「機械学習のための関数解析入門」

サポートページ

（文責：瀬戸）

訂正のページ

(1) p. 13 の下から 9 行目,「零ではないベクトル」→「零ではない複素ベクトル」

(2) p. 49 の下から 6 行目，ここで出てきた $\boldsymbol{x}_k \in \mathcal{M}$ は最初に考えた $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$ とは別のベクトルです．また，その後の計算では $c_j = \langle \boldsymbol{x}, \boldsymbol{e}_j \rangle$ とおいて $c_j^{(k)} \to c_j$ を利用した方が簡単だったかもしれません.

(3) p. 57 の 2 行目から 5 行目，$\dim \mathcal{M}^\perp = 1$ を示すところでもたついているので以下のように差し替えてください.

$\mathbf{0}$ ではない $\boldsymbol{z} \in \mathcal{M}^\perp$ と任意の $\boldsymbol{y} \in \mathcal{M}^\perp$ に対し，$\varphi(\boldsymbol{z}) \neq 0$ であるから，定数 λ を選んで，$\lambda\varphi(\boldsymbol{z}) = \varphi(\boldsymbol{y})$ とできる．このとき，$\varphi(\lambda\boldsymbol{z} - \boldsymbol{y}) = \lambda\varphi(\boldsymbol{z}) - \varphi(\boldsymbol{y}) = 0$ であるから，$\lambda\boldsymbol{z} - \boldsymbol{y} \in \mathcal{M}$ となる．一方，$\lambda\boldsymbol{z} - \boldsymbol{y} \in \mathcal{M}^\perp$ でもある．よって，$\lambda\boldsymbol{z} - \boldsymbol{y} = \mathbf{0}$，すなわち，$\boldsymbol{y} = \lambda\boldsymbol{z}$ である．以上のことから，$\dim \mathcal{M}^\perp = 1$ が得られた.

(4) p. 58 の (ii′) は誤解を招く表現でした．k_x は x だけに依存して定まることが重要です．(ii′) を以下のように訂正してください.

(ii′)　任意の $x \in X$ に対し，次をみたす $k_x \in \mathcal{H}$ が存在する：

$$f(x) = \langle f, k_x \rangle_{\mathcal{H}} \quad (f \in \mathcal{H}) \quad \textbf{（再生核等式）}.$$

(5) p. 63 の下から 7 行目と 8 行目（例 4.1.4 内の 1 行目と 2 行目),「$\mathbb{R}^n$」→「$\mathbb{R}^N$」

(6) p. 79 の解答の後に「この二つの例題は自明なものであったが，問題 $\widetilde{\mathrm{C}}$ を解くにはうまいカーネル関数を見つければよいことが了解できたと思う．」を追加.

(7) p. 80 の定理 4.3.3 では $\Phi(x) = k_x$ を想定しています（p. 148 の議論を念頭においています).

(8) p. 101 の下から 6 行目,「独立な確率変数」→「~~独立な~~確率変数」

(9) p. 107 の下から 5 行目,「$\boldsymbol{x}_* = (x_{n+1}, \ldots, x_m)^\top$」→「$\boldsymbol{x}_* = (x_{n+1}, \ldots, x_{n+m})^\top$」

(10) p. 123 の補題 A.3.1 の前に「また,

$$\begin{pmatrix} A & B \\ C & D \end{pmatrix}^\top = \begin{pmatrix} A^\top & C^\top \\ B^\top & D^\top \end{pmatrix}, \quad \det \begin{pmatrix} A & B \\ O & D \end{pmatrix} = (\det A)(\det D)$$

が成り立つ．」を追加.

(11) p. 140 の下から 9 行目,「$P(\{w_j\}) = 1/6$」→「$P(\{\omega_j\}) = 1/6$」

(12) p. 143 の下から 3 行目,「マージンを達成する」→「最大マージンを達成する」

(13) p. 143 の下から 2 行目,「H」→「H_{max}」

(14) p. 147 の最後の行から p. 148 の最初の行,「Φ を特徴写像とし,」→「特徴写像として $\Phi(x_j) = k_{x_j}$ を選び,」

補足のページ

1　条件付き分布

p. 111 では条件付き分布の確率密度関数を天下り式に定義したが，ここではその定義が妥当なものであることを解説しよう．$X \in [x_0, x_0+h]$ のときに $Y \in A$ となる条件付き確率は

$$P(Y \in A \mid X \in [x_0, x_0+h]) = \frac{P((X,Y) \in [x_0, x_0+h] \times A)}{P((X,Y) \in [x_0, x_0+h] \times \mathbb{R})}$$

