnal Statistics. Available at https://math.mit.edu/~rigollet/PDFs/RigNotes17.pdf(2023/05/17 accessed).

[9] Stiegler, S. M. (1981). Gauss and the invention of least squares. *Annals of Statistics* **9** 465-474.

[10] 新井仁之 (2006). 線形代数　基礎と応用 (第 1 刷). 日本評論社.

[11] 高橋礼二 (2019). 線型代数講義 (第 1 版第 2 刷). 日本評論社.

[12] 永原正章 (2020). スパースモデリング (第 1 版第 4 刷). コロナ社.

[13] 森棟公夫 (2010). 地球の大きさと最小 2 乗法. 社会とマネジメント **8** 111-130.

[14] 柳田英二 (2022). 解析入門 (第 1 版第 1 刷). 裳華房.

注意 14. 補遺では,「~」は確率変数を示す記号としては用いない.

A　補遺: 特異値分解の補足説明

定義 15. $n, p \in \mathbb{N}, n \geq p$ とし, $V \subset \mathbb{R}^n$ をベクトル空間とする. $\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_p \in V$ とする.
(1) $\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_p$ はベクトル空間 V の基底であるとは, 任意の $\boldsymbol{v} \in V$ に対して, 唯一の $a_1, a_2, \ldots, a_p \in \mathbb{R}$ があって,

$$\boldsymbol{v} = a_1\boldsymbol{v}_1 + a_2\boldsymbol{v}_2 + \cdots + a_p\boldsymbol{v}_p$$

とかけるときをいう. この p をベクトル空間 V の**次元**といい, $\mathsf{dim}(V)$ と書く.
(2) $\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_p$ は**線型独立** (1 次独立) であるとは,

$$a_1\boldsymbol{v}_1 + a_2\boldsymbol{v}_2 + \cdots + a_p\boldsymbol{v}_p = \boldsymbol{0}\,(a_1, a_2, \ldots, a_p \in \mathbb{R}) \Longrightarrow a_1 = a_2 = \cdots = a_p = 0$$

が成り立つときをいう. ただし, $\boldsymbol{0} = (0, 0, \ldots, 0)^\top \in \mathbb{R}^n$ である.

定義 16. (1) $n, p \in \mathbb{N}, n \geq p$ とする. ベクトル $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_p \in \mathbb{R}^n$ で張られた $\mathbb{R}^n$ の部分空間 $\mathsf{span}\{\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_p\}$ を

$$\mathsf{span}\{\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_p\} = \big\{a_1\boldsymbol{x}_1 + a_2\boldsymbol{x}_2 + \cdots + a_p\boldsymbol{x}_p;\, a_1, a_2, \ldots, a_p \in \mathbb{R}\big\}$$

と定義する.
(2) $m, n \in \mathbb{N}$ とし, $\boldsymbol{A} \in \mathsf{Mat}(m, n; \mathbb{R})$ とする. $\boldsymbol{A}$ と $\boldsymbol{A}^\top$ の像と核をそれぞれ

$$\begin{aligned}
\mathsf{range}\big(\boldsymbol{A}\big) &= \big\{\boldsymbol{A}\boldsymbol{x};\, \boldsymbol{x} \in \mathbb{R}^n\big\} \subset \mathbb{R}^m, \\
\mathsf{ker}\big(\boldsymbol{A}\big) &= \big\{\boldsymbol{x};\, \boldsymbol{A}\boldsymbol{x} = \boldsymbol{0} \in \mathbb{R}^m\big\} \subset \mathbb{R}^n, \\
\mathsf{range}\big(\boldsymbol{A}^\top\big) &= \big\{\boldsymbol{A}^\top\boldsymbol{y};\, \boldsymbol{y} \in \mathbb{R}^m\big\} \subset \mathbb{R}^n, \\
\mathsf{ker}\big(\boldsymbol{A}^\top\big) &= \big\{\boldsymbol{y};\, \boldsymbol{A}^\top\boldsymbol{y} = \boldsymbol{0} \in \mathbb{R}^n\big\} \subset \mathbb{R}^m
\end{aligned}$$

で定める.

定義 17. $n \in \mathbb{N}$ とする. 行列式は関数 $\det : \mathsf{Mat}(n; \mathbb{R}) \longrightarrow \mathbb{R}$ で以下のように定義する.
(1) 1×1 の行列 (a) の行列式を

$$\det(a) = a.$$

(2) $n \in \mathbb{N}$, $n \geq 2$ とき, 行列

$$\boldsymbol{A} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{pmatrix} \in \mathsf{Mat}(n; \mathbb{R})$$

の行列式を以下のように定める. $i, j = 1, 2, \ldots, n$ に対し $\boldsymbol{A}_{ij}$ を $\boldsymbol{A}$ から i 行と j 列をとった $(n-1) \times (n-1)$ 行列とする. $n \times n$ 行列 $\boldsymbol{A}$ の行列式を

$$\det(\boldsymbol{A}) = \sum_{k=1}^{n} (-1)^{k+1} a_{1k} \det(\boldsymbol{A}_{1k})$$

で定める.

命題 18. $n \times n$ の行列 $\boldsymbol{A} = (\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_n) \in \mathsf{Mat}(n; \mathbb{R})$ に対して, 行列式は以下の性質をみたす.

(1) $k = 1, 2, \ldots, n$ と $d \in \mathbb{R}$ に対し,

$$\det(\boldsymbol{a}_1, \ldots, \lambda \boldsymbol{a}_k, \ldots, \boldsymbol{a}_n) = \lambda \det(\boldsymbol{a}_1, \ldots, \boldsymbol{a}_k, \ldots, \boldsymbol{a}_n).$$

(2) $j \neq k$ に対し,

$$\det(\boldsymbol{a}_1, \ldots, \boldsymbol{a}_k + \lambda \boldsymbol{a}_j, \ldots, \boldsymbol{a}_n) = \lambda \det(\boldsymbol{a}_1, \ldots, \boldsymbol{a}_k, \ldots, \boldsymbol{a}_n).$$

(3) $\det(\boldsymbol{I}_n) = 1$.

(4) $\boldsymbol{A}, \boldsymbol{B} \in \mathsf{Mat}(n; \mathbb{R})$ に対し

$$\det(\boldsymbol{A}\boldsymbol{B}) = \det(\boldsymbol{A}) \det(\boldsymbol{B}).$$

定義 19. $\boldsymbol{A} \in \mathsf{Mat}(n; \mathbb{R})$ に対し, ある $\boldsymbol{B} \in \mathsf{Mat}(n; \mathbb{R})$ が存在し,

$$\boldsymbol{A}\boldsymbol{B} = \boldsymbol{B}\boldsymbol{A} = \boldsymbol{I}_n$$

が成り立つとき, $\boldsymbol{A}$ は可逆であるといい, $\boldsymbol{B}$ を $\boldsymbol{A}$ の**逆行列**といい, これを $\boldsymbol{A}^{-1}$ と記す.