と表すことができる．さらに，P の確率密度関数を p としよう．すなわち，

$$P((X,Y) \in S) = \iint_S p(x,y)\ dxdy$$

を仮定する．このとき，

$$\begin{aligned}
P(Y \in A \mid X \in [x_0, x_0+h]) &= \frac{\iint_{[x_0,x_0+h]\times A} p(x,y)\ dxdy}{\iint_{[x_0,x_0+h]\times \mathbb{R}} p(x,y)\ dxdy} \\
&= \frac{\int_A \left(\int_{x_0}^{x_0+h} p(x,y)\ dx\right)\ dy}{\int_{\mathbb{R}} \left(\int_{x_0}^{x_0+h} p(x,y)\ dx\right)\ dy} \\
&= \frac{\int_A \left(\frac{1}{h}\int_{x_0}^{x_0+h} p(x,y)\ dx\right)\ dy}{\int_{\mathbb{R}} \left(\frac{1}{h}\int_{x_0}^{x_0+h} p(x,y)\ dx\right)\ dy} \\
&\to \frac{\int_A p(x_0,y)\ dy}{\int_{\mathbb{R}} p(x_0,y)\ dy} \quad (h \to 0)
\end{aligned}$$

が成り立つ．このようにして，$X = x_0$ のときの確率密度関数の定義

$$P(Y \in A \mid X = x_0) = \int_A \gamma p(x_0, y)\ dy \quad \left(1/\gamma = \int_{\mathbb{R}} p(x_0,y)\ dy\right)$$

が妥当であることがわかる．参考までに，最後の極限は微分積分学の基本定理から導かれる．その際，極限と積分の順序を交換しているが，ガウス分布のように遠方で急速に 0 に収束する関数だけを考えるのであれば，全く気にしないでよい．ルベーグ積分論がこういったこと保証してくれる．ルベーグ積分論については Rudin [5] を薦める．

2　ガウス過程回帰とカーネル回帰

pp. 112–113 で求めた予測は問題 A の解と完全に対応していることを示そう．まず，p. 112 で $(K_n + \beta^{-1} I_n)^{-1} \widehat{\boldsymbol{z}} = (c_1, \ldots, c_n)^\top$ と定めたが，これは

$$\|(K_n + \beta^{-1} I_n)\boldsymbol{c} - \widehat{\boldsymbol{z}}\|_{\mathbb{R}^n} = 0$$

の解である．次に，δ をデルタ関数

$$\delta(x, y) = \begin{cases} 1 & (x = y) \\ 0 & (x \neq y) \end{cases}$$

とし，$\widetilde{k}(x, y) = k(x, y) + \beta^{-1}\delta(x, y)$ とおく．$\widetilde{k}$ は明らかにカーネル関数である．このとき，$\widehat{\boldsymbol{z}} = (\widehat{z}_1, \ldots, \widehat{z}_n)^\top$ に対し，

$$L(f) = \sum_{j=1}^{n} |f(x_j) - \widehat{z}_j|^2$$

を最小にする $f \in \mathcal{H}_{\widetilde{k}}$ は，pp. 71–72 で求めたように，

$$f = \sum_{j=1}^{n} c_j \widetilde{k}_{x_j}$$

で与えられる．この f と $\ell = 1, \ldots, m$ に対し，

$$\begin{aligned} f(x_{n+\ell}) &= \sum_{j=1}^{n} c_j \widetilde{k}_{x_j}(x_{n+\ell}) \\ &= \sum_{j=1}^{n} c_j (k(x_{n+\ell}, x_j) + \beta^{-1}\delta(x_{n+\ell}, x_j)) \\ &= \sum_{j=1}^{n} c_j k(x_{n+\ell}, x_j) \\ &= \sum_{j=1}^{n} c_j k_{x_j}(x_{n+\ell}) \end{aligned}$$

が成り立つ．これはガウス過程回帰で得られた予測とまったく同じものである．

3 ハーディ空間

第 2 章の最後に，コーシーの積分公式と L^2-内積の関係を紹介した．この関係の背後にある再生核ヒルベルト空間はハーディ空間 H^2 である．以下，$\mathrm{Hol}(\mathbb{D})$ は $\mathbb{D}$ 上で正則な関数の全体とする．任意の $f \in \mathrm{Hol}(\mathbb{D})$ は次のようにべき級数展開される．

$$f(z) = \sum_{n=0}^{\infty} c_n z^n.$$

ハーディ空間 H^2 の定義は次の通り．

$$H^2 = \left\{ f \in \mathrm{Hol}(\mathbb{D}) : \sum_{n=0}^{\infty} |c_n|^2 < \infty \ \left(f(z) = \sum_{n=0}^{\infty} c_n z^n \right) \right\}.$$

H^2 は次の内積により（複素）ヒルベルト空間になる．

$$\langle f, g \rangle_{H^2} = \sum_{j=0}^{\infty} c_n \overline{d_n} \quad \left(f(z) = \sum_{n=0}^{\infty} c_n z^n,\ g(z) = \sum_{n=0}^{\infty} d_n z^n \right).$$

これはフーリエ変換

$$f(z) = \sum_{n=0}^{\infty} c_n z^n \mapsto (c_0, c_1, c_2, \ldots, c_n, \ldots)$$