命題 20. $\boldsymbol{A} \in \mathsf{Mat}(n; \mathbb{R})$ とする. このとき

$$\boldsymbol{A} \text{ は可逆} \iff \det(\boldsymbol{A}) \neq 0.$$

定理 21. (特異値分解) $m,\, n \in \mathbb{N}$ とする．任意の $\boldsymbol{A} \in \mathsf{Mat}(m,\, n;\, \mathbb{R})$ は次の分解をもつ．

$$\boldsymbol{A} = \boldsymbol{U}\boldsymbol{D}\boldsymbol{V}^\top.$$

ただし $\boldsymbol{U} \in \mathsf{Mat}(m;\, \mathbb{R})$, $\boldsymbol{V} \in \mathsf{Mat}(n;\, \mathbb{R})$ は直交行列で, $\boldsymbol{D} \in \mathsf{Mat}(m,\, n;\, \mathbb{R})$ は対角行列で, その対角成分は $d_1 \geq d_2 \geq \cdots \geq d_{\min(m,\, n)} \geq 0$ である．$d_j\,(j = 1,\, 2,\, \ldots,\, \min(m,\, n))$ を行列 $\boldsymbol{A}$ の**特異値**といい, $\{(d_j,\, \boldsymbol{u}_j,\, \boldsymbol{v}_j);\, j = 1,\, 2,\, \ldots,\, \min(n,\, n)\}$ を**特異値系**[8]という．

注意 22. $n \geq m$ のとき

$$\boldsymbol{D} = \begin{pmatrix} d_1 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ 0 & d_2 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & & \ddots & \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & d_m & 0 & \cdots & 0 \end{pmatrix}$$

である．$m \geq n$ のとき,

$$\boldsymbol{D} = \begin{pmatrix} d_1 & 0 & \cdots & 0 \\ 0 & d_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & d_n \\ 0 & 0 & \cdots & 0 \\ \vdots & & \vdots & \\ 0 & 0 & \cdots & 0 \end{pmatrix}$$

である． □

定理 21 の証明 $\boldsymbol{y} \in \mathbb{R}^m$ に対し,

$$|\boldsymbol{y}|_2 = \sqrt{\boldsymbol{y}^\top \boldsymbol{y}}$$

とする．ただし, 次元の異なる Euclid 空間のノルムも同じ記号を流用する．行列 $\boldsymbol{A}$ のノルム

$$\|\boldsymbol{A}\| := \sup_{\boldsymbol{x} \in \mathbb{R}^n} \frac{\left|\boldsymbol{A}\boldsymbol{x}\right|_2}{|\boldsymbol{x}|_2}$$

で定義すると

$$d_1 := \|\boldsymbol{A}\| := \sup_{\boldsymbol{x} \in \mathbb{R}^n,\, |\boldsymbol{x}|_2 = 1} \left|\boldsymbol{A}\boldsymbol{x}\right|_2$$

---

[8] $j = 1,\, 2,\, \ldots,\, \min(m,\, n)$ に対して

$$\boldsymbol{A}\boldsymbol{v}_j = d_j \boldsymbol{u}_j$$

となっている．

となる．$d_1 \neq 0$ と仮定する．そうでなければ, 定理の証明は自明となる．
あるベクトル $\boldsymbol{x} \in \mathbb{R}^n$ ($|\boldsymbol{x}|_2 = 1$) が存在して

$$\left|\boldsymbol{A}\boldsymbol{x}\right|_2 = d_1$$

をみたすとする．

$$\boldsymbol{y} = \frac{1}{d_1}\boldsymbol{A}\boldsymbol{x} \in \mathbb{R}^m$$

とおけば

$$|\boldsymbol{y}|_2^2 = \boldsymbol{y}^\top \boldsymbol{y} = \frac{1}{d_1^2}\boldsymbol{x}^\top \boldsymbol{A}^\top \boldsymbol{A}\boldsymbol{x} = \frac{|\boldsymbol{A}\boldsymbol{x}|_2^2}{d_1^2} = 1$$

より $|\boldsymbol{y}|_2 = 1$ となる．$\boldsymbol{v}_2, \boldsymbol{v}_3, \ldots, \boldsymbol{v}_n \in \mathbb{R}^n$ と $\boldsymbol{u}_2, \boldsymbol{u}_3, \ldots, \boldsymbol{u}_m \in \mathbb{R}^m$ をうまく選んで, $\{\boldsymbol{x}, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_n\}$ と $\{\boldsymbol{y}, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_m\}$ はそれぞれ $\mathbb{R}^n$ と $\mathbb{R}^m$ の正規直交基底となるようにする．

$$\boldsymbol{U}_1 = (\boldsymbol{y}, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_m) \in \mathsf{Mat}(m; \mathbb{R}), \quad \boldsymbol{V}_1 = (\boldsymbol{x}, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_n) \in \mathsf{Mat}(n; \mathbb{R}),$$
$$\boldsymbol{A}_1 = \boldsymbol{U}_1^\top \boldsymbol{A}\boldsymbol{V}_1$$

とおけば,

$$\boldsymbol{A}_1 = \boldsymbol{U}_1^\top \boldsymbol{A}\boldsymbol{V}_1 = \begin{bmatrix} \boldsymbol{y}^\top \\ \boldsymbol{u}_2^\top \\ \vdots \\ \boldsymbol{u}_m^\top \end{bmatrix} \boldsymbol{A}[\boldsymbol{x}, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_n] = \begin{bmatrix} \boldsymbol{y}^\top \\ \boldsymbol{u}_2^\top \\ \vdots \\ \boldsymbol{u}_m^\top \end{bmatrix} [\boldsymbol{A}\boldsymbol{x}, \boldsymbol{A}\boldsymbol{v}_2, \ldots, \boldsymbol{A}\boldsymbol{v}_n]$$
$$= \begin{bmatrix} \boldsymbol{y}^\top \\ \boldsymbol{u}_2^\top \\ \vdots \\ \boldsymbol{u}_m^\top \end{bmatrix} [d_1\boldsymbol{y}, \boldsymbol{A}\boldsymbol{v}_2, \ldots, \boldsymbol{A}\boldsymbol{v}_n] = \begin{bmatrix} d_1\boldsymbol{y}^\top\boldsymbol{y} & \boldsymbol{y}^\top\boldsymbol{A}\boldsymbol{v}_2 & \ldots & \boldsymbol{y}^\top\boldsymbol{A}\boldsymbol{v}_n \\ 0 & \boldsymbol{u}_2^\top\boldsymbol{A}\boldsymbol{v}_2 & \ldots & \boldsymbol{u}_2\boldsymbol{A}\boldsymbol{v}_n \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \boldsymbol{u}_m\boldsymbol{A}\boldsymbol{v}_2 & & \ldots\boldsymbol{u}_m\boldsymbol{A}\boldsymbol{v}_n \end{bmatrix}$$
$$=: \begin{bmatrix} d_1 & \boldsymbol{w}^\top \\ \boldsymbol{0} & \boldsymbol{B} \end{bmatrix}$$