を介して複素 ℓ^2 空間

$$\ell^2_{\mathbb{C}} = \ell^2_{\mathbb{C}}(\mathbb{Z}_{\geq 0}) = \{(c_n) \subset \mathbb{C} : \sum_{n=0}^{\infty} |c_n|^2 < \infty\}$$

を考えることと同じである．特に，p. 37 で与えた k_λ は等比級数の公式により，

$$k_\lambda(z) = \frac{1}{1 - \overline{\lambda} z} = \sum_{n=0}^{\infty} \overline{\lambda}^n z^n \quad (\lambda \in \mathbb{D})$$

と表される．よって，

$$\langle f, k_\lambda \rangle_{H^2} = \sum_{n=0}^{\infty} c_n \lambda^n = f(\lambda) \quad (f \in H^2)$$

が成り立つ．従って，H^2 は再生核ヒルベルト空間である．ハーディ空間 H^2 の基本的な理論については Rudin [5] を参照せよ．制御理論との関係は木村 [4]，ガウス過程との関連は Dym-McKean [2] や飛田・櫃田 [3] を参照せよ．

4　指数カーネル

$\boldsymbol{x} = (x_1, \dots, x_d)^\top \in \mathbb{R}^d$ に対し，

$$\|\boldsymbol{x}\|_1 = |x_1| + \cdots + |x_d|$$

と定める．$\|\boldsymbol{x}\|_1$ は $\boldsymbol{x}$ の ℓ^1-ノルムとよばれる．このとき，任意の $\gamma > 0$ に対し，

$$k(\boldsymbol{x}, \boldsymbol{y}) = \exp(-\gamma\|\boldsymbol{x} - \boldsymbol{y}\|_1) \quad (\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^d)$$

は正定値なカーネル関数であることを示そう．まず，次の補題を用意する．

補題 4.1. i を虚数単位，$\langle \cdot, \cdot \rangle$ を $\mathbb{R}^d$ の通常の内積とする．このとき，異なる N 個の $\boldsymbol{x}_1, \dots, \boldsymbol{x}_N \in \mathbb{R}^d$ に対し，$\{e^{2\pi i \langle \boldsymbol{x}, \boldsymbol{x}_1 \rangle}, \dots, e^{2\pi i \langle \boldsymbol{x}, \boldsymbol{x}_N \rangle}\}$ は線形独立である．

証明． pp. 127–128 と同様に考えて，

$$\prod_{1 \le n < m \le N} (e^{2\pi i \langle \boldsymbol{x}_0, \boldsymbol{x}_n \rangle} - e^{2\pi i \langle \boldsymbol{x}_0, \boldsymbol{x}_m \rangle}) \neq 0$$

をみたす $\boldsymbol{x}_0 \in \mathbb{R}^d$ が存在することを示せばよい．そこで，任意の $\boldsymbol{x} \in \mathbb{R}^d$ に対し，

$$e^{2\pi i \langle \boldsymbol{x}, \boldsymbol{x}_{n(\boldsymbol{x})} \rangle} - e^{2\pi i \langle \boldsymbol{x}, \boldsymbol{x}_{m(\boldsymbol{x})} \rangle} = 0$$

をみたす $1 \le n(\boldsymbol{x}) < m(\boldsymbol{x}) \le N$ が選べたと仮定しよう．このとき，

$$\langle \boldsymbol{x}, \boldsymbol{x}_{n(\boldsymbol{x})} \rangle - \langle \boldsymbol{x}, \boldsymbol{x}_{m(\boldsymbol{x})} \rangle$$

は整数であるから，

$$\mathbb{R}^d = \bigcup_{1 \le n < m \le N} \bigcup_{\ell \in \mathbb{Z}} \left\{ \boldsymbol{x} \in \mathbb{R}^d : \left\langle \boldsymbol{x} - \ell \frac{\boldsymbol{x}_n - \boldsymbol{x}_m}{\|\boldsymbol{x}_n - \boldsymbol{x}_m\|^2}, \boldsymbol{x}_n - \boldsymbol{x}_m \right\rangle = 0 \right\}$$

が成り立つ．しかし，この等式の右辺は超平面の可算和集合であるからその体積は 0 である．従って，これは矛盾である．　□

定理 4.2.

$$k(\boldsymbol{x}, \boldsymbol{y}) = \exp(-\gamma\|\boldsymbol{x} - \boldsymbol{y}\|_1) \quad (\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^d)$$

は正定値なカーネル関数である．

証明．対称性は明らかであるから，正定値性を示す．フーリエ変換の公式

$$e^{-\gamma|x|} = \int_{-\infty}^{\infty} \frac{2\gamma^2}{\gamma^2 + 4\pi^2\xi^2} e^{2\pi i x \xi}\, d\xi$$