と書き直せる．ただし

$$\boldsymbol{w} = \begin{bmatrix} \boldsymbol{y}^\top\boldsymbol{A}\boldsymbol{v}_2 \\ \boldsymbol{y}^\top\boldsymbol{A}\boldsymbol{v}_3 \\ \vdots \\ \boldsymbol{y}^\top\boldsymbol{A}\boldsymbol{v}_n \end{bmatrix} \in \mathbb{R}^{n-1},$$

$$\boldsymbol{B} = \begin{bmatrix} \boldsymbol{u}_2^\top\boldsymbol{A}\boldsymbol{v}_2 & \boldsymbol{u}_2^\top\boldsymbol{A}\boldsymbol{v}_3 & \ldots & \boldsymbol{u}_2^\top\boldsymbol{A}\boldsymbol{v}_n \\ \boldsymbol{u}_3^\top\boldsymbol{A}\boldsymbol{v}_2 & \boldsymbol{u}_3^\top\boldsymbol{A}\boldsymbol{v}3 & \ldots & \boldsymbol{u}_3^\top\boldsymbol{A}\boldsymbol{v}_n \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{u}_m^\top\boldsymbol{A}\boldsymbol{v}_2 & \boldsymbol{u}_m^\top\boldsymbol{A}\boldsymbol{v}_3 & \ldots & \boldsymbol{u}_m^\top\boldsymbol{A}\boldsymbol{v}_n \end{bmatrix} \in \mathsf{Mat}(m-1, n-1; \mathbb{R})$$

である．上記の形に注意して計算すると

$$\boldsymbol{A}_1\begin{bmatrix} d_1 \\ \boldsymbol{w} \end{bmatrix} = \begin{bmatrix} d_1 & \boldsymbol{w}^\top \\ \boldsymbol{0} & \boldsymbol{B} \end{bmatrix}\begin{bmatrix} d_1 \\ \boldsymbol{w} \end{bmatrix} = \begin{bmatrix} d_1^2 + \boldsymbol{w}^\top\boldsymbol{w} \\ \boldsymbol{B}\boldsymbol{w} \end{bmatrix}$$

となる．したがって

$$\left|\boldsymbol{A}_1\begin{bmatrix} d_1 \\ \boldsymbol{w} \end{bmatrix}\right|_2^2 = \left|\begin{bmatrix} d_1^2 + \boldsymbol{w}^\top\boldsymbol{w} \\ \boldsymbol{B}\boldsymbol{w} \end{bmatrix}\right|_2^2 = \{d_1^2+\boldsymbol{w}^\top\boldsymbol{w}\}^2+\boldsymbol{w}^\top\boldsymbol{B}^\top\boldsymbol{B}\boldsymbol{w} \geq \{d_1^2+|\boldsymbol{w}|_2^2\}^2$$

から

$$\left|\boldsymbol{A}_1\begin{bmatrix} d_1 \\ \boldsymbol{w} \end{bmatrix}\right|_2 \geq d_1^2 + |\boldsymbol{w}|_2^2, \quad \left|\frac{1}{\sqrt{d_1^2 + |\boldsymbol{w}|_2}}\begin{bmatrix} d_1 \\ \boldsymbol{w} \end{bmatrix}\right|_2 = 1$$

がわかる．よって

$$\|\boldsymbol{A}_1\| = \sup_{\boldsymbol{x}\in\mathbb{R}^n,\,|\boldsymbol{x}|_2=1}\left|\boldsymbol{A}_1\boldsymbol{x}\right|_2 \geq \frac{1}{\sqrt{d_1^2 + |\boldsymbol{w}|_2^2}}\left|\boldsymbol{A}_1\begin{bmatrix} d_1 \\ \boldsymbol{w} \end{bmatrix}\right|_2 \geq \sqrt{d_1^2 + |\boldsymbol{w}|_2^2}. \tag{25}$$

直交変換に関して $\|\cdot\|$ は不変[9]なので,

$$d_1 = \|\boldsymbol{A}\| = \|\boldsymbol{A}_1\|.$$

である．これと (25) と合わせると

$$d_1 = \|\boldsymbol{A}\| = \|\boldsymbol{A}_1\| \geq \sqrt{d_1^2 + |\boldsymbol{w}|_2^2} \geq d_1$$

となるので,

$$|\boldsymbol{w}|_2 = 0 \Longleftrightarrow \boldsymbol{w} = \boldsymbol{0}.$$

よって

$$\boldsymbol{U}_1^\top\boldsymbol{A}\boldsymbol{V}_1 = \boldsymbol{A}_1 = \begin{bmatrix} d_1 & \boldsymbol{0} \\ \boldsymbol{0}^\top & \boldsymbol{B} \end{bmatrix}$$

と書けることがわかった．

---

[9] $\forall\boldsymbol{x} \in \mathbb{R}^n$ $(|\boldsymbol{x}|_2 = 1)$ に対して $|\boldsymbol{V}_1\boldsymbol{x}|_2 = 1$ に注意すれば,

$$\begin{aligned}\|\boldsymbol{A}\| = \|\boldsymbol{U}_1\boldsymbol{A}_1\boldsymbol{V}_1\| &= \sup_{\boldsymbol{x}\in\mathbb{R}^n;\,|\boldsymbol{x}|_2=1}|\boldsymbol{U}_1\boldsymbol{A}_1\boldsymbol{V}_1\boldsymbol{x}|_2 = \sup_{\boldsymbol{x}\in\mathbb{R}^n;\,|\boldsymbol{x}|_2=1}|\boldsymbol{U}_1\boldsymbol{A}_1\boldsymbol{x}|_2 \\ &= \sup_{\boldsymbol{x}\in\mathbb{R}^n;\,|\boldsymbol{x}|_2=1}\sqrt{\boldsymbol{x}^\top\boldsymbol{A}_1^\top\boldsymbol{U}_1^\top\boldsymbol{U}_1\boldsymbol{A}_1\boldsymbol{x}} = \sup_{\boldsymbol{x}\in\mathbb{R}^n;\,|\boldsymbol{x}|_2=1}\sqrt{\boldsymbol{x}^\top\boldsymbol{A}_1^\top\boldsymbol{A}_1\boldsymbol{x}} \\ &= \sup_{\boldsymbol{x}\in\mathbb{R}^n;\,|\boldsymbol{x}|_2=1}|\boldsymbol{A}_1\boldsymbol{x}|_2 = \|\boldsymbol{A}_1\|\end{aligned}$$