を応用する．ここで，i は虚数単位である．以下，

$$\boldsymbol{x}_n = (x_1^{(n)}, \ldots, x_d^{(n)})^\top, \quad \xi = (\xi_1, \ldots, \xi_d)^\top, \quad F(\xi) = \prod_{j=1}^{d} \frac{2\gamma^2}{\gamma^2 + 4\pi^2\xi_j^2}$$

と略記すれば，

$$\begin{aligned}
\sum_{n,m=1}^{N} c_n c_m k(\boldsymbol{x}_n, \boldsymbol{x}_m) &= \sum_{n,m=1}^{N} c_n c_m \exp(-\gamma\|\boldsymbol{x}_n - \boldsymbol{x}_m\|_1) \\
&= \sum_{n,m=1}^{N} c_n c_m \prod_{j=1}^{d} \exp(-\gamma|x_j^{(n)} - x_j^{(m)}|) \\
&= \sum_{n,m=1}^{N} c_n c_m \prod_{j=1}^{d} \int_{-\infty}^{\infty} \frac{2\gamma^2}{\gamma^2 + 4\pi^2\xi_j^2} e^{2\pi i (x_j^{(n)} - x_j^{(m)})\xi_j}\, d\xi_j \\
&= \sum_{n,m=1}^{N} c_n c_m \int_{\mathbb{R}^d} \left(\prod_{j=1}^{d} e^{2\pi i (x_j^{(n)} - x_j^{(m)})\xi_j} \right) F(\xi)\, d\xi \\
&= \int_{\mathbb{R}^d} \left(\sum_{n,m=1}^{N} c_n c_m e^{2\pi i \langle \boldsymbol{x}_n - \boldsymbol{x}_m, \xi \rangle} \right) F(\xi)\, d\xi \\
&= \int_{\mathbb{R}^d} \left| \sum_{n=1}^{N} c_n e^{2\pi i \langle \boldsymbol{x}_n, \xi \rangle} \right|^2 F(\xi)\, d\xi \geq 0
\end{aligned}$$

を得る．以上の計算により，k は半正定値であり，k がカーネル関数であることがわかった．さらに，$\displaystyle\sum_{n,m=1}^{N} c_n c_m k(\boldsymbol{x}_n, \boldsymbol{x}_m) = 0$ のとき，上の等式から $\displaystyle\sum_{n=1}^{N} c_n e^{2\pi i \langle \boldsymbol{x}_n, \xi \rangle} = 0$ が導かれる．このとき，補題 4.1 により，$c_1 = \cdots = c_N = 0$ を得る．従って，k は正定値である． □

5 ガウス過程

ガウス過程の正確な定義は p. 108 の脚注に述べた．それを完全版ガウス過程ということにすると，pp. 105–108 で構成した $\boldsymbol{y}(\boldsymbol{w}, \boldsymbol{x})$ は簡易版ガウス過程である．本書の範囲で

完全版を使ったとしても，ガウス過程回帰の結果は同じである．以下，完全版と簡易版の違いを解説する．さて，5.3 の準備では，入力 $x_1, x_2, \ldots$ から $\mathcal{H}_k$ のベクトル $\varphi_1, \varphi_2, \ldots$ を次のように構成した．

$$
\begin{aligned}
& x_1 \mapsto k_{x_1} \mapsto \varphi_1 \\
& x_2 \mapsto k_{x_2} \mapsto \varphi_2 \quad (\varphi_2 \in \{\varphi_1\}^\perp = \{k_{x_1}\}^\perp) \\
& \quad \vdots \\
& x_n \mapsto k_{x_n} \mapsto \varphi_n \quad (\varphi_n \in \{\varphi_1, \ldots, \varphi_{n-1}\}^\perp = \{k_{x_1}, \ldots, k_{x_{n-1}}\}^\perp).
\end{aligned}
$$

最初の $\mapsto$ はカーネルトリックであり，次の $\mapsto$ はグラム・シュミットの直交化法を表す．大事なことは，時系列に与えられた $x_1, x_2, \ldots$ から一定の手順で $\varphi_1, \varphi_2, \ldots$ を構成できたことである．このように，$x_1, x_2, \ldots$ が時系列に与えられているとき，$\mathbb{N}$ 上で定義された写像

$$
n \mapsto \boldsymbol{y}_n(\boldsymbol{w}) = \begin{pmatrix} \displaystyle\sum_{j=1}^n \varphi_j(x_1) w_j \\ \vdots \\ \displaystyle\sum_{j=1}^n \varphi_j(x_n) w_j \end{pmatrix} \sim N(\boldsymbol{0}, K(x_1, \ldots, x_n))
$$