からわかる．

次に, $d_2 = \|\boldsymbol{B}\|$ とおけば,

$$
\begin{aligned}
d_2 &= \sup_{\boldsymbol{x}\in\mathbb{R}^{n-1};\,|\boldsymbol{x}|_2=1} |\boldsymbol{B}\boldsymbol{x}|_2 = \sup_{\boldsymbol{x}\in\mathbb{R}^{n-1};\,|\boldsymbol{x}|_2=1} \left| \begin{bmatrix} d_1 & \boldsymbol{0}^\top \\ \boldsymbol{0} & \boldsymbol{B} \end{bmatrix} \begin{bmatrix} 0 \\ \boldsymbol{x} \end{bmatrix} \right|_2 \\
&\le \sup_{\boldsymbol{z}\in\mathbb{R}^{n};\,|\boldsymbol{z}|_2=1} |\boldsymbol{A}_1\boldsymbol{z}|_2 = \|\boldsymbol{A}_1\| = \|\boldsymbol{A}\| = d_1.
\end{aligned}
$$

$d_2 = 0$ のとき, $\boldsymbol{B} = \boldsymbol{0}$ より証明はおわり. $d_2 > 0$ と仮定して議論を進める. 前と同じようにすれば,

$$d_2 = \|\boldsymbol{B}\|$$

とおき,

$$|\boldsymbol{B}\boldsymbol{x}|_2 = 1,\ \boldsymbol{x} \in \mathbb{R}^{n-1}\ \text{s.t.} |\boldsymbol{x}|_2 = 1,\ \boldsymbol{y} = \frac{1}{d_2}\boldsymbol{B}\boldsymbol{x} \in \mathbb{R}^{m-1}$$

と取り, さきほどの議論を繰り返すと $\widetilde{\boldsymbol{U}}_2 \in \mathsf{Mat}(m-1;\,\mathbb{R})$, $\widetilde{\boldsymbol{V}}_2 \in \mathsf{Mat}(n-1;\,\mathbb{R})$ で

$$\widetilde{\boldsymbol{U}}_2^\top \boldsymbol{B} \widetilde{\boldsymbol{V}}_2 = \begin{bmatrix} d_2 & \boldsymbol{0}^\top \\ \boldsymbol{0} & \boldsymbol{C} \end{bmatrix},\ \boldsymbol{C} \in \mathsf{Mat}(m-2,\,n-2;\,\mathbb{R})$$

とできて,

$$\boldsymbol{U}_2 = \begin{bmatrix} 1 & \boldsymbol{0}^\top \\ \boldsymbol{0} & \widetilde{\boldsymbol{U}}_2 \end{bmatrix},\ \boldsymbol{V}_2 = \begin{bmatrix} 1 & \boldsymbol{0}^\top \\ \boldsymbol{0} & \widetilde{\boldsymbol{V}}_2 \end{bmatrix}$$

とおけば, $\boldsymbol{U}_2$, $\boldsymbol{V}_2$ は直交行列で

$$\boldsymbol{U}_2^\top \boldsymbol{U}_1^\top \boldsymbol{A} \boldsymbol{V}_1 \boldsymbol{V}_2 = \begin{bmatrix} d_1 & 0 & 0 \\ 0 & d_2 & 0 \\ 0 & 0 & \boldsymbol{C} \end{bmatrix}$$

となる. さらに $\boldsymbol{U}_1\boldsymbol{U}_2 \in \mathsf{Mat}(m;\,\mathbb{R})$ と $\boldsymbol{V}_1\boldsymbol{V}_2 \in \mathsf{Mat}(n;\,\mathbb{R})$ も直交行列であることに注意する.

以上の操作を繰り返せば, 定理は証明される. □

定理 23. $\boldsymbol{A} = \boldsymbol{U}\boldsymbol{D}\boldsymbol{V}^\top$ と特異値分解されたとする. ただし, $p \le \min(m,\,n)$ に対し

$$d_1 \ge d_2 \ge \cdots \ge d_p > d_{p+1} = d_{p+2} = \cdots = d_{\min(m,\,n)} = 0$$

とする.

$$\boldsymbol{U} = (\boldsymbol{u}_1,\,\boldsymbol{u}_2,\,\ldots,\,\boldsymbol{u}_m),\ \boldsymbol{V} = (\boldsymbol{v}_1,\,\boldsymbol{v}_2,\,\ldots,\,\boldsymbol{v}_n)$$

としたとき,

$$
\begin{aligned}
\mathsf{ker}(\boldsymbol{A}) &= \mathsf{span}\{\boldsymbol{v}_{p+1},\,\boldsymbol{v}_{p+2},\,\ldots,\,\boldsymbol{v}_n\} = \mathsf{range}(\boldsymbol{A})^\perp, \\
\mathsf{ker}(\boldsymbol{A}^\top) &= \mathsf{span}\big\{\boldsymbol{u}_{p+1},\,\boldsymbol{u}_{p+2},\,\ldots,\,\boldsymbol{u}_m\big\} = \mathsf{range}\big(\boldsymbol{A}^\top\big)^\perp.
\end{aligned}
$$

Proof. $\boldsymbol{U}$ の直交性より

$$\boldsymbol{A}\boldsymbol{x} = \boldsymbol{0}\,(\boldsymbol{x} \in \mathbb{R}^n) \iff \boldsymbol{D}\boldsymbol{V}^\top \boldsymbol{x} = \begin{bmatrix} d_1 \boldsymbol{v}_1^\top \boldsymbol{x} \\ d_2 \boldsymbol{v}_2^\top \boldsymbol{x} \\ \vdots \\ d_p \boldsymbol{v}_p^\top \boldsymbol{x} \\ 0 \\ \vdots \\ 0 \end{bmatrix} = \boldsymbol{0}.$$

よって $\mathsf{ker}(\boldsymbol{A})$ の任意の元 $\boldsymbol{x}$ は $\boldsymbol{v}_1,\, \boldsymbol{v},\, \ldots,\, \boldsymbol{v}_p$ と直交するので,

$$\mathsf{ker}(\boldsymbol{A}) = \mathsf{span}\{\boldsymbol{v}_{p+1},\, \boldsymbol{v}_{p+2},\, \ldots,\, \boldsymbol{v}_n\}.$$