が定まる．

さて，完全版では，任意の $n \in \mathbb{N}$ と任意の $x_{i_1}, \ldots, x_{i_n} \in X$ に対し，共分散行列を

$$
K(x_{i_1}, \ldots, x_{i_n}) = \begin{pmatrix} k(x_{i_1}, x_{i_1}) & \cdots & k(x_{i_1}, x_{i_n}) \\ \vdots & \ddots & \vdots \\ k(x_{i_n}, x_{i_1}) & \cdots & k(x_{i_n}, x_{i_n}) \end{pmatrix}
$$

とする多次元ガウス分布に従う確率ベクトル $\boldsymbol{y}(\boldsymbol{w}; x_{i_1}, \ldots, x_{i_n})$ が存在することを主張する．すなわち，写像

$$
(x_{i_1}, \ldots, x_{i_n}) \mapsto \boldsymbol{y}(\boldsymbol{w}; x_{i_1}, \ldots, x_{i_n}) \sim N(\boldsymbol{0}, K(x_{i_1}, \ldots, x_{i_n}))
$$

が存在する．完全版ガウス過程を順序のついた $(x_1, \ldots, x_n)$ に限定したものが簡易版ガウス過程である．完全版が存在することを示すにはコルモゴロフの拡張定理の一般形が必要であり，かなり専門的になる．詳細は飛田・櫃田 [3] を参照せよ．

6　最適化問題と KKT 条件

サポートベクトルマシンに必要な範囲で KKT 条件を導出する．この節の中では，ベクトル $\boldsymbol{x} = (x_1, \ldots, x_n)^\top \in \mathbb{R}^n$ に対し，$x_j \geq 0$ $(j = 1, \ldots, n)$ のとき，$\boldsymbol{x} \geq 0$ と表す．また，$-\boldsymbol{x} \geq 0$ のとき，$\boldsymbol{x} \leq 0$ と表す．

6.1　FJ 条件

行列 A とベクトル $\boldsymbol{b}$ を

$$A = \begin{pmatrix} \boldsymbol{a}_1^\top \\ \vdots \\ \boldsymbol{a}_m^\top \end{pmatrix}, \quad \boldsymbol{a}_k = \begin{pmatrix} a_{k1} \\ \vdots \\ a_{kn} \end{pmatrix} \in \mathbb{R}^n, \quad \boldsymbol{b} = \begin{pmatrix} b_1 \\ \vdots \\ b_m \end{pmatrix} \in \mathbb{R}^m$$

と定め固定する．このとき，次の最適化問題 (P)

$$\begin{aligned} \text{最小化} &: f(\boldsymbol{x}) \\ \text{条件} &: \boldsymbol{x} \in \mathcal{F}_P := \{\boldsymbol{x} \in \mathbb{R}^n : \langle \boldsymbol{a}_k, \boldsymbol{x} \rangle \leq b_k \ (1 \leq k \leq m)\} \end{aligned}$$

を考える．この問題 (P) に解が存在するための必要条件は FJ 条件 (the Fritz John conditions for problem (P)) として知られている．

FJ 条件

$\boldsymbol{x}_P$ が (P) の極小解ならば，以下を満たす $(\lambda, \mu_1, \ldots, \mu_m)^\top \in \mathbb{R}_+^{m+1} \setminus \{\boldsymbol{0}\}$ が存在する．

$$\begin{aligned} &\lambda \nabla f(\boldsymbol{x}_P) + \sum_{k=1}^m \mu_k \boldsymbol{a}_k = \boldsymbol{0}, \\ &\mu_k(\langle \boldsymbol{a}_k, \boldsymbol{x}_P \rangle - b_k) = 0 \quad (1 \leq k \leq m). \end{aligned}$$

任意の $\boldsymbol{x} \in \mathcal{F}_P$ に対し，

$$K(\boldsymbol{x}) = \{1 \leq k \leq m : \langle \boldsymbol{a}_k, \boldsymbol{x} \rangle = b_k\}$$

と定める．また，$\boldsymbol{a}_k^\top$ $(k \in K(\boldsymbol{x}))$ とその他の行を 0 として構成される行列を $B(\boldsymbol{x})$ と表す．例えば，$m = 3$, $K(\boldsymbol{x}) = \{1, 3\}$ のとき，

$$B(\boldsymbol{x}) = \begin{pmatrix} \boldsymbol{a}_1^\top \\ \boldsymbol{0}^\top \\ \boldsymbol{a}_3^\top \end{pmatrix} = \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ 0 & \cdots & 0 \\ a_{31} & \cdots & a_{3n} \end{pmatrix}$$

となる．

補題 6.1 (Birbil-Frenk-Still [1])**.** $\boldsymbol{x}_P$ が (P) の極小解であるならば，$\boldsymbol{d} \in \mathbb{R}^n$ に対し，

$$B(\boldsymbol{x}_P)\boldsymbol{d} \leq \boldsymbol{0} \Rightarrow \langle \nabla f(\boldsymbol{x}_P), \boldsymbol{d} \rangle \geq 0$$

が成り立つ．

証明 . $\boldsymbol{x}_P$ は (P) の極小解であるが，

$$\exists \boldsymbol{d}_0 \in \mathbb{R}^n \quad \text{s.t.} \quad B(\boldsymbol{x}_P)\boldsymbol{d}_0 \leq \boldsymbol{0} \quad \text{and} \quad \langle \nabla f(\boldsymbol{x}_P), \boldsymbol{d}_0 \rangle < 0$$

が成り立つと仮定する．このとき，

$$\langle \nabla f(\boldsymbol{x}_P), \boldsymbol{d}_0 \rangle = \nabla f(\boldsymbol{x}_P)^\top \boldsymbol{d}_0 = \lim_{t \downarrow 0} \frac{f(\boldsymbol{x}_P + t\boldsymbol{d}_0) - f(\boldsymbol{x}_P)}{t}$$

であるから，

$$\exists t_0 > 0 \quad \text{s.t.} \quad f(\boldsymbol{x}_P + t\boldsymbol{d}_0) < f(\boldsymbol{x}_P) \quad (0 < t \leq t_0)$$

が成り立つ．さらに，$B(\boldsymbol{x}_P)$ の定め方と $B(\boldsymbol{x}_P)\boldsymbol{d}_0 \leq \boldsymbol{0}$ により，

$$A(\boldsymbol{x}_P + t\boldsymbol{d}_0) \leq \boldsymbol{b} \quad (0 < t \leq t_0)$$

となるように t_0 を選ぶことができる（例 6.2 を参照）．すなわち，$\boldsymbol{x}_P + t\boldsymbol{d}_0 \in \mathcal{F}_p$ $(0 < t \leq t_0)$ が成り立つ．ところが，これらは $\boldsymbol{x}_P$ が極小解であることに反する． □

例 6.2. $m = 3$, $K(\boldsymbol{x}_0) = \{1, 3\}$ のときを考える．このとき，補題 6.1 の証明に出てきたように

$$B(\boldsymbol{x}_P)\boldsymbol{d}_0 = \begin{pmatrix} \langle \boldsymbol{a}_1, \boldsymbol{d}_0 \rangle \\ 0 \\ \langle \boldsymbol{a}_3, \boldsymbol{d}_0 \rangle \end{pmatrix} \leq \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

を仮定しよう．今，$\langle \boldsymbol{a}_2, \boldsymbol{d}_0 \rangle \geq 0$ かもしれないが，$\langle \boldsymbol{a}_2, \boldsymbol{x}_P \rangle < b_2$ であるから，

$$\langle \boldsymbol{a}_2, \boldsymbol{x}_P \rangle + t\langle \boldsymbol{a}_2, \boldsymbol{d}_0 \rangle \leq b_2 \quad (0 < t \leq t_0)$$

となるように t_0 を選ぶことができる．よって，このとき，

$$A(\boldsymbol{x}_P + t\boldsymbol{d}_0) = \begin{pmatrix} \langle \boldsymbol{a}_1, \boldsymbol{x}_P \rangle + t\langle \boldsymbol{a}_1, \boldsymbol{d}_0 \rangle \\ \langle \boldsymbol{a}_2, \boldsymbol{x}_P \rangle + t\langle \boldsymbol{a}_2, \boldsymbol{d}_0 \rangle \\ \langle \boldsymbol{a}_3, \boldsymbol{x}_P \rangle + t\langle \boldsymbol{a}_3, \boldsymbol{d}_0 \rangle \end{pmatrix} \leq \begin{pmatrix} b_1 \\ b_2 \\ b_3 \end{pmatrix} = \boldsymbol{b} \quad (0 < t \leq t_0)$$

が成り立つ．

補題 6.1 を経由することにより，FJ 条件の導出は次の幾何学的な事実に帰着される．

補題 6.3 (Farkas の補題)**.** $m \times n$ 行列 B と $\boldsymbol{c} \in \mathbb{R}^n$ に対し，次の二条件を考える．

(1) 任意の $\boldsymbol{d} \in \mathbb{R}^n$ に対し，

$$B\boldsymbol{d} \leq \boldsymbol{0} \Rightarrow \langle \boldsymbol{c}, \boldsymbol{d} \rangle \geq 0$$

が成り立つ．

(2) 等式 $\lambda \boldsymbol{c} + B^\top \boldsymbol{\mu} = \boldsymbol{0}$ をみたす $(\lambda, \boldsymbol{\mu}) \in \mathbb{R}_+^{m+1} \setminus \{\boldsymbol{0}\}$ が存在する．

このとき，(1) ならば (2) が成り立つ．

証明． (1) を仮定する．さらに，$\boldsymbol{c} = \boldsymbol{0}$ のときは明らかであるから，$\boldsymbol{c} \neq \boldsymbol{0}$ を仮定する．まず，B を

$$B = \begin{pmatrix} \boldsymbol{b}_1^\top \\ \vdots \\ \boldsymbol{b}_m^\top \end{pmatrix}, \quad \boldsymbol{b}_j \in \mathbb{R}^n$$