さらに $\boldsymbol{A}^\top = \boldsymbol{V}\boldsymbol{D}^\top \boldsymbol{U}^\top$ より

$$\mathsf{ker}\big(\boldsymbol{A}^\top\big) = \mathsf{span}\{\boldsymbol{u}_{p+1},\, \boldsymbol{u}_{p+2},\, \ldots,\, \boldsymbol{u}_m\} = \mathsf{range}\big(\boldsymbol{A}\big)^\perp.$$

一方, $\forall \widetilde{\boldsymbol{x}} \in \mathsf{range}\big(\boldsymbol{A}^\top\big)$ とすれば, ある $\boldsymbol{y} \in \mathbb{R}^m$ があって, $\widetilde{\boldsymbol{x}} = \boldsymbol{A}^\top \boldsymbol{y}$ とかける. このことより,

$$\boldsymbol{x} \in \mathsf{ker}\big(\boldsymbol{A}\big) \iff \boldsymbol{x}^\top \widetilde{\boldsymbol{x}} = \big(\boldsymbol{A}\boldsymbol{x}\big)^\top \boldsymbol{y} = 0$$

より, $\boldsymbol{x}$ は $\mathsf{range}\big(\boldsymbol{A}^\top\big)$ と直交する. 同様に, $\forall \widetilde{\boldsymbol{y}} \in \mathsf{range}\big(\boldsymbol{A}\big)$ をとる. ある $\boldsymbol{x} \in \mathbb{R}^n$ があって, $\widetilde{\boldsymbol{y}} = \boldsymbol{A}\boldsymbol{x}$ とかける. このことより

$$\boldsymbol{y} \in \mathsf{ker}\big(\boldsymbol{A}^\top\big) \iff \boldsymbol{y}^\top \widetilde{\boldsymbol{y}} = \big(\boldsymbol{A}^\top \boldsymbol{y}\big)^\top \boldsymbol{x} = 0$$

より, $\boldsymbol{y}$ は $\mathsf{range}\big(\boldsymbol{A}\big)$ と直交することもわかる. □

注意 24.

$$\boldsymbol{U}_1 = [\boldsymbol{u}_1,\, \boldsymbol{u}_2,\, \ldots,\, \boldsymbol{u}_p],\, \boldsymbol{U}_2 = [\boldsymbol{u}_{p+1},\, \boldsymbol{u}_{p+2},\, \ldots,\, \boldsymbol{u}_m],$$
$$\boldsymbol{V}_1 = [\boldsymbol{v}_1,\, \boldsymbol{v}_2,\, \ldots,\, \boldsymbol{v}_p],\, \boldsymbol{V}_2 = [\boldsymbol{v}_{p+1},\, \boldsymbol{v}_{p+2},\, \ldots,\, \boldsymbol{v}_n]$$

とおく. すなわち

$$\boldsymbol{U} = [\boldsymbol{U}_1,\, \boldsymbol{U}_2], \quad \boldsymbol{V} = [\boldsymbol{V}_1,\, \boldsymbol{V}_2]$$

である. このとき, $\mathbb{R}^n$ から $\mathsf{ker}\big(\boldsymbol{A}\big)$ への直交射影を $\boldsymbol{P}$, $\mathbb{R}^m$ から $\mathsf{ker}\big(\boldsymbol{A}^\top\big)$ への直交射影を $\boldsymbol{Q}$ とおくと

$$\boldsymbol{P} = \boldsymbol{V}_2 \boldsymbol{V}_2^\top, \quad \boldsymbol{Q} = \boldsymbol{U}_2 \boldsymbol{U}_2^\top$$

となる. □

B　補遺: Moore-Penrose の一般化逆行列

$\boldsymbol{A} \in \mathsf{Mat}(m,\, n;\, \mathbb{R})$ で $\boldsymbol{A} \neq \boldsymbol{0}$ とする．定理 21 から $1 \leq p \leq \min(m,\, n)$ があって

$$
\begin{aligned}
&\boldsymbol{A} = \widetilde{\boldsymbol{U}}\widetilde{\boldsymbol{D}}\widetilde{\boldsymbol{V}}^{\top},\\
&\widetilde{\boldsymbol{U}} \in \mathsf{Mat}(m;\, \mathbb{R}),\ \widetilde{\boldsymbol{V}} \in \mathsf{Mat}(n;\, \mathbb{R}) \text{ は直交行列},\\
&\widetilde{\boldsymbol{D}} \in \mathsf{Mat}(m,\, n;\, \mathbb{R}) \text{ は対角行列 (非対角成分は 0) で}\\
&\text{その正の対角成分は } d_1 \geq d_2 \geq \cdots \geq d_p > 0
\end{aligned}
$$

と書ける．ここで $\widetilde{\boldsymbol{U}},\, \widetilde{\boldsymbol{D}},\, \widetilde{\boldsymbol{V}}$ を分解し

$$
\begin{aligned}
&\widetilde{\boldsymbol{U}} = [\boldsymbol{U},\, \boldsymbol{U}^{\perp}],\ \boldsymbol{U} \in \mathsf{Mat}(m,\, p;\, \mathbb{R}),\\
&\widetilde{\boldsymbol{V}} = [\boldsymbol{V},\, \boldsymbol{V}^{\perp}],\ \boldsymbol{V} \in \mathsf{Mat}(n,\, p;\, \mathbb{R}),\\
&\boldsymbol{D} = \begin{bmatrix} d_1 & 0 & \cdots & 0 \\ 0 & d_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & d_p \end{bmatrix} \in \mathsf{Mat}(p;\, \mathbb{R})
\end{aligned}
$$

とする．すると

$$
\begin{aligned}
\widetilde{\boldsymbol{U}}^{\top}\widetilde{\boldsymbol{U}} &= \begin{bmatrix} \boldsymbol{U}^{\top}\boldsymbol{U} & \boldsymbol{U}^{\top}\boldsymbol{U}^{\perp} \\ (\boldsymbol{U}^{\perp})^{\top}\boldsymbol{U} & (\boldsymbol{U}^{\perp})^{\top}\boldsymbol{U}^{\perp} \end{bmatrix} = \boldsymbol{I}_m,\\
\widetilde{\boldsymbol{V}}^{\top}\widetilde{\boldsymbol{V}} &= \begin{bmatrix} \boldsymbol{V}^{\top}\boldsymbol{V} & \boldsymbol{V}^{\top}\boldsymbol{V}^{\perp} \\ (\boldsymbol{V}^{\perp})^{\top}\boldsymbol{V} & (\boldsymbol{V}^{\perp})^{\top}\boldsymbol{V}^{\perp} \end{bmatrix} = \boldsymbol{I}_n
\end{aligned}
$$