と表すと，(1) は $-\boldsymbol{c}$ と $\{\boldsymbol{b}_1, \ldots, \boldsymbol{b}_m\}$ が原点を通る超平面では分離できないことを意味する．よって，$-\boldsymbol{c}$ は $\{\boldsymbol{b}_1, \ldots, \boldsymbol{b}_m\}$ で生成される凸錐に含まれる．すなわち，

$$-\boldsymbol{c} = \mu_1 \boldsymbol{b}_1 + \cdots + \mu_m \boldsymbol{b}_m$$

をみたす $(\mu_1, \ldots, \mu_m) \in \mathbb{R}_+^m \setminus \{\boldsymbol{0}\}$ が存在する．これを B により表せば，(2) が得られる． $\square$

（FJ 条件の導出）

$\boldsymbol{x}_P$ を (P) の極小解とする．このとき，補題 6.1 により，$\boldsymbol{d} \in \mathbb{R}^n$ に対し，

$$B(\boldsymbol{x}_P)\boldsymbol{d} \leq \boldsymbol{0} \Rightarrow \langle \nabla f(\boldsymbol{x}_P), \boldsymbol{d} \rangle \geq 0$$

が成り立つ．これは補題 6.3 の (1) において $B = B(\boldsymbol{x}_P)$, $\boldsymbol{c} = \nabla f(\boldsymbol{x}_P)$ と考えた場合である．よって，補題 6.3 により，

$$\exists (\lambda, \boldsymbol{\mu}) \in \mathbb{R}_+^{m+1} \setminus \{\boldsymbol{0}\} \quad \text{s.t.} \quad \lambda \nabla f(\boldsymbol{x}_P) + B(\boldsymbol{x}_P)^\top \boldsymbol{\mu} = \boldsymbol{0}$$

が成り立つ．ここで，

$$B(\boldsymbol{x}_P)^\top \boldsymbol{\mu} = \sum_{k \in K(\boldsymbol{x}_P)} \mu_k \boldsymbol{a}_k$$

であるから，$k \notin K(\boldsymbol{x}_P)$ に対して，$\mu_k = 0$ と選びなおせば，FJ 条件が得られる．

6.2 KKT 条件

FJ 条件において，$\lambda \neq 0$ のとき，次の KKT 条件 (the Karush-Kuhn-Tucker conditions) が得られる．

KKT 条件

以下を満たす $\boldsymbol{\xi} = (\xi_1, \ldots, \xi_m)^\top \in \mathbb{R}_+^m \setminus \{\boldsymbol{0}\}$ が存在する．

$$\nabla f(\boldsymbol{x}_P) + \sum_{k=1}^m \xi_k \boldsymbol{a}_k = \boldsymbol{0},$$
$$\xi_k(\langle \boldsymbol{a}_k, \boldsymbol{x}_P \rangle - b_k) = 0 \quad (1 \leq k \leq m).$$

KKT 条件に出てきた $\boldsymbol{\xi}$ は KKT ベクトルと呼ばれる．問題 (P) において，いつでも KKT 条件が成り立つとは限らないことに注意する．

定義 6.4. 問題 (P) の極小解 $\boldsymbol{x}_P$ に対し，$\{\boldsymbol{a}_k : k \in K(\boldsymbol{x}_P)\}$ が線形独立であるとき，正規条件が満たされているという．

補題 6.5. 問題 (P) について正規条件が満たされているとき，KKT ベクトルが存在する．

証明． $\boldsymbol{x}_P$ は (P) の極小解であるから，

$$\lambda \nabla f(\boldsymbol{x}_P) + \sum_{k=1}^{m} \mu_k \boldsymbol{a}_k = \mathbf{0}$$

を満たす $(\lambda, \mu_1, \ldots, \mu_m)^\top \in \mathbb{R}_+^{m+1} \setminus \{\mathbf{0}\}$ が存在する（FJ 条件）．今，$\lambda = 0$ ならば，

$$\sum_{k=1}^{m} \mu_k \boldsymbol{a}_k = \mathbf{0}$$

となる．さらに，$k \notin K(\boldsymbol{x}_P)$ に対し，$\mu_k = 0$ と選んでいたので，

$$\sum_{k \in K(\boldsymbol{x}_P)} \mu_k \boldsymbol{a}_k = \mathbf{0}$$

が成り立つ．このとき, 正規条件から $\mu_k = 0\ (k \in K(\boldsymbol{x}_P))$ となり, $(\lambda, \mu_1, \ldots, \mu_m)^\top = \mathbf{0}$ を得る．ところが，これは FJ 条件に反する．よって，$\lambda \neq 0$ が導かれる． □

7　演習問題

問題 1. 2 次の対称行列 $A = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}$ に対する次の問いに答えよ．