から

$$
\boldsymbol{U}^{\top}\boldsymbol{U} = \boldsymbol{I}_p,\ \boldsymbol{U}^{\top}\boldsymbol{U}^{\perp} = \boldsymbol{0},\ \boldsymbol{V}^{\top}\boldsymbol{V} = \boldsymbol{I}_p,\ \boldsymbol{V}^{\top}\boldsymbol{V}^{\perp} = \boldsymbol{0}
$$

となる．よって

$$
\boldsymbol{A} = \widetilde{\boldsymbol{U}}\widetilde{\boldsymbol{D}}\widetilde{\boldsymbol{V}}^{\top} = \boldsymbol{U}\boldsymbol{D}\boldsymbol{V}^{\top} \tag{26}
$$

となる．ただし

$$
\boldsymbol{U}^{\top}\boldsymbol{U} = \boldsymbol{V}^{\top}\boldsymbol{V} = \boldsymbol{I}_p,\ \boldsymbol{D} = \begin{bmatrix} d_1 & 0 & \cdots & 0 \\ 0 & d_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & d_p \end{bmatrix}
$$

となる．

(26) を踏まえて, $\boldsymbol{A}^\dagger \in \mathsf{Mat}(n,\, m;\, \mathbb{R})$ を

$$\boldsymbol{A}^\dagger := \boldsymbol{V}\boldsymbol{D}^{-1}\boldsymbol{U}^\top,\ \boldsymbol{D}^{-1} = \begin{bmatrix} d_1^{-1} & 0 & \cdots & 0 \\ 0 & d_2^{-1} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & d_p^{-1} \end{bmatrix} \tag{27}$$

と定める. $\boldsymbol{A}^\dagger$ を $\boldsymbol{A}$ の Moore-Penrose **の一般逆行列**と呼ぶことにする.
　すると $\boldsymbol{A}^\dagger$ は下記の関係式をみたす $n \times m$ の行列となる.

$$\boldsymbol{A}\boldsymbol{A}^\dagger\boldsymbol{A} = \boldsymbol{A}, \tag{28}$$

$$\boldsymbol{A}^\dagger\boldsymbol{A}\boldsymbol{A}^\dagger = \boldsymbol{A}^\dagger, \tag{29}$$

$$\boldsymbol{A}\boldsymbol{A}^\dagger = \big(\boldsymbol{A}\boldsymbol{A}^\dagger\big)^\top, \tag{30}$$

$$\boldsymbol{A}^\dagger\boldsymbol{A} = \big(\boldsymbol{A}^\dagger\boldsymbol{A}\big)^\top \tag{31}$$

が成立する. 上記 (28) – (31) は (26) と (27) よりわかる. 実際

$$\begin{aligned}
\boldsymbol{A}\boldsymbol{A}^\dagger\boldsymbol{A} &= \boldsymbol{U}\boldsymbol{D}\boldsymbol{V}^\top\boldsymbol{V}\boldsymbol{D}^{-1}\boldsymbol{U}^\top\boldsymbol{U}\boldsymbol{D}\boldsymbol{V}^\top = \boldsymbol{U}\boldsymbol{D}\boldsymbol{V}^\top = \boldsymbol{A}, \\
\boldsymbol{A}^\dagger\boldsymbol{A}\boldsymbol{A}^\dagger &= \boldsymbol{V}\boldsymbol{D}^{-1}\boldsymbol{U}^\top\boldsymbol{U}\boldsymbol{D}\boldsymbol{V}^\top\boldsymbol{V}\boldsymbol{D}^{-1}\boldsymbol{U}^\top = \boldsymbol{V}\boldsymbol{D}^{-1}\boldsymbol{U}^\top = \boldsymbol{A}^\dagger, \\
\boldsymbol{A}\boldsymbol{A}^\dagger &= \boldsymbol{U}\boldsymbol{D}\boldsymbol{V}^\dagger\boldsymbol{V}\boldsymbol{D}^{-1}\boldsymbol{U}^\dagger = \boldsymbol{U}\boldsymbol{U}^\top = \big(\boldsymbol{U}\boldsymbol{U}^\top\big)^\top = \big(\boldsymbol{A}\boldsymbol{A}^\dagger\big)^\top, \\
\boldsymbol{A}^\dagger\boldsymbol{A} &= \boldsymbol{V}\boldsymbol{D}^{-1}\boldsymbol{U}^\top\boldsymbol{U}\boldsymbol{D}\boldsymbol{V}^\top = \boldsymbol{V}\boldsymbol{V}^\top = \big(\boldsymbol{V}\boldsymbol{V}^\top\big)^\top = \big(\boldsymbol{A}^\dagger\boldsymbol{A}\big)^\top
\end{aligned}$$

からわかる. さらに

$$\begin{aligned}
\boldsymbol{P} &= \boldsymbol{A}\boldsymbol{A}^\dagger : \mathbb{R}^m \longrightarrow \mathrm{ran}(\boldsymbol{A} \subset \mathbb{R}^m, \\
\boldsymbol{Q} &= \boldsymbol{A}^\dagger\boldsymbol{A} : \mathbb{R}^n \longrightarrow \mathrm{ran}\big(\boldsymbol{A}^\top\big) \subset \mathbb{R}^n
\end{aligned}$$

は射影行列になっている. すなわち $\boldsymbol{P} \in \mathsf{Mat}(m;\, \mathbb{R})$ と $\boldsymbol{Q} \in \mathsf{Mat}(n;\, \mathbb{R})$

$$\boldsymbol{P}^2 = \boldsymbol{P},\ \boldsymbol{P}^\top = \boldsymbol{P},\ \boldsymbol{Q}^2 = \boldsymbol{Q},\ \boldsymbol{Q}^\top = \boldsymbol{Q}$$

をみたしている.
　最後に $\boldsymbol{A}^\dagger$ の一意性を示す. $\boldsymbol{B}^\dagger$ も $\boldsymbol{A}$ の Moore-Penrose の一般逆行列とする. すると $\boldsymbol{A}^\dagger$ を $\boldsymbol{B}$ と替えた

$$\boldsymbol{A}\boldsymbol{B}\boldsymbol{A} = \boldsymbol{A}, \tag{32}$$

$$\boldsymbol{B}\boldsymbol{A}\boldsymbol{B} = \boldsymbol{B}, \tag{33}$$

$$\boldsymbol{A}\boldsymbol{B} = \big(\boldsymbol{A}\boldsymbol{B}\big)^\top, \tag{34}$$

$$\boldsymbol{B}\boldsymbol{A} = \big(\boldsymbol{B}\boldsymbol{A}\big)^\top \tag{35}$$