(1) A の固有値，固有ベクトルを求めよ．

(2) A のスペクトル分解を求めよ．

(3) A が正定値であることを示せ．

(4) $\|\cdot\|_A$ を A から定まる $\mathbb{R}^2$ 上のノルムとする．このとき，$\|\cdot\|_A$ で計った単位球 $\{\boldsymbol{x} \in \mathbb{R}^2 : \|\boldsymbol{x}\|_A \leq 1\}$ を図示せよ．

問題 2. $A = (a_{ij})$ を n 次の半正定値行列とする．次の問いに答えよ．

(1) $A = B^\top B$ をみたす n 次正方行列 B が存在することを示せ（ヒント：A の対角化を考える）．

(2) (1) で求めた B を $B = (\boldsymbol{b}_1 \ \cdots \ \boldsymbol{b}_n)$ と表す．このとき，$a_{ij} = \langle \boldsymbol{b}_i, \boldsymbol{b}_j \rangle$ を示せ．

(3) $X = \{1, \ldots, n\}$ とおき，$X \times X$ 上の関数 k を $k(i, j) = a_{ij}$ により定める．この k は X 上のカーネル関数であることを示せ（よって，対称行列はカーネル関数である）．

問題 3. 5.3 の準備で構成した $\varphi_1, \ldots, \varphi_n$ に対し，$\Phi = (\varphi_j(x_i))$ は下三角行列であること，すなわち，

$$\Phi = \begin{pmatrix} \varphi_1(x_1) & 0 & \cdots & 0 \\ \varphi_1(x_2) & \varphi_2(x_2) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ \varphi_1(x_n) & \varphi_2(x_n) & \cdots & \varphi_n(x_n) \end{pmatrix}$$

が成り立つことを示せ（よって，$K = \Phi\Phi^\top$ は K のコレスキー分解を与える）．

問題 4. U, V をそれぞれ $\mathbb{R}^n$, $\mathbb{R}^m$ での座標の入れ換えを表す行列とする．例えば，$n = 2$ のときは $U(x_1, x_2)^\top = (x_2, x_1)^\top$ である．このとき，次の問いに答えよ．

(1)

$$K(U\boldsymbol{x}, V\boldsymbol{x}_*) = \begin{pmatrix} U & 0 \\ 0 & V \end{pmatrix} K(\boldsymbol{x}, \boldsymbol{x}_*) \begin{pmatrix} U & 0 \\ 0 & V \end{pmatrix}^\top$$

が成り立つことを示せ.

(2) 多次元ガウス分布 $N(\mathbf{0}, K(U\boldsymbol{x}, V\boldsymbol{x}_*) + \beta^{-1} I_{n+m})$ を考え，入力 $U\boldsymbol{x}$ に対応する出力 $U\widehat{\boldsymbol{z}}$ を得た後の条件付き分布を $p(\boldsymbol{t} \mid U\widehat{\boldsymbol{z}})$ とする．このとき，$p(\boldsymbol{t} \mid U\widehat{\boldsymbol{z}})$ の平均ベクトルは $V\boldsymbol{\mu}(\boldsymbol{x}_*)$, 共分散行列は $V\Sigma(\boldsymbol{x}_*)V^\top$ であることを示せ（よって，条件付き分布は変数の入れ換えに対し整合がとれている）．

問題 5. $\{\boldsymbol{y}_n\}_{n\in\mathbb{N}}$ を簡易版ガウス過程とする．すなわち，$x_1, x_2, \ldots$ が与えられたとき，任意の $n \in \mathbb{N}$ に対し，

$$\boldsymbol{y}_n(\boldsymbol{w}) = \boldsymbol{y}(\boldsymbol{w}; x_1, \ldots, x_n) \sim N(\mathbf{0}, K(x_1, \ldots, x_n))$$

が成り立つことを仮定する．このとき，任意の $B \subset \mathbb{R}^n$ に対し，

$$P(\boldsymbol{y}_{n+m} \in B \times \mathbb{R}^m) = P(\boldsymbol{y}_n \in B)$$

が成り立つことを示せ（ヒント：ブロック行列，平方完成，シューアの補行列）．

問題 6. p. 145 で定めた $\mathcal{C}$ が閉集合であることを示せ.

参考文献

[1] Ş. İ. Birbil, J. B. G. Frenk and G. J. Still, An Elementary Proof of the Fritz-John and Karush-Kuhn-Tucker Conditions in Nonlinear Programming, European J. Oper. Res. 180 (2007), no. 1, 479–484.

[2] H. Dym and H. P. McKean, *Gaussian Processes, Function Theory and the Inverse Spectral Problem*, Dover.

[3] 飛田 武幸, 櫃田 倍之, ガウス過程, 紀伊國屋書店.

[4] 木村 英紀, 古典的補間理論の回路、制御への応用, 京都大学数理解析研究所講究録 653.

[5] W. Rudin, *Real and Complex Analysis*, McGraw-Hill.