も成立する．すると

$$
\begin{aligned}
\boldsymbol{A}^{\dagger} &= \boldsymbol{A}^{\dagger}\boldsymbol{A}\boldsymbol{A}^{\dagger} && (\because (28)) \\
&= \boldsymbol{A}^{\dagger}\big(\boldsymbol{A}\boldsymbol{B}\boldsymbol{A}\big)\boldsymbol{A}^{\dagger} && (\because (32)) \\
&= \boldsymbol{A}^{\dagger}\big(\boldsymbol{A}\boldsymbol{B}\big)^{\top}\big(\boldsymbol{A}\boldsymbol{A}^{\dagger}\big)^{\top} && (\because (30) \text{ と } (34)) \\
&= \boldsymbol{A}^{\dagger}\boldsymbol{B}^{\top}\big(\boldsymbol{A}\boldsymbol{A}^{\dagger}\boldsymbol{A}\big)^{\top} \\
&= \boldsymbol{A}^{\dagger}\boldsymbol{B}^{\top}\boldsymbol{A}^{\top} && (\because (28)) \\
&= \boldsymbol{A}^{\dagger}\big(\boldsymbol{A}\boldsymbol{B}\big)^{\top} \\
&= \boldsymbol{A}^{\dagger}\boldsymbol{A}\boldsymbol{B} && (\because (34)) \\
&= \boldsymbol{A}^{\dagger}\boldsymbol{A}\boldsymbol{B}\boldsymbol{A}\boldsymbol{B} && (\because (32) \\
&= \big(\boldsymbol{A}^{\dagger}\boldsymbol{A}\big)^{\top}\big(\boldsymbol{B}\boldsymbol{A}\big)^{\top}\boldsymbol{B} && (\because (31 \text{ と } (35)) \\
&= \big(\boldsymbol{A}\boldsymbol{A}^{\dagger}\boldsymbol{A}\big)^{\top}\boldsymbol{B}^{\top}\boldsymbol{B} \\
&= \boldsymbol{A}^{\top}\boldsymbol{B}^{\top}\boldsymbol{B} && (\because (28)) \\
&= \big(\boldsymbol{B}\boldsymbol{A}\big)^{\top}\boldsymbol{B} \\
&= \boldsymbol{B}\boldsymbol{A}\boldsymbol{B} && (\because (35)) \\
&= \boldsymbol{B} && (\because (32))
\end{aligned}
$$

からわかる．よって一意性を証明できた.

注意 25. 多くの本では, $\boldsymbol{A} \in \mathsf{Mat}(m, n; \mathbb{R})$ に対して (28) – (31) をみたす $n \times m$ の行列 $\boldsymbol{A}^{\dagger}$ を $\boldsymbol{A}$ の Moore-Penrose の一般逆行列と定義している. □

C　補遺: 凸関数とその性質

$\boldsymbol{x} = (x_1, x_2, \ldots, x_d)^{\top}$, $\boldsymbol{y} = (y_1, y_2, \ldots, y_d)^{\top} \in \mathbb{R}^d$ に対して

$$
(\boldsymbol{x}|\,\boldsymbol{y}) := \sum_{i=1}^{d} x_i y_i, \qquad |\boldsymbol{x}|_2 = \sqrt{(\boldsymbol{x}|\,\boldsymbol{x})}
$$

と定める．すなわち, $(\cdot|\cdot)$ と $|\cdot|_2$ は $\mathbb{R}^d$ の Euclid 内積とノルムである.

部分集合 $C, D \subset \mathbb{R}^d$ と $k \in \mathbb{R}$ に対して

$$
C + D := \{\boldsymbol{x} + \boldsymbol{y};\, \boldsymbol{x} \in C, \boldsymbol{y} \in D\}, \qquad kC := \{k\boldsymbol{x};\, \boldsymbol{x} \in C\}
$$

と定める．さらに

$$
B := \{\boldsymbol{x} \in \mathbb{R}^d;\, |\boldsymbol{x}|_2 = 1\}
$$

と定める.

部分集合 $D \subset \mathbb{R}^d$ に対して, D を含む最小の閉集合を D の閉包といい, $\mathsf{cl}D$ と記す. すなわち

$$\mathsf{cl}D = \bigcap_{F \subset D;\, F \text{ は閉集合}} F$$

である. また, D に含まれる最大の開集合を D の内部といい, D° と記す. すなわち

$$D^\circ = \bigcup_{C \supset D;\, C \text{ は開集合}} C$$

である. さらに D の境界 ∂D を

$$\partial D = \mathsf{cl}D \setminus D^\circ = \mathsf{cl}D \bigcap \left(D^\circ\right)^c$$

で定める. ただし A^c は集合 A の補集合である.

定義 26. 部分集合 $D \subset \mathbb{R}^d$ は**有界**であるとは, ある $k \in \mathbb{R}$ が存在して, $kB \supset D$ とできるときをいう. D は**コンパクト**であるとは, これが閉かつ有界集合のときをいう.

定理 27 (Bolzano-Weierstrass)**.** $\mathbb{R}^d$ の有界点列 $\{\boldsymbol{x}_n\}_{n=1}^\infty$ は収束する部分列 $\{\boldsymbol{x}_{n(k)}\}_{k=1}^\infty$ をもつ.

Proof. たとえば, [14] の p.26 を参照のこと. □

定義 28. $D \subset \mathbb{R}^d$ を空でない部分集合とする. 関数 $f : D \longrightarrow \mathbb{R}$ は点 $\boldsymbol{x} \in D$ で連続であるとは, $\boldsymbol{x}$ に収束する D の任意の点列 $\{\boldsymbol{x}_n\}_{n=1}^\infty$ に対して

$$f(\boldsymbol{x}_n) \longrightarrow f(\boldsymbol{x}) \qquad (n \to \infty)$$

が成り立つときをいう. 任意の $\boldsymbol{x} \in D$ で f は連続のとき, f は D 上で連続であるという.

定理 29. $D \subset \mathbb{R}^d$ を空でない有界閉集合とする. 関数 $f : D \longrightarrow \mathbb{R}$ は D で連続ならば, f は D で最大値と最小値を取る.

Proof. たとえば, [14] の p.85 を参照せよ. □

定義 30. $D \subset \mathbb{R}^d$ を空でない部分集合とする. 関数 $f : D \longrightarrow \mathbb{R}$ と $a \in \mathbb{R}$ に対して, a に対する f の**レベル集合**を

$$\{\boldsymbol{x} \in D;\, f(\boldsymbol{x}) \le a\}$$

で定める.

定義 31. (1) 部分集合 $A \subset \mathbb{R}$ に対して, $b \in \mathbb{R}$ は A の**上界**であるとは, $\forall a \in A$ に対して $a \leq b$ が成り立つときをいう. また, $e \in \mathbb{R}$ は A の**下界**であるとは, $\forall a \in A$ に対して $e \leq a$ が成り立つときをいう.
(2) A の上界すべてがなす集合の最小限 m が存在すれば, m を A の**上限**といい, $\sup A$ と記す. また, A の下界すべてがなす集合の最大限 ℓ が存在すれば, ℓ を A の**下限**といい, $\inf A$ と記す.

定義 32. $D \subset \mathbb{R}^d$ を空でない部分集合とする. 関数 $f : D \longrightarrow \mathbb{R}$ の (大域的) **最小解** $\boldsymbol{x}_0 \in D$ を

$$f(\boldsymbol{x}_0) = \inf\bigl\{f(\boldsymbol{x}); \boldsymbol{x} \in D\bigr\}$$

をみたす点である.

命題 33. $D \subset \mathbb{R}^d$ は空でない閉部分集合とする. 連続関数 $f : D \longrightarrow \mathbb{R}$ のすべてのレベル集合は有界とする. このとき, f は大域的最小解を持つ.

Proof. 定理 29 からわかる. □

定義 34. (1) 部分集合 $C \subset \mathbb{R}^d$ は**凸集合**であるとは, $\forall \boldsymbol{x}, \boldsymbol{y} \in C$ と $0 < \forall \lambda < 1$ に対して

$$\lambda \boldsymbol{x} + (1-\lambda)\boldsymbol{x} \in C$$

が成り立つときをいう.
(2) $C \subset \mathbb{R}^d$ を空でない凸部分集合とする. 関数 $f : C \longrightarrow \mathbb{R}$ は**凸関数**であるとは, $\forall \boldsymbol{x}, \boldsymbol{y} \in C$ と $0 \leq \forall \lambda \leq 1$ に対して

$$f\bigl(\lambda \boldsymbol{x} + (1-\lambda)\boldsymbol{y}\bigr) \leq \lambda f(\boldsymbol{x}) + (1-\lambda) f(\boldsymbol{y})$$

をみたすときをいう.
(3) $C \subset \mathbb{R}^d$ を空でない凸部分集合とする. 関数 $f : C \longrightarrow \mathbb{R}$ は**狭義凸関数**であるとは, $\forall \boldsymbol{x}, \boldsymbol{y} \in C\,(\boldsymbol{x} \neq \boldsymbol{y})$ と $0 < \forall \lambda < 1$ に対して

$$f\bigl(\lambda \boldsymbol{x} + (1-\lambda)\boldsymbol{y}\bigr) < \lambda f(\boldsymbol{x}) + (1-\lambda) f(\boldsymbol{y})$$

をみたすときをいう.

命題 35. $C \subset \mathbb{R}^d$ を空でない凸部分集合とする. 凸関数 $f : C \longrightarrow \mathbb{R}$ が有界なレベル集合を持つための必要十分条件は

$$\lim_{r \to +\infty} \inf\left\{ \frac{|f(\boldsymbol{x})|}{|\boldsymbol{x}|_2}; \boldsymbol{x} \in C \bigcap (rB)^c \right\} > 0 \tag{36}$$

をみたすことである.

Proof. 省略. □

定理 36. $C \subset \mathbb{R}^d$ を空でない凸部分集合とする. 連続な凸関数 $f : C \longrightarrow \mathbb{R}$ が条件 (36) をみたせば, f は C で大域的最小解をもつ.

Proof. 命題 35 と 命題 33 から f は大域的最小解をもつことがわかる. □

注意 37. $C \subset \mathbb{R}^d$ を空でない凸部分集合とする. 狭義凸関数 $f : C \longrightarrow \mathbb{R}$ が条件 (36) をみたせば, f は C で大域的最小解を唯一もつことがわかる.

D　レポート問題

2023 年度数学概論 A(6 月 21 日の回)
レポート問題

- 〆切: 2023 年 <u>7 月 13 日（水）12:00</u>.
- レポート冒頭には学籍番号と名前.
- レポートは $\boxed{\text{問題 1}}$ と $\boxed{\text{問題 2}}$ です.
- pdf ファイルで提出. ファイル名も学生番号と名前とすること.
- 提出先: konno@omu.ac.jp

  レポートの提出後 3 日経過しても受領確認のメイルが来ない場合には, 確認のメイルを再度送信してください.
- レポートは LaTeX で作成せよ. ただし, LaTeX で作成するのを好まないときには, その理由を述べた上で別の方法で作成せよ.
- 問題は p.34 まであります.

$\boxed{\text{問題 1}}$ (a) ～ (f) から 1 題だけを選択して解答せよ.

(a) (8) で与えられる最小化問題の解は (9) であることを証明せよ.

(b) 等式 (18) を証明せよ.

(c) (16) における 2 番目の等号を確認せよ.

(d) $n \geq d\,(n, d \in \mathbb{N})$ と $\boldsymbol{X} \in \mathsf{Mat}(n, d; \mathbb{R})$ とする. $r > 0$ のとき, $\boldsymbol{X}^\top \boldsymbol{X} + r\boldsymbol{I}_d$ は正則であることを示せ. ただし $\boldsymbol{I}_d$ は d 次の単位行列である.

(e) (17) における 2 番目の等号を確認せよ.

(f) $n \geq d\,(n, d \in \mathbb{N})$ と $\boldsymbol{X} \in \mathsf{Mat}(n, d; \mathbb{R})$ とする. このとき, $\mathsf{rank}\boldsymbol{X} = d$ ならば, $\boldsymbol{X}^\top \boldsymbol{X}$ は正則であることを示せ.

なお, 証明でよく教科書に載っているような定理は証明なしで用いてよい. ただし, 用いた定理の主張を書き, 引用元 (どの本の何ページか) を明示すること. 引用文献は著者・書名・出版社・発行年度を書くこと.

$\boxed{\text{問題 2}}$ 資料中の数学的事実の誤り, 不明瞭な箇所, 疑問点, タイプミスを 1 個以上指摘せよ. 不明瞭な箇所や疑問点については考えたことを丁寧に説明せよ. 些細な誤りでもよい. また, 探索範囲は補遺も含む.
=== 訂正箇所の書き方 ===

- Page 3, line 10:「****」は「++++」.
- Page 4, line -5:「****」は「++++」.「line -5」は下から数えて 5 行目という意味です. 頁はそれぞれのページの下にある頁番号を採用していください.

以上.